Dynamics of Phonological Coding in Bilingual Visual Word Perception

Cover illustration: M.C. Escher's “Belvedere” (lithograph, 1958, 46 × 29.5 cm)

Printed by PrintPartners Ipskamp, Enschede.

Dynamics of Phonological Coding in Bilingual Visual Word Perception

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. mr. P.F. van der Heijden
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 9 februari 2005, te 14:00 uur

door

Martin van Leerdam

geboren te Curaçao

Promotor: Prof. dr. A.M.B. de Groot
Co-promotor: Dr. A.M.T. Bosman

Faculteit der Maatschappij- en Gedragswetenschappen
Afdeling Psychologie
Universiteit van Amsterdam

Foreword

This thesis explores the very mental flow in your brain as your eyes cascade along these printed words: The cognitive dynamics of visual word perception. The object was to reveal what actually takes place in our mind during the split second we perceive a written word. Like in many other studies of reading, the focus here is on the visual perception of isolated words. It is widely acknowledged that asking how single words are processed is central to understanding reading, and over the past two decades it has been a major focus of research in one of the most exciting fields of cognitive psychology.

Word reading is an extraordinary cognitive ability. The moment a printed word leaps into our brain, a magnificent piece of machinery is launched that operates both fast and accurately. Indeed, readers make rather short work of written words. To process a word only takes tenths of a second, which allows a skilled reader to digest words at a staggering rate of five words per second or more. To give you an impression, by the time you finish this sentence the current page already plunged 209 words into your head, and for those who will reach the final page of this thesis, it will be 90,000 more. As easy as reading may seem, investigating it certainly is not, for we are dealing with an intricate process. Indeed, one may easily fail to appreciate its true complexity: More than a century of profound scientific inquiry has produced countless interesting facts and theories but, even today, many issues remain unsolved. Nevertheless, a great deal has been learned in the past decades and most researchers working in the field now agree on the basic tenets of visual word processing.

One widely accepted point of view asserts that the dynamics of word perception entail complicated interactions among dimensions of orthography, phonology, and meaning, which exemplifies a complex cognitive system. The dynamics of cognitive systems, however, are hard to track in the anatomy of the living brain, even with contemporary imaging techniques such as computerized axial tomography or magnetic resonance imaging. Fortunately, the branch of experimental psychology yields simple but effective tools with which we can obtain a detailed impression of our mental flow. In the present research, we used a new experimental task to investigate visual word processing. Under controlled laboratory conditions, our research participants were presented with multiple series of experimental trials. In each trial, a printed word was shown on a computer screen, and simultaneously part of a spoken word was played over a headphone. The task of the participants was to decide quickly in each trial whether the visual and auditory stimuli were consistent with one another. The response was recorded, measuring both accuracy and response time. In this study, many thousands of responses were collected (in fact as many as there are words in this thesis), and together they enlighten us about a grand mystery: How our brain succeeds in converting a printed word into a coherent language experience.

Many persons contributed to this project. First, I would like to express my gratitude to my supervisors, Annette de Groot and Anny Bosman. They granted me complete freedom in pursuing my idiosyncratic ideas on visual word processing and in the way these ideas were translated into a research project. Their continuous support in this has been a real privilege, and I am very grateful for that. Special thanks go to Anny, for convincing me many years ago that I could actually be a researcher. In the final phase of this project, Annette de Groot and Anny Bosman provided invaluable comments on preliminary versions of the manuscript. These were always detailed and insightful, and led to many improvements. Furthermore, their positive attitude towards my decision to drastically cut down working time after my children were born was tremendous, and it still means a great deal to me. It was very rewarding to work on my dissertation and at the same time be able to spend so much time with my two precious children. For that I also want to thank the Department of Psychology of the University of Amsterdam for its cutting-edge policy on parenthood leaves. I wish that more institutions would allow their workers to spend *quantity* time with their children.

The Psychonomy Section was a stimulating place to work at. I would like to thank all my colleges for their technical and mental support, especially my roommate Bjørn Dekker. He was the key to all my Macintosh related inquiries and was always willing to here out my lectures on confidence intervals, or any other kind of account. I much appreciated the friendly and reassuring E-mails of Guy Van Orden of the Arizona State University and Marc Brysbaert of the Ghent University. Also encouraging were conversations with Andrea Donker, Mark Rotteveel, Guido Valk, Jan Houtveen, Mark Rietveld, Diane Pecher, Janet van Hell, Susanne Borgwaldt, and Ingrid Christoffels. In addition, many thanks to Bert Molenkamp, Nico Notebaart, and Wim van der Mije, for their technical assistance. Bert manufactured a voice key for me, Nico advised me countless times on sound recording software and equipment, and Wim created a fabulous experiment generator. Judy Kroll of the Pennsylvania State University was so generous to put her experimental lab to our disposal. Credit to Anique Bakker and Sandra Ginder, for testing the American participants. A special word of thanks to Bryony Cooper, who was so kind to help us out with the recording of spoken words. She carefully pronounced hundreds of rather nasty letter strings. Without her assistance, we would never have obtained such fine stimulus materials.

I would not have written this thesis if I was not so fortunate to have such a safe home base. Love to my wonderful wife Loes and to our Roos and Marijn. Finally, I thank my parents for their help during the last three years. Since they always have supported me unconditionally in everything I did, it is to them I dedicate this book.

Martin van Leerdam, December 9, 2004

Contents

1 When MOOD Rhymes with ROAD: Spelling-to-Sound Dynamics in Bilingual Visual Word Perception

Does MOOD rhyme with ROAD? The answer is that it does not. According to the spelling-to-sound correspondence rules of English orthography, the rimes of the words MOOD and ROAD (i.e., *–OOD* and *–OAD*) have dissimilar pronunciations, thus homophony is definitely out of the question. However, to ask whether MOOD and ROAD sound similarly is not as futile as it seems, because, as we will show, within the reader's mind it may actually be possible that processing the word MOOD elicits phonology appropriate to the word ROAD. In general, the present study explores the cognitive dynamics in which processing an English word involves assembly of phonology not commonly associated with the spelling. This phenomenon is not artificially induced by forceful experimental manipulation. On the contrary, we argue that this occurs routinely in the process of visual word perception, that is, for the population of Dutch-English bilinguals when reading English words.

In this study we acknowledge the fact that in writing systems spelling relates to sound in complicated ways, and that these complications exert their influence on the process of visual word perception. Moreover, understanding that the bilingual reader accommodates two languages and therefore has knowledge of more than one way to relate spelling to sound, allows the possibility that inappropriate phonology from the native language emerges jointly with correct phonology in the process of second-language word perception. Returning to the example, bilingual visual processing of MOOD may elicit extraneous phonology rhyming with ROAD because this particular phonology is concordant with knowledge of spelling-to-sound relations in Dutch orthography. Native speakers of Dutch have learned, to asymptotic degrees, that the rime *–OOD* is pronounced consistently /od/, as in ROOD, LOOD, and NOOD (in Dutch meaning "red", "lead", and "distress", respectively)—which, as it happens, matches the rime of ROAD.

A general principle that synthesizes this study is the acknowledgement that cognitive systems tackle complicated relations with the environment by accommodating *manifold relations* between an environmental surface form and multiple cognitive functions (cf. MacWhinney, 1997). This concept of manifold relations is introduced in the next section, followed by a general outline of complications that are apparent in the English writing system. We then consider a theoretical framework that describes how these complications are reflected in monolingual knowledge structures and word processing, and proceed to generalise it to the bilingual domain. The final section clarifies our research goals.

MAPPING FORM TO MULTIPLE FUNCTIONS

Human beings are able to learn the names of one another, even when people carry various names. Suppose "Bart" is the name of a baker you are familiar with. Rumours go that the baker's wife rarely uses that name. Most of the time she calls him "Precious". The employees at the bakery respectfully call him "Mr. Pater". In his absence, however, it is not "Mr. Pater"; it is "Mr. Precious". So the very same person you can refer to as "Bart", "Mr. Pater", "Precious", and "Mr. Precious".

This commonplace example reveals a simple but powerful capacity of the human cognitive system. Through a process of experience, the cognitive system is capable of learning manifold associations between an environmental surface form and multiple cognitive functions (i.e., percepts and actions). The capacity of forming manifold associations implies flexibility: It allows the cognitive system to adapt to environmental variability. Flexibility in associative learning precludes that learning is restricted to singular functional relations only (i.e., single-valued relations, Van Orden, Pennington, & Stone, 1990, see upper panel of Figure 1). Returning to the example, regular visits to the bakery bring about manifold associations between a person's identity and several potential names. Having this ability is of vital importance, for we would be socially challenged if we learned one name for our boss, for instance "Mr. Precious", and were unable to learn another one. Somehow, our cognitive system is capable of learning new facts without being forced to forget others. This relates to what is called the stability-plasticity dilemma, which concerns the problem of: "How does brain processing keep old memories stable and still maintain enough plasticity to learn new things?" (S. Grossberg, 1995, page 438).

For the cognitive system, however, this flexibility comes with a cost. In the process of coupling a stimulus form to a particular function, all previously associated functions are potential candidates. Having more than one candidate induces considerable ambiguity in the system, that is, which of the functions is the relevant one? Ambiguity is resolved by the *contextual dependencies* in which the stimulus form appears. A certain cognitive function is only relevant if the stimulus form occurs in the same circumstances as it was previously experienced. In case there is no disambiguating context, the relative amount of experience (i.e., statistical dominance) with the different associations will determine with which function integration takes place.

Back at the bakery, you were not confined to learning just one name for the owner. But which name do you use to refer to him? Setting foot in the bakery you may like to greet him with "Mr. Pater". It might be inappropriate if you called him "Bart", and even more so if you entered saying: "Hello, Precious". Talking to his wife you could either use "Bart" or "Mr. Pater" to refer to him, but in the presence of his employees you might want to use "Mr. Pater". When speaking to the employees you would probably avoid “Bart” and use the name “Mr. Pater” or perhaps “Mr. Precious”. Finally, if a friend would ask you for the baker’s name you may experience some trouble in coming up quickly with a name, because in this low-context situation “Mr. Pater, “Bart”, or “Mr. Precious” may all be suitable candidates. It is evident that the name you use for the bakery owner depends heavily on the circumstances you find him in.

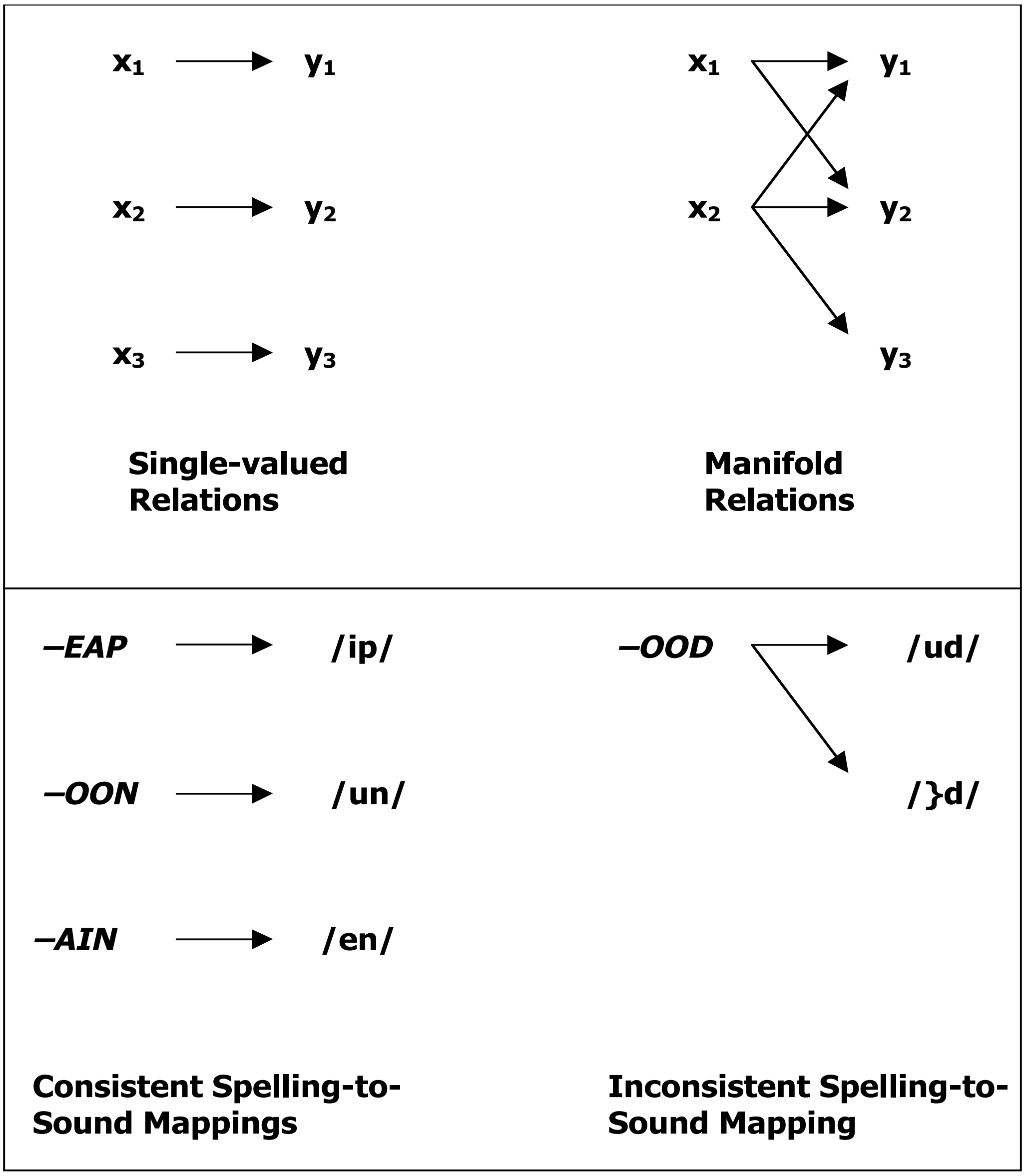

Figure 1. Upper panel: Functional relations can be described in the mathematical sense of single-valued relations. Each *x* of a set of *x*s can be mapped onto a *y* of a set of *y*s, and no *x* can be mapped onto two *y*s. Manifold relations involve violation of the single-valued constraint. Lower panel: For the functional relation with respect to phonological coding, each *x* corresponds to an orthographic string and each *y* corresponds to a phonologic string (see Van Orden et al., 1990). Single-valued relations between orthography and phonology entail consistent spelling-to-sound mappings, whereas manifold relations give rise to inconsistent spelling-to-sound mappings.

The capability of forming manifold associations can be conceived as a general principle of mental functioning. For the cognitive system it is advantageous to encompass a certain rate of plasticity and tolerate inconsistent experience. For example, green apples may be associated with unripe, indigestible fruit. Yet, they may also have been encountered as delicious apples of some “evergreen” species. In that case, contextual information (e.g., relative size or colour of other apples in a tree) provides hints about the kind of fruit one is dealing with. If manifold associations were not tolerated, “evergreen” apples would be off the food chain. This example applies to a broad range of cognitive phenomena. It is beneficial to be able to learn that expressing fear can calm an aggressor—and sometimes evoke it, or that a fire can heat a room but burn the house. In addition, we may augment this general principle. In case of manifold associations between surface form and multiple cognitive functions, the cognitive process of coupling form to function, at least initially, considers *all* previously associated functions (see also Van Heuven, 2000, for a similar discussion). Then, constrained by contextual dependencies and relative statistical dominance of candidates, the system progressively moves to temporarily stable, singular form-function integration. It may also be the case, however, that the system is always entertaining alternative possibilities and never fully settles on a dominant percept or action, with the result that the system is always slightly unstable (i.e., *metastable*, see Van Orden, Jansen op de Haar, & Bosman, 1997).

Manifold Spelling-to-Sound Mappings

The present research is guided by the general principle that when relations are variant, the cognitive system is able to form manifold associations between surface forms and multiple cognitive functions. However, it pays a price for it, since manifold associations imply ambiguity that, in the course of mental processing, must be resolved. In this study, we take a close look at one of the most intricate workings of the human mind: The miraculous skill of reading printed words. In line with contemporary reading research, we put the emphasis on perception of single words, as this is the most predictive aspect of reading skill (e.g., Perfetti, 1985; Rayner & Pollatsek, 1989). We adopt a psycholinguistic approach that captures a set of idiosyncratic relations between a word’s written form and its pronunciation, and investigate experimentally how violations of single-valued relations in a writing system (e.g., the rime *–OOD* pronounced in English as /ud/ in MOOD but also as /}d/ in BLOOD, see lower panel of Figure 1) affect word perception. Manifold spelling-to-sound mappings cause ambiguity of pronunciation, which is assumed to affect reading performance. We take up the case where the reading system is challenged to its utmost with unsystematic spelling-to-sound relations, that is, where spelling relates to sound differently across two languages, such as when the rime *–OOD* is pronounced as /ud/ and /}d/ in the English words MOOD and BLOOD but also as /od/

in the Dutch word LOOD. For that purpose, we turn to the domain of bilingualism and concentrate on bilinguals reading in their second language, using a comprehensive theoretical framework that has been developed for the monolingual field. Before turning to the issue to what extent manifold spelling-to-sound relations have an impact on word perception, we first treat some relevant characteristics of the English writing system.

Inconsistent Relations in Alphabetic Writing Systems

The example below suggests that the English writing system did not evolve with the purpose of promoting literacy acquisition. Consider the following baloney sentences: "H*ave* you h*eard* h*ow* D*ave* *wound* up a w*atch* with a *wound* at the thumb and a *tear* in the b*eard*? D*o* not ever *tear* th*atch* with your bare hands! Didn't he kn*ow* that bl*ood* is n*o* g*ood*?"

Forming manifold associations between form and function turns out to be particularly pertinent in the domain of written language perception, that is, in the human cognitive skill of reading printed words. Making sense of writing involves a highly complex dynamic process of integrating dimensions of spelling, sound, and meaning unfolding over time. Because of irregularities (i.e., complications) in many writing systems, manifold relations among the visual and linguistic dimensions affect both the process of learning to read and reading performance.

Manifold relations between a word's written form (i.e., spelling) and appropriate linguistic functions occur at multiple grain sizes of form-function correspondence. At a *fine-grain* descriptive level it pertains to how individual letters are related to sound segments, for instance the letter *B* corresponding to the pronunciation /b/ in the English word BAKER. Larger units of spelling and sound are associated at an *intermediate-grain* level of correspondence, for instance the letter cluster *–ATCH* corresponding to the pronunciation /cC/ as in WATCH (see Ziegler, Stone, & Jacobs, 1997, for key to phonetic symbols). This level of correspondence is of special concern in this study as shall be clarified later on. Finally, an even coarser level of correspondence is realised in *coarse-grain* (word-size) mappings of spelling to sound and meaning, for instance the spelling BROOD corresponding to the pronunciation /brud/ and its meaning (e.g., "offspring").

Fine-grain spelling-to-sound inconsistencies

A fine-grain level of correspondence is endorsed in alphabetic writing systems. Roman orthographies, used in most western European countries, are considered as basically alphabetic (see Borgwaldt, 2003; Mattingly, 1992, for discussions). The transcription of speech captures the elementary segments of the sound system, namely, the phonemes of the language. In an alphabet, relations between spelling and sound are generally described in terms of systematic links between graphemes and phonemes, where a grapheme is a letter or cluster of letters that corresponds to a single phoneme (e.g., grapheme-phoneme correspondence (GPC) rules, Venezky, 1970; Wijk, 1966). In the word BAKER, for example, the single-letter consonant grapheme *B* corresponds to the phoneme /b/ and the single-letter vowel grapheme *A* corresponds to the phoneme /e/. These grapheme-to-phoneme correspondences are systematic because they also occur in the words BAGEL, BACON, and BASIC.

However, the formation of GPC rules is complicated due to manifold relations between a single grapheme and multiple phonemes. Unlike many other Roman orthographies (e.g., Dutch, Finnish, and Spanish), the English writing system is plagued by a profound *vowel irregularity*. For example, in the word BAKER the grapheme A corresponds to /e/ but in the word WARM it has changed identity and stands for /o/. The grapheme *A* is also pronounced /a/ as in BART, but not in APPLE, here it corresponds to /@/. Thus at a fine-grain level the same grapheme may correspond with multiple phonemes. In fact, relations between vowel spellings and vowel pronunciations in English are always ambiguous; they never are single-valued relations (Van Orden et al., 1990). Manifold grapheme-phoneme correspondences present a challenge for the cognitive system, because in word perception ambiguities obstruct the process of coupling a grapheme to its relevant phoneme. It is assumed that they are resolved by contextual constraints from adjacent letters or letter clusters and the word context.

Intermediate-grain size spelling-to-sound inconsistencies

To characterize the English writing system as an alphabet may be inadequate. According to Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty (1995; see also Kessler & Treiman, 2001), considering orthographic and phonological units larger than single graphemes and phonemes can shed new light on the nature, use, and acquisition of the English writing system. In an extensive linguistic analysis, they showed that in monosyllabic English words with a phonological consonant-vowel-consonant (CVC) structure, letter clusters that correspond to the *rimes* of the spoken syllables play a special role in describing spelling-to-sound relations. At this grain size of description, monosyllabic words (e.g., BREAD) are commonly divided into onset and rime. The onset is the initial sequence of consonants (e.g., *BR*) and the spelling body (or rime) is everything following it (e.g., *–EAD*). We shall dwell on this unit of analysis because the present study focuses entirely on intermediate-grain size spelling-to-sound correspondences. The functional relation between spelling bodies and phonological bodies constitutes a particularly predictive grain size of analysis, because in many English words pronunciations of vowel graphemes are affected by the identity of the consonants that follow. At a fine-grain level, single-letter vowel graphemes often have more than one pronunciation (e.g., *EA* pronounced /E/ as in

BREAD but /i/ as in WHEAT). Yet, at an *intermediate-grain* descriptive level, the final consonant helps to specify the vowel's pronunciation. For example, the vowel *EA* is frequently pronounced as /E/ before /d/, as in the words HEAD, DREAD, and SPREAD, but virtually never so before /p/ (e.g., /ip/ in HEAP, LEAP, CHEAP). This particular relational structure of vowels and final consonants shows the relation between spelling and sound in English to be more systematic at the intermediate-grain level of bodies as compared to the fine-grain level of individual graphemes and phonemes. Put differently, relations between spelling and sound at an intermediate-grain level are more often single-valued (Treiman et al., 1995; Van Orden, Pennington, & Stone, 2001).

Nevertheless, even at an intermediate-grain descriptive level, spelling-to-sound relations in English are often inconsistent. Manifold relations between spelling and phonology arise when the same spelling body has more than one possible pronunciation. For instance, the spelling body *–EAD* is pronounced as /Ed/ in words like BREAD, HEAD, DREAD, and SPREAD, but also as /id/ in words like BEAD and PLEAD (however not for speakers of American English). Following the seminal work of Glushko (1979) and Jared, McRae, and Seidenberg (1990), words like BREAD and PLEAD are traditionally defined spelling-to-sound *inconsistent* because their spelling body maps into more than one pronunciation (e.g., *–EAD* mapping to /Ed/ and /id/) and words like TOAST are defined spelling-to-sound *consistent* because their spelling body has only one pronunciation (e.g., *–OAST* mapping exclusively to /ost/ in TOAST, ROAST, COAST). Words that share a spelling body with other words (e.g., HEAD, BREAD) are called "neighbors". They are classified as *friends* if their spelling bodies are pronounced the same way and *enemies* if they are pronounced differently. For example, the inconsistent word BREAD has many neighbors with identical spelling bodies but they are pronounced in various ways. Thus, HEAD is a friend of BREAD, but *PLEAD* is their enemy. Similarly, BEAD is a friend of PLEAD and now *BREAD* is an enemy. In contrast, a consistent word like TOAST also has many neighbors with identical spelling bodies, but they are all pronounced the same way. Thus, TOAST is accompanied by friends only.

Neighbors of an inconsistent word always include one or several enemies, but some words have more enemies than others. This important fact implies that spelling-to-sound inconsistency is a matter of *degree* (Jared et al., 1990). The relative number and frequency of a word's friends and enemies stipulates how systematic a word's spelling body is mapped to phonology. For example, consider the word BREAD. According to a linguistic database published in Ziegler et al. (1997) this word has 8 friends against only 4 enemies. Also, the summed frequency of occurrence in written language of these friends is considerably higher than that of its enemies (885 and 182 occurrences per million, respectively). Consequently, for all neighbors, the conditional probability of the spelling body *–EAD* mapping to the phonological body /Ed/ (as in BREAD) is higher than in case of mapping to the phonological body /id/

(as in PLEAD). Since the appropriate spelling-to-sound mapping is relatively profound in BREAD (it has more friends then enemies), its degree of inconsistency is fairly modest. This is not the case in the notorious English word PINT. Appendix B of Ziegler et al. (1997) indicates that the spelling body *–INT* corresponds to the phonological bodies /Int/ (as in HINT) and /Ynt/ (as in PINT). The mapping between *–INT* and /Int/ occurs in as many as 9 words, whereas the mapping between *–INT* and /Ynt/ occurs only in PINT. Thus the inconsistent word HINT has 8 friends against a single enemy. In contrast, the word PINT does not have any friends, but faces 9 enemies. Also, the summed frequency of friends of HINT (52 occurrences per million) is higher than that of its enemy PINT (13 occurrences per million). Hence, because the spelling-to-sound mapping in PINT is rather atypical when compared to words with similar spellings, this word is to a large degree inconsistent.

To quantify the degree of inconsistency, counts of number and frequency of friends and enemies can be converted into conditional probabilities, or *consistency ratios* (Ziegler, Jacobs, & Stone, 1996; Ziegler et al., 1997). Consistency ratios are calculated for a particular body by dividing the frequency of each of its correspondences by the total frequency of this particular body. A consistency ratio greater than .5 indicates that a word has stronger friends than enemies (i.e., it contains a typical mapping) and a consistency ratio smaller than .5 indicates that a word has stronger enemies than friends (i.e., it contains an atypical mapping). In case of the inconsistent words BREAD, PLEAD, HINT, and PINT; their consistency ratios are .67, .25, .89, and .00, respectively, for numbers of friends and enemies, and they are .83, .17, .80, and .20, respectively, for summed frequencies. These consistency ratios show that BREAD and HINT have inconsistent but *typical* spelling-to-sound mappings whereas PLEAD and PINT have inconsistent and also *atypical* spelling-to-sound mappings. Typical and atypical spelling-to-sound mappings in inconsistent words reflect asymmetrical manifold mappings between spelling bodies and phonological bodies. Ambiguity in assigning the relevant pronunciation of the spelling body is assumed to be resolved by contextual constraints, here from the onset letter(s). To finish, we have laboured extensively on the subject of intermediate-grain spelling-to-sound correspondences, as this level is critical for the design of the present experiments.

Course-grain size spelling-to-sound inconsistencies

Finally, manifold relations between print and appropriate linguistic functions may occur at a coarse-grain (word-size) descriptive level of correspondence. At this level, the same word-size spelling can be related to more than one possible meaning but all have the same pronunciation (e.g., BEAR, meaning “animal” or “to carry”), or related to more than one possible meaning and more than one possible pronunciation (e.g., TEAR corresponding with “eye fluid” or “to shred” and to pronunciations /tIr/ and

/tEr, respectively). Words with multiple meanings and a single pronunciation are called polysemous, or *homonyms* (Gottlob, Goldinger, Stone, & Van Orden, 1999). The majority of English words are polysemous (e.g., KIND corresponding with “friendly” or “sort”; ROOM corresponding with “chamber” or “space”). Homonyms are homophonic as well because the different meanings of a word share a single pronunciation. Words with multiple meanings that correspond with *different* pronunciations are called *heterophonic homographs* (e.g., TEAR, WOUND, LEAD). Homographs are rather uncommon in English. Yet, despite their rarity, the unique property of homographs makes them an interesting test case for word perception research (cf. Gottlob et al., 1999; Kawamoto, 1993; Kawamoto, Farrar, & Kello, 1994; Kawamoto & Zemblidge, 1992). Specifically, in assigning the relevant pronunciation and meaning to the visual form of a homograph there are no contextual constraints from spelling to resolve ambiguity. For instance, what is the pronunciation and meaning of the English word WIND? Does it rhyme with PINT, or with HINT? And does it refer to “a breeze”, or “to reel”? In case of homographs it is the semantic context provided by ongoing discourse that dictates the relevant identity. Furthermore, as with inconsistency of spelling-to-sound at the intermediate-grain level, inconsistency in the mapping of form to linguistic function is a matter of degree. Homographs may have typical and atypical form-function mappings due to multiple constraints, including meaning dominance, statistical dominance of a mapping in terms of frequency of occurrence, and degree of inconsistency of the spelling body.

FORM-FUNCTION DYNAMICS IN VISUAL WORD PERCEPTION

How does a reader recognise a printed word? For more than a century, dating back to Cattell (1886), Dearborn (1906), Dodge (1900), and other pioneers, this question has been the subject of intense scientific inquiry. One major theoretical issue that has dominated the field for many years concerns the role of phonological information in silent reading (e.g., Gough, 1972; Huey, 1908/1968; Humphreys & Evett, 1985; McCuster, Hillinger, & Bias, 1981; Rayner & Pollatsek, 1989; Rubenstein, Lewis, & Rubenstein, 1971). Although the process of reading ultimately involves the extraction of meaning from print, it is generally acknowledged that phonology plays an important role in it (for reviews, see Berent & Perfetti, 1995; Frost, 1998; Katz & Frost, 1992; Perfetti & Tan, 1998; Perfetti, Zhang, & Berent, 1992; Van Orden & Goldinger, 1994; Van Orden et al., 1990). Moreover, understanding the quality of phonological coding in reading is important, considering that many studies have shown that a deficit in phonological processing is the core cause of poor reading or dyslexia (e.g., Bosman, Van Leerdam, & De Gelder, 2000; Bradley & Bryant, 1978; Wagner & Torgesen, 1987; for a review see Rack, Snowling, & Olson, 1992). In what follows, we describe a comprehensive theoretical account of phonology's fundamental role in visual word perception, and employ it in later sections to assess how the presence of manifold spelling-to-sound mappings is reflected in monolingual and bilingual word processing. As shall be clarified in due course, these spelling-to-sound effects are fully compatible with the significance of phonological coding.

The Role of Phonology in Visual Word Perception

Phonological coding, the process of phonological structure arising from print, is a central and primary constituent of visual word perception. This assertion lies at the heart of the strong phonological theory of word perception (Frost, 1998; see also Carello, Turvey, & Lukatela, 1992; Frost, 1995; Lukatela & Turvey, 1994a, 1994b; Van Orden et al., 1990; Van Orden & Goldinger, 1994). From the perspective of the strong phonological theory, phonology is fundamental to reading in several ways. First, phonological coding is a *mandatory* process, that is, a phonological structure is automatically generated in the process of word perception even though the explicit pronunciation of the phonological structure is not required and may sometimes even hinder task performance. Second, the product of phonological coding is used for word perception. Both at a shallow level of word processing (i.e., "lexical access") and in deeper processing of words (i.e., extraction of meaning) word perception is to a large extent based on a phonological code. Third, phonological coding is an *early* and *primary* source of constraint on word perception. Phonological structures arise very rapidly and mediate the process of word perception by acting as a coherent frame for other ongoing processes (Van Orden & Goldinger, 1994).

This strong claim is based on an abundance of experimental findings that jointly present overwhelming evidence for a leading role of phonology. Cognitive psychologists active in the field have been very creative in devising ingenious experimental techniques to disentangle processes of word perception. For example, evidence for mandatory phonological coding has been found using *the backward masking paradigm* (e.g., Berent & Perfetti, 1995; Gronau & Frost 1997; Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988), *the Stroop paradigm* (Dennis & Newstead, 1981; Tzelgov, Henik, Sneg, & Baruch, 1996), *the letter-search task* (Hooper & Paap, 1997; Ziegler & Jacobs, 1995; Ziegler, Van Orden, & Jacobs, 1997), *the first-letter task* (Bosman & De Groot, 1995; Bosman et al., 2000), and *the proofreading task* (e.g., Bosman & De Groot, 1996; Daneman & Stainton, 1991; Van Orden, 1991; Van Orden et al., 1992).

Furthermore, in numerous studies, phonologic ambiguity of printed words (i.e., manifold mappings between spelling and sound) has been shown to affect performance in a variety of reading tasks. For instance, evidence that word perception is based on a phonological code was suggested by *pseudohomophone effects* in lexical decision (e.g., Coltheart, Davelaar, Jonasson, Besner, 1977; Rubenstein et al., 1971;

Stone & Van Orden, 1993; Van Orden et al., 1992), *cross-alphabet phonological ambiguity effects* in perception of Serbo-Croatian words (e.g., Feldman & Turvey, 1983; Lukatela, Feldman, Turvey, Carello, & Katz, 1989; Lukatela, Turvey Feldman, Carello, & Katz, 1989), and *spelling-to-sound consistency effects* in word naming (e.g., Glushko, 1979; Jared, 1997; Jared et al., 1990).

In addition, Van Orden and his colleagues (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988) demonstrated that phonology is used in reading for meaning by using *the semantic categorisation task* (see also Coltheart, Patterson, & Leahy, 1994; Jared & Seidenberg, 1991). More evidence for phonological constraints on meaning extraction was provided by *homophonic priming effects*, using a fast priming paradigm (e.g., Brysbaert, 2001; Grainger & Ferrand, 1996; Lesch & Pollatsek, 1993; Perfetti & Bell, 1991). This technique was also employed by Lukatela and Turvey (1994a, 1994b), whose priming effects also provided evidence that phonological codes are early sources of constraint in visual word perception. Finally, additional evidence for rapid phonological coding in reading for meaning was found by Peter and Turvey (1994), using the semantic categorisation task.

To give an impression of the compelling nature of phonology effects in experimental paradigms that have been used to study phonological coding in visual word perception, we describe two studies that are frequently cited in the literature. One study that turned out influential was conducted by Van Orden (1987). In this study, participants were asked to decide whether a presented stimulus word belonged to a particular semantic category (e.g., the category “flower”). Van Orden demonstrated that if the stimulus words were homophonic to category exemplars, participants produced large false-positive error rates. For example, if participants were presented with the stimulus word ROWS (which sounds identical to the exemplar word “rose”), they misclassified it more often as a flower than a control word such as ROBS. ROWS is mistaken for a “rose” because homophones share phonology. It is an impressive effect; the pronunciation of the homophone influences performance in a categorization task in which the word is never heard (see also Van Orden, Holden, Podgornik, & Atchison, 1999). Another striking demonstration of phonological coding was presented by Ziegler and Jacobs (1995). These authors used the letter-search task, in which participants are asked to identify a pre-specified target letter in a briefly presented masked letter string. For example, if the participants were presented with a pseudohomophone such as BRANE (which is phonetically identical to the word “brain”), they more often (mis)reported having seen the letter “i” than in case of a control nonword such as BRATE. Presumably, mandatory phonological coding of BRANE suggested that “brain” was presented, causing participants to perceive the letter “i”.

A General Resonance Framework for Word Perception

A central and primary role for phonology is consistent with the *phonological coherence hypothesis* advanced by Van Orden and his colleagues (Van Orden, 1991; Van Orden & Goldinger, 1994; Van Orden et al., 1990). This hypothesis will be introduced below. First, we introduce some key notions of a resonance framework for word perception (Van Orden & Goldinger, 1994), which is essentially based on Grossberg's adaptive resonance theory (e.g., Grossberg, 1995). For a full outline of these notions we refer to the paper of Van Orden and Goldinger (1994). Rooted in dynamic systems theory, the phonological coherence hypothesis is typically expressed in terms of adaptive resonances within a triangular framework consisting of a level of orthographic processing units (i.e., subsymbols or nodes, see Van Orden et al., 1990, 1997; Bosman & Van Orden, 1997), a level of phonologic processing units, and a level of semantic processing units (see also the papers of Grossberg, 1995 and Stone, 1994, for introductory treatments of the resonance approach). Figure 2 portrays the macro dynamics of a recurrent network for word perception, as described in Van Orden and Goldinger (1994). In this general resonance framework, a bidirectional connective matrix links the processing units of all levels into a fully interdependent recurrent network. In the process of word perception, a pattern of activation over orthographic processing units flows forward to create a pattern on all associated phonologic and semantic processing units. In return, these processing units feed activation (i.e., top-down expectations) back to the orthographic units, transforming phonologic and semantic patterns back into an orthographic form.

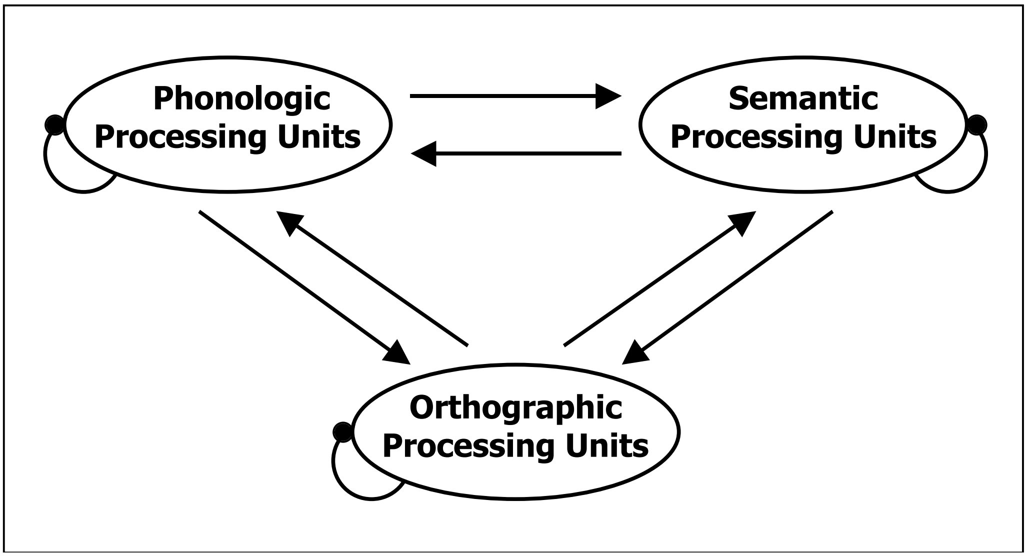

Figure 2. A highly simplified illustration of the macro dynamics of a recurrent network of orthographic, phonologic and semantic processing units as described in Van Orden and Goldinger (1994) and Bosman and Van Orden (1997).

Resonance is achieved when feedforward and feedback sources of interactive activation are mutually reinforcing (e.g., Grossberg, 1980; Grossberg, 1995; Grossberg & Stone, 1986). If the feedback pattern matches the initial feed-forward flow there is conservation of forward activation flowing from orthographic processing units. This enables the orthographic pattern to reactivate the phonologic and semantic patterns which, in turn, have the ability to reactivate the orthographic pattern, and so on. Within limits, this dynamic flow of activation is self-perpetuating, integrating the separate sources of orthographic, phonologic, and semantic information into a coherent perceptual experience (see Gottlob et al., 1999).

The phonological coherence hypothesis

The phonological coherence hypothesis maintains that visual word processing and the extraction of a word's meaning proceeds at a rate that is scaled by the temporal evolution of a unique and stable phonological code leading to resonance (Van Orden & Goldinger, 1994). Within this framework, local orthographic-phonologic resonances cohere before the local phonologic-semantic and orthographic-semantic resonances, providing a foundation for stabilizing other ongoing linguistic codings. A coherent phonological code, emerging instantly in resonance, acts to resolve the competing states of phonologic-semantic and orthographic-semantic dynamics. The orthographic-phonologic activation dynamics stabilises earliest because of *covariant learning* and *self-consistency* (Van Orden & Goldinger, 1994; Van Orden et al., 1990). Spellings of words and their pronunciations covary to a high degree, resulting in strong connection weights between each of a word's orthographic and phonologic processing units. In general, the covariation between orthographic and phonologic structure approximates more closely a single-valued map than the covariation between orthographic and semantic structure, or between phonologic and semantic structure. For example, in words like BAKER, BAGEL, and BACON the word-initial grapheme *B* covaries in a highly systematic way with the phoneme /b/. However, neither the grapheme nor the phoneme has a noticeable semantic relationship with any of thousands words beginning with *B* and /b/. Thus, the covariance statistics dictate the eventual time course of local resonances achieving coherence. Local orthographic-phonologic resonances precede the other ongoing local resonances and thereby provide immediately available constraints on processing (Gottlob et al., 1999; Van Orden & Goldinger, 1994).

The principle of covariant learning explains how a reading system is developmentally configured by tracking (i.e., learning) nested correlations of spelling forms and their linguistic functions, and the specific consequences for reading performance (Van Orden & Goldinger, 1994). The principle of self-consistency describes how this covariance predicts reading performance. In the resonance framework, a matrix of weighted connections codes covariations between orthographic and phonologic patterns at various levels of grain size. The adjustments in the weights that develop with experience will reflect the consistency of orthographic-phonologic mappings. When the same spelling-to-sound correspondence (e.g., *B* - /b/) is shared across a neighborhood of words, the weights are altered to promote local (fine-grain) resonance of the respective orthographic and phonologic processing units. Such strong-rule resonance exemplifies local dynamics exhibiting high self-consistency.

Using another construct of dynamic systems theory, manifold form-function associations can be described in a topology of *attractors* in a state space (e.g., Abraham, Abraham, & Shaw, 1991; Port & Van Gelder, 1995; Van Orden & Goldinger, 1994). The state space of word perception may be distributed along spelling, sound, and meaning dimensions, corresponding to the processing units of spelling, sound, and meaning. Points in the state space are unique combinations of these units. The time course of progressing from one point to another corresponds to temporal properties of reading performance. The dynamic leading to local or global resonance in word processing may be described as movement between a point approximating the *initial conditions* and a point attractor reflecting expected patterns from previous experience. Initial conditions include both appropriate and inappropriate unit dimensions. That is, the initial conditions of word processing include all phonologic and semantic units previously associated to an orthographic form, each activated in proportion to its statistical dominance. Following this initial state, cooperative-competitive dynamics activate appropriate dimensions and inhibit the inappropriate ones. Activation of a unit is movement toward a point that includes that dimension, and inhibition of a unit is movement toward a point that lacks that dimension. This results in a trajectory to an attractor point that includes the unit dimensions that eventually come into resonance. Each attractor in state space is bounded by a separatrix that circumscribes an attractor basin. Within the attractor basin, dynamics move encodings toward the respective attractor point. However, beyond the separatrix, encodings fall in the basin of some other attractor; adjoining attractor basins often share combinations of dimensions. Encodings that fall into false-positive attractor basins can be conceived as reading errors (e.g., mispronouncing BEAD to rhyme with BREAD).

Prototypical processing of an inconsistent word

We will illustrate the word dynamics in a resonance framework by trailing a strongly simplified prototypical process of reading an inconsistent word. To reiterate, spelling-to-sound consistency effects in word perception are assumed to reflect phonological mediation, because if the reading system is sensitive to ambiguity of pronunciation we can accept that the process of phonological coding is relevant.

In this section we consider the cases of the English words MOOD and BLOOD. At a descriptive level of English orthography, the spelling body *–OOD* supports more than one possible pronunciation, in MOOD it corresponds to the phonological body /ud/, and in BLOOD it corresponds to the phonological body /}d/. Since *MOOD* has stronger friends (e.g., FOOD, BROOD, and SNOOD) than enemies (e.g., BLOOD and FLOOD) and, conversely, *BLOOD* has stronger enemies (e.g., MOOD, FOOD, BROOD, and SNOOD) than friends (e.g., FLOOD), the intermediate-grain level spelling-to-sound relation of MOOD is typical and that of BLOOD is atypical. Embodied in a recurrent network, the intermediate-grain size spelling-to-sound associations in MOOD and BLOOD, which are coded in the network's connection weights, differ in self-consistency. For MOOD it is relatively high because its intermediate-grain size correspondence (i.e., [*–OOD* - /ud/]) is shared by many words. The word BLOOD is at odds with this strong association, resulting in lower self-consistency. The relative number and summed frequency of friends and enemies indicates the relative self-consistency of the competing local, intermediate-grain size associations. The time course and outcome of competition are predicted by these self-consistencies (i.e., of [*–OOD* - /ud/] and of [*–OOD* - /}d/]). In terms of an attractor topology, *MOOD* has a strong local (i.e., componential) attractor promoting fast, correct reading performance. The friendly local attractor pulls orthographic-phonologic dynamics toward correct phonology (i.e., /ud/), whereas the local enemy attractor pulls dynamics toward incorrect phonology (i.e., /}d/ as in BLOOD). When a friendly attractor is strong and an enemy attractor is weak, performance is fast and correct. Compared to MOOD, the word *BLOOD* has a relatively weak local friendly attractor to pull orthographic-phonologic dynamics toward correct phonology (i.e., /}d/) while a strong local enemy attractor engages in pulling dynamics toward incorrect phonology (i.e., /ud/ as in MOOD).

Progressing to the reading process, presentation of the word MOOD builds a pattern of activation among orthographic processing units, that flows forward to create a diffuse pattern on all previously associated phonologic and semantic processing units. Since both /ud/ and /}d/ are previously associated to the spelling body *–OOD*, these initial conditions of perception include both (body) phonology of MOOD as well as (body) phonology of BLOOD, each structure emerging in proportion to its statistical dominance. Likewise, the initial conditions of the perception of BLOOD also include multiple phonological structures. This is due to *multistabilty* (e.g., Hock, Schöner, & Giese, 2003; Ploeger, Van der Maas, & Hartelman, 2002; Van Orden et al., 1997); the same stimulus supports multiple percepts. In the present context, multistability is caused by ambiguity of pronunciation (i.e., spelling-to-sound inconsistency).

In the course of cooperative-competitive ("clean-up") dynamics, the typical, highly self-consistent association *–OOD* - /ud/ competes with the atypical, low self-consistent association [*–OOD* - /}d/] to be absorbed in global resonance. Over time, depending on contextual constraints and feedback from the semantic level (i.e., "lexical shaping"), dynamics pull MOOD and BLOOD towards correct phonology. Nevertheless, the competition between local orthographic-phonologic resonances needs to be resolved, obstructing the process of word perception. In case of MOOD, local enemy attractors are relatively weak. Therefore, the cost in processing of MOOD, who has a strong friendly attractor, may be small, resulting in fast and correct reading. In the case of BLOOD, however, strong local enemy attractors obstruct the trajectory to the attractor point that corresponds to correct phonology, because this friendly attractor is low in strength. Consequently, processing BLOOD suffers a sizable delay plus a heightened risk of erroneous reading, such as mispronouncing BLOOD to rhyme with MOOD. As orthographic-phonologic dynamics eventually coalesce into an attractor state that best agrees with emerging phonologic-semantic dynamics, these phonologic-semantic dynamics continue toward resonance. Ultimately, orthographic-phonologic-semantic dynamics all settle into a (temporarily) coherent global resonance.

Influence of Manifold Relations on Visual Word Perception

What we know about visual word perception is largely based upon empirical studies that investigate how reading performance is influenced by characteristics of printed word stimuli. Hundreds of psycholinguistic studies have been conducted to demonstrate the influence of physical factors (e.g., degrading and masking), lexical characteristics (e.g., printed word frequency, bigram frequency and spelling-to-sound consistency), or relational factors (e.g., prime-target similarity) on visual word processing. In this section we focus on research that has accumulated experimental evidence for effects of manifold spelling-to-sound relations at multiple-grain sizes on monolingual reading of English words. In succession, we review fine-grain *regularity effects*, intermediate-grain *consistency effects,* and coarse-grain *homograph disadvantage effects*, to illustrate that manifold relations at multiple-grain sizes are indeed reflected in visual word perception.

Regularity effects in reading printed words

Researchers have shown repeatedly that the regularity of fine-grain grapheme-phoneme correspondences has an impact on word-naming time. Regular words (e.g., LIKE) that obey pronunciation rules (i.e., grapheme-phoneme-correspondence (GPC) rules, Venezky, 1970) are found to be named faster than irregular, or exception words (e.g., SAID, PINT) that do not obey these rules (e.g., Andrews, 1982; Baron & Strawson, 1976; Gough & Cosky, 1977; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Stanovich & Bauer, 1978). Regularity effects are expected because of fine-grain self-consistency of grapheme-phoneme relations. Since graphemes and phonemes covary together at a high-frequent rate, fine-grain grapheme-phoneme resonances cohere quickly, providing early constraints on more global resonances. If a word contains a grapheme-phoneme relation that corresponds to a weak pronunciation rule, global resonances emerge more slowly.

Spelling-to-sound consistency effects in reading printed words

The relative self-consistency of intermediate-grain spelling-to-sound associations predicts how long it takes to initiate word pronunciation. Numerous empirical demonstrations suggest that inconsistent words (e.g., PINT, MINT, BLOOD, MOOD) produce longer word-naming latencies than consistent words (e.g., HEAP, TOAST), particularly when inconsistent words have atypical spelling-to-sound mappings (e.g., PINT, BLOOD) and are low in printed-word frequency (e.g., Andrews, 1982; Brown & Watson, 1994; Glushko, 1979; Jared, 2002; Jared, 1997; Jared et al., 1990; Kay & Bishop, 1987; Seidenberg et al., 1984). For instance, in a representative study of spelling-sound consistency effects in word naming, Jared et al. (1990, Experiment 1) compared naming latencies for consistent words (e.g., CAPE) and two groups of inconsistent words, one with low summed frequency of enemies (e.g., CRUSH) and one with high summed frequency of enemies (e.g., BEAD). These word types correspond to words with consistent, typical, and atypical mappings (i.e., CM, TM, and AM), respectively. In line with a rank order of self-consistency, participants were slower in naming inconsistent words (i.e., with typical and atypical mappings) than in naming consistent words (AM & TM > CM). Specifically, words with atypical mappings were named slower than words with typical mappings (AM > TM). Similar effects were found in subsequent work (Jared, 2002, Experiment 1). Again, the rank order of self-consistency correctly predicted the differences AM > TM, AM > CM, and TM > CM. Within the resonance framework of Van Orden and Goldinger (1994), this effect of spelling-to-sound inconsistency is caused by simultaneous codings of multiple phonological structures, leading to time-consuming competition processes. In addition, the relative self-consistency of mappings predicted the chance of a mispronunciation. It was found that readers occasionally made errors in naming words, primarily in words that contain atypical mappings (e.g., Jared, 2002; Jared et al., 1990). Interestingly, skilled readers frequently produce regularisation errors in naming, such as incorrectly pronouncing PINT to rhyme with MINT (e.g., Jared, 2002; Monsell, Patterson, Graham, Hughes, & Milroy, 1992).

Homograph disadvantage effects in reading printed words

Although many English words have inconsistent spelling-to-sound relations (e.g., Jared et al., 1990; Ziegler at al., 1997), most words maintain global consistency, as their whole-word spellings denote singular pronunciations. The inconsistent word

PINT, for example, harbours a hideous atypical spelling-sound cluster (/Ynt/), which strongly attracts regularisation. Nevertheless, the pronunciation of PINT is always /pYnt/. By contrast, homographs entail separate pronunciations as well as separate meanings (e.g., TEAR corresponding with “eye fluid” or “to shred” and the pronunciations /tIr/ and /tEr, respectively). In terms of a state space, homographs support two coarse-grain attractor states (i.e., potential resonances), so competition occurs at all grain sizes of spelling-to-sound correspondence (e.g., Gottlob et al., 1999; Kawamoto, 1993; Van Orden et al., 1990). Hence, the dynamic processes that drive homograph perception toward a final attractor have to overcome multiple constraints. This lack of global self-consistency predicts exaggerated naming times. Empirical evidence for this idea was provided by Kawamoto and Zemblidge (1992). They compared naming latencies for inconsistent, atypical words (e.g., PINT) and homographs (e.g., LEAD). In terms of an attractor topology, processing of both types of words involves a strong enemy local attractor engaging in pulling dynamics toward incorrect phonology. Because homographs in addition have multiple, competing global-size orthographic-phonologic attractors, naming should be slower for homographs than for atypical, inconsistent words. In accordance with this reasoning, Kawamoto and Zemblidge found substantially longer naming latencies for words like LEAD than for words like PINT.

To conclude, manifold relations between spelling and sound occur at multiple grain sizes in the English orthography and these carry an effect on the cognitive process of visual word perception. If a spelling (i.e., grapheme, spelling body, or a word-size spelling) has multiple pronunciations across words, then processing of a word containing this spelling involves ambiguity of pronunciation, which must be resolved.

Mapping Form to Function Across Languages

Of all human cognitive abilities, the use of speech is certainly the most impressive. The language system that human beings use to communicate is, more than anything else, responsible for the current advanced state of human civilisation (Anderson, 1990). The cognitive system’s capability to accommodate the faculty of language is truly impressive, but now consider the cognitive system of the bilingual human being, which harbours *two* languages. For more than two decades, the nature of this coexistence has been a source of dispute among linguists and cognitive psychologists (e.g., Altenberg & Cairns, 1983; Macnamara & Kushnir, 1971; Schwanenflugel & Rey, 1986; Van Heuven, 2000; Van Heuven, Dijkstra, & Grainger, 1998; for reviews see De Groot & Kroll, 1997; Dijkstra & Van Heuven, 2002; Grainger, 1993; Harris, 1992; Kroll & De Groot, in press, Van Hell, 1998; Van Heuven, 2000).

Theoretical and Empirical Issues in Bilingual Visual Word Perception

One major issue in early bilingual research concerned the mental organisation of the two languages. Are the knowledge structures underlying the bilingual's languages stored in two independent systems or are they arranged in a single integrated system? An analogous question pertains to the way words from the different languages are processed. Is word processing in one language influenced by knowledge of words from the other language? Evidence for the independence hypothesis can be appreciated given the fact that fluent bilinguals seem to be able to use a single language at will (Van Heuven, 2000). According to the independence hypothesis, bilinguals are believed to have separate language systems that can be switched on and off by means of a control mechanism, keeping interference from the non-target language at a minimum level (e.g., Macnamara & Kushnir, 1971). In this view, word processing is essentially *language selective*. That is, the process of coupling a word's surface form (i.e., spelling or speech) to a relevant linguistic function takes only linguistic knowledge into account that involves the contextually appropriate language system, which is determined by language set information.

The language non-selective access hypothesis

Recent insights, however, hold that bilinguals cannot completely suppress the non-target language. In brief, a large number of psycholinguistic studies have supported the theoretical position that bilingualism entails a single integrated language system that processes words from the less dominant language (i.e., second language) basically *language non-selectively* (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & Van Hell, 2003; Dijkstra & Van Heuven, 2002; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; De Groot, Delmaar, & Lupker, 2000; Nas, 1983; Van Heuven et al., 1998; Van Hell & Dijkstra, 2002). For example, in a number of these studies it was demonstrated that in Dutch-English and French-English bilinguals performing lexical decisions in English ("is the stimulus an English word or a nonword?"), word knowledge from the irrelevant language was somehow involved. The bilingual participants in these studies knew, however, that words from their dominant language were included in the experiment, for instance as distracters or as prime words.

Interlingual homographs. Many of the aforementioned researchers used the case of interlingual homographs as their primary tool. A good example of an interlingual homograph is the English word ROOM, which is also a Dutch word meaning "cream". In general, interlingual homographs are letter strings that have different pronunciations and meanings across languages. Take notice that, in the present setting, interlingual homographs are exemplary to coarse-grain, *cross-language* manifold relations between form and linguistic functions. To give an impression of the compelling nature of typical experimental findings, the rationale of employing interlingual homographs will be outlined in more detail, followed by a description of a representative study by Dijkstra et al. (1998).

According to the language non-selective access hypothesis, visual word processing in one language engages word knowledge from the other language. To some, this conception of bilingual visual word perception may be at odds with common-sense ideas of bilingualism. For instance, the language of written discourse inherently provides an unequivocal language context, rendering total irrelevance for the other language. Moreover, bilingual readers seldom express that they suffer cross-language interference problems. In recent years, these considerations have been empirically challenged. Blessed with the phenomenal case of interlingual homographs, psycholinguistic researchers have produced sensational experimental effects supporting the language non-selective view. Similar to English homographs, interlingual homographs present a fruitful psycholinguistic test case in this matter.

A pair of words that may typify Dutch-English interlingual homographs is BAKER and BROOD. In English, the word BAKER is pronounced /bekR/ and usually refers to a man who bakes for his profession—bread, for instance. BAKER is also a Dutch word, but pronounced roughly similar to the word "biker" (i.e., /bikR/) and meaning "maternity nurse", it only shares the visual form with the English identity. Now consider the word BROOD. In English it is pronounced as /brud/ and usually means "offspring". In Dutch, however, BROOD is pronounced /brod/ and refers to a baker's most important merchandise: bread. Importantly, one general feature distinguishes the exemplars BAKER and BROOD. The English and Dutch readings of these words may differ in frequency of occurrence in written language (i.e., word frequency). Inspecting the Dutch and English corpus type lexicons of the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993), we find that the Dutch readings of BAKER and BROOD occur 67 and 2616 times, respectively, whereas the English readings occur 269 and 68 times, respectively. Apparently, the relative word frequencies of the two readings differ markedly, and hence their familiarity, which may affect bilingual word processing.

Regarding processing of interlingual homographs, language-selective and language non-selective views produce different lines of expectation. If bilingual word processing proceeds *language selectively*, presentation of an interlingual homograph should lead to the extraction of meaning that corresponds only to the reading of the target language. Thus, for example, if a native Dutch bilingual reads in English, an interlingual homograph such as BAKER should be recognised solely as an English word meaning "a person who bakes bread". In contrast, if bilingual word processing proceeds essentially *language non-selectively*, presentation of an interlingual homograph should elicit multiple meanings, that correspond to both the English and the Dutch reading. Thus in case of the word BAKER, a Dutch bilingual should extract meanings corresponding to both “a person who bakes bread” and “maternity nurse”. As a result, the extraction of multiple meanings may be expected to lead to differential performance in bilingual reading tasks for interlingual homographs at the one hand and single-language (non-homographic) control words at the other hand. As a matter of fact, this has been shown many times, although the exact pattern of findings is rather mixed (see Dijkstra et al., 2000, for a review). Interlingual homographs may be recognised faster than, slower than, or as fast as control words (e.g., Dijkstra et al., 1998). How can this be understood?

The language mode hypothesis. According to Grosjean (1997, 2001), the degree in which bilingual word processing proceeds language non-selectively may be a function of various factors, and this may influence performance on interlingual homographs. Thus the degree in which bilingual word processing proceeds language non-selectively in an experiment may explain the mixed results. One such factor is the relative *proficiency* of bilingual participants in their second language. If for example a Dutch bilingual participant is highly proficient in English, it can be expected that English readings of interlingual homographs are more dominant than in case of a less proficient participant. Another factor pertains to aspects of *stimulus-list composition*, such as whether words from one language or from both languages are included in the experiment. This may affect the *language mode* of the participant (i.e., the language mode hypothesis, see Dijkstra & Van Hell, 2003; Dijkstra & Van Heuven, 2002, for thorough discussions and alternative explanations). If words from the non-target language are present in the experiment, the relative prominence of the non-target language system may increase. For example, in case of a Dutch bilingual reading English words, the presence of Dutch filler words may increase the prominence of the Dutch language system, potentially accelerating the Dutch reading of an interlingual homograph.

In a series of studies, Dijkstra and his colleagues have systematically examined the effects of such factors and concluded that the homograph effect generally depends on task demands and stimulus-list composition (e.g., Dijkstra & Van Heuven, 1998; Dijkstra, Grainger, & Van Heuven, 1998; Dijkstra, Grainger, & Van Heuven, 1999; Dijkstra et al., 1998; Van Heuven, Dijkstra, & Grainger, 1998; see Dijkstra & Van Heuven, 2002, for a review). For example, in a representative study of bilingual visual word perception, Dijkstra et al. (1998; see also De Groot et al., 2000) provided striking evidence for language non-selective processing. In their study, Dutch bilingual participants performed an English lexical decision task including interlingual (English/Dutch) homographs (e.g., BAD, meaning “unsound” in English, and “bath” in Dutch) and matched single-language control words (e.g., LOW). When the stimulus list also included Dutch words, which required a “no” response, participants produced considerably slower response latencies for the homographs relative to the control words. This was not the case, however, when the stimulus list did not include Dutch words (see also Dijkstra et al., 1999, for further discussion).

Evidently, responses to the English reading of interlingual homographs were strongly inhibited by the presence of Dutch readings. Moreover, the size of the homograph effect depended directly on the relative difference in frequency of occurrence between the two readings of the homograph. If the frequency of the Dutch (non-target) reading was relatively high, response latencies to the homograph were even slower. Dijkstra et al.'s bilingual participants also performed a generalised lexical decision task, in which they had to respond with "yes" if a word from either language was presented, and respond with "no" if a nonword appeared. Under these experimental conditions the homograph effect reversed: Participants produced faster response latencies for the homographs relative to the control words. Again, the homograph effect depended on the relative difference in frequency of the two readings. Apparently, participants took advantage from the double readings of interlingual homographs to improve their lexical decision performance.

The Bilingual Interactive Activation model. In order to explain the available empirical evidence on language non-selective processing, Dijkstra and his colleagues built a computation model for bilingual word recognition, the Bilingual Interactive Activation (BIA) model (e.g., Dijkstra & Van Heuven, 1998; Van Heuven et al., 1998). Essentially, the BIA model is an extension of the Interactive Activation (IA) model (e.g., McClelland & Rumelhart, 1981). Identical to the original model, the BIA architecture consists of two lower level layers of nodes that represent features and letters, and a higher level layer of nodes that represent words. Unlike the IA model, however, the BIA version has an integrated lexicon for both Dutch and English words, and an additional higher level layer containing two "language nodes". In the BIA model, the language nodes fulfil different functions. Serving as linguistic representations, the language nodes represent language tags and also collect activation from word representations within a language. Interactive activation dynamics in the network are governed by a combination of excitatory and inhibitory sets of connections between nodes of adjacent layers, and lateral inhibitions between word nodes. In general, the implemented BIA model has been quite successful in accounting for several experimental effects in bilingual word recognition. Recently, Dijkstra and Van Heuven (2002) proposed a new theoretical model, which they called the BIA+ model. This model extends the earlier version by appending phonological and semantic lexical representations. Furthermore, it assigns a somewhat different role to the language nodes in the model (see Dijkstra & Van Heuven, 2002, for an extensive discussion). The BIA+ model has recently been implemented in a localist connectionist model, the Semantic, Orthographic, and PHonological Interactive Activation model (SOPHIA, Van Heuven & Dijkstra, in preparation).

The Role of Phonology in Bilingual Visual Word Perception

Research on bilingual word perception has yielded strong evidence that processing words in the less dominant language proceeds essentially language non-selectively, that is, second-language reading can engage knowledge of the dominant language. Thus, for example, if a Dutch bilingual reads English words, Dutch word knowledge may come into play. In the studies on bilingualism discussed so far, visual word perception was advantageously looked upon as a straightforward process of mapping spelling to meaning. In this approach, the concept of word knowledge refers to *word meaning* exclusively, ignoring the interplay with a phonological dimension. However, words do have sounds, and since word perception research in the monolingual domain has convincingly demonstrated that phonological coding is a central and primary constituent of visual word perception, current models of bilingual word perception should accommodate this process. Until recently, the role of phonological coding was surprisingly absent in accounts of bilingual word perception. This state of affairs was accurately described by Dijkstra et al. (1999) as "the neglected role of phonology".

Manifold interlingual spelling-to-sound relations

If a bilingual's two languages are transcribed in writing systems that follow the same alphabetic principle it is possible that spellings relate to sounds differently across languages. For the combination of the English and Dutch orthographies, this is in fact the case. The letter string BROOD, for example, is pronounced as /brud/ in English, but in Dutch it is pronounced as /brod/. As a matter of fact, the majority of English neighbor words with the spelling body *–OOD* is pronounced to rhyme with /ud/, whereas all Dutch words ending with *–OOD* are pronounced to rhyme with /od/. An interesting question that builds on the idea of language non-selective reading concerns whether bilinguals engage knowledge of spelling-to-sound relations from one or both languages when reading in either one of them. Thus, the phenomenon of a spelling that relates differently to sounds in two (or more) different languages can be thought of as a challenging case of manifold relations between form and function. In the case of an interlingual homograph such as BROOD we are dealing with coarse-grain, interlingual manifold relations between a single spelling (i.e., BROOD) and several pronunciations (i.e., /brud/ and /brod/). In sum, the existence of interlingual homographs verifies bilingual readers' capability to assign different, language-specific pronunciations to the same spelling.

Language non-selective phonological coding

The phonological coherence hypothesis states that phonological codes arise very rapidly in word processing, thereby acting as a mediating structure for other ongoing cognitive processes. When a Dutch bilingual reads an English word (e.g., ROOM, an interlingual homograph), the process of phonological coding, driven by English spelling-to-sound knowledge, causes a phonological structure corresponding to the English reading of the word (i.e., /rum/). The concept of language non-selectivity gives rise to the related question whether the phonological structure corresponding to the Dutch reading of the word (i.e., /rom/), driven by Dutch spelling-to-sound knowledge, is also established. Again, this refers to coarse-grain, interlingual manifold spelling-to-sound mappings.

A similar question can be posed regarding the bilingual processing of an inconsistent English word such as MOOD. Phonological coding, driven by knowledge of English spelling-to-sound relations, initially launches multiple intermediate-grain phonological structures corresponding to all English pronunciations that were previously assigned to the spelling bodies of its neighbors (i.e., /ud/ and /}d/). Eventually, dynamics are drawn to correct global, word-size phonology. The presently pertinent question is whether a phonological structure driven by knowledge of Dutch spelling-to-sound relations (i.e., /od/), is initially coded as well. That is, does second-language word processing involve simultaneous cross-language phonological coding? If it is acknowledged that in Dutch readers the spelling body *–OOD* associates very strongly with the pronunciation /od/, in fact more strongly than with any other English pronunciation of *–OOD*, we may expect this to be the case. Again, take notice that here we are concerned with intermediate-grain size, manifold interlingual relations between a single spelling body (i.e., *–OOD*) and several phonological bodies (i.e., /ud/, /}d/, and /od/), which is the unit of analysis in all experiments of the present study.

Influence of cross-language spelling-to-sound knowledge

The aforementioned considerations lead us to the more general question whether phonological coding is fundamental to bilingual visual word perception as it is to monolingual visual word perception. If we were to observe language non-selective spelling-to-sound effects in a bilingual reading task it would be reasonable to assume that the notion of a primary role for phonology in word perception, as maintained by a strong phonological model, can be extended to the bilingual domain. Until now, there has only been a hand-full of studies that investigated the role of phonology in bilingual visual word perception (Brysbaert, Van Dyck, & Van der Poel, 1999; Dijkstra et al., 1999; Gollan, Forster, & Frost, 1997; Jared & Kroll, 2001; Lam, Perfetti, & Bell, 1991; Nas, 1983; Tzelgov et al., 1996; Van Wijnendaele & Brysbaert, 2002).

An early demonstration of cross-language effects of spelling-to-sound knowledge came from a simple but ingenious experiment of Nas (1983, Experiment 2). In this experiment, Dutch-English bilinguals performed an English lexical decision task ("is the letter string an English word?"). The critical manipulation concerned the type of nonwords that participants had to reject. Specifically, half of the nonwords were interlingual pseudohomophones, letter strings that do not occur in English or Dutch orthography, but sound like Dutch words if processed as English words (e.g., SLOCK, BOAM, DOOL, MOOST, HEALP, when pronounced mean respectively "draught", "tree", "goal", "had to", and "helped", respectively). Nas created these stimuli by taking common Dutch words (e.g., SNEE, in Dutch pronounced as /sne/ and meaning "cut") and change the spellings of the words such (e.g., SNAY) that, according to knowledge of English, not Dutch, spelling-to-sound correspondences, the resulting letter string sounded like a Dutch word. In other respects, the interlingual pseudohomophones were identical to the other nonwords in the experiment. Nas observed that interlingual pseudohomophones were rejected slower than control nonwords, and they were also more often misclassified. Apparently, for the Dutch-English bilingual participants, reading English involved English-based phonological coding of letter strings, the product of which gave rise to Dutch meaning extraction. Apart from a striking demonstration of mandatory phonological coding in (bilingual) visual word processing, this study showed that second-language reading may engage word knowledge of the dominant language (i.e., Dutch phonological structures) even if it hurts performance. A similar conclusion was reached by Dijkstra et al. (1999, Experiment 2). They also asked Dutch-English bilinguals to perform an English lexical decision task. Dijkstra et al., however, focused on the English target words, some of which were homophonic with a Dutch word (e.g., the English word COW sounds like the Dutch word KOU, meaning "cold"). In comparison with English control words, these interlingual homophones produced slower "yes" latencies, indicating that Dutch meaning extraction caused delay in English word processing.

Brysbaert et al. (1999, Experiment 2) also used homophones and pseudohomophones to investigate phonological coding in bilinguals processing words in the second language. These researchers employed a masked priming paradigm, previously used by Perfetti and Bell (1991) who were first to succeed in showing priming effects with pseudohomophones. In their first experiment, replicating the findings of Grainger and Ferrand (1996) but using bilingual participants, Brysbaert et al. asked Dutch-French bilinguals to identify briefly (i.e., tachistoscopically) presented French words. The critical manipulation involved masked phonological priming. Each French target word (e.g., FAIM, which means "hunger") was preceded by a briefly presented, masked nonword prime that was either a French (intralingual) pseudohomophone of the target word with a *large* orthographic overlap (e.g., *FAIN* -

FAIM), a French pseudohomophone of the target word but with a *small* orthographic overlap (e.g., *FINT* - FAIM), or a graphemic control that had a large orthographic overlap with the target word but only a small phonological overlap (e.g., *FAIC* - FAIM). The relevant contrast here is the comparison between the FAIN and FAIC prime conditions. Brysbaert et al. observed that participants recognised more French target words preceded by a pseudohomophone such as FAIN than preceded by a graphemic control such as FAIC. This finding was later replicated by Van Wijnendaele and Brysbaert (in preparation) with Dutch stimuli and French-Dutch bilinguals. These phonological priming effects clearly indicate that word processing in the second language is sensitive to non-native intralingual phonological information.

In further pursuit, Brysbaert et al. (1999, Experiment 1) predicted that, given the importance of phonological coding, French words might also be primed by interlingual homophonic Dutch words of the participants' primary language—Dutch words that strongly resemble French words in phonologic form. This prediction was tested by contrasting three types of Dutch word primes: Dutch homophonic words (e.g., the Dutch word *WIE* for the French target word OUI, which sound rather similar and mean "who" and "yes", respectively), Dutch graphemic controls (e.g., the Dutch word *JIJ* for the French target word OUI, the Dutch word meaning "you"), and unrelated Dutch words (e.g., the Dutch word *DAG* for the French target word OUI, the Dutch word meaning "day"). Brysbaert et al. only used Dutch interlingual homophones that were homophonic according to Dutch, but not to French, spelling-to-sound knowledge. Again, they observed phonological priming effects. The Dutch-French participants recognised more French target words when they were preceded by an interlingual homophone such as the Dutch word WIE than when preceded by a graphemic control such as the Dutch word JIJ.

Similar results were found in Experiment 2 of Brysbaert et al. (1999), using Dutch interlingual pseudohomophones as nonword primes. These interlingual pseudohomophones are letter strings that do not occur in French or Dutch orthography, but sound like French words if processed as Dutch words. They used three types of nonword primes, interlingual pseudohomophones (e.g., SOUR for the French target word SOURD, the French word meaning "deaf"), graphemic controls (e.g., SIARD for the target word SOURD), and unrelated nonwords (e.g., CHANE for the target word SOURD). Brysbaert et al. found that performance on French target words improved when primed phonologically with an interlingual pseudohomophone.

The findings obtained with interlingual homophones and pseudohomophones agree with those of Nas (1983) and Dijkstra et al. (1999) and support the hypothesis for a leading role of phonology in bilingual visual word perception. They first show that processing words in the second language involves mandatory phonological coding according to (non-native) intralingual spelling-to-sound knowledge, and, secondly, that processing of words in the second language can at the same time deploy phonological codes arising from native-language intralingual spelling-to-sound knowledge. Thus, bilingual readers processing words in the second language are able to apply knowledge of spelling-to-sound of both their languages in parallel. Framed in the language non-selective view, second-language reading not only engages spelling-to-sound knowledge of the second language, but also of the dominant language.

Simultaneous Cross-Language Phonological Coding

Even though the importance of phonological coding in bilingual word perception has been demonstrated (e.g., Nas, 1983; Dijkstra et al., 1999; Brysbaert et al., 1999; see also Gollan et al., 1997; Tzelgov et al., 1996; Van Wijnendaele & Brysbaert, 2002), it has not been established that processing a second-language word involves the *simultaneous* coding of multiple, cross-language phonological structures in a single perceptual event (Jared & Kroll, 2001). Thus, although the observation of phonological priming effects with interlingual pseudohomophones (Brysbaert et al., 1999) indicates that bilinguals reading in the second language can code phonological structures according to knowledge of both languages, the evidence comes from performance on two separate stimulus events, prime and target. Jared and Kroll (2001) correctly noted that these phonological priming effects do not tell us whether processing of the *prime* involves simultaneous cross-language phonological coding.

To address the specific question whether bilingual word perception involves simultaneous coding of cross-language phonology, Jared and Kroll (2001) conducted a very interesting word-naming study with French-English bilinguals. The technique they used to reveal such coding processes was to examine the influence of enemy neighbors on visual word perception. This paradigm was previously used by Jared and colleagues (e.g., Jared et al., 1990) to study spelling-to-sound consistency effects in monolingual visual word processing. To recapitulate, processing of a word such as PINT or BEAD appears to be hindered by the existence of enemy neighbors such as MINT or HEAD. In the resonance framework of Van Orden and Goldinger (1994), this effect of spelling-to-sound inconsistency is caused by simultaneous codings of multiple phonological structures, leading to time-consuming competition processes. In order to study *cross-language* phonological coding in naming words from the second language, Jared and Kroll (2001, Experiment 3) examined the influence of *interlingual* enemy neighbors in French-English bilinguals. For example, the English word BAIT has a spelling body (*–AIT*) that also occurs in French words (e.g., FAIT, meaning “act”). BAIT and FAIT share the same spelling body, but they are pronounced differently (i.e., /et/ vs. /@/) which makes FAIT an enemy of BAIT. Therefore, since BAIT’s spelling body relates differently to sounds across languages, BAIT is an *interlingual inconsistent word*.

In their Experiment 3, Jared and Kroll asked whether second-language naming of an English word such as BAIT is slowed by French spelling-to-sound knowledge due to French enemies (e.g., FAIT), analogously to the monolingual case where naming of a word such as PINT is slowed by spelling-to-sound knowledge due to English enemies such as MINT. To address this question, Jared and Kroll asked French-English bilinguals to perform an English word-naming task. In this study, three word types were contrasted. One group of words had no enemies in either language (e.g., BUMP). Hence, these words were both intralingually and interlingually spelling-to-sound consistent. A second group consisted also of words that were intralingually consistent, but, in contrast to the words of the first group, these were interlingually inconsistent (e.g., BAIT). That is, they had French enemies (e.g., FAIT) but no English enemies. In the third group of words this was reversed. The words in this group were interlingually consistent and intralingually inconsistent (e.g., BEAD), that is, they had English enemies (e.g., HEAD) but no French enemies because its spelling bodies do not occur in French. Jared and Kroll reasoned that if knowledge of spelling-to-sound correspondences affects English word naming then intralingual consistent words with only French enemies (e.g., BAIT) should be named more slowly than matching words with no French enemies (e.g., BUMP). Unlike their expectation, Jared and Kroll (2001, Experiment 3) found little evidence of a negative effect of French enemies on English naming latencies (7 ms) although there was some indication of it in the error data. Even in the condition in which the participants performed the English naming task after naming a block of French words no evidence was found for a negative effect of French enemies (2 ms). In accordance with monolingual naming studies, however, English enemies were found to have a large negative impact on naming performance, with words like BEAD named more slowly than words like BUMP and BAIT (respectively 33 ms and 26 ms).

In a subsequent experiment, Jared and Kroll (2001, Experiment 4) used French-English bilinguals that were relatively more proficient in their dominant (French) language than those who participated in Experiment 3. These participants turned out more similar to those tested by, for instance, Brysbaert et al. (1999) and Dijkstra et al. (1999). The underlying rationale for this change was that the participants in the previous experiment might have weak French spelling-to-sound knowledge, which may have reduced the impact of French enemies. Indeed, with this new group of more proficient participants, Jared and Kroll observed a clear effect of French enemies on English word naming. English words with only French enemies were named 8 ms slower than words without French enemies (e.g., BAIT vs. BUMP). This effect increased to 26 ms when participants first named a block of French words (a manipulation of relative language activity). Further, English word naming was somewhat faster (by 32 ms) for English words with only French enemies than for words with only English enemies (e.g., BAIT vs. BEAD) albeit statistical support for this effect was not very strong. In contrast, error rates were substantially lower for

English words with only French enemies (e.g., BAIT), than for words with only English enemies (e.g., BEAD). Finally, English enemies again had a large negative impact on naming performance, with words like BEAD named 42 ms slower than words like BUMP.

In sum, in line with the conclusions reached by Brysbaert et al. (1999), the interlingual consistency effects of Jared and Kroll (2001) provide evidence for the hypothesis that processing of a second-language word involves simultaneous, mandatory coding of multiple, cross-language phonological structures. This conclusion agrees with the general idea that forming manifold form-function associations, here at the intermediate-grain level of spelling-to-sound correspondences, has implications for cognitive processing. Thus if a French-English bilingual reads an English word such as BAIT, multiple phonological structures are coded (e.g., /et/ and /@/) that are driven by language-specific knowledge of spelling-to-sound relations. Interestingly, the results of Jared and Kroll also show that the interlingual consistency effect (e.g., BAIT named slower than BUMP) observed in English word naming is *smaller* than the intralingual consistency effect (e.g., BEAD named slower than BUMP and BAIT). The implication of this is exciting and is that the impact of (relatively weak) English enemy spelling-to-sound knowledge of the non-native language on word naming is *larger* than that of the (relatively strong) French enemy spelling-to-sound knowledge of the native, dominant language. Thus, the relatively weak English enemy HEAD had a greater impact on the processing of BEAD than the relatively strong French enemy FAIT had on the processing of BAIT. Evidently, in the word-naming task, bilinguals are able to restrict the influence of strong native spelling-to-sound knowledge. Finally, the effect of first naming a block of French words prior to the main task showed that bilingual word-naming performance is sensitive to the relative prominence of the two languages. This result extends the finding of an influence of language intermixing on the coarse-grain size interlingual homograph effect (e.g., Dijkstra et al. 1998) to the intermediate-grain size of the interlingual spelling-to-sound consistency effect.

A resonance account of interlingual consistency effects

Once more, we illustrate the dynamics in a resonance framework by trailing a simplified prototypical process of reading an English word but this time for a Dutch-English bilingual. We will consider two words that are both intralingual and interlingual inconsistent, MOOD and BLOOD. At a descriptive level, the spelling body *–OOD* has more than one possible pronunciation. In English it is associated with the phonological bodies /ud/ and /}d/ (as in MOOD and BLOOD), whereas in Dutch it is associated with the phonological body /od/ (as in LOOD). Within the English orthography, the intermediate-grain size spelling-to-sound relation of MOOD is typical and that of BLOOD is atypical. In contrast, within the Dutch orthography, the intermediate-grain size spelling-to-sound relation of LOOD is highly consistent.

In a recurrent network, the competing intralingual and interlingual intermediate-grain size spelling-to-sound associations in MOOD and BLOOD differ in self-consistency, resulting from relative experience with friendly and enemy neighbors. In case of interlingual inconsistent words, these neighbors emanate from both languages. For a Dutch literate, the spelling-to-sound association [*–OOD* - /od/] is highly self-consistent in Dutch, because all words that contain this rime are pronounced as /od/. This self-consistent association may be a great deal stronger than any of the English spelling-to-sound associations because of extensive reading experience in the native language. Hence, assuming the hypothesis of simultaneous cross-language phonological coding, for a Dutch-English bilingual, processing of an English word that contains the spelling body *–OOD* involves *three* competing local associations, [*–OOD* - /ud/], [*–OOD* - /}d/)], and [*–OOD* - /od/]. The relative self-consistencies of these competing local associations predict the time course and outcome of competition. Again, cooperative-competitive dynamics can be framed in terms of an attractor topology. The local friendly attractors that pull orthographic-phonologic dynamics toward correct phonology are strong for MOOD but for BLOOD they are weak. In addition, the intralingual and interlingual local enemy attractors that pull dynamics toward incorrect phonology differ in strength as well. In case of processing *MOOD*, the English local enemy attractor that developed due to experience with words like BLOOD, is relatively weak, since BLOOD has few friends and lots of enemies. In contrast, the Dutch local enemy attractor that developed due to experience with Dutch words that contain the spelling body *–OOD,* has great strength, since Dutch words like LOOD have many native friends and only distant cross-language enemies such as MOOD and BLOOD that evolved late in reading development. In case of *BLOOD*, the local English enemy attractor that developed due to experience with words like MOOD is relatively strong, since MOOD has many friends and only few enemies. Furthermore, as with processing MOOD, the local Dutch enemy attractor has strong potential to pull dynamics toward incorrect phonology (i.e., /od/ as in LOOD).

Prototypical bilingual processing of an inconsistent word

Progressing to the bilingual reading process, presentation of the word *MOOD* builds a pattern of activation among orthographic processing units, which flows forward to create a diffuse pattern on all previously associated phonologic and semantic processing units. As both /ud/ and /}d/, and also /od/ have previously become associated with the spelling body *–OOD*, these initial conditions of perception include both body phonology of the English words MOOD and BLOOD as well as that of the Dutch word LOOD, each established in proportion to its statistical dominance. Likewise, the initial conditions of perception of the word *BLOOD* also includes multiple phonological structures. In the course of cooperative-competitive dynamics, the typical, highly self-consistent association [–*OOD* - /ud/], the atypical, low self-consistent association [–*OOD* - /}d/], and the highly self-consistent association [–*OOD* - /od/] compete to be absorbed in global resonance. Over time, depending on contextual constraints and feedback from the semantic level (i.e., phonologic-semantic dynamics), dynamics pull MOOD toward correct phonology. The competition that arises between local orthographic-phonologic resonances needs to be resolved, however, which may affect the process of word perception. In the case of MOOD, the English local friendly attractor is relatively strong and the English local enemy attractor is relatively weak, whereas the Dutch local enemy attractor is very strong. In the case of BLOOD, however, the English local friendly attractor is relatively weak, the English local enemy attractor is relatively strong, and the Dutch local enemy attractor is very strong. Because the English friendly local attractor in BLOOD is relatively weak, strong English and Dutch enemy local attractors have a large potential to obstruct the trajectory to the attractor point that corresponds to correct phonology. This may be so to a lesser degree in the case of MOOD because here the English local friendly attractor has sufficient strength to tolerate these attempts of obstruction. As orthographic-phonologic dynamics eventually coalesce into an attractor state that best agrees with emerging phonologic-semantic dynamics, these phonologic-semantic dynamics process toward resonance. Ultimately, orthographic-phonologic-semantic dynamics all settle into a coherent global resonance.

Language expectations

The element of strategic control on the influence of language-specific knowledge may be conceived within the resonance framework using the idea of *language expectations*, which is central in Grosjean's (1997, 2001) concept of language modes. If a bilingual reader is in a monolingual language mode and expects words from one language only, feedback from the semantic level (i.e., phonologic-semantic dynamics), which contains knowledge of the language that words belong to, may be tuned towards language-specific expectations of phonological codings. Conceivably, in such a monolingual mode, phonologic-semantic dynamics relevant to words from the non-target language are inhibited. Therefore, inappropriate phonological codings with respect to the irrelevant, non-target language are not supported by expectations generated by the semantic level. As a result, these inappropriate orthographic-phonologic resonances emerging from the initial conditions degrade progressively. The rate at which these resonances degrade is faster than for the inappropriate intralingual orthographic-phonologic resonances, because these receive support from the semantic level. On the other hand, if a reader is in a bilingual language mode and expects words from either language, feedback from the semantic level may be tuned towards language non-specific expectations of phonological codings. In this situation, inappropriate phonological codings with respect to the non-target language may also be supported by expectations from the semantic level, and hence the rate at which these inappropriate resonances degrade is more similar to that of the inappropriate intralingual orthographic-phonologic resonances.

THE PRESENT STUDY

This investigation explores phonological coding processes of Dutch-English bilinguals and native speakers of English reading English words. To aid this research, we adopt a general resonance framework developed in the domain of monolingual visual word perception (Van Orden & Goldinger, 1994). Our starting point is the hypothesis that phonology plays a central and primary role in visual word perception (e.g., Carello et al., 1992; Frost, 1998; Lukatela & Turvey, 1994a, 1994b; Van Orden et al., 1990), which raises the question whether the process of bilingual word perception also involves mandatory phonological coding. This notion, examined in Chapters 3 and 4 of this study, is accordant with evidence obtained in studies with French and Dutch bilinguals (e.g., Brysbaert et al., 1999). Furthermore, the hypothesis that bilingual word perception proceeds essentially language non-selectively brings forth the question whether Dutch-English bilinguals engage spelling-to-sound knowledge from one or both languages. We address this question in Chapter 5, where we examine the specific issue of simultaneous cross-language phonological coding in second-language visual word perception. In the present study, cross-language phonological coding concerns a *collateral* process of extraneous native Dutch phonology emerging simultaneously with appropriate (and inappropriate) non-native English phonology. Thus presentation of the (second-language) English word MOOD launches phonological coding processes according to knowledge of spelling-to-sound correspondences of both languages, which bring forth both the English phonological structures /ud/ (as in MOOD) and /}d/ (as in BLOOD), but also the Dutch phonological structure /od/ (as in LOOD). This collateral process of cross-language phonological coding can be expected for English spellings that are pronounced differently in English and Dutch: Since Dutch-English bilinguals have knowledge of spelling-to-sound correspondences of both languages, phonological structures can potentially arise from either source (see also Jared & Kroll, 2001).

Intralingual and Interlingual Intermediate-grain Consistency

In this study we focus on the intermediate-grain level of spelling-to-sound correspondence, the level of correspondence that concerns spelling bodies and phonological bodies (e.g., Glushko, 1979; Jared & Kroll, 2001; Jared et al., 1990, Ziegler et al., 1997). Our experimental materials consist of English words that contain spelling bodies that, within English orthography, have a consistent mapping (no enemies, e.g., MOON), a typical mapping (more friends than enemies, e.g., MOOD), or an atypical mapping (more enemies than friends, e.g., BLOOD). In other words, these words vary in spelling-to-sound consistency due to manifold intralingual relations between spelling and sound: The same spelling body has more than one possible pronunciation in one language. Importantly, the spelling bodies of these English words also occur in Dutch words where they are associated with a different pronunciation. Hence, these words are interlingually inconsistent due to manifold cross-language relations between spelling and sound: The same spelling body has more than one possible pronunciation across languages.

Word Perception as a Continuous Process

Our main research goal is to study the process of intralingual and interlingual phonological coding before *and* after it is apprehended by global coherence of orthographic-phonologic-semantic activation dynamics—a transient integration of spelling, sound, and meaning information that can be utilised to launch word pronunciation or for word identification. In this dynamic view it is assumed that word perception is a *continuous* process (e.g., Goldinger, Azuma, Abramson, & Jain, 1997; Van Orden, Atchison, & Podgornik, 1996, Van Orden et al., 1999), a process that does not stop when the system reaches a state that enables a reader to identify a word, to name it, or to classify it as familiar or not. This view implies the idea of *metastability*, that is, a system never settles fully in a dominant attractor, and is therefore more flexible (Van Orden et al., 1997). The stability in attractors is thought to be affected by network oscillatory activity (i.e., noise) that allows a system to jump from one state to another (e.g., Ploeger et al., 2002).

The reason we take this dynamic approach is that it may be difficult if not impossible to detect phonological coding processes in tasks that demand a "cognitive moment of identification" (Perfetti & Tan, 1998). For example, in word naming or in reading for meaning, global processes of word perception (e.g., phonologic-semantic dynamics) may operate highly efficiently to such a degree that they conceal the more local processes. Consequently, a bilingual reader may manage to read a second-language word both fast and correctly, and show no or weak evidence of cross-language interference (cf. Jared & Kroll, 2001, Experiments 3 and 4). In sum, the aforementioned considerations assume a dynamic systems view of word perception, where pre- and post-identification "stages" of word processing actually reflect a common system of which dynamics progress continuously over time. This continuous progress is perturbed by the demands of a laboratory task, "in which an obligatory response at a certain point of time may be like a static snap shot of a continuously varying system" (Goldinger et al., 1997; Van Orden et al., 1996).

The Print-to-Speech Correspondence Task

To accommodate our requirement that experimental observations reflect continuously progressing phonological coding, we looked for a reading task that does not involve explicit word identification or reading aloud. This reading task should preferably be hypersensitive to detect processes originating from the initial conditions of word perception; the early phases wherein the system launches codings of all phonological structures that have previously been associated with a particular spelling body. For that purpose, we devised a bimodal reading task that might carry these qualifications. In this so-called *print-to-speech correspondence task*, two stimuli are presented simultaneously to the visual and auditory modalities. One of the stimuli is a visually presented printed word (e.g., MOOD), and the other is an auditory presented unit of speech, which may or may not be the spoken rime (i.e., the phonological body) of the word. The printed word is presented for approximately 200 ms, which, according to estimates of Rayner and Pollatsek (1989; see also Van Orden et al., 1999), is a sufficient amount of time for the reading system to reach a state that can be utilised to launch word pronunciation or for recognition.

The task of the participant is to judge whether the printed word and the spoken rime correspond to one another. For example, the participant may be presented with the English word MOOD accompanied by the spoken rime /ud/ (derived from the spoken word MOOD). In this case, the two stimuli are congruent with each other and a "yes" response would be appropriate. If however the word MOOD is accompanied by an unrelated spoken rime, for example /Yd/ (as in BRIDE), the two stimuli are not congruent with each other and a "no" response is required. We assume that the participants can perform this task by comparing in each trial the phonological structure that emerges from print with the coding generated by the speech unit.

Match and no-match trials

If the print-to-speech correspondence task indeed reflects phonological coding processes, the spelling-to-sound consistency of the printed words may influence performance on "match" trials that require a "yes" response but also on "no-match" trials that require a "no" response. Fast and accurate "yes" responses may be elicited by a word with a consistent spelling-to-sound mapping, such as MOON, presented simultaneously with a matching spoken rime (i.e., /un/, derived from the spoken word MOON), because orthographic-phonologic dynamics cohere quickly in consistent words. If correct phonology emerges quickly, participants can perceive a match or mismatch between the two stimuli relatively early. However, performance on an inconsistent word with a typical mapping, such as MOOD, presented simultaneously with the spoken rime /ud/ (derived from the spoken word MOOD) may be slower and more error-prone. For such a word, orthographic-phonologic dynamics cohere slower because of competition between local resonances. This competition is induced by spelling-to-sound knowledge of enemy words (e.g., BLOOD). Furthermore, for an inconsistent word with an atypical mapping, such as BLOOD, presented simultaneously with the spoken rime /}d/ (derived from the spoken word BLOOD), performance may be even worse. Here, the competition between local resonances needs more time to resolve because friendly attractors are weak and enemy attractors are strong, the latter resulting from knowledge of strong enemy words such as MOOD. These enemy attractors have large potential to disrupt orthographic-phonologic coherence. For "no" trials, we may expect an analogous pattern of observations. As with "yes" trials, the self-consistency of spelling-to-sound associations predicts the time course of local resonances achieving coherence. The longer it takes for a local resonance to cohere, the longer it takes a participant to perceive a mismatch between print and speech, such as when BLOOD and MOOD are accompanied by the unrelated spoken rime /Yd/ (as in BRIDE), and MOON is accompanied by the unrelated spoken rime /en/ (as in VEIN).

Catch trials

Because there is no discrete point where processing is supposed to terminate, a multistable dynamic process of word perception that is assumed to progress in a continuous fashion may be sensitive to external influences throughout all points in time. In view of our purposes, one special feature of the print-to-speech correspondence task may be of particular interest. This feature involves the possibility to accompany a printed inconsistent word (e.g., MOOD) with the spoken rime of its *enemy* (e.g., /}d/, derived from BLOOD). In such a trial, which we designate as a "catch trial", the appropriate phonological structure that emerges from spelling is obviously not congruent with the spoken rime. Yet, for an inconsistent word, we may expect inappropriate local orthographic-phonologic resonances, that have been inhibited in the course of word processing (e.g., the inappropriate association [*–OOD* - /}d/] in MOOD), to be restored at full strength if they are fostered by external form-similar codings. If such an external influence consists of a unit of speech (e.g., the spoken rime /}d/ of the enemy BLOOD) presented in the print-to-speech correspondence task, we may be able to mark local orthographic-phonologic resonances that were launched at the start of word presentation and continue to lie dormant. Essentially, this is the approach we take in Chapters 4 and 5 of this study.

Thus if we assume that processing an inconsistent English word (e.g., MOOD) also involves coding of inappropriate, enemy phonology (e.g., proper to BLOOD), we can make it observable by presenting to a participant the printed word MOOD simultaneously with an auditory presented spoken rime that is derived from its enemy neighbor BLOOD (i.e., /}d/), and ask the participant to judge whether they correspond. Of course, in this example, a spoken rime derived from BLOOD and one derived from BRIDE both represent incorrect, unrelated phonology in relation to the printed word MOOD, and both call for a "no" response. However, if inappropriate phonology (i.e., /}d/) is part of the initial conditions of perception of MOOD, auditory presentation of a rime derived from BLOOD may foster the degraded *–OOD* - /}d/ resonance to such a degree that it regains sufficient capacity to undermine the appropriate orthographic-phonologic resonance. If the appropriate resonance indeed suffers destabilisation, participants may take more time to perceive a mismatch between correct phonology of MOOD and the structure generated by the spoken rime of BLOOD than with the structure generated by the spoken rime of BRIDE. Destabilisation may also take catastrophic proportions if the inappropriate orthographic-phonologic *–OOD* - /}d/ resonance is restored to a level that is comparable with MOOD's correct orthographic-phonologic *–OOD* - /ud/ resonance. In that case, participants may not be able to perceive a mismatch, and respond accordingly by incorrectly pressing the "yes" button, thus indicating that they perceived MOOD's phonology to rhyme with the rime of BLOOD.

In conclusion, with the print-to-speech correspondence task we may be able to track the cooperative-competitive dynamics of competing local orthographic-phonologic associations in word perception. Specifically, it enables us to study the competing orthographic-phonologic dynamics in a fairly straightforward way, that is, the obtained experimental observations may inform us directly of the competing orthographic-phonologic dynamics without relying on the ultimate outcome of these competitions in terms of naming or identification performance.

To take this a step further, we may distinguish between inconsistent words with a typical spelling-to-sound mapping (e.g., MOOD) and those with an atypical spelling-to-sound mapping (e.g., BLOOD). Local, inappropriate strong-rule orthographic-phonologic resonances that have been inhibited in the course of processing may be more readily restored than less strong ones, especially when the appropriate local orthographic-phonologic association is weak. This implies that catch-trial performance on an inconsistent word with a typical mapping (weak enemy local attractor) is different from that of an inconsistent word with an atypical mapping (strong enemy local attractor). Specifically, for a catch trial with the atypical BLOOD together with the spoken rime /ud/ (derived from MOOD) it may be more difficult to press the "no" button than for a catch trial with the typical MOOD together with the spoken rime /}d/ (derived from BLOOD). The former will lead to relatively large false-positive error rates, and exaggerated correct-no latencies.

In Chapter 5 we take this line of argument another step further by hypothesising that processing of an English inconsistent word such as MOOD also involves coding of the phonology proper to its *interlingual enemy* LOOD (Dutch for "lead"). We might demonstrate this by presenting a Dutch-English bilingual participant the printed word MOOD simultaneously with a spoken rime that is derived from its Dutch enemy neighbor LOOD (i.e., /od/). In bilingual processing of an atypical word such as BLOOD, the highly self-consistent Dutch spelling-to-sound association [*–OOD* - /od/] concerns a very powerful local enemy attractor. This combination of a weak friendly local attractor that tries to pull BLOOD toward correct phonology and this extremely strong local enemy attractor that tries to pull it away from it, predicts large false-positive error rates and extremely long correct-no latencies. Again, destabilisation may take catastrophic proportions if the inappropriate orthographic-phonologic [*–OOD* - /od/] resonance is restored to a level that is comparable with MOOD's correct, global spelling-to-sound resonance. In that case, participants may not be able to perceive a mismatch, and respond accordingly by incorrectly pressing the "yes" button, thus indicating that they perceived MOOD's phonology to rhyme with the rime of ROAD. Take notice that for performance on the print-to-speech correspondence task we predict a *large* impact of knowledge of native Dutch enemy words, one that is larger than the impact of non-native knowledge of English enemy words. This contrasts with Jared and Kroll's (2001) key observation that, relative to the effect of interlingual enemies, the effect of interlingual enemies on second-language word naming is only moderate.

Time-Course Analysis of Bilingual Spelling-to-Sound Dynamics

In one version of the print-to-speech correspondence task, two stimuli are displayed simultaneously over the visual and auditory modalities. In this study, we also explored variations in the time duration between the visually presented word and the auditory presented unit of speech, that is, we manipulated stimulus-onset asynchrony (SOA) between the two stimuli. In Experiments 5 and 7 we contrasted three SOA's. In one SOA condition, print and speech were initiated simultaneously. In the second SOA condition, speech was initiated first and print followed after an interval of approximately 500 ms. In a third SOA condition the order of events was reversed, print was initiated first and speech followed after an interval of approximately 500 ms. This SOA manipulation was performed to find out whether the readiness of a diminished local orthographic-phonologic resonance to be reinstated by a fostering speech unit from the auditory modality, is changed if there is a substantial gap between presentation of the printed word and presentation of the spoken rime. This procedure may inform us on time-critical orthographic-phonologic dynamics in relative advanced phases of word processing.

Effect of Stimulus-List Composition

Second-language reading can engage knowledge of words of the dominant language. Bilingual-reading research has shown that this occurs primarily when the reading task involves words from both languages, a condition that affects the relative prominence of the two languages. The idea is that if the non-target language is also active it may accelerate coding of word knowledge of the non-target language. For example, lexical decisions of Dutch-English participants on interlingual homographs (e.g., BAD) take more time than on single-language control words (e.g., LOW) when Dutch filler words are added to the stimulus list. If the stimulus list does not contain any Dutch words, performance on words like BAD and LOW is more or less the same (e.g., Dijkstra et al., 1998). The same holds for the use of spelling-to-sound knowledge from both languages in bilingual naming (Jared & Kroll, 2001).

Effects of stimulus-list composition on bilingual word processing are assumed to originate from global-level linguistic codings. These codings encompass some degree of integration of meaning with linguistic surface forms, and possibly include language type coding (i.e., "language tags"). In Chapter 5 of the present study we explored the effects of language intermixing on performing the print-to-speech correspondence task. As stated above, our general focus is on the ballistic processes of phonological coding and we expect to detect all codings that originate from the initial conditions of word perception, where all phonological structures that have previously been associated with a spelling body are launched. Because all associated phonological structures are launched irrespective of the relative prominence of the non-target language we may anticipate minor effects of stimulus-list composition. Thus, it is possible that rejecting a trial in which, for example, the printed word BLOOD is accompanied by /od/ (derived from the Dutch word LOOD), is as difficult when the stimulus list is composed of a mix of English and Dutch words as when it is entirely composed of English words.

Relevance of this Study

If one defines bilingualism as the ability to understand a text written in a non-native language, it can be estimated that more than half of the literate world population is bilingual (Brysbaert, in press; see also De Groot & Kroll, 1997). Hence, bilinguals are *not* a minority and therefore any comprehensive model of visual word perception should generalize beyond the case of monolingual reading. The present study is an effort to contribute to an understanding of bilingual visual word perception. Guided by a simple general principle of manifold form-function relations and conceived within the resonance framework of Van Orden and Goldinger (1994), it brings together two central theoretical issues in bilingual reading research: The role of phonology in bilingual word perception and the language non-selective view of second-language processing. Building on related work of Brysbaert et al. (1999), Dijkstra et al. (1999), and Jared and Kroll (2001), this study adds significant insight in the use of spelling-to-sound knowledge in the processing of second-language words. Foremost, it extends the work of Jared and Kroll (2001) on simultaneous cross-language phonological coding, by investigating processes of interlingual phonological coding that may be obscured if word processing were to be observed in naming or recognition responses. The present study distinguishes itself from that of Jared and Kroll (2001) in that we make an effort to detect the actual competing phonological codings originating from the initial conditions of word perception—competing codings that commonly are assumed from inferred patterns in word-naming data. The task we developed, the print-to-speech correspondence task, enables us to perform a time-course analysis of orthographic-phonologic dynamics in the processing of intralingual (for monolingual and bilingual participants) and interlingual (for bilingual participants) spelling-to-sound inconsistent words. Furthermore, studying cross-language phonological codings that emerge in the initial conditions of word perception enables us to address the interesting question of how these are influenced by the relative prominence of the bilingual's two languages. This will extend our understanding of the influence of language intermixing as observed for global-level processing of second-language words (e.g., naming of interlingual inconsistent words; lexical decisions on interlingual homographs) to more local-level orthographic-phonologic processing.

Summary of Research Questions

The primary object of the present study is to explore phonological coding processes of Dutch-English bilinguals reading English words, for the purpose of which we developed a bimodal print-to-speech correspondence task. In a series of eight experiments we addressed four general questions:

(1) Does the process of English word perception in Dutch-English bilinguals involve mandatory intralingual phonological coding? This question is addressed by investigating the intralingual consistency effect. Specifically, it is concerned with simultaneous, competing phonological codings that result from manifold *intralingual* spelling-to-sound mappings in inconsistent words.

(2) Does the process of bilingual word perception involve simultaneous cross-language phonological coding? Specifically, this question is concerned with competing phonological codings that result from manifold *interlingual* spelling-to-sound mappings in inconsistent words.

(3) Is, in a catch trial, the readiness of a diminished local orthographic-phonologic resonance to be reinstated by a fostering speech unit from the

auditory modality changed if there is a substantial gap between presentation of the printed word and presentation of the spoken rime?

(4) Does the relative prominence of a bilingual's two languages have an impact on simultaneous cross-language phonological coding?

Plan of Research

The empirical work is divided into three sections. In Chapter 3 we report an English word-naming study (Experiment 1) with participants from two different nationalities. We tested a group of monolingual native English speakers and a group of bilingual native Dutch speakers, with English as their second language. This naming study has two goals. One, validating the differences in spelling-to-sound consistency of the English experimental materials, a goal that can be accomplished by running a word-naming study with native English speakers to see whether it replicates effects reported in the literature (e.g., Jared, 2002; Jared, 1997; Jared et al., 1990). This is an important first step in this investigation, because the experimental materials will be used in the experiments presented in Chapters 4 and 5, in which the print-to-speech correspondence task is employed. Two, addressing the question whether English naming performance of Dutch-English bilinguals is sensitive to differences in spelling-to-sound consistency. Because the groups of native English speakers and Dutch-English bilinguals were tested on the same set of experimental materials, we can directly compare the effects of spelling-to-sound consistency. Critically, we compared two lists of English words. Both lists consisted of English words that either contained a consistent mapping, a typical mapping, or an atypical mapping. The primary difference between the lists was that in one list the English words had Dutch neighbors (e.g., HOOK, SEEM, and MOOD) whereas in the other list the English words did not have Dutch neighbors (e.g., THROUGH, PULL, and TEACH). These two lists were used as different stimulus sets in Experiments 2 and 3 in Chapter 3.

Chapter 4 investigates the *intralingual* spelling-to-sound consistency effect or, more specifically, simultaneous intralingual phonological coding using the print-to-speech correspondence task, and considers only *English enemy neighbors*. In Experiments 2, 3, 4, and 5, English printed words were presented along with spoken rimes to create three types of trials. In *match trials* the printed word and spoken rime are congruent with each other and, therefore, a "yes" response is required. An example of such a trial is when the printed word BLOOD is presented along with the spoken rime /}d/, which is derived from this word. In contrast, there are two types of mismatch trials that require a "no" response, *no-match trials* and *catch trials*. In no-match trials, the printed word and spoken rime are not congruent with each other. An example of such a trial is when the printed word BLOOD is presented along with the unrelated spoken rime /Yd/, which is derived from the word BRIDE. Likewise, in catch trials the printed word and spoken rime do not correspond either. In contrast to no-match trials, however, the spoken rime for a catch trial is not derived from an unrelated word but derived from an English enemy neighbor of the printed word. An example of such a trial is when the printed word BLOOD is presented along with the spoken rime /ud/, which is derived from the English enemy word MOOD.

Experiment 2 examines no-match and catch trials for English words that do not have Dutch neighbors (e.g., THROUGH, PULL, and TEACH), whereas Experiments 3-5 (and likewise Experiments 6-8 in Chapter 4) examine those for English words that do have Dutch neighbors (e.g., HOOK, SEEM, and MOOD). We allocated an unusually large number of research participants (respectively 60 and 80 Dutch-English participants) to Experiments 2 and 3, which is considerably more than we assigned to the other experiments. The main reason for this strategy is that in these two experiments we introduced the print-to-speech correspondence task, and as such served as a foundation and point of reference for subsequent experiments. Experiment 4 involves preparatory work for the SOA manipulation in Experiments 5 and 7. Experiment 5 contrasts a group of monolingual native English speakers with a group of bilingual native Dutch speakers.

Chapter 5 investigates the *interlingual* spelling-to-sound consistency effect with Dutch-English bilingual participants, using the print-to-speech correspondence task. In this chapter only *Dutch enemy neighbors* are considered. In Experiments 6, 7, and 8 English printed words are presented along with spoken rimes, again to create three types of trials. In *"match trials"* the printed word and spoken rime are congruent with each other, requiring a "yes" response. Two other types of mismatch trials require a "no" response. In no-match trials, the printed word and spoken rime are not congruent with each other, for example when the printed word BLOOD is presented along with the unrelated spoken rime /Yd/. In catch trials the printed word and spoken rime are not congruent with each other either. However, the spoken rime is not unrelated to the printed word, but derived from a *Dutch enemy neighbor* of this word. An example of such a trial is the printed word BLOOD presented with the spoken rime /od/, which is derived from the Dutch word LOOD (meaning "lead"). Experiment 6 examines no-match trials and catch trials (derived from Dutch enemy neighbors) for English words, and includes a manipulation of stimulus-list composition. Two groups of Dutch-English participants are contrasted. For one group the stimulus list contains 25% of additional English filler trials (i.e., match and no-match trials using English printed words), whereas for the other the stimulus list contains 25% of additional Dutch filler trials (i.e., match and no-match trials using Dutch printed words). Experiment 7 is similar to Experiment 6 but includes an SOA manipulation as well. Finally, Experiment 8 is a replication of Experiment 6 but with filler language not varied between but within participants.

2 General Method and Statistical Deliberation

The next sections deal with general issues of participants, stimulus collection, and data analysis, and pertain to all experiments that follow. The section on statistical data analysis is meant as supplementary and can therefore be skipped or consulted at a later point in time.

PARTICIPANTS AND MATERIALS

Participants

A total of 322 Dutch first-year psychology students from the University of Amsterdam, The Netherlands, participated in Experiments 1-8 for either course credit or a small financial compensation. All were native speakers of Dutch with normal or corrected-to-normal vision. In addition, 54 native English speakers, also with normal or corrected-to-normal vision, volunteered to participate in either Experiment 1 (n = 30) or Experiment 5 (n = 24). The native English speakers were psychology students from Pennsylvania State University, State College, USA.

The Dutch participants were unbalanced Dutch-English bilinguals, with Dutch as their native language and English as their strongest foreign language. All were fairly fluent in their second language: They had learned English at school for about 3-4 hours a week, starting in the last two years of primary school (around the age of ten), and until the end of secondary school. Their education at the psychology department required them to read mainly in English. Outside school they had been exposed to English in various settings, like watching American and British television shows, and listening to English lyrics in music.

Each Dutch participant took part in one of a total of 8 experiments, according to when he or she arrived at the laboratory. None participated in more than one experiment. A total of 30 students participated in Experiment 1, 60 in Experiment 2, 80 in Experiment 3, 20 in Experiment 4, 36 in Experiment 5, 60 in Experiment 6, 24 in Experiment 7, and 12 students participated in Experiment 8.

Selection of Printed Word Stimuli

The printed word stimuli used in Experiments 1-8 consisted of 120 monosyllabic English words. These were extracted from the linguistic database of Ziegler et al. (1997) which is based on 2,694 monosyllabic, monomorphemic words found in Kucera and Francis (1967) and provides a statistical analysis of the bidirectional consistency of spelling and sound in English. This database presents consistent and inconsistent correspondences for orthographic and phonologic rimes (spelling bodies and phonological bodies) and provides information on the number and the frequency of enemies and friends (Jared et al., 1990).

Initially, the Ziegler et al. database was sectioned to reveal spelling bodies that English words share with Dutch words (e.g., *–OON* in MOON and ZOON, meaning "son" in Dutch). In this preparatory phase all 609 spelling bodies of the database were divided over two lists. The first list contained 284 spelling bodies that also occur in one or more Dutch words. For example, the spelling bodies *–OOK*, *–EEM*, and *–OOD* not only occur in the English words HOOK, SEEM, and MOOD, but also in the Dutch words ROOK, ZEEM, and ROOD (in Dutch meaning "smoke", "wash leather", and "red", respectively). Hence, English words that contain these spelling bodies have Dutch neighbours. The second list contained 325 spelling bodies that do not occur in Dutch words. Thus English words with these spelling bodies do not have Dutch neighbours. Examples of these bodies are *–OUGH*, *–ULL*, and *–EACH* as in THROUGH, PULL, and TEACH. The division of spelling bodies over the two lists proceeded by entering the spellings in the Dutch corpus type lexicon (version N3.1) of the Centre for Lexical Information (CELEX) in Nijmegen, The Netherlands, using a text editor. The CELEX Dutch database contains information on 381,292 present-day Dutch word forms, corresponding to 124,136 lemmata (see Baayen, Piepenbrock, & Van Rijn, 1993; Burnage, 1990). If one or more Dutch words in the lexicon contained the spelling body (only in the rime position), the spelling body was assigned to the first list (bodies with Dutch neighbours); if not it was assigned to the second list (bodies without Dutch neighbours). The two lists were further divided into sublists with spelling-to-sound consistent and spelling-to-sound inconsistent bodies. Ziegler et al. classified a spelling body as spelling-to-sound consistent if it mapped onto one and only one phonological body. Hence, all words containing this body were consistent. A spelling body was classified as spelling-to-sound inconsistent if it mapped onto more than one phonological body, and all words containing this body were therefore inconsistent. For example, the spelling body *–OOD* has more than one phonological body, /ud/ as in MOOD, but also /}d/ as in BLOOD. This subdivision resulted in 214 consistent and 70 inconsistent spelling bodies for the first list (bodies with Dutch neighbours), and 281 consistent and 44 inconsistent spelling bodies for the second list (bodies without Dutch neighbours).

The four extracted spelling-body sublists formed the basis to select printed word stimuli. Two experimental word lists were created. One list consisted of 60 English words that have Dutch neighbours ("with Dutch neighbours", e.g., HOOK, SEEM, MOOD), and the other consisted of 60 English words that do not have Dutch neighbours ("no Dutch neighbours", e.g., THROUGH, PULL, TEACH). Homophones, homographs, and interlingual cognates were avoided in these lists, and so were words that were expected to be unfamiliar to the bilingual Dutch participants. These restrictions severely limited the number of candidate word stimuli. As a consequence, it was not possible to exceed the number of 20 word stimuli per experimental condition. Printed word frequencies of the word stimuli extended a full range. Kucera and Francis (1967) frequency counts ranged from 2 to 7289 with an average of 311, and the 10th, 25th, 50th, 75th, and 90th percentiles were respectively 5, 20, 74, 230, and 760. On the log scale these frequency counts ranged from 0.30 to 3.86 with an average of 1.86, and the 10th, 25th, 50th, 75th, and 90th percentiles were respectively 0.70, 1.31, 1.88, 2.36, and 2.88. Printed word frequency counts were also collected from the English corpus type lexicon of the CELEX database (Release E2.5), which contains 52,446 lemmata representing 160,594 word forms. The CELEX log frequency counts ranged from 0.40 to 3.86 with an average of 1.91, and the 10th, 25th, 50th, 75th, and 90th percentiles were respectively 0.88, 1.38, 1.87, 2.38, and 2.92. The CELEX log frequency counts correlated highly with the log-transformed frequency counts of Kucera and Francis (1967), $r = .96$ ($n = 120$), with a 95% confidence interval (95% CI) of .94 to .97. A similar high correlation was observed with another, unrelated list of 160 English words, $r = .94$ ($n = 160$), with a 95% CI of .93 to .96. This indicates that the different databases provide very similar estimates of printed word frequency.

In both lists three word types were contrasted that differed in the degree of (in)consistency of spelling-to-sound mappings. In each list 20 words had *consistent* spelling-to-sound mappings, 20 words had *typical* spelling-to-sound mappings, and 20 words had *atypical* spelling-to-sound mappings. The words with consistent mappings had spelling bodies that mapped onto a single phonological body (e.g., MOON). The words with typical and atypical mappings, on the other hand, were all inconsistent words and had spelling bodies that mapped onto more than one phonological body (e.g., MOOD, BLOOD). The latter word types differed in the number and frequency of friends and enemies. Words with typical mappings (e.g., MOOD) had a larger number and higher frequency of friends (e.g., FOOD, BROOD, and SNOOD) relative to the number and frequency of enemies (e.g., BLOOD and FLOOD). In contrast, words with atypical mappings (e.g., BLOOD) had a larger number and higher frequency of enemies (e.g., MOOD, FOOD, BROOD, and SNOOD) relative to the number and frequency of friends (e.g., FLOOD). In order to achieve that the two groups of inconsistent words were similar in visual form, both contained the same set of 20 spelling bodies (e.g., *–OOD* in MOOD-BLOOD; *–EAR* in DEAR-BEAR). The complete list of word stimuli is provided in Appendix A.

The groups of word stimuli representing the three word types varied systematically with regard to number and frequency of English friends and enemies. At the same time, they were carefully matched for a set of linguistic dimensions that are known to affect visual word perception. The statistics for the relevant variables for each experimental word list are presented in Table 1. This table provides measures of central tendency and variation in terms of mean (M) and standard deviation (SD), as well as in robust median (Mdn) and interquartile range (IQR).

Table 1.
Characteristics of the English Words in Experiments 1-3. (CM = consistent spelling-to-sound mappings, TM = typical spelling-to-sound mappings, AM = atypical spelling-to-sound mappings).

	No Dutch Neighbors				With Dutch Neighbors		
	CM	TM	AM		CM	TM	AM
Number of words	20	20	20		20	20	20
Number of letters							
M	4.65	4.80	4.80		4.10	3.95	4.05
SD	0.88	0.83	0.77		0.72	0.60	0.83
Mdn	5	5	5		4	4	4
IQR	1	1	1		0	0	0
CELEX Log Frequency							
M	1.89	1.82	1.83		1.95	1.98	1.96
SD	0.79	0.76	0.52		0.85	0.78	0.83
Mdn	1.84	1.87	1.74		1.84	1.99	2.08
IQR	1.13	0.97	0.49		0.87	1.17	1.25
K&F Log Frequency							
M	1.88	1.72	1.78		1.85	1.96	1.94
SD	0.78	0.86	0.56		0.91	0.74	0.85
Mdn	1.68	1.76	1.84		1.86	1.98	2.06
IQR	1.08	1.13	0.65		1.14	.95	1.35
Log Bigram Frequency							
M	5.86	6.01	5.94		6.10	6.06	6.01
SD	0.29	0.29	0.27		0.44	0.25	0.34
Mdn	5.97	6.02	5.91		5.95	6.04	6.11
IQR	0.30	0.30	0.35		0.51	0.31	0.36
Familiarity							
M	565	556	562		540	560	565
SD	36	38	30		68	52	49
Mdn	564	565	556		546	584	572
IQR	51	46	42		92	85	42

Table 1 (*continued*)

	No Dutch Neighbors			With Dutch Neighbors		
	CM	TM	AM	CM	TM	AM
Imagability						
M	503	481	477	418	485	468
SD	99	111	121	116	104	126
Mdn	517	514	507	444	510	491
IQR	148	178	176	128	161	228
Number of Friends						
M	5.60	5.55	1.40	4.55	4.50	1.85
SD	5.52	5.23	1.88	3.20	4.02	2.58
Mdn	3.50	3.50	0.50	3.50	3.00	1.00
IQR	7.50	8.00	3.00	4.50	4.00	2.00
□ Frequency of Friends						
M	589	653	178	1101	1186	346
SD	621	598	286	1217	1234	417
Mdn	416	370	83	886	968	250
IQR	527	740	96	1116	1276	418
Log □ Frequency of Friends						
M	2.59	2.65	1.97	2.76	2.79	2.05
SD	0.40	0.40	0.47	0.58	0.61	0.85
Mdn	2.62	2.57	1.92	2.94	2.98	2.39
IQR	0.54	0.60	0.45	0.74	0.78	1.20
Consistency Ratio						
M	1.00	0.79	0.21	1.00	0.79	0.21
SD	0.00	0.13	0.13	0.00	0.13	0.13
Mdn	1.00	0.82	0.18	1.00	0.76	0.24
IQR	0.00	0.20	0.20	0.00	0.20	0.20

As can be seen in Table 1, words with consistent mappings (CM) and inconsistent words with typical mappings (TM) were closely matched on the number and log summed frequency of friends. Further, both groups of words had a larger number of friends and greater summed frequency of friends than the group of inconsistent words with atypical mappings (AM). Table 2 shows the mean (standardised) differences (Hedges *g*; see Rosnow & Rosenthal, 2003) in number of friends and log summed frequency of friends between the three word types for each experimental word list. Inspection of this table confirms that the means of the distributions of number of friends for the inconsistent words with atypical mappings versus the other word types deviate in the range of 0.8 to 1.1 *SD* units. These standardised differences correspond to a nonoverlap of distributions of 47.4 to 58.9% (see Cohen, 1988). Likewise, the means of log summed frequency of friends distributions deviate in the range of 1.0 to 1.6 *SD* units, corresponding to a nonoverlap of distributions of 55.4 to 73.1%. Thus although the distributions have some overlap, there is substantial variation in number and frequency of English friends across the three word types.

Table 2.
Mean differences in number of friends and log summed frequency of friends between CM-words and TM-words, between CM-words and AM-words, and between TM-words and AM-words in original units (frequency counts), *SD* units (standardized difference: Hedges *g*), and percent of nonoverlap for the words in Experiments 1-3. (CM = Consistent Mappings; TM = Typical Mappings; AM = Atypical Mappings.)

Number of friends

No Dutch Neighbors	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.05	4.20	4.15
Standardized Difference (*g*)	0.01	1.02	1.06
Percent of nonoverlap	0	55.4	58.9
With Dutch Neighbors	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.05	2.70	2.65
Standardized Difference (*g*)	0.01	0.93	0.78
Percent of nonoverlap	0	51.6	47.4

Log summed frequency of friends

No Dutch Neighbors	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.06	0.62	0.68
Standardized Difference (*g*)	0.15	1.42	1.56
Percent of nonoverlap	7.7	68.1	73.1
With Dutch Neighbors	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.03	0.71	0.74
Standardized Difference (*g*)	0.05	0.98	1.00
Percent of nonoverlap	0	55.4	55.4

Note. *SD* units are pooled *SDs* ($\sqrt{MSw}$) of the contrasting word lists (e.g., Rosnow & Rosenthal, 2003).

Table 1 also provides consistency ratios for the three word types. The consistency ratio varies between 0 and 1 and reflects the degree of (in)consistency. It is determined as the summed frequency of a word's friends relative to the summed frequency of a word's friends plus enemies. A consistency ratio greater than .5 indicates that a word has stronger friends than enemies. Conversely, a consistency ratio smaller than .5 indicates that a word has stronger enemies than friends. In calculating the ratios for words with *typical* mappings, the summed frequency of friends of the words with atypical mappings were utilised here as summed frequency of enemies, and for words with *atypical* mappings the summed frequency of friends of the words with typical mappings were utilised as summed frequency of enemies.

As can be seen in Table 1, in both experimental word lists inconsistent words with typical mappings had a consistency ratio of approximately .8 and inconsistent words with atypical mappings had a consistency ratio of approximately .2. This confirms that the words with atypical mappings had stronger enemies than the words with typical mappings.

Further inspection of Table 1 verifies that the groups of word stimuli representing the three word types were matched on a number of variables that are known to affect visual word perception: Number of letters, printed word frequency, bigram frequency, familiarity, and imageability. Table 3 presents the associations (□^{2}) between word type and the set of linguistic dimensions for both experimental word lists, augmented with values of *F*. This table indicates that there is no notable relation between word type and any of the linguistic dimensions.

Table 3.
Associations between Word Type and key linguistic dimensions as expressed in □^{2} for the words in Experiments 1-3.

	No Dutch Neighbors			*With Dutch Neighbors*	
	□^{2}	*F*		□^{2}	*F*
Number of Letters	-0.03	0.22		-0.03	0.22
CELEX Log Frequency	-0.01	0.82		-0.03	0.15
Log Bigram Frequency	-0.02	0.37		-0.02	0.32
Familiarity	-0.03	0.25		-0.01	0.83
Imagability	-0.03	0.23		0.02	1.36
Number of Friends	0.14	5.69		0.10	4.33
Log □ Frequency of Friends	0.33	15.89		0.17	7.26

Word frequency estimates were taken from the English corpus type lexicon of the CELEX database and also from the Kucera and Francis (1967) database for comparison. Bigram estimates were collected by means of the computer program LexStat (version 2.28; van Heuven, 2000) and using the Kucera and Francis (1967) database. Familiarity and imageability ratings were obtained from the Medical Research Council (MRC) psycholinguistic database (Coltheart, 1981), a machine usable dictionary file containing 150837 English words. This database provides familiarity ratings for 9392 words and imageability ratings for 9240 words, on a scale from 100 to 700. The imageability ratings were derived from a merging of the Colorado norms (Toglia & Battig, 1978), the Paivio norms (unpublished, these are an expansion of the norms of Paivio, Yuille, & Madigan, 1968), and the Gilhooly-Logie norms (Gilhooly & Logie, 1980). Details of merging are given in Coltheart (1981). The lowest and highest imageability rating in the database are 129 and 669, respectively, with a mean of 450 and a standard deviation of 108. The familiarity values were derived from merging three sets of familiarity norms: Paivio (unpublished), Toglia and Battig (1978), and Gilhooly and Logie (1980). The method by which these three sets of norms were merged is described in detail in Appendix 2 of the MRC Psycholinguistic Database User Manual (Coltheart, 1981). The highest familiarity rating is 657, with a mean of 488 and a standard deviation of 99. Notice that integer values are used here (in the original norms the equivalent range was 1.00 to 7.00). The familiarity ratings correlated highly with the CELEX log frequency counts, $r = .62$ ($n = 96$), with a 95% CI of .48 to .73. A similar high correlation was observed with another, unrelated list of 160 English words, $r = .69$ ($n = 147$), with a 95% CI of .59 to .77.

The two experimental word lists could not be fully equated due to numerous restrictions in stimulus selection. The most notable differences between the lists are that, on average, words with Dutch neighbours are 0.72 letter shorter ($g = 0.94$, $F = 26.28$), 0.13 units higher in log bigram frequency ($g = 0.40$, $F = 4.53$), 0.12 units higher in CELEX log frequency ($g = 0.16$, $F = 1.85$), 0.13 units higher in log summed frequency of friends ($g = 0.20$, $F = 1.21$), but 28.47 units lower on the imageability scale ($g = 0.25$, $F = 3.15$). Taken together, the words of the list “with Dutch neighbours” are more familiar in form and orthographically less complex than the list “no Dutch neighbours”. Therefore, because the list “with Dutch neighbours” may be more easy to process, it cannot be unequivocally contrasted to the list “without Dutch neighbours” in a statistical analysis.

Thus, in brief, the printed word stimuli used in Experiments 1-8 consisted of 120 English words assigned to lists of AM, TM, and CM words such as PAID, SAID, and STAIN for the set “no Dutch neighbors”, and lists of AM, TM, and CM words such as BLOOD, MOOD, and MOON for the set “with Dutch neighbors”.

STATISTICAL DATA ANALYSIS

The intention of the statistical analyses performed in this study was to determine how much faith one can put in the observed pattern of sample means as reflecting the associated set of population means they are estimating (Loftus & Masson, 1994). For sake of clarity, in this section we explicate the procedures, outcomes and assumptions of the statistical data analyses we performed. Additionally, we also introduce the meaning and intended use of *confidence intervals*. The reason we do so is that recommendations of statisticians to include interval estimates in research reports (e.g., Bakan, 1966; Cohen, 1990; Hunter, 1997; Kirk, 1996; Schmidt, 1996) have not been widely adopted by experimental psychologists. This may be because that, in contrast to most ANOVA procedures provided by current statistical packages, interval-estimation procedures are not transparent for complex designs, such as ones that for example include blocking variables. Furthermore, they are often presented secondary in courses and prominent books on statistics (cf. Hays, 1994; Kirk, 1995; Myers &

Well, 1995; Winer, Brown, & Michels, 1991). Recently, however, the American psychological Association's (APA) Task Force on Statistical Inference (TFSI) emphasised the importance of confidence intervals (CIs) in reporting the results of psychological studies (Wilkinson & TFSI, 1999; see also Cumming & Finch, 2001; Thompson, 2002). The APA Task Force recommended, for example, that CIs should be reported along with, or instead of, hypothesis test results (cf. Loftus, 1991, 1993, 1995, 1996). Furthermore, in the latest version of the APA *Publication Manual* it is suggested that CIs represent "in general, the best reporting strategy. The use of confidence intervals is therefore strongly recommended" (p. 22). Rosnow and Rosenthal (2003) further comment that "the expectation is that guidelines promulgated by the APA Task Force will be absorbed into the mainstream of psychological experimentation and reflected in statistical training practises" (p. 221). Our approach to data analysis is intended to reflect this statistical reform.

Statistical Assumptions

In all univariate (mixed-model) repeated-measures ANOVAs performed in this study (with Type III sum of squares), Huynh-Feldt epsilon-adjusted degrees of freedom were used for omnibus tests if the sphericity assumption was not met (e.g., Kirk, 1995; Maxwell & Delaney, 1990). Distributions of participant latency means and error percentages, and relevant difference scores of means and percentages were plotted and visually inspected for outliers and general departures of normality. ANOVAs were performed on participant latency means and participant error percentages (i.e., F_1 analyses with participants as the random variable). We did not perform analyses over item means (i.e., F_2 analyses), because the selected word items consisted of a non-random and exhaustive selection from the item population (cf. Dijkstra et al., 1999; Jared & Kroll, 2001; see also Clark, 1973; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999).

Because square-root or arc-sine transformed error data (e.g., Myers, 1979) produced virtually identical results as untransformed data, we only present analyses on untransformed error data. Nevertheless, given that error rates have distributional properties that may deviate considerably from normality (i.e., in certain conditions the majority of participants may perform errorless), primary parametric analyses (i.e., ANOVA) of the error data were confirmed with nonparametric tests to check robustness of the ANOVAs against possible violations of the normality assumption. For one-way comparisons of independent groups of participants we used the Kruskal-Wallis test, and for unpaired two-group comparisons the Mann-Whithey *U* test. For one-way repeated-measures comparisons we applied Friedman's test and for paired comparisons the sign test. Split-plot ANOVAs (e.g., Kirk, 1995; Maxwell & Delaney, 1990) for interaction effects were nonparametrically 'mimicked' by performing either the Kruskal-Wallis or the Mann-Whithey *U* test on the relevant difference scores.

Such a procedure is similar to a one-way (between-subjects) ANOVA on difference scores, which yields identical values of F and p as the actual test of the interaction effect. In like manner, interaction effects of multiway repeated-measures ANOVAs were nonparametrically mimicked by performing either the Friedman test or the sign test on the relevant difference scores. This procedure resembles a repeated-measures ANOVA on difference scores, which yields the same values of F and p as for the actual interaction effect.

Confidence Intervals

Procedures of null hypothesis significance testing (NHST) were augmented with interval estimates of principal mean differences (i.e., contrasts), which entail the technique of constructing *95% confidence intervals*. A confidence interval (CI) for a mean difference shows the magnitude of a difference, the point estimate (e.g., "20 ms"), and the *precision* with which it is estimated (e.g., "from 15 to 25"). The width of an interval, bounded by its confidence limits (i.e., "15 ms" and "25 ms"), is a direct measure of the estimate's precision (i.e., the degree of error associated with it). For example, a CI that ranges from 15 to 25 is more precise than one that ranges from 2 to 38, which, in turn, is more precise than an interval that ranges from -30 to 70. In general, the narrower the CI, the greater the precision of the sample mean difference as an estimate of the population value. The variance of the sample scores and sample size are directly related to the width of a CI, in the same way as they are related to the size of the p-value.

Accuracy and precision

A 95% CI is calculated from the data and provides a range of plausible values for the *population* mean difference that we wish to estimate. The degree of our confidence that the population mean difference is captured by the CI is represented by the confidence *coefficient*, which is conventionally set to 1-alpha. For the same data set, an 80% CI is narrower than a 99% CI (due to the larger t or F statistic). However, it is associated with a smaller degree of confidence (for the 80% CI we expect only in 8 out of 10 experiments that our interval includes the population value). In general, the chosen level of confidence determines the *accuracy* of a CI, that is, whether it includes the population value or not. A 95% CI rather than a 90% CI increases the accuracy of the CI, but decreases its precision (see Sim & Reid, 1999).

The essential meaning of a 95% CI can be expressed as follows. In statistical terms, a 95% CI means that if a series of identical studies were carried out repeatedly on different samples from the same populations, and each time a 95% CI for the mean difference was calculated, then, in the long run, 95% of these intervals would include the population mean difference (e.g., Altman, Machin, Bryant, & Gardner, 2000;

Estes, 1997; Gardner & Altman, 1986; Kirk, 1995; Loftus, 2002; Loftus & Masson, 1994; Masson & Loftus, 2003; Sim & Reid, 1999; Tryon, 2001).

Construction of ANOVA-based confidence intervals

Confidence intervals can be calculated using the same estimate of error variance as in the corresponding ANOVAs, and thus provide mathematically identical outcomes (e.g., Loftus & Masson, 1994; Masson & Loftus, 2003; Maxwell & Delaney, 1990). In a good number of statistical analyses performed in this study, error variance is estimated for a Latin square design (i.e., using the treatments × participants(group) interaction sum of squares; see Cotton, 1989; Maxwell & Delaney, 1990; Myers & Well, 1995; Pollatsek & Well, 1995; Reese, 1997). To ensure correspondence between the appropriate ANOVA and a CI, the same estimate of error variance is used for both.

ANOVA-based CIs are scarcely treated in common textbooks on statistics and many researchers seem unfamiliar with it. Moreover, the CIs we have calculated for our Latin square designs are, as far as we can tell, not discussed in textbooks. Therefore, we will next present the derivation of a CI of a difference between two sample means that is based on (repeated-measures) ANOVA and incorporates the efficiency of a Latin square design. For that we will adopt the notation used in Kirk (1995).

A 100(1-alpha)% CI for a difference between two paired sample means (e.g., from a repeated-measures design) is usually given by

$$\bar{Y}_1 - \bar{Y}_2 - SE\, t_{\alpha/2,\, df} < \psi < \bar{Y}_1 - \bar{Y}_2 + SE\, t_{\alpha/2,\, df}$$

which can also be expressed as

$$\bar{Y}_1 - \bar{Y}_2 \pm SE\, t_{\alpha/2,\, df}$$

where *SE* is the appropriate (paired-samples) standard error and *t* is the (two-tailed) critical value in Student's *t* distribution for the desired alpha level and with the appropriate degrees of freedom (*df*). Expressed as a contrast or comparison among means, the CI has the more general form (e.g., Kirk, 1995)

$$\hat{\psi} \pm \hat{\sigma}_{\hat{\psi}} t_{\alpha/2,\, df}$$

where, as the reader may want to verify, for the specific case of *two* sample means the contrast and its standard error simply reduce to

$$\hat{\psi} = \sum_{j=1}^{p} c_j \overline{Y}_{\cdot j} = 1\overline{Y}_{\cdot 1} + (-1)\overline{Y}_{\cdot 2} = \overline{Y}_{\cdot 1} - \overline{Y}_{\cdot 2} \text{ and}$$

$$\hat{\sigma}_{\hat{\psi}} = \sqrt{\sum_{j=1}^{p} \frac{c_j^2}{n_j} MS_{\text{error}}} = \sqrt{\frac{2}{n} MS_{\text{error}}} = SE$$

After some simple algebra it follows that a 100(1-alpha)% CI for the difference between two paired sample means can be formulated as

$$\overline{Y}_1 - \overline{Y}_2 \pm \sqrt{F_{\alpha,\ 1,\ df_{\text{error}}}} \sqrt{\frac{2}{n} MS_{\text{error}}}$$

where, under the standard assumptions of repeated-measures ANOVA, F is the critical value in the F distribution for the desired alpha level and in which the appropriate df_{error} and MS_{error} are (e.g., Pollatsek & Well, 1995)

$$MS_{\text{error}} = MS_{\text{Treatments*Participants}},\ df_{\text{error}} = n - 1 \qquad \text{for repeated - measures analysis}$$
$$MS_{\text{error}} = MS_{\text{Treatments*Participants(Groups)}},\ df_{\text{error}} = n - a \quad \text{for Latin square analysis}$$

Example. Suppose we obtain two repeated measures (conditions A and B) from each of 10 participants. The sample means are 650 ms (SD = 163) and 705 ms (SD = 221), respectively, and the difference between the two means is 55 ms (SE = 30). We may first consider the use of standard NHST. A simple paired t-test shows that the difference of 55 ms is not statistically significant ($t(9) = 1.83$, $p = .11$). An identical result is obtained if we perform a repeated-measures ANOVA on these data, because with only two sample means the square root of F is equal to t ($F(1,9) = 3.35$, MSE = 4516.67, $p = .11$). Obviously, the temporal position in which conditions A and B were administered to the participants was counterbalanced, using a single Latin square: Half the participants received the order AB and for the other half it was reversed. By adding subgroup (AB versus BA) as a between-participants factor in the repeated-measures ANOVA, the statistical analysis reflects our experimental design (Pollatsek & Well, 1995). This may reduce the estimate of error variance, thereby improving statistical power. Indeed, in our example, by running the appropriate analysis the value of MSE reduces considerably ($F(1,8) = 7.83$, MSE = 1930.63, $p = .02$), and in addition we have reached the desired statistical significance. In this fictional example, the use of the appropriate estimate of error variance pays of rather well: The relative efficiency (RE, e.g., Kirk, 1995; Myers & Well, 1995) of the Latin square design over the regular repeated-measures design is approximately 2.3, which indicates that in order to reach a MSE of 1930.63 in a regular repeated-measures ANOVA, we would require about twice the number of participants. Positively, this advantage should also be reflected in an interval estimate of the difference between conditions A and B. Using the regular formula for paired samples, as shown above, a 95% CI is calculated from the estimated difference of 55 ms, $SE = 30$, $df = 9$, and $t = 2.26$, which gives an interval from -13 ms to 123 ms. Note that this interval includes the value 0, which agrees with the outcome of the *t*-test (i.e., $p > .05$). The same CI is obtained with the ANOVA-based formula. With 9 degrees of freedom for the error term, a critical value of $F = 5.12$ and $MSE = 4516.67$, the 95% CI also runs from -13 ms to 123 ms. However, with the appropriate ANOVA, *MSE* reduces to 1930.63 and by loosing one degree of freedom the critical value is $F = 5.32$. This results into a more precise CI of 10 ms to 100 ms, one that does *not* include the value 0. Thus, constructing CIs with the appropriate variance estimates provided by ANOVA gives a result that reflects the efficiency of the experimental design and which is fully analogous to the outcome of the ANOVA in terms of *p*-values.

NHST and interval estimation

Because they are built on the same statistical theory, a CI and a null hypothesis test of a mean difference contain the same inferential information. That is, for the specific case of comparisons among sample means (assuming *t* or *F* sampling distributions), a CI can be employed to test a null hypothesis. This relation is discussed in introductory books on statistics. Specifically, if for a sample data set a 95% CI includes the value 0 (e.g., “from -10 to 50”), which may be considered as a plausible value of the population mean difference, the corresponding ANOVA produces a value of $p > .05$ (i.e., not statistically significant). In particular, for a narrow interval (e.g., “from -1 to 3”) that is associated with a value of $p > .05$, high precision and statistical power of the experiment is indicated and, although there is no statistical significance, it suggests here that a zero or almost zero effect is a plausible value of the population mean difference (i.e., *statistical equivalence*, e.g., Seaman & Serlin, 1998; Tryon, 2001). Thus, a CI provides an indication of (post-hoc) statistical power and statistical equivalence of condition means (e.g., Loftus, 2002).

In turn, if a 95% CI does *not* include the value 0 (e.g., “from 15 to 25”), the corresponding ANOVA produces a value of $p < .05$ (i.e., a statistically significant result). In that case, the value 0, and any other value outside the confidence interval, is not considered a plausible value of the population mean. Specifically, if one of the confidence limits has the *exact* value of 0 (e.g., “from 0 to 40”), the corresponding ANOVA produces a value of *p* that is *exactly* .05 (i.e., a test of scientific integrity). As a matter of fact, a 95% CI is equivalent to a hypothesis test for not just one population value (i.e., a difference of zero) but a *range* of possible population values. In other words, a 95% CI can serve to reject not only a conventional null hypothesis based on a population difference of zero, but also a null hypothesis based on any population difference outside the limits of the confidence interval. For example, if a mean difference of 20 ms is observed and the 95% CI ranges from 15 to 25, all null hypothesis tests that are based on values < 15 and > 25 will be rejected at the .05 level. Thus, in this example, an ANOVA will reject the null hypothesis that the population mean difference is zero, but it also rejects the null hypotheses that the population mean differences are, for instance, 5, 14, 26 or 70 ms.

Furthermore, the information conveyed by a confidence interval clarifies why "not significant" does not imply evidence for a null effect. For example, if a small difference is observed (e.g., "2 ms") and a significance test produces a p-value of .894, which is not a statistically significant result, it cannot be unequivocally asserted that the population value is probably close to zero (i.e., the fallacy of accepting the null hypothesis). The width of a corresponding confidence interval shows us why this is true. If the observed non-significant difference is associated with a rather wide confidence interval (e.g., "from -66 to 70") we are able to tell that the experiment has low precision (and statistical power), in which a difference that is not statistically significant may actually be consistent with a rather huge population value (e.g., "70 ms"), which clearly is a non-zero value. A confidence interval provides crucial information if we are dealing with a low-power study that estimates a population mean difference with a large value but of which a NHST procedure yields a p-value > .05, which strongly invites us to believe that there is no real difference in the population. Too much reliance on p-values may cause us to prematurely discard interesting and promising patterns of data (e.g., Carver, 1978; Schmidt, 1996).

Multiple Comparisons

Omnibus ANOVAs were followed by one-degree-of-freedom contrast tests (i.e., tests of differences) that, in the present study, involved elementary pairwise comparisons among means. This general recommended strategy is fully compatible with the construction of confidence intervals, since the selected F statistic that is used to construct a confidence interval also has one degree of freedom in the numerator (e.g., Masson & Loftus, 2003; Loftus, 2002). Each contrast was tested with ANOVA and estimated with a confidence interval using a *separate variance estimates* approach (e.g., Maxwell & Delaney, 1990), which allows each contrast to have its own specific error term. Thus, when two particular condition means are compared, the data of the other conditions are omitted from the analysis. This procedure evaded the issue of the sphericity assumption in the repeated-measures analyses, but did not result in a substantially reduced power or precision.

For all sets of pairwise comparisons, familywise Type I error rates were controlled by the Bonferroni procedure (i.e., Dunn's procedure), for the null hypothesis tests as well as for the confidence intervals (see Kirk, 1995; Maxwell & Delaney, 1990). For the null hypothesis tests, this involved restricting the per-comparison alpha level, to keep overall Type I error rate at a nominal level of 5 percent, and for interval estimation construction of *simultaneous* 95% confidence intervals (95% SCI, e.g., see Maxwell & Delaney, 1990). For example, with three pairwise comparisons, the per-comparison alpha level is divided by three and set to .017, and the per-comparison confidence interval is actually a 98.3% confidence interval.

Explication of Meaning of *P*-Values

The values of p reported in the present ANOVA tables and elsewhere in this study result from traditional NHST and refer to the probability of obtaining a value of the test statistic (i.e., F) as large as, or larger than, the one obtained—conditional on the null hypothesis being true (see Kirk, 1995; Nickerson, 2000). Put differently, it refers to the probability of finding an experimental effect given that only chance factors (i.e., random sampling variation) operate on the observations. Notice that here we simply use the *conventional* meaning of the p-value. The alpha level was set to .05. If, for a null hypothesis test, the value of p was $< .05$, we declared an experimental effect statistically significant. In that case we reasoned that, assuming that H0 is true, the observed effect may have a low chance of occurrence (i.e., we expect chance to yield it in less than 5 of 100 experiments), which lends more credence to the alternative hypothesis which states that it was not sampling error, but actually our experimental manipulation that caused the effect. In other words, we obtained evidence for an effect of the independent variable. If, on the other hand, a value of p was $> .05$, we designated an experimental effect as not statistically significant. This means that we did not obtain evidence for an effect of the independent variable, nor however did we obtain evidence for a zero effect. For a large value of p, random sampling variation is suspected to yield an effect of this particular size a little too often, which decreased our confidence that there is a true effect in the population. If such was the case, the corresponding 95% CI supplied the necessary information to asses the statistical power and precision of the experiment, and the actual degree of statistical equivalence (i.e., whether there is a narrow confidence interval centred around the value of 0).

3 The Intralingual Consistency Effect in Monolingual and Bilingual Word Naming Performance

This study examined whether manifold intralingual spelling-to-sound relations have an impact on English word-naming performance of monolingual native English speakers and bilingual native Dutch speakers with English as their second language. Essentially, it investigated whether naming of an English word (e.g., MOOD) is influenced by spelling-to-sound knowledge of English enemy neighbors (e.g., BLOOD), which suggests competition between phonological codings. The ratio of two kinds of neighbors, friends (similar pronunciations) and enemies (dissimilar pronunciations), determines the degree of (in)consistency of a spelling body mapping to a phonological body. This, in turn, predicts word-naming performance (i.e., naming latency and error rate). It is expected that when a word has English enemies, word-naming performance is worse than when it has no enemies. Furthermore, when a word has more enemies than friends naming performance is expected to be worse than in the opposite case, when a word has more friends than enemies. These predictions can also (in a different order) be stated as: AM > TM, AM > CM, and TM > CM.

In addition, for the Dutch bilingual participants, Experiment 1 also explored whether manifold *interlingual* spelling-to-sound relations have an impact on English word-naming performance, that is, whether naming of an English word (e.g., MOOD) is influenced by spelling-to-sound knowledge of Dutch neighbors (e.g., ROOD, LOOD, and NOOD). We were not sure whether an interlingual consistency effect was to be expected here, because we did not manage to equate the two word lists on a number of linguistic variables that are known to affect word-naming performance. In fact, as described in the General Method section, the English words with Dutch neighbors are relatively more familiar in form and orthographically less complex. Also, unfortunately, the Dutch neighbors were not all interlingual enemies. In approximately half of the cases, Dutch pronunciations of spelling bodies in English words were fairly similar to the English pronunciations. For example, the phonologic bodies of the English words THEM and YARD are very similar to that of the Dutch words REM and HARD (in Dutch meaning “brake” and “fast”), which yields a more or less friendly neighborhood. In any case, if Dutch neighbors cause interference, potentially poor word-naming performance on the list of English words with Dutch neighbors (e.g., MOOD) must first overcome a small inherent word familiarity advantage.

EXPERIMENT 1

Method

Participants and materials

A group of 30 native English speakers (USA participants) and a group of 30 Dutch-English bilinguals (Dutch participants) took part in the naming experiment. They were presented with the 120 English words described in the General Method section.

Experimental design

In this experiment, three basic word types were contrasted. One group of consistent words had spelling bodies that are pronounced the same across all of its English neighbors (e.g., MOON, STAIN). Hence, the words of this group contained *consistent mappings* (CM). A second group of words had spelling bodies that are not pronounced the same across all of its English neighbors. These words had both friends and enemies. For this group of inconsistent words, the number and summed frequency of friends was larger than that of enemies (e.g., MOOD, SAID). Hence, these words had *typical mappings* (TM). A third group of words also had friends and enemy English neighbors. However, for this group, the number and summed frequency of friends was smaller than that of enemies. Therefore, these inconsistent words had *atypical mappings* (AM). They are unusual when compared to words with identical spelling bodies (e.g., BLOOD, PAID). Furthermore, English words that either had consistent, typical, or atypical mappings were grouped in two separate word lists. In the first list the English words had Dutch neighbors (e.g., MOON, MOOD, and BLOOD) and in the second list they did not have Dutch neighbors (e.g., STAIN, SAID, and PAID). Both word lists contained an equal number of words for each group of words representing one of the three word types.

Thus, by combining differences in interlingual and intralingual neighborhood, this experiment had a two by three factorial design with Interlingual Neighborhood (With Dutch Neighbors vs. No Dutch Neighbors) and Word Type (AM vs. TM vs. CM) as the independent variables. Both naming latencies and error rates were measured to assess main and interaction effects. Each participant named all of the 20 words of each group of words. Thus, the participants were observed under each of the six combinations formed by Interlingual Neighborhood and Word Type. That is, repeated measures were obtained on the participants, which is a common practise in psycholinguistic research that uses word materials, and is intended to increase the statistical power and precision of the experiment by isolating residual variance due to individual differences between participants in response latencies and error rates.

In this experiment and the following ones, the spelling body of an English word with a typical mapping (e.g., MOOD) also appeared in an English word with an atypical mapping (e.g., BLOOD). Thus, the *same* spelling body was used to create both a typical and an atypical English word. There were several reasons why we used the same spelling body across different word types. One main reason, which is of relevance for the print-to-speech correspondence task (see Chapters 4 and 5), is that it required a substantially smaller number of recordings of spoken rimes. For example, the recording of the spoken rime of the word MOOD could be used to create both an auditory stimulus for a match trial with the printed word MOOD (i.e., MOOD - /ud/) and for a catch trial with the printed word BLOOD (i.e., BLOOD - /ud/). Another reason for using the same spelling body across word types is that, as a result, the words with typical and atypical mappings were as closely as possible matched in visual form (e.g., MOOD – BLOOD). It also had the effect that the consistency ratios of the typical and atypical words were inversely related: The larger the consistency ratio of a typical word, the smaller it was for its atypical counterpart. This substantially simplified the process of stimulus contrasting and matching.

Thus, participants were presented twice with the same spelling body. To prevent intralist-priming effects of spelling bodies, the two words containing the same spelling body were presented in two separate blocks of trials. These blocks were administered in two separate experimental sessions conducted one week apart, a precautionary procedure that was also used by Jared (1997). Hence, participants named words in two separate blocks of trials, A and B. Table 4 presents the layout of the experimental design. The blocks A and B contained equal numbers of words from all three word types. This was accomplished by separating each list of 20 words comprising one of the three word types (CM, TM, and AM) into two sub word-lists (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2). For example, if the typical word MOOD (from TM_1) was presented in Trial Block A, then the atypical word BLOOD (from AM_1) was presented approximately one week later in Trial Block B. Conversely, if the atypical word HOME (from AM_2) was presented in Trial Block A, then the typical word COME (from TM_2) was presented approximately one week later in Trial Block B. In general terms, if the sub word-list CM_1 was presented in Trial Block A, then the sub word-list CM_2 was presented in Trial Block B. Further, if the sub word-list TM_1 (e.g., with the word MOOD) was presented in Trial Block A, then the sub word-list TM_2 (e.g., with the word COME) was presented in Trial Block B. Consequently, in case the sub word-list TM_1 (e.g., with the word MOOD) was presented in Trial Block A, the sub word-list AM_1 (e.g., with the word BLOOD) was to be presented in Trial Block B. If, conversely, the sub word-list TM_2 (e.g., with the word COME) was presented in Trial Block B, then the sub word-list AM_2 (e.g., with the word HOME) was to be presented in Trial Block A. In sum, both for the list of

English words with Dutch neighbors and for the list without Dutch neighbors, Trial Block A comprised sublists CM_1, TM_1, and AM_2, and Trial Block B comprised sublists CM_2, TM_2, and AM_1 (see Table 4).

Table 4.
Experimental design for Experiment 1 (Word Naming: Dutch Participants vs. USA Participants). The word list comprising each word type is separated into two sub word-lists (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2). These sub word-lists are systematically distributed over two trial blocks (A and B). The temporal order of trial block is Latin-square counterbalanced across two different participant groups (Participant Group 1: Sequence A-B; Participant Group 2: Sequence B-A). (CM = Consistent Mappings; TM = Typical Mappings; AM = Atypical Mappings.)

No Dutch Neighbors

Trial Block	A	B
Sub Word-List	CM_1	CM_2
Example	***CAPE***	***STAIN***
Sub Word-List	TM_1	TM_2
Example	***GREY***	***SAID***
Sub Word-List	AM_2	AM_1
Example	***PAID***	***KEY***

With Dutch Neighbors

Trial Block	A	B
Sub Word-List	CM_1	CM_2
Example	***HOOK***	***MOON***
Sub Word-List	TM_1	TM_2
Example	***COME***	***MOOD***
Sub Word-List	AM_2	AM_1
Example	***BLOOD***	***HOME***

The temporal order of the two trial blocks containing different sub word-lists was counterbalanced across two different participant groups according to a single Latin square (participant group 1: Sequence A-B; participant group 2: Sequence B-A). Participants were randomly assigned to the different sequences. The counterbalancing procedure was intended to disentangle the effect of temporal position of the procedural ("nuisance") variable Trial Block from the effects of the independent variables (i.e., sub word-lists nested within trial blocks). Running this procedure is essential, because it may occur that experience from naming words in the first block of trials (either A or B) results in an advantage for naming performance in the second block (i.e., a positive transfer). Foremost, the counterbalancing procedure established that, within each temporal position, different participants named different lists of words, although all participants eventually named all of the words. That is, within each temporal position, one group of participants named the words of sub word-lists CM_1, TM_1, and AM_2 (Trial Block A) and another group of participants named the words of sub word-lists CM_2, TM_2, and AM_1 (Trial Block B). These differences between participants and word items possibly induce additional variance in the statistical model. However, this potential source of variance can, just as the variance due to individual differences between participants, be isolated and removed from the estimate of error variance, which may improve the efficiency of the design in terms of statistical power and precision. Procedures for this are provided by for example Kirk (1995) and Myers and Well (1995; see also Pollatsek & Well, 1995), and in the present case involves simply adding participant group (Sequence A-B vs. Sequence B-A) as a between-subjects variable in an analysis of variance (ANOVA), and testing the effects against the resulting treatments×participants(group) error term. With respect to the terminology, in this study we use the classic partition of sum of squares approach to describe an univariate (mixed-model) repeated-measures ANOVA. Nevertheless, the procedure may be unnecessary altogether, because all participants received the other trial block one week later in the second experimental session.

Analyses that included participant group as a between-subjects variable were considered if the main effect of participant group (i.e., the Sequence effect) did not explain a substantial percentage of variance. The relevance of this criterion is that, in general, sequence effects also reflect the interaction between temporal position and treatment effects. Thus, a notable sequence effect may suggest asymmetrical transfer across positions, which indicates that treatment effects are contaminated with the effect of temporal position.

The data from the monolingual native English speakers and the bilingual native Dutch speakers were analysed separately, the reason being that the data from the two language groups were collected separately in two different countries (the USA and The Netherlands) and in different time periods. However, in both cases the same type of computer with identical system software was used, together with the very same voice key. Thus we nevertheless present a number of analyses that involve comparisons across language groups.

In sum, this experimental design contrasted naming performance on lists of AM-, TM-, and CM-words such as PAID, SAID, and STAIN (for the set without Dutch neighbors), and also naming performance on lists of AM-, TM-, and CM-words such as BLOOD, MOOD, and MOON (for the set with Dutch neighbors).

Procedure

The words were displayed in lowercase letters in the centre of the computer screen of an Apple Macintosh PowerPC 4400/200, using a standard Macintosh font (Geneva, size 18). The monitor (Apple Multiple Scan15 Display) was set to a refresh-rate of 75-Hz. Naming latencies were collected with a microphone connected to a voice key interfaced to the computer, of which the real-time clock timed latencies in milliseconds from the appearance of the word to the onset of the participant's naming response. Stimulus presentation and data recording were controlled by the computer program fLexi (version 3.3.8), an experiment generator developed at the Department of Psychology of the University of Amsterdam. Prior to data collection, the software was extensively tested within the experimental set-up of the naming task.

Participants were tested individually in a quiet and normally lit room. They were seated at approximately 50 cm in front of the computer screen and were given verbal instructions. Words were presented one at a time and remained on the screen until the participant began to speak into the microphone. The order of trials was randomised for each participant and for each trial block. The participants were instructed to read each word aloud as quickly and accurately as possible. The experimenter recorded mispronunciation and voice key errors into a note book. Experimental sessions were also recorded on audio tape. A trial block started with a block of 20 practice trials. None of the practise words contained spelling bodies found in the experimental words. The participants named the two blocks of trials (A and B) in two separate experimental sessions that spanned approximately one week.

Results

Each participant named 20 words of each of the six groups representing combinations of Interlingual Neighborhood and Word Type. The correct naming latencies within these groups were averaged for each participant. In addition, for each participant, a percentage of naming errors was calculated for each of the six word groups. Hence, the data that entered the statistical analyses consisted, for each participant, of a set of six latency means and a set of six percentages of errors (e.g., for PAID, SAID, and STAIN; BLOOD, MOOD, and MOON).

Data filtering

Naming latencies of trials where the word had been pronounced incorrectly were excluded from the latency analyses. This resulted in a rejection of a total of 2.2% and 4.9% trials for the USA and Dutch participants, respectively. Further, a total of 0.6% of the naming trials of the USA participants and 2.1% of the naming trials of the

Dutch participants were excluded from all analyses because the voice key failed to detect the participant's voice or because of apparatus failure. Also, 0.1% of the trials of all participants were excluded from the latency analyses because the naming latency was shorter than 200 ms. Furthermore, following recommendations of Ulrich and Miller (1994), less than 0.5% of the correct naming latencies were classified to be outliers. A cut-off procedure (i.e., truncation) was used that rejected all latencies greater than 1000 ms, which was suggested by a visual inspection of plotted naming latencies. We avoided the "restricted means" procedure that is commonly used in the psycholinguistic literature, which rejects all latencies beyond a criterion number of standard deviations (typically 2-3 *SD*s) from the mean of each experimental cell (either within or across participants). Monte Carlo simulations of Miller (1991) and Van Selst and Jolicoeur (1994) have shown that this procedure introduces a bias: A mean latency that is computed this way underestimates the true population average, the size of which is a function of item sample size and the skewness of the latency distribution. This may cause systematic bias if experimental conditions are compared that differ on these variables. In Experiment 1, the cut-off procedure resulted in rejection of 0.2% of the correct naming latencies for the USA participants and 0.3% for the Dutch participants.

Data of USA participants

The data of two participants were lost, resulting in a sample that consisted of 28 native English speakers. The mean percentages of naming errors and mean naming latencies for the USA (and Dutch) participants are presented in Figures 3 and 4, respectively. We first present the analyses of the *error data*. It will be recalled that, for English word naming, the ratio of friends and enemies is expected to affect the rate of naming errors. Specifically, words with atypical mappings (e.g., BLOOD) should produce more naming errors than words with typical mappings (e.g., MOOD) and consistent mappings (e.g., MOON). We also expect more naming errors for words with typical mappings than for words with consistent mappings. As can be seen in Figure 3, the ratio of friends and enemies was indeed associated with the number of naming errors. USA participants produced more naming errors on inconsistent words that have stronger enemies than friends (e.g., BLOOD and PAID), than on inconsistent words that have stronger friends than enemies (e.g., MOOD and SAID). The number of naming errors for consistent words (e.g., MOON and STAIN), however, was not markedly lower compared to the words with typical mappings. On the contrary, consistent words with Dutch neighbors (e.g., MOON) were associated with higher error rates.

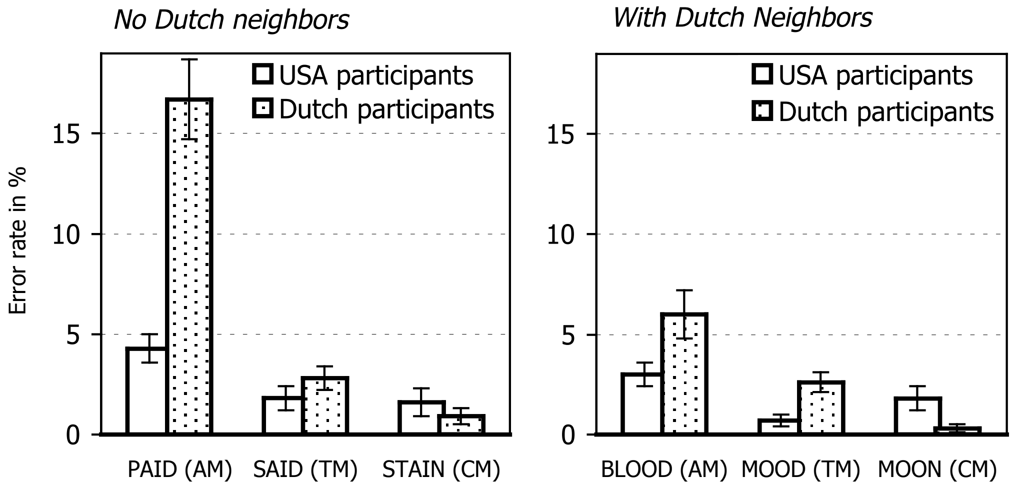

Figure 3. Mean percentages of naming errors of USA and Dutch participants as a function of Word Type (AM-words vs. TM-words vs. CM-words) for English words without Dutch neighbors (left panel) and English words with Dutch neighbors (right panel) in Experiment 1. Error bars represent the standard error of the mean.

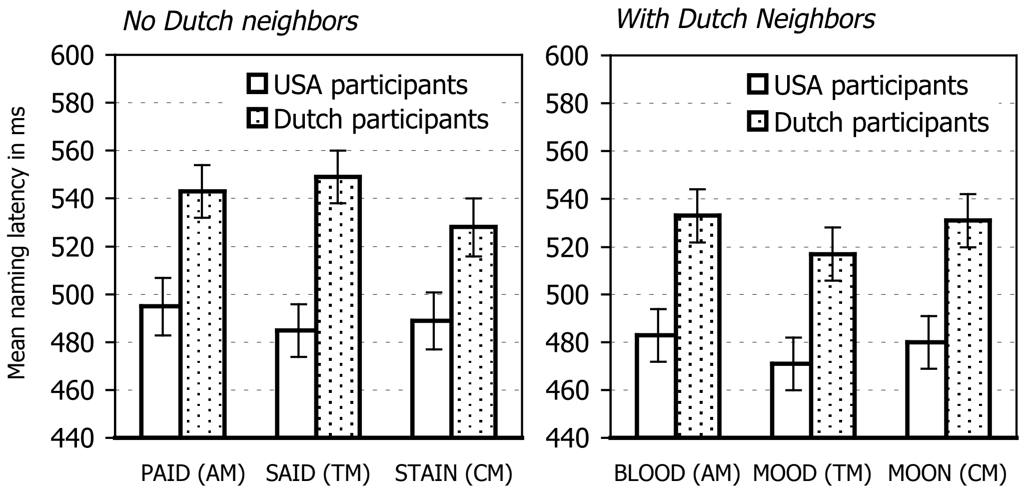

Figure 4. Mean naming latencies of USA and Dutch participants as a function of Word Type (AM-words vs. TM-words vs. CM-words) for English words without Dutch neighbors (left panel) and English words with Dutch neighbors (right panel) in Experiment 1. Error bars represent the standard error of the mean.

Omnibus analysis of variance. Table 5 presents the results of a repeated-measures ANOVA. The table also provides values of partial eta squared as a measure of association strength and the results of non-parametric tests. Preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect or an effect of temporal position of trial block. Therefore, adding participant group (Sequence A-B vs. Sequence B-A) as a between-subjects variable was not expected to improve statistical power nor precision because a small reduction in error variance might not weigh against the loss in degrees of freedom due to this additional variable in the analysis. Hence, for the significance tests and the accompanying simultaneous confidence intervals, the regular (repeated-measures) treatments × participants interaction sum of squares was used to estimate error variance.

Table 5.
Analysis of variance on error percentages for Experiment 1 (USA participants).

Source of variance	*SS*	□	*df*	*MS*	*F*	*p (F|H0)*	p^a	$□_p{}^2$
• Block Position	9.72		1	9.72	3.20	.085	.096	.106
Block Position × Participant	81.94		27	3.04				
• Sequence	.07		1	.07	.01	.943	.815	.000
Participant(Group)	376.71		26	14.49				
• Neighborhood	21.43		1	21.43	2.02	.167	.134	.069
Neighborhood × Participant	286.91		27	10.63				
• Word Type	184.23	1.0	2.00	92.11	9.06	< .001	.002	.251
Word Type × Participant	549.11	1.0	54.00	10.17				
• Neighborhood × Word Type	16.96	1.0	2.00	8.48	1.02	.368	.089	.036
Neighborhood × Word Type × Participant	449.70	1.0	54.00	8.33				

Note. Type III Sum of Squares with Huynh-Veldt □-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

Table 6.
Analysis of variance on error percentages for Experiment 1 (Dutch participants).

Source of variance	*SS*	□	*df*	*MS*	*F*	*p (F|H0)*	p^a	$□_p{}^2$
• Block Position	9.39		1	9.39	1.54	.226	.286	.052
Block Position × Participant	171.17		28	6.11				
• Sequence	58.34		1	58.34	1.37	.252	.345	.048
Participant(Group)	1147.70		27	42.51				
• Neighborhood	625.86		1	625.86	29.25	< .001	< .001	.511
Neighborhood × Participant	599.14		28	21.40				
• Word Type	3793.39	.59	1.17	3231.50	47.77	< .001	< .001	.630
Word Type × Participant	2223.28	.59	32.87	67.64				
• Neighborhood × Word Type	1035.35	.64	1.27	814.96	28.57	< .001	< .001	.505
Neighborhood × Word Type × Participant	1014.66	.64	35.57	28.52				

Note. Type III Sum of Squares with Huynh-Veldt □-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

Planned contrasts. As can be confirmed in Table 5, there was a statistically significant main effect of Word Type, which accounted for a considerable amount of variance. This overall effect was further inspected with three (Bonferroni-adjusted) pairwise comparisons, which kept familywise Type I errors at 5%. Hence, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals (95% SCI). The pairwise comparisons involved three planned contrasts that evaluated whether error rates for words like BLOOD and PAID were higher than for words like MOOD and SAID (AM > TM) and for words like MOON and STAIN (AM > CM), and higher for words like MOOD and SAID than for words like MOON and STAIN (TM > CM). For the **AM > TM** contrast there was a statistically significant difference of **2.4** percentage points, with a 95% SCI of 0.9 to 3.9 ($F(1,27) = 16.36$, $MSE = 4.97$, $p < .001$). The **2.0** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 0.4 to 3.6 ($F(1,27) = 9.99$, $MSE = 5.41$, $p = .004$). Finally, the **TM > CM** contrast showed a difference in opposite direction, **-0.4** percentage points, with a 95% SCI of -2.0 to 1.1, that was not statistically significant ($F(1,27) = .57$, $MSE = 4.87$, $p = .456$).

Table 5 further shows that the main effect of Interlingual Neighborhood was not statistically significant. Participants made fewer naming errors on words with Dutch neighbors (e.g., MOON) than on words without Dutch neighbors (e.g., STAIN). The difference was 0.7 percentage points, with a 95% CI of -0.3 to 1.7 ($F(1,27) = 2.02$, $MSE = 3.54$, $p = .167$). The interaction effect of Interlingual Neighborhood and Word Type was not statistically significant either, and accounted for a very small percentage of variance. Thus the ANOVA indicated no evidence for differential error patterns of Word Type across Interlingual Neighborhood. The same (Bonferroni-adjusted) planned contrasts were performed separately for the words with (e.g., MOON) and the words without Dutch neighbors (e.g., STAIN). For words like MOON, the **AM > TM**, **AM > CM**, and **TM > CM** contrasts gave differences of **2.3** (95% SCI 0.7 to 4.0), **1.2** (95% SCI -0.4 to 2.9), and **-1.1** (95% SCI -2.6 to 0.4) percentage points, respectively, $F(1,27) = 12.57$, $MSE = 6.00$, $p = .002$; $F(1,27) = 3.57$, $MSE = 6.13$, $p = .070$; $F(1,27) = 3.24$, $MSE = 4.96$, $p = .083$, respectively. For words like STAIN the same contrasts gave differences of **2.5** (95% SCI 0.1 to 4.9), **2.7** (95% SCI 0.2 to 5.2), and **0.2** (95% SCI -2.2 to 2.6) percentage points, respectively, $F(1,27) = 7.00$, $MSE = 12.50$, $p = .013$; $F(1,27) = 7.49$, $MSE = 13.41$, $p = .011$; $F(1,27) = .04$, $MSE = 12.48$, $p = .851$, respectively.

Turning to the analyses of the *latency data*, Figure 4 shows the mean naming latencies of the USA (and Dutch) participants. By comparing Figures 3 and 4 it can be verified that for the USA participants the patterns of naming latencies and error rates were fairly alike. Naming latencies were longer for inconsistent words that have stronger enemies than friends (e.g., BLOOD and PAID), than for inconsistent words that have stronger friends than enemies (e.g., MOOD and SAID). Naming latencies for consistent words (e.g., MOON and STAIN) were not markedly shorter compared to the words with typical mappings. On the contrary, consistent words with Dutch neighbors (e.g., MOON) produced longer naming latencies.

Omnibus analysis of variance. Table 7 presents the results of a repeated-measures ANOVA. Preliminary analyses on the procedural variable again indicated no evidence for a Sequence or temporal position effect. Therefore, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Table 7.
Analysis of variance on naming latencies for Experiment 1 (USA participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F\H0)*	η_p^2
• Block Position	407.16		1	407.16	.51	.483	.018
Block Position × Participant	21723.34		27	804.57			
• Sequence	29.80		1	29.80	.00	.970	.000
Participant(Group)	551394.49		26	21207.48			
• Neighborhood	5417.36		1	5417.36	38.23	< .001	.586
Neighborhood × Participant	3825.98		27	141.70			
• Word Type	3369.33	1.0	2.00	1684.67	8.45	.001	.238
Word Type × Participant	10771.00	1.0	54.00	199.46			
• Neighborhood × Word Type	219.00	.96	1.92	114.16	.56	.566	.020
Neighborhood × Word Type × Participant	10502.67	.96	51.80	202.77			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Table 8.
Analysis of variance on naming latencies for Experiment 1 (Dutch participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F\H0)*	η_p^2
• Block Position	2028.43		1	2028.43	2.98	.095	.096
Block Position × Participant	19073.07		28	681.18			
• Sequence	110250.90		1	110250.90	6.37	.018	.191
Participant(Group)	467111.14		27	17300.41			
• Neighborhood	7325.52		1	7325.52	41.26	< .001	.596
Neighborhood × Participant	4970.98		28	177.54			
• Word Type	2024.01	.94	1.87	1081.84	4.07	.025	.127
Word Type × Participant	13928.66	.94	52.39	265.89			
• Neighborhood × Word Type	9166.22	.90	1.80	5083.57	18.42	< .001	.397
Neighborhood × Word Type × Participant	13933.78	.90	50.49	275.99			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Planned contrasts. As can be confirmed in Table 7, there was a statistically significant main effect of Word Type, which accounted for a considerable amount of variance. This overall effect was further inspected with three (Bonferroni-adjusted) pairwise comparisons. The pairwise comparisons involved the same three planned contrasts as for the error data. The alpha level was again set to .017.

For the **AM > TM** contrast there was a statistically significant difference of **11** ms, with a 95% SCI of 4 to 18 ($F(1,27) = 14.29$, $MSE = 118.13$, $p < .001$). The difference for the **AM > CM** contrast was **5** ms, with a 95% SCI of -2 to 11, and was not statistically significant ($F(1,27) = 3.40$, $MSE = 91.84$, $p = .076$). Finally, the **TM > CM** contrast showed a difference in opposite direction, **-6** ms, with a 95% SCI of -13 to 0, that was not statistically significant ($F(1,27) = 6.14$, $MSE = 89.39$, $p = .020$).

Table 7 further shows that, in contrast with the error data, here the main effect of Interlingual Neighborhood was statistically significant. Participants had shorter naming latencies for words with Dutch neighbors (e.g., MOON) than for words without Dutch neighbors (e.g., STAIN). The difference was 11 ms, with a 95% CI of 8 to 15 ($F(1,27) = 37.72$, $MSE = 47.93$, $p < .001$). The interaction effect of Interlingual Neighborhood and Word Type was again not statistically significant, and accounted for a very small percentage of variance. Thus the ANOVA provided no evidence for differential latency patterns of Word Type across Interlingual Neighborhood. Again, (Bonferroni-adjusted) planned contrasts were performed separately for the words with (e.g., MOON) and those without Dutch neighbors (e.g., STAIN). For words like MOON the **AM > TM**, **AM > CM**, and **TM > CM** contrasts gave differences of **12** ms (95% SCI 3 to 21), **3** ms (95% SCI -6 to 12), and **-9** ms (95% SCI -17 to -1), respectively, $F(1,27) = 11.57$, $MSE = 178.91$, $p = .002$; $F(1,27) = .78$, $MSE = 164.96$, $p = .384$; $F(1,27) = 7.94$, $MSE = 146.63$, $p = .009$, respectively. For words like STAIN the same contrasts gave differences of **10** ms (95% SCI -1 to 21), **6** ms (95% SCI -5 to 17), and **-3** ms (95% SCI -12 to 5), respectively, $F(1,27) = 5.07$, $MSE = 265.53$, $p = .033$; $F(1,27) = 2.19$, $MSE = 262.98$, $p = .151$; $F(1,27) = .97$, $MSE = 166.25$, $p = .333$, respectively.

Data of Dutch participants

Data of one participant was lost, resulting in a sample that consisted of 29 Dutch-English bilinguals. The mean percentages of naming errors and mean naming latencies for the Dutch participants are also presented in Figures 3 and 4. The analyses of the *error data* are presented first. Figure 3 shows that the ratio of friends and enemies was indeed associated with the number of naming errors. Similar to the USA participants, Dutch participants produced more naming errors on inconsistent words that have stronger enemies than friends, than on inconsistent words that have stronger friends than enemies. Contrary to the USA participants, however, the rate of naming errors for consistent words was lower compared to the words with typical mappings.

Omnibus analysis of variance. Table 6 presents the results of a repeated-measures ANOVA. Preliminary analyses on the procedural variable again indicated no evidence for a Sequence or temporal position effect. Therefore, the regular (repeated-measures) treatments × participants interaction sum of squares was used to estimate error variance.

Planned contrasts. As can be confirmed in Table 6, there was a statistically significant main effect of Word Type, which accounted for a large percentage of variance. This overall effect was further inspected with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals. For the **AM > TM** contrast there was a statistically significant difference of **8.7** percentage points, with a 95% SCI of 5.2 to 12.2 ($F(1,28) = 40.16$, $MSE = 27.37$, $p < .001$). The **10.8** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 7.2 to 14.4 ($F(1,28) = 57.54$, $MSE = 29.26$, $p < .001$). Finally, for the **TM > CM** contrast there was a statistically significant difference of **2.1** percentage points, with a 95% SCI of 0.9 to 3.2 ($F(1,28) = 21.28$, $MSE = 2.92$, $p < .001$).

Table 6 further shows that, contrary to the USA participants, the main effect of Interlingual Neighborhood was also statistically significant and accounted for a considerable amount of variance. Dutch participants made fewer naming errors on words with Dutch neighbors (e.g., MOON) than on words without Dutch neighbors (e.g., STAIN). The difference was 3.8 percentage points, with a 95% CI of 2.4 to 5.2 ($F(1,28) = 29.25$, $MSE = 7.13$, $p < .001$). Also in contrast with the USA participants, the interaction effect of Interlingual Neighborhood and Word Type was statistically significant, and accounted for a large percentage of variance. Thus here the ANOVA provided evidence for differential error patterns of Word Type across Interlingual Neighborhood. The source of this interaction can be appreciated by inspection of Figure 3, which shows that Dutch participants produced an excessive rate of naming errors on words like PAID that have atypical mappings and no Dutch neighbors. Simple effects comparing words like PAID with words like BLOOD with atypical mappings, comparing words like SAID with words like MOOD with typical mappings, and comparing words like STAIN with words like MOON with consistent mappings showed that only the first comparison was statistically significant, $F(1,28) = 36.59$, $MSE = 45.29$, $p < .001$, $F(1,28) = .05$, $MSE = 9.36$, $p = .832$, and $F(1,28) = 1.30$, $MSE = 2.99$, $p = .264$, respectively. Further, (Bonferroni-adjusted) planned contrasts were performed separately for the words with (e.g., MOON) and those without Dutch neighbors (e.g., STAIN). For words like MOON, the **AM > TM**, **AM > CM**, and **TM > CM** contrasts gave differences of **3.4** (95% SCI 0.2 to 6.7), **5.7** (95% SCI 2.5 to 8.8), and **2.2** (95% SCI 0.9 to 3.6) percentage points, respectively, $F(1,28) = 7.12$, $MSE = 24.20$, $p = .013$; $F(1,28) = 21.26$, $MSE = 22.08$, $p < .001$; $F(1,28) = 17.79$, $MSE = 4.09$, $p < .001$, respectively. For words like STAIN the same contrasts gave differences of **14.0** (95% SCI 8.9 to 19.0), **15.9** (95% SCI 10.6 to

21.1), and **1.9** (95% SCI 0.3 to 3.5) percentage points, respectively, $F(1,28) = 49.97$, $MSE = 56.59$, $p < .001$; $F(1,28) = 60.03$, $MSE = 60.78$, $p < .001$; $F(1,28) = 9.11$, $MSE = 5.73$, $p = .005$, respectively.

Turning to the analyses of the *latency data*, Figure 4 shows the mean naming latencies of the Dutch participants. Inspection of Figure 4 indicates that Dutch and USA participants produced a similar pattern of naming latencies on the words with Dutch neighbors, with naming latencies longer for inconsistent words that have stronger enemies than friends (e.g., BLOOD), than for inconsistent words that have stronger friends than enemies (e.g., MOOD). Furthermore, latencies for consistent words were longer than for the inconsistent words with typical mappings. With regard to the words without Dutch neighbors, naming latencies of Dutch participants for atypical inconsistent words (e.g., PAID) were not longer than the ones with typical mappings (e.g., SAID), as was observed with the USA participants, but slightly shorter. Furthermore, in contrast with the USA participants, naming latencies of Dutch participants for consistent words (e.g., STAIN) were markedly shorter than for inconsistent words with typical mappings (e.g., SAID).

Omnibus analysis of variance. Table 8 presents the results of a repeated-measures ANOVA. Preliminary analyses on the procedural variable indicated that temporal position of trial block explained a small percentage of variance. In addition, a statistically significant but moderate sequence effect was observed that possibly reflected some interaction of trial block and temporal position of trial block. Hence, because these effects were very small, participant group was not added as a between-subjects variable. Thus, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Planned contrasts. As can be confirmed by Table 8, there was a statistically significant main effect of Word Type, which accounted for a moderate percentage of variance. This overall effect was further inspected with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was set to .017. For the **AM > TM** contrast the difference of **5** ms was not statistically significant, with a 95% SCI of -3 to 13 ($F(1,28) = 2.17$, $MSE = 151.66$, $p = .152$). The difference for the **AM > CM** contrast was **8** ms, with a 95% SCI of 0 to 16, and was statistically significant ($F(1,28) = 6.98$, $MSE = 143.70$, $p = .013$). Finally, for the **TM > CM** contrast there was a difference of **4** ms, with a 95% SCI of -2 to 9, that was not statistically significant ($F(1,28) = 2.33$, $MSE = 78.88$, $p = .138$).

Table 8 further shows that, in agreement with the error data, the main effect of Interlingual Neighborhood was statistically significant. This effect on naming latency was also found in the data of the USA participants. Dutch participants had shorter naming latencies for words with Dutch neighbors (e.g., MOON) than for words without Dutch neighbors (e.g., STAIN). The difference was 13 ms, with a 95% CI of 9 to 17 ($F(1,28) = 37.72$, $MSE = 47.93$, $p < .001$). Also in agreement with the error data, the interaction effect of Interlingual Neighborhood and Word Type was statistically significant, and accounted for a considerable percentage of variance. Thus the ANOVA provided evidence for differential latency patterns of Word Type across Interlingual Neighborhood. The source of this interaction can be appreciated by inspection of Figure 4, which shows that the pattern of naming latencies across Word Type was essentially reversed for the words with and the words without Dutch neighbors. Naming latencies for words like PAID that have atypical mappings and no Dutch neighbors, on which Dutch participants produced an excessive rate of naming errors, were, unexpectedly, not longer than for words like SAID. However, in this case we should consider the possibility that the overall naming latency for words like PAID was deflated because the trials of words that potentially produce long latencies were excluded due to mispronunciation. Again, (Bonferroni-adjusted) planned contrasts were performed separately for the words with (e.g., MOON) and those without Dutch neighbors (e.g., STAIN). For words like MOON the **AM > TM**, **AM > CM**, and **TM > CM** contrasts gave differences of **16** ms (95% SCI 7 to 24), **2** ms (95% SCI -8 to 11), and **-14** ms (95% SCI -23 to -5), respectively, $F(1,28) = 22.77$, $MSE = 159.33$, $p < .001$; $F(1,28) = .22$, $MSE = 209.29$, $p = .640$; $F(1,28) = 15.22$, $MSE = 187.32$, $p < .001$, respectively. For words like STAIN the same contrasts gave differences of **-6** ms (95% SCI -18 to 5), **15** ms (95% SCI 1 to 28), and **21** ms (95% SCI 11 to 31), respectively, $F(1,28) = 1.94$, $MSE = 296.91$, $p = .175$; $F(1,28) = 7.80$, $MSE = 409.13$, $p = .009$; $F(1,28) = 27.21$, $MSE = 238.00$, $p < .001$, respectively.

Analyses comparing USA and Dutch participants

Overall, Dutch participants produced more naming errors and longer naming latencies on the English words than the native, English speaking USA participants. The statistically significant differences in group means for error rates and latencies were 2.7 percentage points (95% CI 1.5 to 3.9) and 49 ms (95% CI 18 to 81), respectively, $F(1,55) = 21.38$, $MSE = 4.8$, $p < .001$ (Mann-Whithey U, $p < .001$) and $F(1,55) = 10.18$, $MSE = 3422.30$, $p = .002$, respectively. We also examined the difference in main effects of Interlingual Neighborhood (With Dutch Neighbors vs. No Dutch Neighbors) for the two participant groups, which is equivalent to the interaction effect of participant group and Interlingual Neighborhood. Comparison of the groups showed that the difference in error rates between words like MOON and words like STAIN were larger for the Dutch participants (3.8%) than for the USA participants (0.7%). The difference between these effects was 3.1 percentage points (95% CI 1.3 to 4.8) and statistically significant, $F(1,55) = 12.57$, $MSE = 10.74$, $p < .001$ (Mann-Whithey U, $p < .001$). Further, for the latency analyses, the difference in means between words like MOON and words like STAIN was also larger for the Dutch participants (13 ms) than for the USA participants (11 ms). However, the 2 ms difference (95% CI -4 to 7) was not statistically significant, $F(1,55) = .33$, $MSE = 106.89$, $p = .571$. Finally, the two language groups were compared on the AM > TM contrast for the list of words without Dutch neighbors (e.g., PAID vs. SAID). We explored this pattern because for the Dutch participants particular high error rates were observed for words like PAID. For Dutch participants the difference for the AM > TM contrast was 14.0 percentage points and for the USA participants it was 2.5 percentage points. The difference between these effects was 11.5 percentage points (95% CI 7.0 to 15.9) and statistically significant, $F(1,55) = 26.80$, $MSE = 69.89$, $p < .001$ (Mann-Whithey U, $p < .001$).

Discussion

Experiment 1 investigated the intralingual consistency effect in English word naming. As expected for the native speakers of English, longer naming latencies and more errors were observed for inconsistent words with relatively many English enemies (e.g., BLOOD and PAID) than for inconsistent words with relatively few English enemies (e.g., MOOD and SAID) and for consistent words with no English enemies (e.g., MOON and STAIN). However, contrary to expectation, we found no evidence that naming performance was worse for words with typical mappings than for words with consistent mappings. Overall, the results of the native English speakers support the idea that naming of an English word is influenced by spelling-to-sound knowledge of other English words containing the same spelling body. This replicates previous work on the consistency effect (e.g., Jared et al., 1990), and validates that the inconsistent spelling-to-sound mappings of this specific selection of word materials are reflected in the process of word perception.

Turning to the Dutch-English bilinguals, also for them more errors were observed for inconsistent words with relatively many English enemies than for inconsistent words with relatively few English enemies and for consistent words with no English enemies. Contrary to the native English speakers but in line with expectation, however, for Dutch-English bilinguals more naming errors were observed for words with typical mappings than for words with consistent mappings. Therefore, also for Dutch-English bilinguals word naming is influenced by spelling-to-sound knowledge of English neighbors. This result is in accordance with the observation that French-English bilinguals reading in their second language produce more naming errors for inconsistent English words such as BEAD that have many enemies than for consistent words such as BUMP without enemies (Jared & Kroll, 2001). Thus, in terms of error rates, the native English speakers and the Dutch-English bilinguals performed fairly similar on the word-naming task. There was one notable difference. Compared to the native English speakers, Dutch-English bilinguals had particularly high rates of naming errors for words like PAID that have atypical mappings and no Dutch neighbors. Evidently, for Dutch-English bilinguals, knowledge of spelling-to-sound mappings of inconsistent words that have many English but no Dutch enemy neighbors is fairly poor.

Also similar to the native English speakers, for Dutch-English bilinguals, longer naming latencies were observed for AM-words like BLOOD than for TM-words like MOOD. However, for AM-words like PAID we did not observe longer naming latencies than for TM-words like SAID. This finding, however, should not induce any stringent conclusions. Recall that Dutch-English bilinguals had particularly high rates of naming errors on AM-words like PAID. This may have caused that the more difficult AM items, the ones that are expected to yield long naming latencies, were omitted from the latency analysis. As a result, the average naming latency for AM-words such as PAID may have been deflated. Further, longer naming latencies were observed for AM-words like PAID than for CM-words like STAIN, but there was no evidence for such a difference in case of AM- and CM-words like BLOOD and MOON. These findings are consistent with the observation that reading of French-English bilinguals in their second language produces longer naming latencies for inconsistent English words such as BEAD than for consistent words such as BUMP (Jared & Kroll, 2001). Finally, for Dutch-English bilinguals, the same as for native English speakers, naming latencies for TM-words such as MOOD were unexpectedly shorter than for CM-words such as MOON. It may be the case that, even though the words of the different word types were matched as carefully as possible, the TM-words nevertheless were easier to name due to linguistic factors we did not account for. In sum, although the findings for the Dutch-English bilinguals were not entirely conform predictions, there was, both in terms of naming latencies and in terms of naming errors, strong evidence of an influence of spelling-to-sound knowledge of other English words.

We also explored whether interlingual neighborhood (e.g., words like BLOOD, MOOD, and MOON vs. words like PAID, DAID, and STAIN) had an impact on naming performance of the monolingual English speakers and the Dutch-English bilinguals, albeit the two word lists representing this variable were not properly equated on various word characteristics (see Chapter 2, Table 1). For the native English speakers there was no evidence for such an impact on the number of naming errors, although shorter naming latencies were observed for words with Dutch neighbors. For the Dutch-English bilinguals, however, interlingual neighborhood did have an effect on the number of naming errors, and it was notably larger than for the native English speakers, although the effect can be entirely attributed to the excessive number of naming errors for AM-words such as PAID. Furthermore, for Dutch-English bilinguals, shorter naming latencies were observed for words with Dutch neighbors. Thus both the language groups showed evidence that words with Dutch neighbors were named faster than words without Dutch neighbors. There was however no evidence that the effect was different for the two language groups. Hence, this effect may reflect an advantage that results from a higher overall familiarity of these words or from other linguistic sources. In sum, no evidence was found that the effect of interlingual neighborhood was different for the two language groups. Even in the case of the Dutch-English bilinguals there was no reason to expect this effect, since Jared and Kroll (2001) only obtained evidence for an effect of interlingual neighborhood on second-language English word naming when their participants first named a block of words from their native language.

Finally, for the native English speakers, we found no evidence of differential effects of intralingual consistency effects on naming performance across the two lists containing words with and words without Dutch neighbors. Indeed, this was not expected for the native English speakers because they have no knowledge of Dutch spelling-to-sound relations. In contrast, for Dutch-English bilinguals differential effects of intralingual consistency effects on the number of naming errors were observed. However, this interaction can be entirely attributed to exorbitantly poor naming performance on AM-words like PAID. Also, for the differential effects on naming latencies, the difficult to interpret PAID cell and the unexpectedly poor performance on CM-words like MOON, which was also observed for the native English speakers, precludes a valid interpretation of the interaction.

To conclude, manifold intralingual spelling-to-sound relations have an impact on English word-naming performance of monolingual native English speakers and also of Dutch-English bilinguals. For both language groups, spelling-to-sound knowledge of enemy neighbors hampers naming performance. This observation is consistent with the idea that, in general, manifold associations implicate ambiguity that must be resolved by a time-consuming process. In a resonance framework, this ambiguity involves a competition between local orthographic-phonologic resonances (i.e., local friendly and local enemy attractors). The outcome of this competition depends on the relative self-consistency of intermediate-grain spelling-to-sound associations, which are determined by the relative number and summed frequency of friends and enemies. The results of the Dutch-English bilinguals are also in agreement with the phonological coherence hypothesis, and support a strong phonological view that phonological coding is fundamental not only to monolingual but also to bilingual visual word perception (Brysbaert et al., 1999).

4 When MOOD Rhymes with BLOOD: Intralingual Phonological Coding in English Visual Word Perception

The present study's first objective was to investigate whether the process of English word perception in Dutch-English bilinguals involves mandatory, *intralingual* phonological coding. Evidence for this was found in Chapter 3 (Experiment 1), where an intralingual consistency effect was observed in second-language English word naming. This consistency effect indicates that spelling-to-sound knowledge of English enemy neighbors hinders reading performance, presumably because emerging inappropriate phonology is engaged in competition with appropriate phonology. In Chapter 4 we make an attempt to track the competing dynamics of local intralingual orthographic-phonologic associations in English word perception of Dutch-English bilinguals (Experiments 2-5) and of native speakers of English (Experiment 5). With respect to the initial conditions of perception of inconsistent words, *simultaneous* coding of appropriate and inappropriate phonological structures is assumed to result from manifold intermediate-grain-size spelling-to-sound mappings, and the competition between these codings is assumed to underlie the intralingual consistency effect. For example, in English, the spelling body *–OOD* is pronounced as /ud/ in MOOD but also as /}d/ in BLOOD. Since both /ud/ and /}d/ are previously associated to this spelling body, the initial conditions of perception of *MOOD* is assumed to include appropriate body phonology of MOOD but also inappropriate body phonology of BLOOD. For correct reading, the resulting competition between these phonological codings must be resolved, with the outcome that reading performance may be worse for an inconsistent word such as MOOD than for a consistent word such as MOON.

To recapitulate, the print-to-speech correspondence task does not require explicit word-naming or identification responses, thus the obtained data do not rely on the ultimate outcome of a competition between phonological codings, in terms of naming or identification performance. In the present study, we used the print-to-speech correspondence task to address the first research question whether, both for Dutch-English bilinguals and for native speakers of English, perception of an inconsistent word is influenced by spelling-to-sound knowledge of English enemy words (i.e., the consistency effect) or, stated differently, whether processing of such a word involves competition between appropriate (i.e., *–OOD* - /ud/) and inappropriate (i.e., *–OOD* - /}d/) phonological codings. Such a competition process may be inferred if we can actually demonstrate that processing an inconsistent word involves simultaneous coding of appropriate and inappropriate phonological structures. For this specific purpose, the use of *catch trials* is of critical importance. In a catch trial, an inconsistent word (e.g., MOOD) is simultaneously presented with a spoken rime that is derived from an *enemy* of the word (e.g., /}d/, derived from BLOOD). The catch trial distinguishes itself from the other trial types in that it may enable us to mark local phonological codings that were launched at the start of word presentation, but have been inhibited in the course of word processing. Thus if, in a catch trial, MOOD is accompanied with the spoken rime /}d/ (derived from BLOOD), and if inappropriate phonology (i.e., /}d/) is actually part of the initial conditions of perception of MOOD, this spoken rime may restore the degraded, inappropriate coding to such a degree that participants may find it difficult or are actually unable to perceive a mismatch. That is, the spoken rime may put the degraded, inappropriate coding back into a full-blown competition. Consequently, participants may occasionally react with an incorrect "yes" response (i.e., a false-positive), thus indicating that they perceived MOOD's phonology to rhyme with the rime of BLOOD. In general, evidence for simultaneous phonological coding provides strong support for the view that the process of English word perception in Dutch-English bilinguals involves mandatory phonological coding.

In Experiment 2, the list of English words with Dutch neighbors (e.g., MOON) was used for match trials and the list of English words without Dutch neighbors (e.g., STAIN) was used for no-match and catch trials. In all experiments that followed this was reversed. In Experiment 3-8, the list of English words without Dutch neighbors (e.g., STAIN) was used for the match trials and the list of English words with Dutch neighbors (e.g., MOON) was used for no-match and catch trials.

GENERAL METHOD

Additional Materials for Experiments 2-8

Selection of sound stimuli

The sound stimuli used in Experiments 2-8 consisted of 320 speech units (i.e., spoken rimes) with a vowel-consonant (VC) structure (e.g., /ud/). These were obtained by having a native English speaker read aloud lists of words and pseudowords. The resulting pronunciations were all recorded, and subsequently modified such that of each letter string's spoken form (e.g., /mud/) only the *rime* remained (e.g., /ud/). Note that we used only the rimes of the spoken words and not the entire words. The reason for this was that in the print-to-speech correspondence task, printed words could be accompanied with either a congruent (i.e., correctly pronounced) or an incongruent (i.e., incorrectly pronounced) sound stimulus. This implies that if entire spoken words would be used, match trials would contain word pronunciations while mismatch trials would contain nonword pronunciations.

Consequently, if such was the case, participants would be able to base their binary "yes" or "no" decision entirely on the lexicality of the sound stimulus, that is, whether the aurally presented speech was a word or not. Therefore, the sound stimulus should not provide hints for correct responding on the basis of its lexicality. Using only the rimes of the spoken words accomplished this goal.

Coupling printed words to spoken rimes

For the print-to-speech correspondence task, three types of trials were created: Match trials, no-match trials, and catch trials. All trial types involved presentation of both a printed word and a spoken rime. Table 9 presents typical examples of the three trial types. *Match trials* consisted of a printed word and a spoken rime that were congruent with each other (and hence required a "yes" response). This congruency was accomplished by using the actual rime of the printed word's spoken form. An example of such a trial is when the target word MOOD is accompanied by the spoken rime /ud/ (derived from the word MOOD). No-match trials and catch trials consisted of a printed word and a spoken rime that were not congruent with each other (and hence required a "no" response). For *no-match trials* this incongruency was accomplished by using the rime of an unrelated spoken word. Specifically, this unrelated word (e.g., BRIDE) shared only the coda (i.e., final consonant) with the target word (e.g., /mud/ – /brYd/). An example of such a trial is when the target word MOOD is accompanied by the spoken rime /Yd/ (derived from the word BRIDE). Finally, incongruency was accomplished for *catch trials* by using the rime of the spoken form of an *enemy* of the target word (e.g., BLOOD). Likewise, this enemy word shared only the coda with the target word (e.g., /mud/ – /bl}d/). An example of such a trial is when the target word MOOD is accompanied by the spoken rime /}d/ (derived from the word BLOOD).

Table 9.
Examples of trial types for Experiments 2-5 (print-to-speech correspondence task). (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

Yes-trials				**No-trials**			
Word Type	*Example*	Match Sound	Analogy	No-Match Sound	Analogy	Catch Sound	Analogy
CM	*MOON*	/un/	"moon"	/en/	"vein"	-	-
TM	*MOOD*	/ud/	"mood"	/Yd/	"bride"	/}d/	"blood"
AM	*BLOOD*	/}d/	"blood"	/Yd/	"bride"	/ud/	"mood"

Recording and modification of sound stimuli

The basis for the sound stimuli (i.e., spoken rimes) of Experiments 2-8 were words and pseudowords spoken by a female British-English native speaker and digitally recorded on an Apple Macintosh PowerPC 4400/200 using an EAGLE GMM 400 mixer and a YOGA EM240 microphone. Speech was digitised at a sampling rate of 44.1 kHz and 16-bit analog-to-digital conversion with the SoundEdit™ 16 software. The English native speaker was born and raised in the United Kingdom near London and had immigrated to The Netherlands as an adult. At the time of the recordings she had resided in The Netherlands for almost 25 years, and as such she had substantial fluency in Dutch. This was a prerequisite for pronouncing some of the letter strings. Recordings took place in an acoustically shielded room. The English speaker was seated behind a desk and presented with 16 printed lists that consisted of letter-string sequences (e.g., MOOD – BLOOD; MOOD – BRIDE – MIDE). She was instructed to read aloud the letter strings in “common” (i.e., British) English. In all, 968 letter strings were pronounced and recorded.

For one set of lists, each sequence consisted of two English words in a row with identical spelling bodies but with dissimilar phonological bodies (e.g., MOOD – BLOOD). With instructions to keep intonation and pace steady across sequences and recording sessions, the English speaker read aloud the first word and then the second word *to rhyme with the first*. Thus, the second word was purposely mispronounced analogously to the first word. For example, the first word of the sequence MOOD – BLOOD was properly pronounced /mud/, and the second word faulty pronounced /blud/. In a following recording session, the order of words was reversed so that eventually for both words a correct and an incorrect pronunciation was obtained. The correct pronunciation was used for match trials (e.g., MOOD - /ud/) and the incorrect pronunciation was used for catch trials (e.g., MOOD - /}d/).

For another set of lists, each sequence consisted of a row of three letter strings; two English words and one pseudoword (e.g., MOOD – BRIDE – MIDE). The first two words differed in orthographic and phonologic form but shared the coda (i.e., the final consonant, /d/). In contrast, the pseudoword and the first word of the sequence (e.g., MOOD, MIDE) were equivalent in onset (/m/) and coda (/d/), but differed in the nucleus (/u/ vs. /Y/). Again, the English speaker held intonation and pace steady and read aloud the first word followed by the second, and then the pseudoword analogously to the second word. For the first word a correct pronunciation was obtained. The second word served as a model for pronouncing the pseudoword. Hence, reading aloud the pseudoword created a mispronunciation of the *first* word that was analogous to a specific, unrelated word. For example, the first word of the sequence MOOD - BRIDE - MIDE was properly pronounced /mud/, the second word was properly pronounced /brYd/, and the pseudoword was pronounced /mYd/, thus yielding a mispronounced MOOD, analogously to the word BRIDE. Again, the correct pronunciation was used for match trials (e.g., MOOD - /ud/) and the incorrect pronunciation was used for no-match trials (e.g., MOOD - /Y}d/). In sum, all spoken rimes that were used in match trials originated from words that were pronounced correctly, and all spoken rimes that were used in mismatch trials originated from words and pseudowords that were purposely mispronounced.

Ultimately, 716 accurately pronounced letter strings served as basis for the sound stimuli. These were all manually modified using the waveform editor of SoundEdit™ 16 and a set of SONY MDR-V100 headphones. Of each spoken word the onset was removed. This was accomplished by successively removing small sections (varying in the range 1 - 20 ms) of a word's speech signal until the onset was inaudible and a distinct spoken rime (i.e., phonological body) remained. Finally, each spoken rime was converted into an Apple Macintosh System 7 sound file, which was an appropriate format for the experiment generator.

Apparatus and Procedure for Experiments 2-8

The printed word stimuli were displayed in lowercase letters in the centre of the computer screen of an Apple Macintosh PowerPC 4400/200, using a standard Macintosh font (Geneva, size 14). The screen (Apple Multiple Scan 15 Display) was set to a refresh-rate of 75-Hz. Presentation of the visual stimuli was synchronised with the refresh cycle of the screen. The sound stimuli (i.e., spoken rimes) were presented at a comfortable level through a conventional set of headphones. Error rates and response latencies were collected by means of two button boxes interfaced to the serial ports of the computer. Stimulus presentation and data recording were controlled by the computer program fLexi (version 3.3.8). Prior to data collection, the software was extensively tested within the experimental set-up of the print-to-speech correspondence task.

Participants were tested individually in a quiet and normally lit room. They were seated at approximately 50 cm in front of the computer screen and were given verbal and written instructions, followed by a block of practise trials. After having completed the practise trials, the participants were given the opportunity to ask questions concerning the experimental task. When they indicated that the procedure was clear, the experimental session was initiated. An experimental session involved a series of experimental trial blocks. The first two trial blocks were preceded by, depending on the specific experiment, a block of 10 (Experiments 4, 5, and 7), or 30 (Experiments 2, 3,6, and 8) practise trials. The order of trials was randomised for each participant and for each trial block.

Figure 5 illustrates the typical sequence of events for a trial. Each trial started with a fixation region that was displayed in the centre of the computer screen. The fixation region was composed of two horizontal dashes that extended approximately 2.1 cm. It remained on the screen for approximately 500 ms and was then immediately replaced by a printed word. The printed word was displayed for approximately 200 ms and was in turn immediately replaced by a pattern mask that remained on the screen for another 1.8 sec, or until the participant responded. The pattern mask consisted of rows of chequered black and white squares, extending a plane of approximately 2.5 by 1.0 cm that completely covered the printed word.

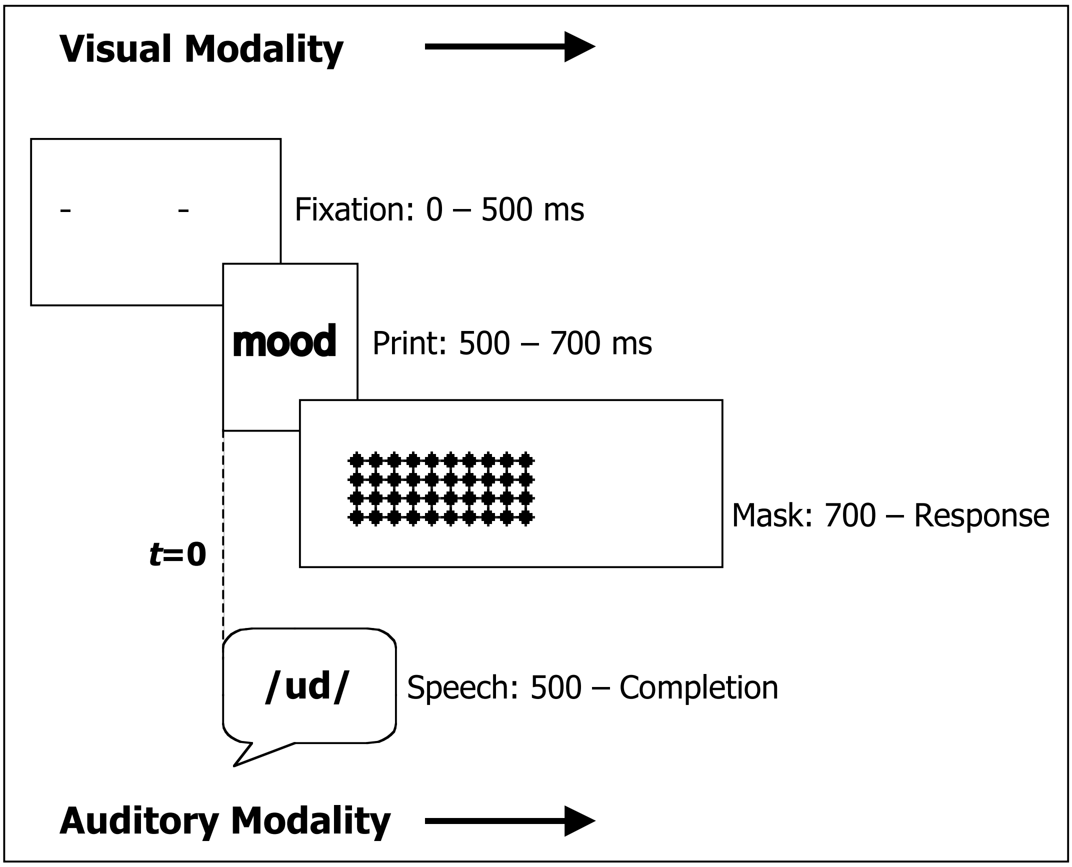

Figure 5. Sequence of events of a typical trial in the print-to-speech correspondence task.

Sound stimuli were presented simultaneously with the onset of the printed word stimuli in one series of experiments (Experiments 2, 3, 4, 6, and 8) and in another series (Experiments 5 and 7) either before (SOA = -500 ms), simultaneously (SOA = 0 ms), or after the onset (SOA = 500 ms). The printed word stimuli were coupled to different sound stimuli to create two basic kinds of trials, "yes" trials (i.e., match trials) and "no" trials (i.e., no-match and catch trials). In "yes" trials the phonological body of the visually presented printed word corresponded with the aurally presented spoken rime (e.g., MOOD – /ud/), and in "no" trials they did not correspond (e.g., MOOD – /Yd/; MOOD – /}d/). Participants were instructed to decide as quickly and accurately as possible whether print and speech were congruent with each other by pressing either the "yes" or "no" button, using the index fingers. They all used the index finger of the dominant hand for "yes" responses. After the participant made a response, immediate visual feedback was provided which replaced the pattern mask. The feedback was displayed for approximately 750 ms. If the participant pressed the correct button the phrase “correctly responded” appeared approximately 2.8 cm to the right of the centre of the screen; if not, the phrase “wrong” appeared. If the participant failed to respond within 2000 ms after onset of the printed word, the trial was cancelled and the word “slowly…” appeared. Irrespective of the participants’ response, the total duration of a trial was approximately 2500 ms. The next trial was initiated after an interstimulus interval of approximately 1000 ms.

EXPERIMENT 2

MISMATCH-TRIAL PERFORMANCE ON ENGLISH WORDS WITHOUT DUTCH NEIGHBORS

In Experiment 2 we used the print-to-speech correspondence task to investigate whether in Dutch-English bilinguals processing of an inconsistent English word (e.g., SAID) involves simultaneous coding of appropriate (i.e., /Ed/) and inappropriate (i.e., /ed/, as in PAID) intermediate-grain-size phonological structures. In word perception, multiple phonological codings are assumed to give rise to competition, of which the outcome may be observed in an intralingual consistency effect.

Foremost, Experiment 2 intended to find out whether, in the print-to-speech correspondence task, manifold intralingual spelling-to-sound relations have an impact on match-trial and no-match-trial performance. That is, it examined whether perceiving a match or a mismatch between an English word and a spoken rime is influenced by spelling-to-sound knowledge of English enemy neighbors. The ratio of friends and enemies determines the degree of (in)consistency of a spelling body mapping onto a phonological body, which, in turn, predicts performance on match trials and no-match trials. If on a trial a printed word causes a correct phonological structure to emerge quickly, participants may be able to perceive relatively early a match or mismatch between the printed word and the spoken rime.

With respect to the three word types (AM, TM, and CM) the predictions for the match and no-match trials have the same direction as the predictions for the word-naming task. For match trials, task performance is measured in correct “yes” latencies and percentages of errors (i.e., false-negatives), whereas for no-match trials, task performance is measured in correct “no” latencies and percentages of false-positives. Again, it is expected that when a word has English enemies, task performance is worse than when it has no enemies. Furthermore, when a word has more enemies than friends task performance is expected to be worse than in the opposite case, when a word has more friends than enemies. These predictions can also be stated as: AM > TM, AM > CM, and TM > CM.

In general, we do not expect that participants will produce large numbers of false-negative and false-positive errors on match trials and no-match trials, respectively, because the task at hand appears not too difficult. That is, participants should be quite capable perceiving a match or mismatch between a printed word and a spoken rime. Nevertheless, we expect participants to make errors occasionally, in particular for AM-words such as BLOOD and PAID, because spelling-to-sound associations of AM-words are assumed to be unstable, resulting in noisy phonological codings. Therefore, for AM-words, congruent and incongruent spoken rimes may not readily be accepted and rejected, respectively. On the contrary, if a phonological coding persists in staying unstable and takes various forms, in particular for non-native speakers of English, congruent speech (e.g., /}d/, for BLOOD) may occasionally not be perceived as such, and incongruent speech (e.g., /Yd/, for BLOOD) may, occasionally, actually be considered as congruent.

The specific question whether processing of an inconsistent word involves simultaneous coding of appropriate and inappropriate phonological structures is addressed by comparing performance on catch trials and no-match trials, separately for both types (AM and TM) of inconsistent words (e.g., PAID and SAID). In all analyses that included catch trials, consistent words were disregarded because for this type of words it is not possible to create a catch trial. Evidence for simultaneous coding of appropriate and inappropriate phonological structures may be found if in catch trials the spoken rimes appear to foster the degraded, inappropriate local phonological codings to such a degree that participants are relatively slow to perceive a mismatch and/or produce a relatively large number of false-positives. In no-match trials, inappropriate phonological codings are not fostered by the spoken rimes, because these sound stimuli are derived from unrelated words. Therefore, it is expected that, in terms of correct “no” latencies and percentages of false-positives, performance on catch trials is worse than performance on no-match trials, which can also be stated as: catch > no-match (i.e., a Trial Type effect).

Again, we distinguish between inconsistent words with atypical mappings and inconsistent words with typical mappings, AM-words and TM-words. A local, inappropriate strong-rule phonological coding (e.g., /Ed/) that has been inhibited for an AM-word such as PAID may be more readily restored by a fostering sound stimulus than a less strong local, inappropriate phonological coding (e.g., /ed/) for a TM-word such as SAID, especially when the correct local orthographic-phonologic association is weak (e.g., [*–AID* - /ed/] in PAID). This implies that catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping. This Word Type effect on catch-trial performance can also be stated as: catch (AM-words) > catch (TM-words). Furthermore, this Word Type effect may be larger than in case of the contrast no-match (AM-words) > no-match (TM-words), because, for no-match trials, the inappropriate phonological codings are not fostered by spoken rimes, whereas in catch trials this is expected to occur, in particular for AM-words. Therefore, the Trial Type effect (i.e., catch > no-match) is expected to be larger for AM-words than for TM-words. This can also be stated as: Trial Type effect (AM-words) > Trial Type effect (TM-words). Notice that this expectation of an interaction effect can also be stated in terms of a difference in Word Type effect for catch trials and no-match trials. In either case, in terms of a resonance framework, this interaction effect may be expected if a spoken rime that is derived from a strong enemy indeed has the capacity to restore an inappropriate orthographic-phonologic resonance in a word such as PAID to such a degree that dynamics pull PAID's phonology close to that of a word such as SAID. These strong dynamics are not expected to occur in the case of a no-match trial, because the unrelated spoken rime does not have such a strong capability to foster inappropriate codings. In no-match trials, dynamics pull PAID's phonology rhyming to that of SAID only in the early phases of word processing, where competition is solved relatively fast and incorrect phonology is quickly inhibited and subdued by correct phonology.

In sum, for *match trials* we expected the contrasts AM > TM, AM > CM, and TM > CM. In other words, we expected that performance on a trial like BLOOD - /}d/ is worse than on a trial like MOOD - /ud/, of which the performance is worse than on a trial like MOON - /un/. For *no-match trials* we expected the same contrasts, thus performance on a trial like PAID - /id/ (with the spoken rime derived from the unrelated word BEAD) is worse than on a trial like SAID - /id/, of which the performance is worse than on a trial like STAIN - /}n/ (with the spoken rime derived from the unrelated word SUN). Furthermore, for the comparisons of *no-match trials* and *catch trials* we expected the contrast catch > no-match. Thus, performance on trials like PAID - /Ed/ and SAID - /ed/ is worse than on trials like PAID - /id/ and SAID - /id/. We also expected the contrast catch (AM-words) > catch (TM-words), or, in other words, performance on trials like PAID - /Ed/ is worse than on trials like SAID - /ed/. Finally, we expected a similar contrast: Trial Type effect (AM-words) > Trial Type effect (TM-words). Thus, the difference in performance on trials like PAID - /Ed/ and PAID - /id/ is larger than the difference in performance on trials like SAID - /ed/ and SAID - /id/. This set of planned contrasts is shown graphically in Figure 6.

Method

Participants and materials

A group of 60 Dutch-English bilinguals served as participants (see Chapter 2). They were presented with the 120 printed English words described in Chapter 2 and the Method section of Chapter 3, and the 120 spoken rimes described in the General Method section of Chapter 4.

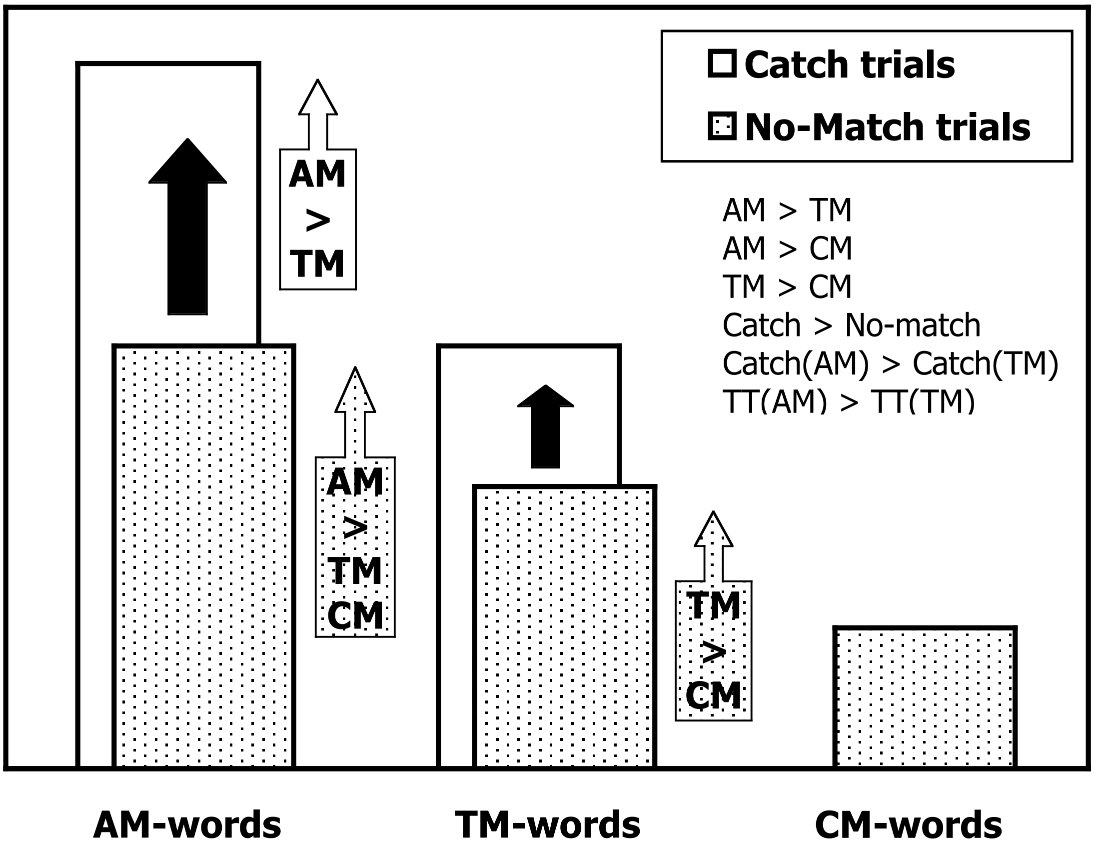

Figure 6. Graphical presentation of planned contrasts for Experiments 2-8

Experimental design

In Experiment 2 and further, three basic types of English words were contrasted: AM-words, TM-words, and CM-words. These words were grouped in two separate word lists. In the first list, the English words had Dutch neighbors (e.g., BLOOD, MOOD, and MOON) whereas in the second list they did not have Dutch neighbors (e.g., PAID, SAID, and STAIN). Both word lists contained an equal number of words for each group of words representing one of the three word types. In Experiment 2, the words of the first list (e.g., MOON) were used for match trials, whereas the words of the second list (e.g., STAIN) were used for no-match and catch trials, and in the Experiments 3-8 this was the other way round. Eight groups of trials were created that represented specific combinations of Trial Type and Word Type. These eight combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM).

In this experiment and the following ones, the spelling body of an English word with a typical mapping (e.g., SAID, MOOD) also appeared in an English word with an atypical mapping (e.g., PAID, BLOOD). Thus, participants were presented with the same spelling body twice. To prevent intralist-priming effects of spelling bodies, the two words containing the same spelling body were presented in separate blocks of trials.

For the match trials, printed words (e.g., BLOOD, MOOD, and MOON) were coupled to spoken rimes that were congruent with it. Because in this trial type each printed word could only be coupled to one type of spoken rime (e.g., MOOD - /ud/), each match trial contained a unique printed word. This, however, was not the case for the set of trials that required a "no" response, that is, the no-match and catch trials. For these mismatch trials, inconsistent words (e.g., PAID, and SAID) were coupled to spoken rimes that were incongruent with it. Specifically, each inconsistent word could be coupled to a spoken rime to create a no-match trial (e.g., SAID - /id/), but also to a spoken rime to create a catch trial (e.g., SAID - /ed/). In Experiments 2 and 3, each inconsistent word that was assigned to the mismatch trials was used once in a no-match trial and once in a catch trial.

To control for repetition-priming effects of printed words, the two trial types containing the same printed word were presented in separate blocks of trials. As a result, there were two constraints that required trials to appear in separate blocks of trials. One, spelling bodies were not to be presented within the same trial block. Two, identical printed words were not to be presented within the same trial block. Consequently, each spelling body (e.g., *–AID*) that was part of a particular inconsistent word (e.g., PAID) appeared in four blocks of trials, Trial Blocks A, B, C, and D. For example, in block A this spelling body appeared in the word PAID, together with the spoken rime /id/ to create a no-match trial. In block B it appeared in the word SAID, together with the spoken rime /ed/ to create a catch trial. In block C, it appeared again in the word PAID, together with the spoken rime /Ed/, and this time to create a catch trial. Finally, in block D, it appeared again in the word SAID, together with the spoken rime /id/, and this time to create a no-match trial. Thus, in short, the spelling body *–AID* appeared in the trial blocks A, B, C, and D respectively as: PAID (no-match trial), SAID (catch trial), PAID (catch trial), and SAID (no-match trial).

Table 10 presents the layout of the experimental design. The four trial blocks contained equal numbers of words from all three word types. Again, this was accomplished by separating each list of 20 words comprising one of the three word types (CM, TM, and AM) into two sub word-lists, each containing 10 words (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2). For example, if the typical word GREY (from TM_1) was presented in Trial Block A (in a catch trial), then the atypical word KEY (from AM_1) was presented in Trial Block B (in a no-match trial). The typical word GREY (from TM_1) also appeared in Trial Block C (in a no-match trial) and the atypical word KEY (from AM_1) also appeared in Trial Block D (in a catch trial). Conversely, if the atypical word PAID (from AM_2) was presented in Trial Block A (in a no-match trial), then the typical word SAID (from TM_2) was presented in Trial

Block B (in a catch trial). The atypical word PAID (from AM_2) also appeared in Trial Block C (in a catch trial) and the typical word SAID (from TM_2) also appeared in Trial Block D (in a no-match trial). Therefore, in each mismatch block, there were 10 catch trials and 10 no-match trials for inconsistent words, and 10 no-match trials for consistent words. In general terms, Trial Block A and C each comprised sub-lists CM_1, TM_1, and AM_2, and Trial Block B and D each comprised sub-lists CM_2, TM_2, and AM_1. If the words from a sub-list were used to create no-match trials in block A, then the same words were used in block C to create catch trials. Conversely, if the words from a sub-list were used to create catch trials in block A, then the same words were used in block C to create no-match trials. Further, if the words from a sub-list were used to create no-match trials in block B, then the same words were used in block D to create catch trials. Conversely, if the words from a sub-list were used to create catch trials in block B, then the same words were used in block D to create no-match trials.

In sum, for the word-list without Dutch neighbors (e.g., SAID), used for the mismatch trials, each of the blocks A, B, C, and D contained 10 words of each word type (AM, TM, and CM) from either the sub-lists containing, for instance, KEY, SAID, and STAIN, or the sub-lists containing, for instance, PAID, GREY, and CAPE. These sub-lists appeared in the trial blocks A, B, C, and D; Trial Block A: PAID (no-match trial), GREY (catch trial), CAPE (no-match trial); Trial Block B: KEY (no-match trial), SAID (catch trial), STAIN (no-match trial); Trial Block C: PAID (catch trial), GREY (no-match trial), CAPE (no-match trial); Trial Block D: KEY (catch trial), SAID (no-match trial), STAIN (no-match trial).

For the match trials (using words such as BLOOD, MOOD, and MOON), Trial Block A comprised sub-lists CM_1, TM_1, and AM_2 and Trial Block B comprised sub-lists CM_2, TM_2, and AM_1 (see Table 10). Trial Blocks C and D were identical to Trial Blocks A and B, respectively.

In Experiments 2 and 3, each participant was presented with each of the four trial blocks. Hence, for each participant data was obtained for each possible combination of trial type and word type, using each word twice. That is, repeated measures were obtained for the participants, intended to increase the statistical power and precision of the experiment by isolating residual variance due to individual differences between participants in response latencies and error rates. The temporal order of all four trial blocks was counterbalanced across four different participant groups according to a single Latin square (participant group 1: Sequence A-B-C-D; participant group 2: Sequence B-A-D-C; participant group 3: Sequence C-D-A-B; participant group 4: Sequence D-C-B-A). Participants were randomly assigned to the different sequences (see Table 11).

Table 10 shows that words used in the blocks A and B also appeared in the blocks C and D, but, for mismatch trials, they were coupled to different spoken rimes. Thus, across the two sets of blocks, the no-match trials (e.g., PAID - /id/) were converted into catch trials (e.g., PAID - /Ed/), and the catch trials (e.g., GREY - /i/) were converted into no-match trials (e.g., GREY - /o/), with the result that all possible combinations of printed word and spoken rime were realised. Therefore, data obtained from performance on blocks C and D can be conceived as both a between-participants and a within-participants replication of data obtained from performance on blocks A and B, and visa versa. However, because presenting the same printed words twice to the participants may affect the data in complicated ways, we also analysed the data separately for the blocks presented in the first two temporal positions (Positions 1 and 2, *primary test*), and the blocks presented in the second two temporal positions (Positions 3 and 4, *replication*). The primary (i.e., original) test comprised blocks A and B for participant groups 1 and 2, and blocks C and D for participant groups 3 and 4. Conversely, the replication comprised blocks C and D for participant groups 1 and 2, and blocks A and B for participant groups 3 and 4. The data obtained from Positions 1 and 2 were considered the most valid, because the trial blocks in these positions contained unique words, although we also performed analyses using the full data set. Furthermore, we used the same precautionary procedure as in Experiment 1 to prevent intralist-priming effects of spelling bodies. The two separate blocks of trials containing the same spelling bodies (blocks A and B for participant groups 1 and 2, and blocks C and D for participant groups 3 and 4) were administered in two separate experimental sessions conducted one week apart. The other two trial blocks, administered in Positions 3 and 4, immediately followed the trial block administered in Position 2 (see Table 12).

In Positions 1 and 2, the blocks A and B (for participant groups 1 and 2), and the blocks C and D (for participant groups 3 and 4) contained different stimulus couples that were processed by different participants. These differences between participants and stimulus couples possibly induce additional variance in the statistical model. However, this potential source of variance can be isolated and removed from the estimate of error variance, which may improve the efficiency of the design. Likewise, In Positions 3 and 4, the blocks C and D (for participant groups 1 and 2), and the blocks A and B (for participant groups 3 and 4) also contained different stimulus couples that were processed by different participants. This potential source of variance can also be isolated and removed from the estimate of error variance. Again, this was accomplished by adding participant group (Sequence A-B-C-D vs. Sequence B-A-D-C vs. Sequence C-D-A-B vs. Sequence D-C-B-A) as a between-subjects variable in an ANOVA, and testing the effects against the resulting treatments×participants(group) error term. Nevertheless, the procedure may be unnecessary for analyses involving the full data set, because all participants received the other two (replication) trial blocks in the second experimental session.

Table 10.
Experimental design for Experiments 2 and 3 (print-to-speech correspondence task). In Experiment 2 English words with Dutch neighbors are used for Yes-trials, and English words without Dutch neighbors for No-trials. Conversely, in Experiment 3 English words with Dutch neighbors are used for No-trials, and those without Dutch neighbors for Yes-trials. The word lists comprising each word type are separated into two sub word-lists (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2). Combinations of word type and trial type (Match, No-Match, English Catch) are systematically distributed over four trial blocks (A, B, C, and D). (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

No-trials

Trial Block	A	B	C	D
Sub Word-List	CM_1	CM_2	CM_1	CM_2
Trial Type	No-Match	No-Match	No-Match	No-Match
Example Word	***MOON***	***SEEM***	***MOON***	***SEEM***
Sub Word-List	TM_1	TM_2	TM_1	TM_2
Trial Type	English Catch	No-Match	No-Match	English Catch
Example Word	***MOOD***	***SCAN***	***MOOD***	***SCAN***
Sub Word-List	AM_2	AM_1	AM_2	AM_1
Trial Type	No-Match	English Catch	English Catch	No-Match
Example Word	***SWAN***	***BLOOD***	***SWAN***	***BLOOD***

Yes-trials

Trial Block	A	B	C	D
Sub Word-List	CM_1	CM_2	CM_1	CM_2
Trial Type	Match	Match	Match	Match
Example Word	***MOON***	***SEEM***	***MOON***	***SEEM***
Sub Word-List	TM_1	TM_2	TM_1	TM_2
Trial Type	Match	Match	Match	Match
Example Word	***MOOD***	***SCAN***	***MOOD***	***SCAN***
Sub Word-List	AM_2	AM_1	AM_2	AM_1
Trial Type	Match	Match	Match	Match
Example Word	***SWAN***	***BLOOD***	***SWAN***	***BLOOD***

Table 11.
Trial block ordered by block position according to a single standard Latin square for Experiments 2 and 3 (print-to-speech correspondence task). The temporal order of trial block is Latin-square counterbalanced across four different participant groups (Participant Group 1: Sequence A-B-C-D; Participant Group 2: Sequence B-A-D-C; Participant Group 3: Sequence C-D-A-B; Participant Group 4: Sequence D-C-B-A).

	Temporal Position of Trial Block			
Sequence	P1	P2	P3	P4
S1	A	B	C	D
S2	B	A	D	C
S3	C	D	A	B
S4	D	C	B	A

Table 12.
Trial block ordered by block position for No-trials for Experiments 2 and 3 (print-sound matching). (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

Participant Group 1: Sequence A-B-C-D

Block Position	P1	P2	P3	P4
Sub Word-List	CM_1	CM_2	CM_1	CM_2
Trial Type	No-Match	No-Match	No-Match	No-Match
Sub Word-List	TM_1	TM_2	TM_1	TM_2
Trial Type	English Catch	No-Match	No-Match	English Catch
Sub Word-List	AM_2	AM_1	AM_2	AM_1
Trial Type	No-Match	English Catch	English Catch	No-Match

Participant Group 2: Sequence B-A-D-C

Block Position	P1	P2	P3	P4
Sub Word-List	CM_2	CM_1	CM_2	CM_1
Trial Type	No-Match	No-Match	No-Match	No-Match
Sub Word-List	TM_2	TM_1	TM_2	TM_1
Trial Type	No-Match	English Catch	English Catch	No-Match
Sub Word-List	AM_1	AM_2	AM_1	AM_2
Trial Type	English Catch	No-Match	No-Match	English Catch

Table 12. (*continued*)

Participant Group 3: Sequence C-D-A-B

Block Position	P1	P2	P3	P4
Sub Word-List	CM_1	CM_2	CM_1	CM_2
Trial Type	No-Match	No-Match	No-Match	No-Match
Sub Word-List	TM_1	TM_2	TM_1	TM_2
Trial Type	No-Match	English Catch	English Catch	No-Match
Sub Word-List	AM_2	AM_1	AM_2	AM_1
Trial Type	English Catch	No-Match	No-Match	English Catch

Participant Group 4: Sequence D-C-B-A

Block Position	P1	P2	P3	P4
Sub Word-List	CM_2	CM_1	CM_2	CM_1
Trial Type	No-Match	No-Match	No-Match	No-Match
Sub Word-List	TM_2	TM_1	TM_2	TM_1
Trial Type	English Catch	No-Match	No-Match	English Catch
Sub Word-List	AM_1	AM_2	AM_1	AM_2
Trial Type	No-Match	English Catch	English Catch	No-Match

Table 13.
Primary and secondary combinations of print and sound over successive trial blocks for Experiments 2 and 3 (print-to-speech correspondence task). (P = period; PSC = print-sound combination.)

First test session (Period 1)		Second test session (Periods 2, 3, 4)			
Primary presentation of word stimuli		Secondary presentation of word stimuli (new print-sound combination)			
Period	P1	P2		P3	P4
PSC1	A	B	PSC2	C	D
PSC1	B	A	PSC2	D	C
PSC2	C	D	PSC1	A	B
PSC2	D	C	PSC1	B	A

Results

The Results sections of the experiments presented in Chapters 4 and 5 are all build around the same basic narrative structure. This was done to improve comparisons across experiments but also to advance multi-level reading. That is, the Results sections can be read at different levels of description and statistical inference. A typical Results section first explains how participants' summary statistics were calculated across levels of the independent and procedural variables. Next, the filtering criteria for excluding erroneous responses and spurious observations are described. The section continues with presenting the relevant descriptive statistics in tables (sample means with accompanying standard errors), separately for match trials and mismatch trials, both for error rates and response latencies. Inquisitive readers can inspect these tables in Appendix C. At this point, the reader is referred to one or more figures that provide a comprehensive graphical summary of the primary findings. Each figure is discussed with respect to a set of predictions that is listed in each experiment's introductory section. The reader is further referred to the accompanying ANOVA table that provides omnibus null hypothesis significance tests of main and interaction effects, augmented with effect sizes (partial eta squared, η_p^2). The main predictions are then evaluated by means of sets of pairwise comparisons that involve planned contrasts. The estimated differences (with confidence intervals) for these principal comparisons are printed in boldface.

We now carry on with the results of Experiment 2. Participants were, as mentioned earlier, presented with eight groups of trials representing specific combinations of Trial Type and Word Type. These eight combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). Each participant responded to 60 match trials in the first two blocks (i.e., the primary test) and 60 identical match trials in the second two blocks (i.e., the replication). These 60 trials consisted of 20 AM-words (e.g., BLOOD), 20 TM-words (e.g., MOOD), and 20 CM-words (e.g., MOON). In addition, each participant responded to 40 no-match trials and 20 catch trials in the first two blocks (i.e., the primary test) and to 40 no-match trials and 20 catch trials in the second two blocks (i.e., the replication). The no-match trials consisted of 10 AM-words (e.g., PAID), 10 TM-words (e.g., SAID), and 20 CM-words (e.g., STAIN), and the catch trials consisted of 10 AM-words (e.g., PAID) and 10 TM-words (e.g., SAID). For each participant, the correct response latencies within these eight groups were averaged, separately for the first two blocks (i.e., primary test) and for the second two blocks (i.e., replication). In addition, for each participant, percentage of errors was calculated for each of the eight groups, also separately for the first two blocks and for the second two blocks. These error rates concerned percentages of false-negatives (i.e., incorrectly pressing the "no" button) for the match trials and percentages of false-positives (i.e., incorrectly pressing the "yes" button) for the mismatch trials. Hence, the data that entered the statistical analyses consisted, for each participant and separately for the first and second two blocks, of three latency means and three percentages of false-negatives for match trials (e.g., for BLOOD, MOOD, and MOON), three latency means and three percentages of false-positives for no-match trials (e.g., for PAID, SAID, and STAIN), and two latency means and two percentages of false-positives for catch trials (e.g., for PAID and SAID).

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. This resulted in a rejection of a total of 30.3% for the catch trials, 6.5% for the no-match trials, and 3.3% for the match trials. Furthermore, less than 0.1% of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. A trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. In the analyses, this experimental procedure resulted in a cut-off that rejected all latencies greater than 2000 ms. We did not consider further truncation, because the procedure resulted in rejection of 0.3% of the correct response latencies and, following recommendations of Ulrich and Miller (1994), it is undesirable to exceed this percentage.

Error data of match trials

A participant produced an error (i.e., a false-negative) in a match trial when he or she pressed the "no" button when presented with a printed word and a spoken rime that were actually congruent with each other. The mean percentages of false-negatives as a function of Word Type, separately for each trial block and participant group, are presented in Table 1 of Appendix C. Figure 7 shows the mean percentages of false-negatives for the AM-words, TM-words, and CM-words, both for the primary test and the (within-participants) replication. These mean percentages of false-negatives were collapsed over the procedural variables trial block and participant group.

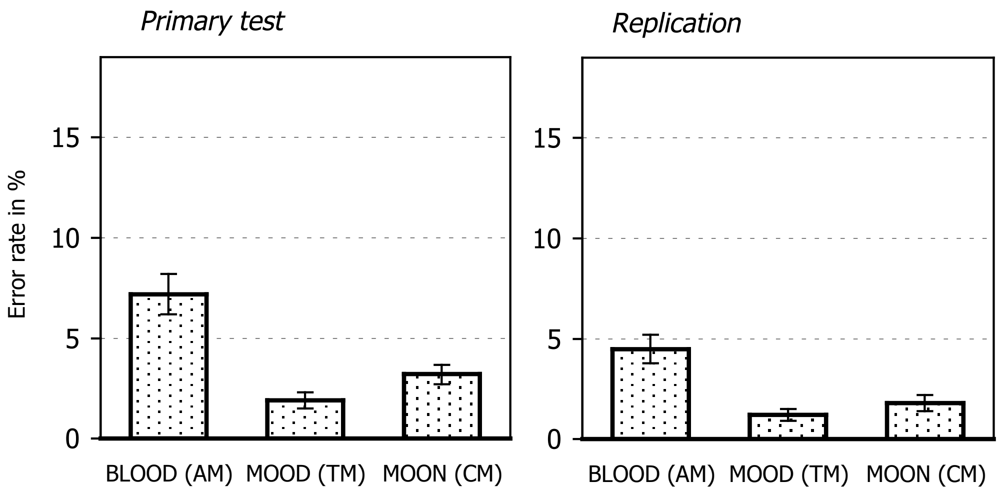

Figure 7. Mean percentages of false-negatives as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 2. Error bars represent the standard error of the mean.

Recall that for inconsistent words such as BLOOD and MOOD, simultaneous coding of appropriate and inappropriate phonological structures is assumed to result from manifold intermediate-grain-size spelling-to-sound mappings, and these local phonological codings are assumed to give rise to competition. The time course and outcome of this competition is predicted by the relative self-consistency of these codings, which is indicated by the ratio of English friends and enemies. Thus, match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. Specifically, words with atypical mappings (e.g., BLOOD) should induce more errors than words with typical mappings (e.g., MOOD) and consistent mappings (e.g., MOON). We also expected more errors for words with typical mappings than for words with consistent mappings. As can be seen in Figure 7, the ratio of friends and enemies was indeed associated with the number of false-negatives. Both for the primary test and the replication, participants made more errors on AM-words such as BLOOD than on TM -and CM-words such as MOOD and MOON. The number of errors for CM-words, however, was not markedly lower than for the words with typical mappings. On the contrary, they were associated with higher error rates.

Omnibus analysis of variance. The six mean percentages of false-negatives obtained for combinations of Word Type and Replication were subjected to statistical analysis. Table 14 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. The table also provides values of partial eta squared (η_p^2) as a measure of association strength and the results of non-parametric tests.

Table 14.
Analysis of variance on percentages of false-negatives for Experiment 2.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Block Position	171.46	.83	2.50	68.55	6.69	.001	< .001	.102
Block Position × Participant	1511.64	.83	147.56	10.24				
• Sequence	8.89		3	2.96	.05	.986	.986	.003
Participant(Group)	3465.00		56	61.88				
• Word Type	1225.97	.71	1.43	859.78	27.22	< .001	< .001	.316
Word Type × Participant	2657.36	.71	84.13	31.59				
• Replication	233.61		1	233.61	30.09	< .001	< .001	.338
Replication × Participant	458.06		59	8.13				
• Word Type × Replication	56.81	.94	1.89	30.14	3.11	.051	.122	.050
Word Type × Replication × Participant	1076.53	.94	111.22	9.68				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

As is confirmed by looking at Table 14, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced fewer errors in the replication (M = 2.5%) than in the primary test (M = 4.1%). There was no evidence for a substantial interaction effect of Word Type and Replication. Further, adding participant group (Sequence A-B-C-D vs. Sequence B-A-D-C vs. Sequence C-D-A-B vs. Sequence D-C-B-A) as a between-subjects variable did not improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Planned contrasts. Returning to the results presented in Figure 7, there was a statistically significant main effect of Word Type (Table 14), which accounted for a considerable percentage of variance. This overall effect was further inspected, separately for the primary test and the replication, with three (Bonferroni-adjusted) pairwise comparisons, which kept familywise Type I errors at 5%. Hence, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals (95% SCI). The pairwise comparisons involved three planned contrasts that evaluated whether (false-negative) error rates for words like BLOOD were higher than for words like MOOD (AM > TM) and for words like MOON (AM > CM), and higher for words like MOOD than for words like MOON (TM > CM). Starting with the results of the *primary test,* for the **AM > TM** contrast there was a statistically significant difference of **5.3** percentage points, with a 95% SCI of 2.9 to 7.6 ($F(1,59) = 30.77$, $MSE = 26.88$, $p < .001$). The **4.0** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 1.5 to 6.5 ($F(1,59) = 15.56$, $MSE = 30.85$, $p < .001$). Finally, the **TM >**

CM contrast showed a difference in opposite direction, **-1.3** percentage points, with a 95% SCI of -2.6 to 0.1, that was not statistically significant ($F(1,59) = 4.89$, $MSE = 9.59$, $p = .031$). Turning to the results of the *replication*, the same (Bonferroni-adjusted) planned contrasts gave differences of **3.3** (95% SCI 1.8 to 4.9), **2.8** (95% SCI 1.3 to 4.2), and **-0.6** (95% SCI -1.6 to 0.4) percentage points, respectively, $F(1,59) = 27.44$, $MSE = 12.15$, $p < .001$; $F(1,59) = 21.06$, $MSE = 10.77$, $p < .001$; $F(1,59) = 2.17$, $MSE = 4.70$, $p = .146$, respectively.

Latency data of match trials

The mean correct yes-response latencies as a function of Word Type, separately for each trial block and participant group, are presented in Table 2 of Appendix C. Figure 8 shows the mean correct yes-response latencies for the AM-words, TM-words, and CM-words, both for the primary test and the (within-participants) replication. These mean correct yes-response latencies were collapsed over trial block and participant group.

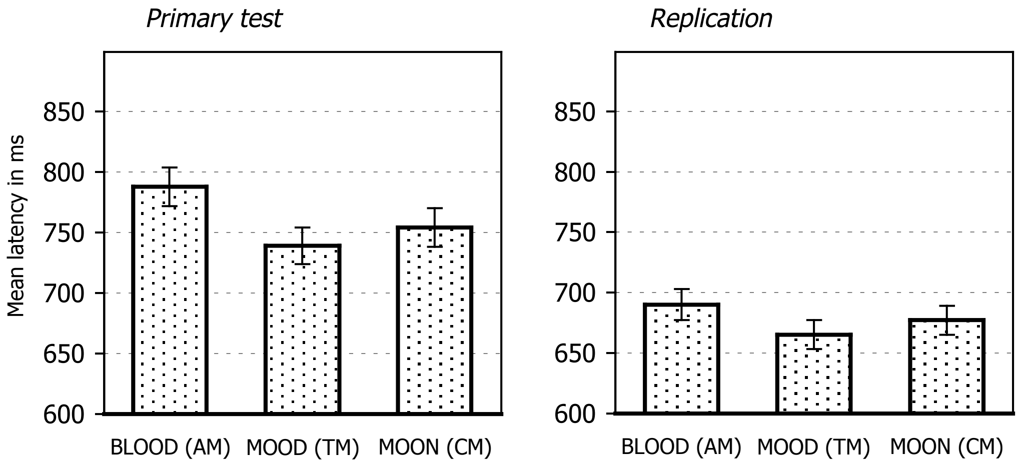

Figure 8. Mean correct yes-response latencies as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 2. Error bars represent the standard error of the mean.

Again, the ratio of English friends and enemies is expected to affect match-trial performance and, as can be seen in Figure 8, such influence was in fact observed. Moreover, comparing Figures 7 and 8 indicates that the patterns of yes-response latencies and false-negative error rates were nearly identical. Specifically, response latencies were longer for AM-words (e.g., BLOOD) than for TM-words (e.g., MOOD) and CM-words (e.g., MOON). Response latencies for CM-words were, however, not markedly shorter as compared to the TM-words. On the contrary, they were associated with longer naming latencies.

Omnibus analysis of variance. The six mean correct yes-response latencies obtained for combinations of Word Type and replication were subjected to statistical analysis. Table 15 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. Preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect (Table 15), but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced shorter response latencies in the replication (M = 677 ms) than in the primary test (M = 760 ms). Again, adding participant group as a between-subjects variable did not improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Table 15.
Analysis of variance on correct yes-response latencies for Experiment 2.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	493211.35	.73	2.20	224523.46	46.64	< .001	.441
Block Position × Participant	623986.40	.73	129.61	4814.51			
• Sequence	101201.34		3	33733.78	.55	.648	.029
Participant(Group)	3414368.36		56	60970.86			
• Word Type	84095.44	.92	1.84	45803.07	19.52	< .001	.249
Word Type × Participant	254232.23	.92	108.33	2346.93			
• Replication	620923.34		1	620923.34	98.81	< .001	.626
Replication × Participant	370766.50		59	6284.18			
• Word Type × Replication	9731.21	.98	1.97	4943.47	5.12	.008	.080
Word Type × Replication × Participant	112228.46	.98	116.14	966.31			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Returning to the results presented in Figure 8, there was a statistically significant main effect of Word Type, which accounted for a considerable percentage of variance (Table 15). In contrast to the error data, this main effect was embedded in a statistically significant interaction effect of Word Type and Replication. Inspection of Figure 8 reveals the source of this interaction effect. Compared with the data of the primary test, response latencies were, in particular for AM-words, substantially shorter in the replication.

Planned contrasts. The overall Word Type effect was further inspected, separately for the primary test and the replication, with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was again set to .017. Starting with the results of the *primary test*, for the **AM > TM** contrast there was a statistically significant difference of **49** ms, with a 95% SCI of 31 to 67 (F(1,59) = 45.32, MSE = 1569.06, p < .001). The **33** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 11 to 56 ($F(1,59) = 13.05$, $MSE = 2553.79$, $p < .001$). Finally, the **TM > CM** contrast showed a difference in opposite direction, **-15** ms, with a 95% SCI of -33 to 2, that was not statistically significant ($F(1,59) = 4.61$, $MSE = 1534.15$, $p = .036$). Turning to the *replication* results, the same (Bonferroni-adjusted) planned contrasts gave differences of **25** ms (95% SCI 10 to 41), **13** ms (95% SCI -4 to 30), and **-12** ms (95% SCI -26 to 2), respectively, $F(1,59) = 15.58$, $MSE = 1237.15$, $p < .001$; $F(1,59) = 3.51$, $MSE = 1466.20$, $p = .066$; $F(1,59) = 4.66$, $MSE = 964.19$, $p = .035$, respectively.

Error data of mismatch trials

A participant produced an error (i.e., a false-positive) in a *mismatch* trial when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other (e.g., PAID - /id/, PAID - /Ed/, SAID - /id/, SAID - /ed/, and STAIN - /}n/). The mean percentages of false-positives as a function of Trial Type and Word Type, separately for each trial block and participant group, are presented in Table 3 of Appendix C. Figure 9 shows the mean percentages of false-positives for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials), both for the primary test and the replication. These mean percentages of false-positives were collapsed over trial block and participant group.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. Specifically, for no-match trials, words with atypical mappings (e.g., PAID) should induce more false-positive errors than words with typical mappings (e.g., SAID) and consistent mappings (e.g., STAIN). We also expected more errors for words with typical mappings than for words with consistent mappings. Looking at the data of the no-match trials in Figure 9, the ratio of friends and enemies was indeed associated with the number of false-positives. Both for the primary test and the replication, participants made more errors on AM-words such as PAID than on TM-words such as SAID. The number of errors for CM-words such as STAIN, however, was not markedly lower than for inconsistent words. On the contrary, they were associated with higher error rates, especially in the replication.

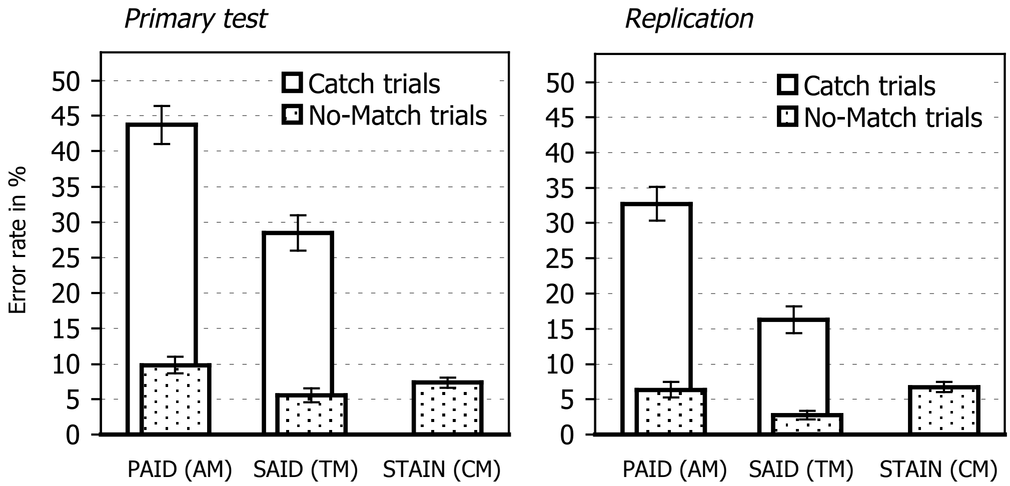

Figure 9. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 2. Error bars represent the standard error of the mean.

Omnibus analysis of variance. The mean percentages of false-positives were subjected to statistical analysis. Table 16 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. The table also provides the results of non-parametric tests. Preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect (Table 16), but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced fewer errors in the replication (M = 14.5%) than in the primary test (M = 21.9%). Adding participant group as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Returning to the data of the mismatch trials, there was a statistically significant main effect of Word Type (AM vs. TM), which accounted for a considerable percentage of variance (Table 16). There was no evidence for a substantial Word Type by Replication interaction effect, or for a substantial three-way Trial Type by Word Type by Replication interaction effect.

Planned contrasts. The overall Word Type effect, analyzed separately for the no-match trials, was further inspected with three (Bonferroni-adjusted) pairwise comparisons, which included CM-words. Again, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals.

Table 16.
Analysis of variance on percentages of false-positives for Experiment 2.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F∣H0)*	p^a	η_p^2
• Block Position	1930.37	1.0	3.00	643.46	20.23	< .001	< .001	.265
Block Position × Participant(Group)	5344.03	1.0	168.00	31.81				
• Sequence	1118.96		3	372.99	.61	.608	.506	.032
Participant(Group)	33991.67		56	606.99				
• TT	70325.21		1	70325.21	234.53	< .001	< .001	.807
TT × Participant(Group)	16791.67		56	299.85				
• WT	11701.88		1	11701.88	115.61	< .001	< .001	.674
WT × Participant(Group)	5668.33		56	101.22				
• TT × WT	4141.88		1	4141.88	31.37	< .001	< .001	.359
TT × WT × Participant(Group)	7395.00		56	132.05				
• Replication	6526.88		1	6526.88	69.25	< .001	< .001	.553
Replication × Participant(Group)	5278.33		56	94.26				
• TT × Replication	2125.21		1	2125.21	23.22	< .001	< .001	.293
TT × Replication × Participant(Group)	5125.00		56	91.52				
• WT × Replication	1.88		1	1.88	.02	.884	.999	.000
WT × Replication × Participant(Group)	4908.33		56	87.65				
• TT × WT × Replication	25.21		1	25.21	.26	.612	.689	.005
TT × WT × Replication × Participant(Group)	5428.33		56	96.94				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom. WT = Word Type; TT = Trial Type.
[a] *P*-value of nonparametric test, *p (□∣H0)* for Kruskal-Wallis test, *p* (χ^2*∣H0)* for Friedman test and *p (□∣H0)* for sign test.

The pairwise comparisons involved three planned contrasts that evaluated whether, collapsed over primary test and replication, (false-positive) error rates for words like PAID were higher than for words like SAID (AM > TM) and for words like STAIN (AM > CM), and higher for words like SAID than for words like STAIN (TM > CM). For the **AM > TM** contrast there was a statistically significant difference of **4.0** percentage points, with a 95% SCI of 2.3 to 5.7 ($F(1,56) = 33.39$, $MSE = 14.38$, $p < .001$). The **1.4** difference in percentage points for the **AM > CM** contrast was not statistically significant, with a 95% SCI of -0.7 to 3.5 ($F(1,56) = 2.81$, $MSE = 21.43$, $p = .099$). Finally, the **TM > CM** contrast showed a statistically significant difference in opposite direction, **-2.6** percentage points, with a 95% SCI of -4.4 to -0.8 ($F(1,56) = 12.76$, $MSE = 15.68$, $p < .001$). The same three (Bonferroni-adjusted) planned contrasts were also performed separately for the primary test and the replication. For the *primary test*, these contrasts gave differences of **4.3** (95% SCI 1.7 to 6.9), **2.5** (95% SCI -0.4 to 5.4), and **-1.8** (95% SCI -4.4 to 0.8) percentage points, respectively, $F(1,56) = 16.96$, $MSE = 33.21$, $p < .001$; $F(1,56) = 4.68$, $MSE = 40.09$, $p = .035$; $F(1,56) = 3.06$, $MSE = 32.95$ $p = .086$, respectively. For the *replication*, the differences were **3.7** (95% SCI 1.6 to 5.8), **-0.3** (95% SCI -2.6 to 2.0), and **-4.0** (95% SCI -5.9 to -2.1) percentage points, respectively, $F(1,56) = 18.41$, $MSE = 21.90$, $p < .001$; $F(1,56) = 0.13$, $MSE = 25.71$, $p = .720$; $F(1,56) = 26.35$, $MSE = 18.21$, $p < .001$, respectively.

Turning to the primary analyses that contrasted no-match-trial and catch-trial performance, recall that simultaneous coding of appropriate and inappropriate phonological structures may be indicated when in catch trials spoken rimes (derived from enemy words) appear to foster degraded, inappropriate local phonological codings to such a degree that participants are relatively slow to perceive a mismatch and/or produce a relatively large number of false-positives. If indeed this is the case, then participants may, for example, perceive SAID's phonology to rhyme with that of PAID. In contrast, in no-match trials, inappropriate phonological codings are not restored by the spoken rimes. Thus, mismatch-trial performance should be worse for catch trials than for no-match trials. This expectation was stated in the contrast catch > no-match. Furthermore, perceiving PAID's phonology to rhyme with that of SAID may be more likely to occur than perceiving SAID's phonology to rhyme with that of PAID. Thus, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping. This Word Type effect on catch-trial performance was stated as catch (AM-words) > catch (TM-words). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect (see introduction of Chapter 4), which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

It turned out that Trial Type was indeed associated with different numbers of false-positives. Figure 9 shows that error rates for catch trials reached strikingly high levels, up to 44% for AM-words. Overall, participants made more errors on catch trials ($M = 30.3\%$) than on no-match trials ($M = 6.1$), in which the **catch > no-match** contrast gave a huge overall difference of **24.2** percentage points ($MSE = 74.96$, 95% CI 21.0 to 27.4). Table 16 indicates that this statistically significant main effect accounted for a considerable percentage of variance. However, as Figure 9 shows, and supported by a statistically significant Trial Type by Replication interaction effect (which accounted for a considerable percentage of variance, see Table 16), the Trial Type effect was larger in the primary test (**28.4** percentage points, 95% CI 24.5 to 32.4, $F(1,56) = 207.49$, $MSE = 116.76$, $p < .001$) than in the replication (**20.0** percentage points, 95% CI 16.8 to 23.2, $F(1,56) = 152.04$, $MSE = 78.93$, $p < .001$), with a difference of 8.4 percentage points ($MSE = 132.05$, 95% CI 4.9 to 11.9).

Figure 9 also suggests that catch-trial performance was generally worse for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a large statistically significant difference of **15.8** percentage points with a 95% CI of 12.1 to 19.4 ($F(1,56) = 72.77$, $MSE = 102.26$, $p < .001$). The same contrast was performed separately for the primary test and replication. For the primary test, the contrast gave a difference of **15.2** percentage points with a 95% CI of 9.6 to 20.7 ($F(1,56) = 30.29$, $MSE = 227.86$, $p < .001$), and for the replication it was **16.3** percentage points with a 95% CI of 12.1 to 20.6 ($F(1,56) = 59.34$, $MSE = 134.88$, $p < .001$).

Finally, with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was indeed larger for AM-words (30.1 percentage points, 95% SCI 25.3 to 34.9, $F(1,56) = 209.76$, $MSE = 129.43$, $p < .001$) than for TM-words (18.3 percentage points (see Figure 9), 95% SCI 14.2 to 22.2, $F(1,56) = 116.55$, $MSE = 86.52$, $p < .001$), with a large difference of **11.8** percentage points ($MSE = 132.05$, 95% CI 7.5 to 16.0). Table 16 indicates that this interaction effect was statistically significant and accounted for a considerable percentage of variance. There was no evidence for a substantial three-way interaction effect of Trial Type, Word Type, and Replication. The same contrast was performed separately for the primary test and replication. For the *primary test*, the Trial Type effect was larger for AM-words (33.8 percentage points, 95% SCI 27.3 to 40.4, $F(1,56) = 142.66$, $MSE = 240.71$, $p < .001$) than for TM-words (23.0 percentage points, 95% SCI 17.9 to 28.1, $F(1,56) = 109.99$, $MSE = 144.29$, $p < .001$), with a difference of **10.8** percentage points (95% CI 4.5 to 17.2, $F(1,56) = 11.62$, $MSE = 302.98$, $p = .001$). For the *replication*, the Trial Type effect was also larger for AM-words (26.3 percentage points, 95% SCI 21.2 to 31.5, $F(1,56) = 1139.13$, $MSE = 149.52$, $p < .001$) than for TM-words (13.7 percentage points, 95% SCI 9.8 to 17.6, $F(1,56) = 65.28$, $MSE = 85.83$, $p < .001$), with a difference of **12.7** percentage points (95% CI 8.1 to 17.2, $F(1,56) = 31.05$, $MSE = 155.00$, $p < .001$).

Latency data of mismatch trials

The mean correct no-response latencies as a function of Trial Type and Word Type, separately for each trial block and participant group, are presented in Table 4 of Appendix C. Figure 10 shows the mean correct no-response latencies for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials), both for the primary test and the replication. These mean correct no-response latencies were collapsed over trial block and participant group.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. However, as can be seen in Figure 10, both in the primary test and the replication, no-response latencies were not longer for AM-words (e.g., PAID) than for TM-words (e.g., SAID) or CM-words (e.g., STAIN). As a matter of fact, they were associated with slightly shorter naming latencies. Furthermore, response latencies for CM-words were not markedly shorter than for the TM-words.

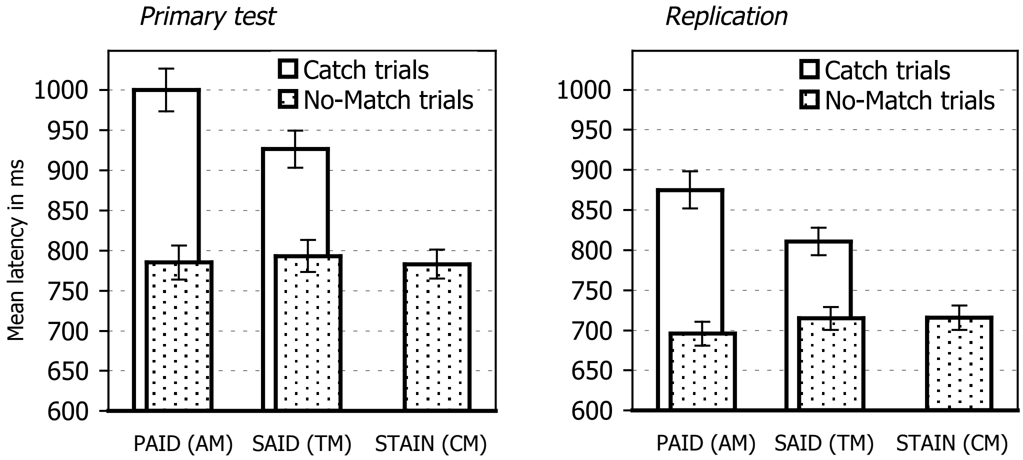

Figure 10. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 2. Error bars represent the standard error of the mean.

Omnibus analysis of variance. The mean no-response latencies were subjected to statistical analysis. Table 17 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. As can be confirmed by looking at Table 17, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced shorter response latencies in the replication (M = 774 ms) than in the primary test (M = 876 ms). Adding participant group as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments$\times$participants(group) interaction sum of squares was used to estimate error variance.

Returning to the data of the no-match trials, there was a statistically significant main effect of Word Type (AM vs. TM), which accounted for a moderate percentage of variance (Table 17). There was no evidence for a substantial Word Type by Replication interaction effect, or for a substantial three-way Trial Type by Word Type by Replication interaction effect.

Planned contrasts. The overall Word Type effect, separately for the no-match trials, was further inspected with three (Bonferroni-adjusted) pairwise comparisons, which included CM-words. The alpha level was again set to .017.

Table 17.
Analysis of variance on correct no-response latencies for Experiment 2.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	557708.28	.82	2.46	226359.84	40.18	< .001	.418
Block Position × Participant(Group)	777236.90	.82	137.97	5633.23			
• Sequence	389592.97		3	129864.32	1.068	.370	.054
Participant(Group)	6808087.41		56	121572.99			
• TT	2913916.59		1	2913916.59	283.59	< .001	.835
TT × Participant(Group)	575411.59		56	10275.21			
• WT	93956.56		1	93956.56	11.41	.001	.169
WT × Participant(Group)	461221.60		56	8236.10			
• TT × WT	204918.69		1	204918.69	14.36	< .001	.204
TT × WT × Participant(Group)	799321.94		56	14273.61			
• Replication	1231983.32		1	1231983.32	81.01	< .001	.591
Replication × Participant(Group)	851665.30		56	15208.31			
• TT × Replication	38932.76		1	38932.76	7.07	.010	.112
TT × Replication × Participant(Group)	308362.27		56	5506.47			
• WT × Replication	3540.31		1	3540.31	.44	.510	.008
WT × Replication × Participant(Group)	451673.59		56	8065.60			
• TT × WT × Replication	12.77		1	12.77	.00	.971	.000
TT × WT × Replication × Participant(Group)	547018.34		56	9768.19			

Note. Type III Sum of Squares with Huynh-Veldt ε adjusted degrees of freedom. WT = Word Type; TT = Trial Type.

For the **AM > TM** contrast there was a difference (in opposite direction) of **-13** ms, with a 95% SCI of -34 to 7, that was not statistically significant (F(1,56) = 2.57, MSE = 2078.43, p = .115). The **-9** ms difference (in opposite direction) (95% SCI -27 to 8) for the **AM > CM** contrast was not statistically significant either (F(1,56) = 1.69, MSE = 1501.02, p = .199), nor was the **4** ms difference for the **TM > CM** contrast (95% SCI -15 to 23, F(1,56) = 0.30, MSE = 1719.42, p = .590). The same three (Bonferroni-adjusted) planned contrasts were also performed separately for the primary test and the replication. For the *primary test*, these contrasts gave differences of **-8** ms (95% SCI -43 to 28), **2** ms (95% SCI -25 to 29), and **9** ms (95% SCI -16 to 35), respectively, F(1,56) = 0.28, MSE = 6213.30, p = .600; F(1,56) = 0.03, MSE = 3664.74, p = .864; F(1,56) = 0.84, MSE = 3203.54, p = .363, respectively. For the *replication*, these differences were **-19** ms (95% SCI -42 to 3), **-20** ms (95% SCI -39 to -2), and **-1** ms (95% SCI -21 to 19), respectively, F(1,56) = 4.41, MSE = 2483.23, p = .040; F(1,56) = 7.32, MSE = 1689.90, p = .009; F(1,56) = 0.02, MSE = 1921.31, p = .881, respectively.

Turning to the analyses that contrasted no-match-trial and catch-trial performance, recall that mismatch-trial performance should be worse for catch trials than for no-match trials. This expectation was stated in the contrast catch > no-match. Furthermore, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping. This Word Type effect on catch-trial performance was stated as catch (AM-words) > catch (TM-words). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect, which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 10 suggests, Trial Type indeed had an effect on no-response latencies. Participants produced longer response latencies on catch trials (M = 903 ms) than on no-match trials (M = 747 ms), in which the **catch > no-match** contrast gave a huge overall difference of **156** ms (MSE = 2568.80, 95% CI 137 to 174). Table 17 indicates that this statistically significant main effect accounted for a considerable percentage of variance. However, as suggested by Figure 10, and as is supported by a statistically significant Trial Type by Replication interaction effect (which accounted for a moderate percentage of variance, see Table 17), the Trial Type effect was larger in the *primary test* (**174** ms, 95% CI 149 to 199, $F(1,56) = 194.68$, $MSE = 4656.98$, $p < .001$) than in the *replication* (**138** ms, 95% CI 117 to 159, $F(1,56) = 176.20$, $MSE = 3233.86$, $p < .001$), with a difference of 36 ms (MSE = 5506.47, 95% CI 9 to 63).

Furthermore, catch-trial performance was indeed worse for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a large statistically significant difference of **67** ms, with a 95% CI of 34 to 104 ($F(1,56) = 15.70$, $MSE = 9176.42$, $p < .001$). The same contrast was performed separately for the primary test and replication. For the *primary test*, this contrast gave a difference of **74** ms, with a 95% CI of 20 to 129 ($F(1,56) = 7.52$, $MSE = 22078.32$, $p = .008$), and for the *replication* it was **64** ms, with a 95% CI of 28 to 100 ($F(1,56) = 12.92$, $MSE = 9568.63$, $p < .001$).

Finally, Figure 10 suggests that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was indeed larger for AM-words (197 ms, 95% SCI 161 to 233, $F(1,56) = 161.91$, $MSE = 7202.05$, $p < .001$) than for TM-words (115 ms, 95% SCI 85 to 144, $F(1,56) = 77.55$, $MSE = 5072.36$, $p < .001$), with a large difference of **83** ms (MSE = 14273.61, 95% CI 39 to 126). Table 17 indicates that this interaction effect was statistically significant and accounted for a moderate percentage of variance. The same contrast was performed separately for the primary test and replication. For the *primary test*, the Trial Type effect was larger for AM-words (215 ms, 95% SCI 162 to 267, $F(1,56) = 89.35$, $MSE = 15497.05$, $p < .001$) than for TM-words (133 ms, 95% SCI 89 to 177, $F(1,56) = 47.69$, $MSE = 11102.14$, $p < .001$), with a difference of **82** ms (95% CI 14 to 150, $F(1,56) = 5.83$, $MSE = 34570.46$, $p = .019$). For the *replication*, the Trial Type effect was also larger for AM-words (179 ms, 95% SCI 143 to 216, $F(1,56) = 125.56$, $MSE = 7695.60$, $p < .001$) than for TM-words (96 ms, 95% SCI 65 to 127, $F(1,56) =$

50.18, *MSE* = 5528.68, $p < .001$), with a difference of **83** ms (95% CI 41 to 126, $F(1,56) = 15.40$, *MSE* = 13513.12, $p < .001$).

The discussion of the results of Experiment 2 is presented simultaneously with the results of the next experiment. To anticipate, the findings indicate that processing of an inconsistent English word involves auxiliary coding of inappropriate phonology, resulting from spelling-to-sound knowledge of enemy neighbors.

EXPERIMENT 3

MISMATCH-TRIAL PERFORMANCE ON ENGLISH WORDS WITH DUTCH NEIGHBORS

In Experiment 3 (and in all experiments that follow), the list of English words without Dutch neighbors (e.g., PAID, SAID, and STAIN) was used for the match trials while the list of English words with Dutch neighbors (e.g., BLOOD, MOOD, and MOON) was used for no-match and catch trials. Notice that in the previous experiment this was reversed, with words like STAIN used for no-match and catch trials and words like MOON used for match trials. Except for this change, the purpose and design of Experiment 3 were identical to those of Experiment 2. Thus, the critical difference with Experiment 2 is that in order to create no-match and catch trials, the present experiment used words such as MOOD and BLOOD, English words that have as enemy neighbor Dutch words such as LOOD. As explained in the Introduction section, this particular type of English words is suitable to investigate cross-language spelling-to-sound consistency effects with Dutch-English bilingual participants. With this type of words, Experiments 3-5 explored the impact of knowledge of *English* enemy words on English word perception. In Experiments 6-8 (presented in Chapter 5), the same words were used to study the impact of knowledge of *Dutch* enemy words on English word perception.

In sum, in Experiment 3 we used the print-to-speech correspondence task to investigate whether, in Dutch-English bilinguals, processing of an inconsistent English word (e.g., MOOD) involves simultaneous coding of appropriate (i.e., /ud/) and inappropriate (i.e., /}d/, as in BLOOD) intermediate-grain-size phonological structures. To recapitulate, the predictions are summarised as follows. For *match trials* we expected the contrasts AM > TM, AM > CM, and TM > CM, in other words, that performance on a trial like PAID - /ed/ is worse than on a trial like SAID - /Ed/, of which the performance is worse than on a trial like STAIN - /en/. Likewise, for *no-match trials* we expected the contrasts AM > TM, AM > CM, and TM > CM. Thus, performance on a trial like BLOOD - /Yd/ (with /Yd/ derived from the unrelated word BRIDE) is worse than on a trial like MOOD - /Yd/, of which the performance is worse than on a trial like MOON - /en/ (with /en/ derived from the unrelated word VEIN). Furthermore, for the comparisons of *no-match trials* and *catch* *trials* we expected the contrast catch > no-match. Thus, performance on trials like BLOOD - /ud/ and MOOD - /}d/ is worse than on trials like BLOOD - /Yd/ and MOOD - /Yd/. We also expected the contrast catch (AM-words) > catch (TM-words). Thus, performance on trials like BLOOD - /ud/ is worse than on trials like MOOD - /}d/. Finally, we expected the following contrast: Trial Type effect (AM-words) > Trial Type effect (TM-words), in other words, that the difference in performance on trials like BLOOD - /ud/ and BLOOD - /Yd/ is larger than the difference in performance on trials like MOOD - /}d/ and MOOD - /Yd/.

Method

Participants, materials, and experimental design

A group of 80 Dutch-English bilinguals served as participants. They were presented with the 120 printed English words described in Chapter 2, and the 120 spoken rimes described in the General Method section of Chapter 3. The experimental design (see Table 10) was identical to Experiment 2, except for the reversed assignment of lists.

Results

Presentation of the results proceeds in exactly the same way as in Experiment 2. Accordingly, since Experiments 2 and 3 have the same design, the primary effects on error rates and response latencies for match and mismatch trials can be evaluated more swiftly by tracking the same planned contrasts as performed in Experiment 2. Again, these comparisons are printed in boldface.

As in Experiment 2, participants were presented with eight groups of trials representing specific combinations of Trial Type and Word Type: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). For each participant, the correct response latencies within these eight groups were averaged, separately for the first two blocks (i.e., primary test) and for the second two blocks (i.e., replication). In addition, for each participant percentage of errors was calculated for each of the eight groups, again separately for the first two blocks and for the second two blocks. Hence, the data that entered the statistical analyses consisted, for each participant and separately for the first and second two blocks, of three latency means and three percentages of false-negatives for match trials (e.g., for PAID, SAID, and STAIN), three latency means and three percentages of false-positives for no-match trials (e.g., for BLOOD, MOOD, and MOON), and two latency means and two percentages of false-positives for catch trials (e.g., for BLOOD and MOOD).

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. This resulted in a rejection of a total of 28.1% for the catch trials, 4.0% for the no-match trials, and 4.3% for the match trials. Furthermore, less than 0.1% of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. A trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. In the analyses, this resulted in a cut-off that rejected all latencies greater than 2000 ms. We did not consider further truncation, because the procedure resulted in rejection of 0.4% of the correct response latencies, a percentage that was not to be exceeded (Ulrich & Miller, 1994).

Error data of match trials

A participant produced an error (i.e., a false-negative) in a match trial when he or she pressed the “no” button when presented with a printed word and a spoken rime that were actually congruent with each other (e.g., PAID - /ed/, SAID - /Ed/, and STAIN - /en/). The mean percentages of false-negatives as a function of Word Type, separately for each trial block and participant group, are presented in Table 5 of Appendix C. Figure 11 shows the mean percentages of false-negatives for the AM-words, TM-words, and CM-words, both for the primary test and the (within-participants) replication. These mean percentages of false-negatives were collapsed over trial block and participant group.

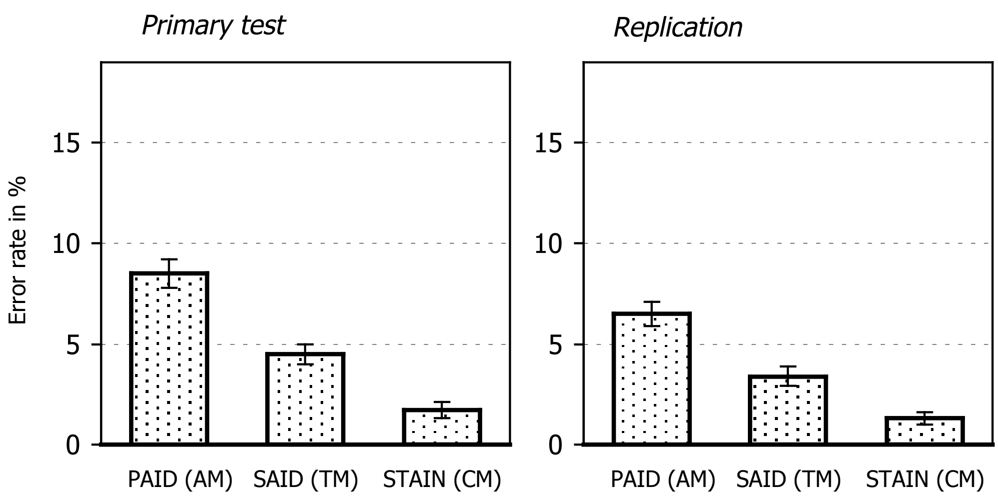

Figure 11. Mean percentages of false-negatives as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 3. Error bars represent the standard error of the mean.

Recall, match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 11, the ratio of friends and enemies was indeed associated with the number of false-negatives. Both for the primary test and the replication, participants made more errors on AM-words such as PAID than on TM-words such as SAID. Furthermore, participants made fewer errors on CM-words such as STAIN than on words with typical mappings.

Omnibus analysis of variance. The six mean percentages of false-negatives obtained for combinations of Word Type and replication were subjected to statistical analysis. Table 18 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. The table also provides the results of non-parametric tests. As can be confirmed by looking at Table 18, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced fewer errors in the replication (M = 3.7%) than in the primary test (M = 4.9%). There was no evidence for a substantial interaction effect of Word Type and Replication. Further, adding participant group (Sequence A-B-C-D vs. Sequence B-A-D-C vs. Sequence C-D-A-B vs. Sequence D-C-B-A) as a between-subjects variable did not improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Table 18.
Analysis of variance on percentages of false-negatives for Experiment 3.

Source of variance	*SS*	□	*df*	*MS*	*F*	*p (F|H0)*	p^a	$□_p{}^2$
• Block Position	127.15	.98	2.96	42.95	3.36	.020	.027	.041
Block Position × Participant	2993.19	.98	233.86	12.80				
• Sequence	71.41		3	23.80	.70	.555	.502	.027
Participant(Group)	2582.29		76	33.98				
• Word Type	2941.98	.95	1.89	1554.28	61.85	< .001	< .001	.439
Word Type × Participant	3758.02	.95	149.53	25.13				
• Replication	169.22		1	169.22	8.90	.004	.017	.101
Replication × Participant	1501.62		79	19.01				
• Word Type × Replication	49.06	.87	1.74	28.13	1.88	.162	.275	.023
Word Type × Replication × Participant	2067.60	.87	137.79	15.01				

Note. Type III Sum of Squares with Huynh-Veldt □-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

Planned contrasts. Returning to the results presented in Figure 11, there was a statistically significant main effect of Word Type, which accounted for a considerable percentage of variance (Table 18). This overall effect was further inspected, separately for the primary test and the replication, with three (Bonferroni-adjusted) pairwise comparisons, which kept familywise Type I errors at 5%. Hence, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals (95% SCI). The pairwise comparisons involved three planned contrasts that evaluated whether (false-negative) error rates for words like PAID were higher than for words like SAID (AM > TM) and for words like STAIN (AM > CM), and also higher for words like SAID than for words like STAIN (TM > CM). Starting with the results of the *primary test,* for the **AM > TM** contrast there was a statistically significant difference of **4.0** percentage points, with a 95% SCI of 2.1 to 5.9 ($F(1,79) = 27.55$, $MSE = 23.23$, $p < .001$). The **6.8** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 4.7 to 8.9 ($F(1,79) = 62.91$, $MSE = 29.51$, $p < .001$), and so was the **2.8** difference in percentage points for the **TM > CM** contrast, with a 95% SCI of 1.4 to 4.2 ($F(1,79) = 25.09$, $MSE = 12.61$, $p < .001$). Turning to the results of the *replication*, the same (Bonferroni-adjusted) planned contrasts gave differences of **3.1** (95% SCI 1.6 to 4.7), **5.2** (95% SCI 3.7 to 6.8), and **2.1** (95% SCI 0.8 to 3.5) percentage points, respectively, $F(1,79) = 23.57$, $MSE = 16.57$, $p < .001$; $F(1,79) = 68.45$, $MSE = 16.11$, $p < .001$; $F(1,79) = 14.35$, $MSE = 12.59$, $p < .001$, respectively.

Latency data of match trials

The mean correct yes-response latencies as a function of Word Type, separately for each trial block and participant group, are presented in Table 6 of Appendix C. Figure 12 shows the mean correct yes-response latencies for the AM-words, TM-words, and CM-words, both for the primary test and the (within-participants) replication. These mean correct yes-response latencies were collapsed over trial block and participant group.

Again, the ratio of English friends and enemies is expected to influence match-trial performance and, as can be seen in Figure 12, such an effect was indeed observed. Moreover, by comparing Figures 11 and 12 it can be verified that the patterns of yes-response latencies and false-negative error rates were nearly identical. Specifically, response latencies were longer for AM-words (e.g., PAID), than for TM-words (e.g., SAID). Also, response latencies for CM-words (e.g., STAIN) were shorter than for TM-words.

Omnibus analysis of variance. The six mean correct yes-response latencies obtained for combinations of Word Type and replication were subjected to statistical analysis. Table 19 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA.

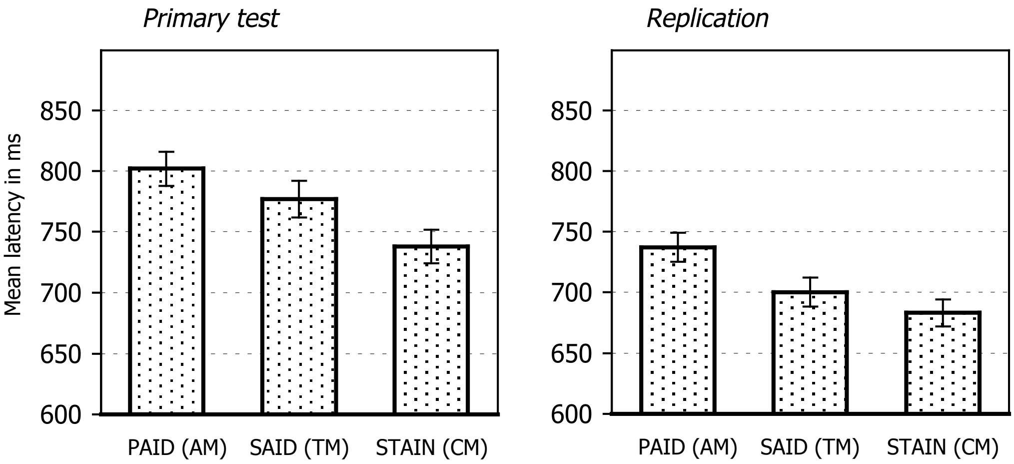

Figure 12. Mean correct yes-response latencies as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 3. Error bars represent the standard error of the mean.

As can be confirmed by inspection of Table 19, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced shorter response latencies in the replication (M = 707 ms) than in the primary test (M = 772 ms). Again, adding participant group as a between-subjects variable did not improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the regular (repeated-measures) treatments×participants interaction sum of squares was used to estimate error variance.

Table 19.
Analysis of variance on correct yes-response latencies for Experiment 3.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	441206.12	.79	2.36	187218.22	37.65	< .001	.323
Block Position × Participant	925861.90	.79	186.18	4973.08			
• Sequence	114529.97		3	38176.66	.55	.648	.021
Participant(Group)	5246719.23		76	69035.78			
• Word Type	279562.92	.92	1.83	152427.09	63.05	< .001	.444
Word Type × Participant	350294.87	.92	144.89	2417.63			
• Replication	511268.40		1	511268.40	67.27	< .001	.460
Replication × Participant	600418.53		79	7600.24			
• Word Type × Replication	9394.69	.98	1.96	4805.44	2.88	.060	.035
Word Type × Replication × Participant	257415.47	.98	154.45	1666.70			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Returning to the results presented in Figure 12, there was a statistically significant main effect of Word Type, which accounted for a considerable percentage of variance (Table 19). There was no indication of a substantial interaction effect of Word Type and Replication.

Planned contrasts. The overall Word Type effect was further inspected, separately for the primary test and the replication, with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was again set to .017. Starting with the results of the primary test, for the **AM > TM** contrast there was a statistically significant difference of **26** ms, with a 95% SCI of 6 to 45 ($F(1,79) = 10.65$, $MSE = 2484.34$, $p = .002$). The **64** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 44 to 85 ($F(1,79) = 59.72$, $MSE = 2772.10$, $p < .001$), and so was the **39** ms difference for the **TM > CM** contrast, with a 95% SCI of 22 to 55 ($F(1,79) = 33.73$, $MSE = 1767.78$, $p < .001$). Turning to the replication results, the same (Bonferroni-adjusted) planned contrasts gave differences of **37** ms (95% SCI 20 to 53), **54** ms (95% SCI 38 to 69), and **17** ms (95% SCI 4 to 30), respectively, $F(1,79) = 29.74$, $MSE = 1829.44$, $p < .001$; $F(1,79) = 72.33$, $MSE = 1602.06$, $p < .001$; $F(1,79) = 10.60$, $MSE = 1083.09$, $p = .002$, respectively.

Error data of mismatch trials

A participant produced an error (i.e., a false-positive) in a mismatch trial when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other (e.g., BLOOD - /Yd/, BLOOD - /ud/, MOOD - /Yd/, MOOD - /}d/, and MOON - /en/). The mean percentages of false-positives as a function of Trial Type and Word Type, separately for each trial block and participant group, are presented in Table 7 of Appendix C. Figure 13 shows the mean percentages of false-positives for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials), both for the primary test and replication. These mean percentages of false-positives were collapsed over trial block and participant group.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. Looking at the data of the no-match trials in Figure 13, the ratio of friends and enemies was indeed associated with the number of false-positives. Both for the primary test and the replication, participants made more errors on AM-words such as BLOOD than on TM-words such as MOOD. The number of errors for CM-words such as MOON, however, was not lower than for the words with typical mappings.

Omnibus analysis of variance. The mean percentages of false-positives were subjected to statistical analysis. Table 20 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. The table also provides the results of non-parametric tests.

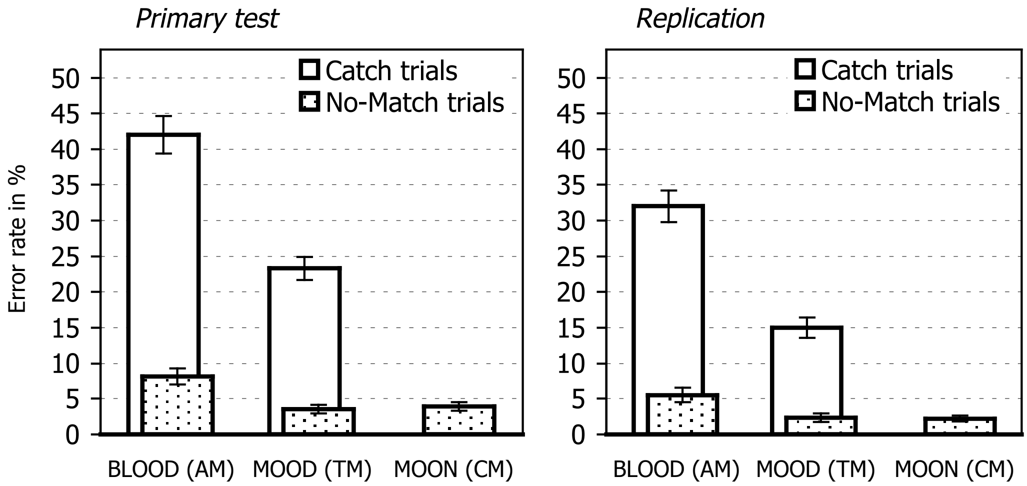

Figure 13. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 3. Error bars represent the standard error of the mean.

As is confirmed by looking at Table 20, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced fewer errors in the replication (M = 13.7%) than in the primary test (M = 19.2%). Adding participant group as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Returning to the data of the mismatch trials, there was a statistically significant main effect of Word Type (AM vs. TM), which accounted for a considerable percentage of variance (Table 20). There was no evidence for a substantial Word Type by Replication interaction effect, or of a substantial Trial Type by Word Type by Replication (three-way) interaction effect.

Planned contrasts. The overall Word Type effect for the no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons, which included CM-words. The alpha level was set to .017. The pairwise comparisons involved three planned contrasts that evaluated whether, collapsed over primary test and replication, (false-positive) error rates for words like BLOOD were higher than for words like MOOD (AM > TM) and for words like MOON (AM > CM), and higher for words like MOOD than for words like MOON (TM > CM). For the **AM > TM** contrast there was a statistically significant difference of **3.9** percentage points, with a 95% SCI of 2.2 to 5.6 (F(1,76) = 32.21, MSE = 19.25, p < .001).

Table 20.
Analysis of variance on percentages of false-positives for Experiment 3.

Source of variance	*SS*	□	*df*	*MS*	*F*	*p (F|H0)*	p^a	$□_p^2$
• Block Position	1696.27	1.0	3.00	565.42	25.89	< .001	< .001	.254
Block Position × Participant(Group)	4979.51	1.0	228.00	21.84				
• Sequence	1774.22		3	591.41	1.40	.248	.253	.052
Participant(Group)	32036.88		76	421.54				
• TT	86257.66		1	86257.66	424.88	< .001	< .001	.848
TT × Participant(Group)	15429.38		76	203.02				
• WT	19031.41		1	19031.41	256.37	< .001	< .001	.771
WT × Participant(Group)	5641.88		76	74.24				
• TT × WT	7770.16		1	7770.16	94.04	< .001	< .001	.553
TT × WT × Participant(Group)	6279.38		76	82.62				
• Replication	4895.16		1	4895.16	54.42	< .001	< .001	.417
Replication × Participant(Group)	6836.88		76	89.96				
• TT × Replication	2066.41		1	2066.41	19.21	< .001	.047	.202
TT × Replication × Participant(Group)	8174.38		76	107.56				
• WT × Replication	97.66		1	97.66	1.30	.259	.125	.017
WT × Replication × Participant(Group)	5731.88		76	75.42				
• TT × WT × Replication	1.41		1	1.41	.019	.892	.728	.000
TT × WT × Replication × Participant(Group)	5744.38		76	75.58				

Note. Type III Sum of Squares with Huynh-Veldt □adjusted degrees of freedom. WT = Word Type; TT = Trial Type.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

The **3.8** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 2.4 to 5.1 ($F(1,76) = 43.40$, $MSE = 12.96$, $p < .001$). Finally, the **TM > CM** contrast showed a difference in opposite direction, **-0.2** percentage points, with a 95% SCI of -1.5 to 1.1, that was not statistically significant ($F(1,76) = 0.13$, $MSE = 10.88$, $p = .720$). The same three (Bonferroni-adjusted) planned contrasts were also performed separately for the primary test and the replication. For the *primary test*, these contrasts gave differences of **4.6** (95% SCI 2.2 to 7.1), **4.2** (95% SCI 2.0 to 6.4), and **-0.4** (95% SCI -2.4 to 1.5) percentage points, respectively, $F(1,76) = 21.48$, $MSE = 39.84$, $p < .001$; $F(1,76) = 22.38$, $MSE = 31.34$, $p < .001$; $F(1,76) = 0.30$, $MSE = 25.12$, $p = .583$, respectively. For the *replication*, these differences were **3.2** (95% SCI 1.1 to 5.4), **3.3** (95% SCI 1.5 to 5.1), and **0.1** (95% SCI -1.5 to 1.6) percentage points, respectively, $F(1,76) = 14.27$, $MSE = 29.61$, $p < .001$; $F(1,76) = 20.19$, $MSE = 21.74$, $p < .001$; $F(1,76) = .01$, $MSE = 15.91$, $p = .921$, respectively.

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, recall that in a catch trial participants may, for example, perceive

MOOD's phonology to rhyme with that of BLOOD. In contrast, in no-match trials, inappropriate phonological codings are not restored by the spoken rimes. Thus, mismatch-trial performance should be worse for catch trials than for no-match trials (catch > no-match). Further, perceiving BLOOD's phonology to rhyme with that of MOOD may be more likely to occur than perceiving MOOD's phonology to rhyme that of BLOOD. Thus, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping (catch (AM-words) > catch (TM-words)). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect (see introduction of Chapter 4), which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 13 shows, Trial Type was indeed associated with different numbers of false-positives. As in Experiment 2, error rates for catch trials reached stunningly high levels, up to 42% for AM-words. Overall, participants made more errors on catch trials (M = 28.1%) than on no-match trials (M = 4.8%), in which the **catch > no-match** contrast gave a huge overall difference of **23.2** percentage points (MSE = 50.75, 95% CI 21.0 to 25.5). Table 20 indicates that this statistically significant main effect accounted for a considerable percentage of variance. However, as Figure 13 suggests, and supported by a statistically significant Trial Type by Replication interaction effect (which accounted for a moderate percentage of variance, see Table 20), the Trial Type effect was larger in the primary test (**26.8** percentage points, 95% CI 24.0 to 29.6, $F(1,76) = 356.78$, $MSE = 80.60$, $p < .001$) than in the replication (**19.6** percentage points, 95% CI 16.9 to 22.3, $F(1,76) = 206.27$, $MSE = 74.69$, $p < .001$), with a difference of 7.2 percentage points (MSE = 107.56, 95% CI 3.9 to 10.5).

Figure 13 also suggests that catch-trial performance was generally worse for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a large statistically significant difference of **17.9** percentage points with a 95% CI of 15.5 to 20.3 ($F(1,76) = 215.97$, $MSE = 59.18$, $p < .001$). The same contrast was performed separately for the primary test and replication. For the primary test, this contrast gave a difference of **18.8** percentage points with a 95% CI of 15.5 to 22.0 ($F(1,76) = 128.84$, $MSE = 109.14$, $p < .001$), and for the replication it was **17.0** percentage points with a 95% CI of 13.4 to 20.6 ($F(1,76) = 89.42$, $MSE = 129.28$, $p < .001$).

Finally, Figure 13 suggests that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was indeed larger for AM-words (30.2 percentage points, 95% SCI 27.0 to 33.4, $F(1,76) = 462.63$, $MSE = 78.79$, $p < .001$) than for TM-words (16.2 percentage points, 95% SCI 13.4 to 19.1, $F(1,76) = 164.96$, $MSE = 64.03$, $p < .001$), with a large difference of **13.9** percentage points (MSE = 82.62, 95% CI 11.1 to 16.8). Table 20 indicates that this interaction effect was statistically significant and accounted for a considerable percentage of variance. There was no evidence for a substantial three-way interaction effect of Trial Type, Word Type and Replication. The same contrast was performed separately for the primary test and replication. For the *primary test*, the Trial Type effect was larger for AM-words (33.9 percentage points, 95% SCI 29.5 to 38.2, $F(1,76) = 316.77$, $MSE = 144.90$, $p < .001$) than for TM-words (19.8 percentage points, 95% SCI 16.2 to 23.3, $F(1,76) = 158.00$, $MSE = 98.75$, $p < .001$), with a difference of **14.1** percentage points (95% CI 10.1 to 18.2, $F(1,76) = 48.40$, $MSE = 164.90$, $p < .001$). For the *replication*, the Trial Type effect was also larger for AM-words (26.5 percentage points, 95% SCI 22.2 to 30.8, $F(1,76) = 199.42$, $MSE = 140.86$, $p < .001$) than for TM-words (12.8 percentage points, 95% SCI 9.4 to 16.1, $F(1,76) = 77.16$, $MSE = 84.28$, $p < .001$), with a difference of **13.8** percentage points (95% CI 9.9 to 17.6, $F(1,76) = 49.91$, $MSE = 151.51$, $p < .001$).

Latency data of mismatch trials

The mean correct no-response latencies as a function of Trial Type and Word Type, separately for each trial block and participant group, are presented in Table 8 of Appendix C. Figure 14 shows the mean correct no-response latencies for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials), both for the primary test and replication. These mean correct no-response latencies were collapsed over trial block and participant group.

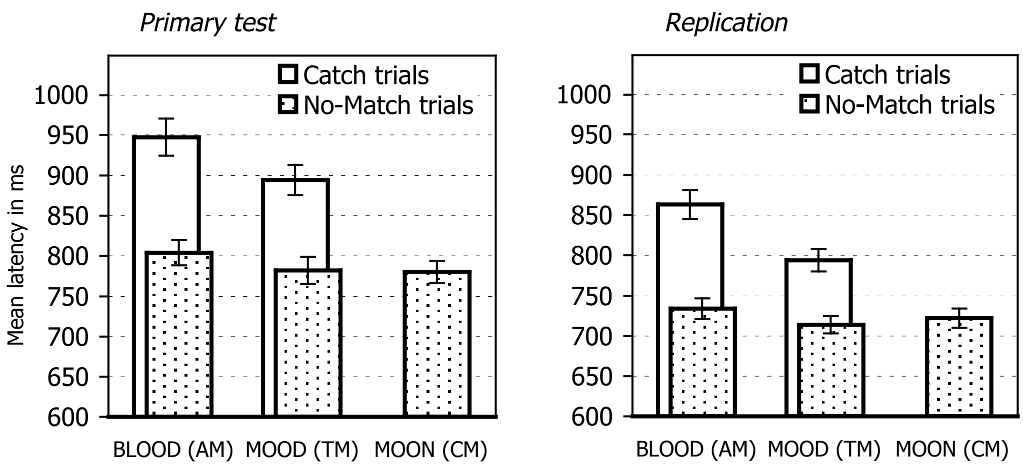

Figure 14. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test (left panel) and replication (right panel) in Experiment 3. Error bars represent the standard error of the mean.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. As can be seen in Figure 14, such influence was indeed observed. For both the primary test and the replication, response latencies were longer for AM-words (e.g., BLOOD), than for TM-words (e.g., MOOD). Response latencies for CM- words (e.g., MOON) were not markedly shorter than for TM-words. On the contrary, for the replication they were in fact longer.

Omnibus analysis of variance. The mean no-response latencies were subjected to statistical analysis. Table 21 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 2 (Replication: primary test vs. replication) repeated-measures ANOVA. As is confirmed by inspecting Table 21, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position). Specifically, participants produced shorter response latencies in the replication (M = 776 ms) than in the primary test (M = 856 ms). Adding participant group as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Table 21.
Analysis of variance on correct no-response latencies for Experiment 3.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	632401.38	.83	2.49	253647.79	65.13	< .001	.462
Block Position × Participant(Group)	737904.95	.83	189.49	3894.26			
• Sequence	195034.59		3	65011.53	.58	.629	.022
Participant(Group)	8504262.97		76	111898.20			
• TT	2152380.04		1	2152380.04	282.58	< .001	.788
TT × Participant(Group)	578893.89		76	7617.03			
• WT	268099.69		1	268099.69	63.60	< .001	.456
WT × Participant(Group)	320384.89		76	4215.59			
• TT × WT	65185.44		1	65185.44	8.65	.004	.102
TT × WT × Participant(Group)	572875.87		76	7537.84			
• Replication	1032417.23		1	1032417.23	102.37	< .001	.574
Replication × Participant(Group)	766510.34		76	10085.66			
• TT × Replication	21125.51		1	21125.51	3.05	.085	.039
TT × Replication × Participant(Group)	525917.62		76	6919.97			
• WT × Replication	2186.70		1	2186.70	.45	.507	.006
WT × Replication × Participant(Group)	373750.02		76	4917.76			
• TT × WT × Replication	3895.69		1	3895.69	.63	.431	.008
TT × WT × Replication × Participant(Group)	472737.74		76	6220.23			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom. WT = Word Type; TT = Trial Type.

Returning to the data of the no-match trials, there was a statistically significant main effect of Word Type (AM vs. TM), which accounted for a considerable percentage of variance (Table 21). There was no evidence for a substantial Word Type by Replication interaction effect, or of a substantial three-way Trial Type by Word Type by Replication interaction effect.

Planned contrasts. The overall Word Type effect for the no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons (which included CM-words). The alpha level was again set to .017. For the **AM > TM** contrast there was a statistically significant difference of **21** ms, with a 95% SCI of 4 to 37 ($F(1,76) = 9.52$, $MSE = 1809.56$, $p = .003$). The **18** ms difference (95% SCI 5 to 30) for the **AM > CM** contrast was also statistically significant ($F(1,76) = 12.29$, $MSE = 1017.68$, $p < .001$), but the **-3** ms difference (in opposite direction) for the **TM > CM** contrast (95% SCI -16 to 9) was not ($F(1,76) = 0.36$, $MSE = 1043.33$, $p = .550$). The same three (Bonferroni-adjusted) planned contrasts were also performed for the primary test and the replication. For the *primary test*, these contrasts gave differences of **22** ms (95% SCI -6 to 50), **24** ms (95% SCI 5 to 42), and **2** ms (95% SCI -17 to 20), respectively, $F(1,76) = 3.81$, $MSE = 5077.59$, $p = .055$; $F(1,76) = 9.87$, $MSE = 2277.29$, $p = .002$; $F(1,76) = 0.05$, $MSE = 2297.04$, $p = .822$, respectively. For the *replication*, these differences were **20** ms (95% SCI 3 to 36), **12** ms (95% SCI -5 to 28), and **-8** ms (95% SCI -25 to 9), respectively, $F(1,76) = 8.45$, $MSE = 1801.31$, $p = .005$; $F(1,76) = 3.11$, $MSE = 1749.88$, $p = .082$; $F(1,76) = 1.26$, $MSE = 1960.03$, $p = .266$, respectively.

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, recall that mismatch-trial performance should be worse for catch trials than for no-match trials (catch > no-match). Furthermore, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping (catch (AM-words) > catch (TM-words)). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect, which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 14 shows, Trial Type indeed had an effect on no-response latencies. Participants produced longer response latencies on catch trials ($M = 874$ ms) than on no-match trials ($M = 758$ ms), in which the **catch > no-match** contrast gave a huge overall difference of **116** ms ($MSE = 1904.26$, 95% CI 102 to 130). Table 21 indicates that this statistically significant main effect accounted for a considerable percentage of variance. Although suggested by Figure 14, there was no evidence for a substantial Trial Type by Replication interaction effect. In the primary test the contrast gave a difference of **127** ms (95% CI 108 to 147, $F(1,76) = 166.10$, $MSE = 3913.33$, $p < .001$) and in the replication **104** ms (95% CI 86 to 123, $F(1,76) = 130.17$, $MSE = 3355.17$, $p < .001$), with a difference of 23 ms that was not statistically significant ($MSE = 6919.97$, 95% CI -3 to 49).

Figure 14 also shows that, overall, catch-trial performance was indeed worse for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a large statistically significant difference of **61** ms, with a 95% CI of 41 to 81 ($F(1,76) = 36.74$, $MSE = 4067.16$, $p < .001$). The same contrast was performed separately for the primary test and replication. For the primary test, this contrast gave a difference of **52** ms, with a 95% CI of 21 to 84 ($F(1,76) = 11.15$, $MSE = 9883.83$, $p = .001$), and for the replication it was **70** ms, with a 95% CI of 45 to 94 ($F(1,76) = 31.75$, $MSE = 6128.69$, $p < .001$).

Finally, Figure 14 shows that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was indeed larger for AM-words (136 ms, 95% SCI 112 to 160, $F(1,76) = 168.69$, $MSE = 4396.62$, $p < .001$) than for TM-words (96 ms, 95% SCI 75 to 116, $F(1,76) = 115.41$, $MSE = 3180.81$, $p < .001$), with a difference of **40** ms ($MSE = 7537.84$, 95% CI 13 to 68). Table 21 indicates that this interaction effect was statistically significant and accounted for a moderate percentage of variance. The same contrast was performed separately for the primary test and replication. For the *primary test*, the Trial Type effect was larger for AM-words (143 ms, 95% SCI 108 to 177, $F(1,76) = 89.91$, $MSE = 9062.30$, $p < .001$) than for TM-words (112 ms, 95% SCI 80 to 145, $F(1,76) = 62.41$, $MSE = 8072.04$, $p < .001$), with a difference of **30** ms (95% CI -12 to 73, $F(1,76) = 2.00$, $MSE = 18615.37$, $p = .162$). For the *replication*, the Trial Type effect was also larger for AM-words (130 ms, 95% SCI 99 to 160, $F(1,76) = 96.18$, $MSE = 6986.39$, $p < .001$) than for TM-words (79 ms, 95% SCI 56 to 103, $F(1,76) = 60.37$, $MSE = 4174.33$, $p < .001$), with a difference of **50** ms (95% CI 21 to 80, $F(1,76) = 11.34$, $MSE = 8900.77$, $p = .002$).

Discussion

Experiments 2 and 3 successfully demonstrated intralingual spelling-to-sound consistency effects in Dutch-English bilinguals performing the print-to-speech correspondence task to English words. Evidence was obtained for the assumption that perceiving a match or mismatch between an English printed word and a spoken rime is affected by knowledge of English enemy neighbors. Hence, Experiments 2 and 3 extend the intralingual spelling-to-sound consistency effects observed in English word naming with Dutch-English (Experiment 1) and French-English bilinguals (Jared & Kroll, 2001) to the print-to-speech correspondence task. In general, match-trial and no-match-trial performance was worse for inconsistent words with atypical mappings (stronger enemies than friends, e.g., PAID, BLOOD) than for inconsistent words with typical mappings (stronger friends than enemies, e.g., SAID, MOOD). These findings support the hypothesis that processing of an English inconsistent word (e.g., MOOD) by Dutch-English bilinguals involves simultaneous coding of appropriate (e.g., /ud/) and inappropriate (e.g., /}d/, as in BLOOD) intermediate-grain size phonological structures that compete with each other. The outcome of this competition depends on the relative self-consistency of appropriate and inappropriate intermediate-grain spelling-to-sound associations, that, in terms of the resonance framework, correspond to the relative strength of friendly and enemy local attractors pulling orthographic-phonologic dynamics toward correct and incorrect phonology, respectively.

Evidence was obtained for the specific hypothesis that perception of an inconsistent word involves auxiliary coding of inappropriate intermediate-grain size phonology. For Dutch-English bilinguals, rejecting a catch trial that consisted of an inconsistent printed word (e.g., BLOOD) and a spoken rime that was derived from an enemy of the word (e.g., /ud/, as in MOOD) appears to demand exceptional effort. Participants frequently responded with false-positives, thus indicating that they, for instance, perceived BLOOD's phonology to rhyme with the rime of MOOD. The effect of presenting a spoken rime derived from an enemy neighbor was truly impressive. With regard to the primary test, false-positive error rates (averaged over Experiments 2 and 3) for TM-words such as SAID and MOOD were already as high as 26%, but for AM-words such as PAID and BLOOD they reached the extraordinarily high level of 43%. In comparison, for no-match trials that contained spoken rimes derived from unrelated words, the false-positive error rates for TM- and AM-words reached moderate levels of 5% and 9%, respectively. A large effect of trial type was also apparent in the latency data, even after removal of a substantial proportion of trials due to false-positive errors. To reiterate, for the primary test of Experiment 2, there was a large difference of 133 ms between catch trials and no-match trials for TM-words but the difference actually reached 215 ms for AM-words. For the primary test of Experiment 3, these differences were somewhat smaller, namely 112 ms and 143 ms, respectively.

The observed effects of Trial Type indicate that, in catch trials, spoken rimes restore the degraded, inappropriate phonological codings to such a degree that competition between appropriate and inappropriate codings is resumed at full strength. Moreover, the Trial Type effect was considerably larger for AM-words than for TM-words. That is, in a catch trial, BLOOD's phonology is more readily perceived to rhyme with that of MOOD than in the reversed case, thus perceiving MOOD's phonology to rhyme with that of BLOOD. It therefore seems that degraded, inappropriate phonological codings corresponding to highly self-consistent mappings are more readily restored than codings corresponding to less self-consistent mappings. Taken together, the finding that rejecting a catch trial takes extraordinary effort strongly indicates that, in Dutch-English bilinguals, processing of an inconsistent English word (e.g., MOOD) involves auxiliary coding of an inappropriate phonological structure (e.g., /}d/). This coding results from spelling-to-sound knowledge of enemy neighbors, and probably emerges simultaneously with correct phonology (e.g., /ud/). This supports the view that perception of inconsistent English words by Dutch-English bilinguals involves mandatory phonological coding, resulting in multiple, competing phonological structures. In terms of the resonance framework, manifold relations between spelling and sound, operationalised as inconsistency or phonological ambiguity, imply multistable local orthographic-phonologic resonances that are resolved through successive cycles of cooperative and competitive interactive activation.

EXPERIMENT 4

PILOT STUDY FOR EXPERIMENTS 5-8

The purpose of Experiment 4 was to set the stage for Experiments 5-8, in particular for Experiment 5 and 7 (both involving an SOA manipulation; see below). Anticipating the design of the experiments of Chapter 5, consistent and inconsistent English words were marked to be suitable for the creation of “Dutch catch trials”, that is, English words were selected that contained spelling bodies which have distinct pronunciations in English and Dutch. The spelling body *–OOD*, for example, is pronounced /ud/ in MOOD but it is pronounced /od/ in Dutch (rhyming with ROAD, as in the Dutch words ROOD, LOOD, and NOOD). These types of words were needed for the experiments of Chapter 5 to create catch trials that consisted of a printed English word and a spoken rime derived from a *Dutch enemy neighbor*. Due to restrictions of stimuli selection (see the General Method section), only half of the AM-words, TM-words, and CM-words were suitable for this purpose. For Experiments 4-8, these specific words were used to create catch trials (with spoken rimes derived from either English or Dutch enemy neighbors), and the other inconsistent words were used to create no-match trials. Thus, basically, Experiment 4 was identical to Experiment 3 in that the list of English words without Dutch neighbors was used for match-trials and the list of English words with Dutch neighbors for no-match trials. The main difference between the two experiments was that in Experiment 4, due to design constraints, inconsistent words were coupled to a single spoken rime. Thus, an inconsistent word was used to create either a no-match trial or a catch trial. Consequently, because printed words and spoken rimes now formed exclusive couples, there was, in contrast to Experiments 2 and 3, no need for using more than two trial blocks. In sum, Experiment 4 was a modified, strongly abbreviated version of Experiment 3, of which its basic form was employed in all subsequent experiments. Because of the similarity with Experiment 3 it was expected that Experiment 4 would replicate the basic results of the primary test of Experiment 3.

Furthermore, in Experiment 4, participants performed the entire experimental procedure three times. The purpose of this was to assess carry-over effects when participants go over the same experimental procedure more than one time, which, in fact, is the procedure of Experiments 5 and 7 involving an SOA manipulation. In Experiments 5 and 7, SOA (SOA1 vs. SOA2 vs. SOA3) was not, as it usually is, varied between but *within* participants. If SOA was planned as a between-participants variable (which generally requires a great deal more participants), we would have been unable to find sufficient numbers of participants to form three independent groups. Repeating the same experimental procedure twice should provide the necessary information whether stable effects of Trial Type and Word Type can be observed across repetitions. If the results turn out to be acceptable, we can proceed to apply this procedure in Experiments 5 and 7.

Method

Participants

A group of 24 Dutch-English bilinguals served as participants. They were presented with the 120 printed English words described in the Method section of Experiment 4 and a subset of the 120 spoken rimes described in the General Method section of Chapter 4.

Selection of printed word stimuli

The same printed word stimuli were used as in Experiments 1-3. Again, one list consisted of 60 English words with Dutch neighbours (e.g., MOON), and the other consisted of 60 English words without Dutch neighbours (e.g., STAIN). As in Experiment 3, the list of English words without Dutch neighbors was used for match-trials and the list of English words with Dutch neighbors for no-match trials. This allocation applied to each of the Experiments 4-8.

There was however one important difference with Experiment 3. To recapitulate, in Experiment 3, each of the 40 inconsistent word stimuli of the word list "with Dutch neighbours" was coupled to both of the available sound stimuli, one to create a no-match trial and one to create a catch trial (e.g., MOOD - /Yd/, MOOD - /}d/). Thus, in Experiment 3, Trial Type (no-match trial vs. catch trial) was not linked to a one set of word stimuli. In Experiments 4-8, however, the 40 inconsistent word stimuli were coupled to only one of the available sound stimuli, thus creating either a no-match trial or a catch trial. Thus, in Experiments 4-8, Trial Type was linked to a specific set of word stimuli, and, therefore, estimates of Trial Type effects may be contaminated with the idiosyncratic characteristics of the word stimuli. Both for the group of 20 words with typical mappings and the group of 20 words with atypical mappings, 10 served to create no-match trials and 10 served to create catch trials. Note that the two groups of words have the same spelling bodies.

For creating the catch trials, only words were considered of which the spelling body was dissimilarly pronounced in a corresponding Dutch word. For example, the spelling body of the English words MOOD and BLOOD is pronounced /od/ in the Dutch word ROOD (rhyming with the English word BOAT), which is clearly distinct from the English pronunciation. So, these words served as candidates for Dutch catch trials in Experiments 6, 7, and 8 of Chapter 5. In such a catch trial, a participant may be presented with a printed word such as MOOD along with an incongruent spoken rime /od/, which requires a "no" response. Less distinct, however, are the English and Dutch pronunciations of the spelling body *–EN* in the English and Dutch words THEN and DEN (in Dutch meaning "fir"), or the spelling body *–AT* in WHAT and NAT (in Dutch meaning "wet"). The shared spelling bodies of these English and Dutch words have approximately the same pronunciation. Therefore, these words are not appropriate candidates, because using the Dutch pronunciation of the spelling body does not achieve the required incongruence between print and sound.

Table 22 shows the statistics for the relevant variables separately for the word stimuli allocated to no-match trials and for the word stimuli allocated to catch trials. It seems that this new arrangement of word stimuli did not result in a critical unevenness on the set of linguistic dimensions. Inspection of Table 22 verifies that the groups of word stimuli representing the three word types were matched on number of letters, printed word frequency, bigram frequency, familiarity, and imageability. However, it turned out that the groups of word stimuli representing the three types of words for the catch trials did lose some efficacy with regard to systematic (independent) variation of number and frequency of English friends and enemies. Table 23 shows the mean (standardised) differences in number of friends and log summed frequency of friends between the three word types for the word stimuli allocated to no-match trials and for the word stimuli allocated to catch trials. As can be seen in Tables 22 and 23, the differences in number and frequency of English friends between all three word types for those stimuli allocated to catch trials were reduced but not totally eliminated. Words with typical mappings not only had a larger number of friends and greater summed frequency of friends than words with atypical mappings, but the consistency ratio was also larger. This confirms that for the word stimuli allocated to catch trials, words with atypical mappings had stronger enemies than the words with typical mappings. For the word stimuli allocated to no-match trials this new arrangement of word stimuli did not affect the consistency ratios, which were similar to the ones in Experiments 1-3.

Table 22.
Characteristics of the English Words in Experiments 4-8. (CM = consistent spelling-to-sound mappings, TM = typical spelling-to-sound mappings, AM = atypical spelling-to-sound mappings).

	No Dutch Neighbors				With Dutch Neighbors		
	CM	TM	AM		CM	TM	AM
Number of words	10	10	10		10	10	10
Number of letters							
M	4.30	3.70	3.90		3.90	4.20	4.20
SD	0.95	0.67	1.10		0.32	0.42	0.42
Mdn	4	4	4		4	4	4
IQR	1	1	1		0	0	0
CELEX Log Frequency							
M	2.00	1.86	1.76		1.90	2.11	2.16
SD	0.80	0.84	0.94		0.94	0.73	0.69
Mdn	1.84	1.85	1.57		1.92	2.16	2.30
IQR	0.65	1.24	1.46		1.43	0.99	0.53
K&F Log Frequency							
M	1.90	1.94	1.74		1.79	1.99	2.13
SD	0.84	0.74	0.95		1.02	0.79	0.75
Mdn	1.84	1.91	1.64		1.86	2.14	2.24
IQR	0.93	0.75	1.65		1.46	1.03	0.83
Log Bigram Frequency							
M	6.23	6.01	6.01		5.98	6.12	6.01
SD	0.52	0.22	0.40		0.32	0.29	0.29
Mdn	6.15	6.04	6.20		5.92	6.04	6.08
IQR	0.93	0.49	0.59		0.23	0.32	0.28
Familiarity							
M	533	547	557		546	569	572
SD	86	52	66		56	53	29
Mdn	542	541	582		549	590	568
IQR	90	82	90		83	67	33

Table 22 (*continued*)

	No Dutch Neighbors				With Dutch Neighbors		
	CM	TM	AM		CM	TM	AM
Imagability							
M	400	542	448		431	440	486
SD	108	60	136		127	112	122
Mdn	416	564	418		474	417	510
IQR	84	95	226		179	169	199
Number of Friends							
M	6.00	6.30	0.90		3.10	2.70	2.80
SD	3.46	4.92	0.99		2.23	1.64	3.33
Mdn	5.00	5.00	0.50		2.00	2.00	1.50
IQR	7.00	8.00	2.00		4.00	1.00	2.00
Frequency of Friends							
M	1184	1342	332		1018	1031	360
SD	877	1675	561		1530	599	229
Mdn	1026	515	32		276	1109	359
IQR	1091	2232	453		956	859	203
Log Frequency of Friends							
M	2.95	2.74	1.78		2.57	2.83	2.33
SD	0.37	0.67	0.91		0.69	0.58	0.74
Mdn	3.01	2.66	1.50		2.44	3.04	2.55
IQR	0.51	1.08	1.63		0.81	0.40	0.28
Consistency Ratio							
M	1.00	0.83	0.17		1.00	0.74	0.26
SD	0.00	0.13	0.13		0.00	0.13	0.13
Mdn	1.00	0.85	0.15		1.00	0.74	0.26
IQR	0.00	0.25	0.25		0.00	0.14	0.14

Table 23.
Mean differences in number of friends and log summed frequency of friends between CM-words and TM-words, between CM-words and AM-words, and between TM-words and AM-words in original units (frequency counts), *SD* units (standardized difference: Hedges *g*), and percent of nonoverlap for the words in Experiments 4-8. (CM = Consistent Mappings; TM = Typical Mappings; AM = Atypical Mappings.)

Number of friends

No-Match Trials	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.30	5.10	5.40
Standardized Difference (*g*)	0.07	2.00	1.52
Percent of nonoverlap	7.7	81.1	70.7
Catch Trials	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.40	0.30	0.10
Standardized Difference (*g*)	0.20	0.11	0.04
Percent of nonoverlap	14.7	7.7	0

Log summed frequency of friends

No-Match Trials	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.21	1.17	0.96
Standardized Difference (*g*)	0.39	1.68	1.20
Percent of nonoverlap	27.4	75.4	62.2
Catch Trials	CM vs. TM	CM vs. AM	TM vs. AM
Difference in original units	0.26	0.24	0.50
Standardized Difference (*g*)	0.41	0.34	0.75
Percent of nonoverlap	27.4	21.3	47.4

Note. *SD* units are pooled *SDs* ($\sqrt{MSw}$) of the contrasting word lists (e.g., Rosnow & Rosenthal, 2003).

Experimental design

As in the previous experiments, three basic word types were contrasted, AM-words, TM-words, and CM-words. The words were grouped in two separate word lists, a list of English words with Dutch neighbors (e.g., BLOOD, MOOD, and MOON; used for mismatch trials) and a list of English words without Dutch neighbors (e.g., PAID, SAID, and STAIN; used for match trials). Both word lists contained an equal number of words for each group of words representing one of the three word types. Again, eight groups of trials were created that represented specific combinations of Trial Type and Word Type. These eight combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). As in the previous experiments, the spelling body of an English word with a typical mapping (e.g., MOOD) also appeared in an English word with an atypical mapping (e.g., BLOOD). Thus, participants were presented twice with the same spelling body. To prevent intralist-priming effects of spelling bodies, the two words containing the same spelling body were presented in two separate blocks of trials. We did not use exactly the same precautionary procedure as in Experiments 1-3 to prevent intralist-priming effects. In Experiments 4 (and also in Experiments 5-8), the two separate blocks of trials containing the same spelling bodies were administered in one experimental session, interrupted by a short break.

In the mismatch trials of Experiment 4 (and also of Experiments 5-8), inconsistent words (e.g., BLOOD and MOOD) were coupled to spoken rimes that were incongruent with it. Specifically, each inconsistent word was either coupled to a spoken rime to create a no-match trial (e.g., MOOD - /Yd/), or to a spoken rime to create a catch trial (e.g., MOOD - /}d/). As a result, there was only one constraint that required trials to appear in separate blocks of trials, namely, spelling bodies were not to be presented within the same trial block. Consequently, each spelling body (e.g., –*OOD*) that was part of a particular inconsistent word (e.g., MOOD) appeared in two blocks of trials, A and B. For example, in block A the spelling body appeared in the word BLOOD, together with the spoken rime /ud/ to create a catch trial. In block B it appeared in the word MOOD, together with the spoken rime /}d/, again to create a catch trial.

Table 24 presents the layout of the experimental design. The two trial blocks contained equal numbers of words from all three word types. Again, this was accomplished by separating each list of 20 words comprising one of the three word types (CM, TM, and AM) anew into two sub word-lists, each containing 10 words (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2). Note that this is a different subdivision than the one used in Experiments 2 and 3. For example, if the typical word MOOD (from TM_1) was presented in Trial Block A (in a catch trial), then the atypical word BLOOD (from AM_1) was presented afterwards in Trial Block B (also in a catch trial). Conversely, if the atypical word WARD (from AM_2) was presented in Trial Block A (in a no-match trial), then the typical word YARD (from TM_2) was presented afterwards in Trial Block B (also in a no-match trial). Therefore, in each mismatch block, there were 10 catch trials and 10 no-match trials for inconsistent words, and 10 no-match trials for consistent words. Thus, Trial Block A comprised sub-lists CM_1, TM_1, and AM_2, and Trial Block B comprised sub-lists CM_2, TM_2, and AM_1. In sum, for the word-list with Dutch neighbors (e.g., MOOD), used for the mismatch trials, each of the blocks A and B contained 10 words of each word type (AM, TM, and CM) from either the sub-lists containing, for instance, WARD, MOOD, and MOON, or the sub-lists containing, for instance, BLOOD, YARD, and THEFT.

Table 24.

Experimental design for Experiments 4 (pilot study) and 5 (alternating-SOA study: Dutch participants vs. USA participants). English words with Dutch neighbors are used for No-trials, and English words without Dutch neighbors for Yes-trials. The word lists comprising each word type are split up in two distinct sub word-lists (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2), according to whether or not a word can be used to create a Dutch catch trial. For the new sub word-lists CM_1, TM_1, and AM_1 Dutch catch trials were created, but not for the sub word-lists CM_2, TM_2, and AM_2. Combinations of word type and trial type (Match, No-Match, English Catch) are systematically distributed over two trial blocks (A and B). The temporal order of trial block is Latin-square counterbalanced across two different participant groups (Participant Group 1: Sequence A-B; Participant Group 2: Sequence B-A). (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

No-trials

Trial Block	A	B
Sub Word-List	CM_1	CM_2
Trial Type	No-Match	No-Match
Example Word	***MOON***	***SEEM***
Sub Word-List	TM_1	TM_2
Trial Type	English Catch	No-Match
Example Word	***MOOD***	***SCAN***
Sub Word-List	AM_2	AM_1
Trial Type	No-Match	English Catch
Example Word	***SWAN***	***BLOOD***

Yes-trials

Trial Block	A	B
Sub Word-List	CM_1	CM_2
Trial Type	Match	Match
Example Word	***MOON***	***SEEM***
Sub Word-List	TM_1	TM_2
Trial Type	Match	Match
Example Word	***MOOD***	***SCAN***
Sub Word-List	AM_2	AM_1
Trial Type	*Match*	*Match*
Example Word	**SWAN**	**BLOOD**

These sub-lists appeared in the trial blocks A and B as follows. Trial Block A: BLOOD (catch trial), YARD (no-match trial), THEFT (no-match trial); Trial Block B: WARD (no-match trial), MOOD (catch trial), MOON (no-match trial). For the match trials, using words such as PAID, SAID, and STAIN, Trial Block A comprised sub-lists CM_1, TM_1, and AM_2 and Trial Block B comprised sub-lists CM_2, TM_2, and AM_1.

In Experiment 4, each participant was presented with each of the two trial blocks. Hence, for each participant data was obtained for each available combination of trial type and word type, using every word only once. That is, repeated measures were obtained on the participants. The temporal order of the two trial blocks was counterbalanced across two different participant groups according to a single Latin square (participant group 1: Sequence A-B; participant group 2: Sequence B-A). Participants were randomly assigned to the different sequences. Again, the counterbalancing procedure was intended to disentangle the effect of temporal position of the procedural variable Trial Block from the effects of the independent variables.

Results

As in Experiments 2 and 3, participants were presented with eight different groups of trials representing specific combinations of Trial Type and Word Type: match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). Each participant responded three times (i.e., in the primary test, repetition 1, and repetition 2) to 60 match trials which consisted of 20 AM-words (e.g., PAID), 20 TM-words (e.g., SAID), and 20 CM-words (e.g., STAIN). In addition, each participant responded three times to 40 no-match trials and 20 catch trials. The no-match trials consisted of 10 AM-words (e.g., BLOOD), 10 TM-words (e.g., MOOD), and 20 CM-words (e.g., MOON), and the catch trials consisted of 10 AM-words (e.g., BLOOD) and 10 TM-words (e.g., MOOD). For each participant, and separately for each test session, the correct response latencies within these eight groups were averaged and percentage of errors was calculated for each of the eight groups. Hence, the data that entered the statistical analyses consisted, for each participant, of nine latency means and nine percentages of false-negatives for match trials (e.g., for PAID, SAID, and STAIN), nine latency means and nine percentages of false-positives for no-match trials (e.g., for BLOOD, MOOD, and MOON), and six latency means and six percentages of false-positives for catch trials (e.g., for BLOOD and MOOD).

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. This resulted in a rejection of a total of 17.2% for the catch trials, 4.6% for the no-match trials, and 5.2% for the match trials. Further, less than 0.2% of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. A trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. This resulted in a cut-off that rejected all latencies greater than 2000 ms. Because the procedure resulted in rejection of 0.5% of the correct response latencies we did not consider further truncation (Ulrich & Miller, 1994).

Error data of match trials

A participant produced an error (i.e., a false-negative) in a match trial when he or she pressed the "no" button when presented with a printed word and a spoken rime that were actually congruent with each other (e.g., PAID - /ed/, SAID - /Ed/, and STAIN - /en/). The mean percentages of false-negatives as a function of Word Type and Repetition, separately for each trial block, are presented in Table 9 of Appendix C. Figure 15 shows the mean percentages of false-negatives for the AM-words, TM-words, and CM-words for the primary test, repetition 1, and repetition 2. These mean percentages of false-negatives were collapsed over trial block and participant group.

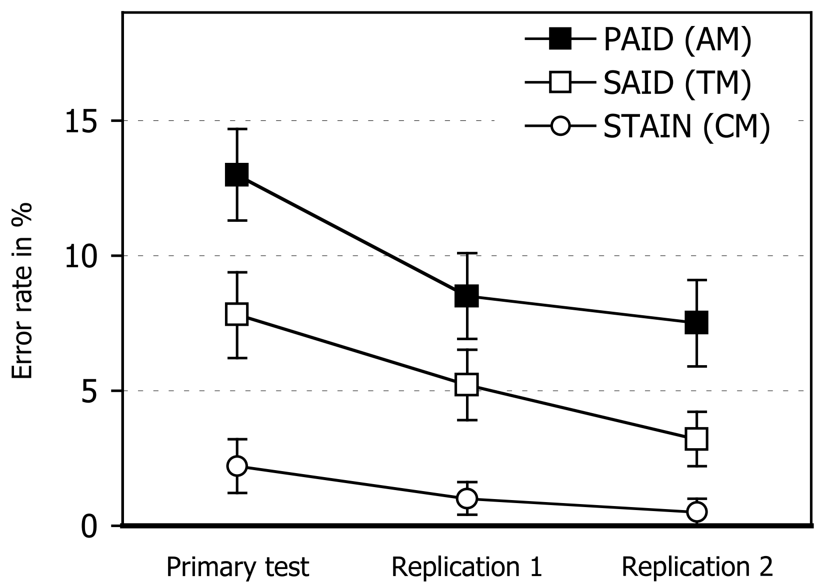

Figure 15. Mean percentages of false-negatives as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test, replication 1 and replication 2 in Experiment 4. Error bars represent the standard error of the mean.

As in Experiment 3, match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 15, the ratio of friends and enemies was indeed associated with the number of false-negatives. For the primary test, participants made more errors on AM-words such as PAID than on TM-words such as SAID. Furthermore, participants made fewer errors on CM-words such as STAIN than on words with typical mappings. This pattern of error rates was nearly identical to the pattern observed in Experiment 3.

Omnibus analysis of variance. The nine mean percentages of false-negatives obtained for combinations of Word Type and Repetition were subjected to statistical analysis. Table 25 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 3 (Repetition: primary test vs. repetition 1 vs. repetition 2) repeated-measures ANOVA. The table also provides the results of non-parametric tests. As can be confirmed by looking at Table 25, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position) and in particular of an effect of Repetition. Specifically, participants produced more errors in the primary test (M = 7.7%) than in repetition 1 (M = 4.9%) and repetition 2 (M = 3.8%). There was no evidence for a substantial interaction effect of Word Type and Repetition.

Table 25.
Analysis of variance on percentages of false-negatives for Experiment 4.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Repetition	161.83	1.0	2.00	80.91	10.95	< .001	< .001	.366
Repetition × Participant	280.87	1.0	38.00	7.39				
• Block Position	120.00		1	120.00	4.49	.047	.096	.191
Block Position × Participant	507.75		19	26.72				
• Block Position × Repetition	13.78	.91	1.83	7.55	.62	.532	.174	.031
Block Position × Repetition × Participant	425.84	.91	34.68	12.28				
• Sequence	.56		1	.56	.01	.928	.621	.000
Participant(Group)	1202.78		18	66.82				
• Sequence × Repetition	8.41	1.0	2.00	4.21	.56	.578		.030
Sequence × Repetition × Participant(Group)	272.45	1.0	36.00	7.57				
• Word Type	2125.28	1.0	2.00	1062.64	27.49	< .001	< .001	.591
Word Type × Participant	1469.17	1.0	38.00	38.66				
• Word Type × Repetition	93.89	.71	2.83	33.14	.87	.456		.044
Word Type × Repetition × Participant	2045.00	.71	53.83	37.99				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (U|H0)* for Mann-Whitney test, *p* (χ^2*|H0)* for Friedman test and *p* (□*|H0)* for sign test.

Planned contrasts. Returning to the results presented in Figure 15, there was a statistically significant main effect of Word Type, which, as can be seen in Table 25, accounted for a considerable percentage of variance. This overall effect was further inspected, exclusively for the *primary test*, with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was set to .017. The pairwise comparisons involved three planned contrasts that evaluated whether (false-negative) error rates for words like PAID were higher than for words like SAID (AM > TM) and for words like STAIN (AM > CM), and higher for words like SAID than for words like STAIN (TM > CM). For the **AM > TM** contrast there was a difference of **5.2** percentage points, with a 95% SCI of -0.7 to 11.2, that was not statistically significant ($F(1,19) = 5.44$, $MSE = 50.63$, $p = .031$). The **10.8** difference in percentage points for the **AM > CM** contrast was statistically significant, with a 95% SCI of 5.5 to 16.0 ($F(1,19) = 29.01$, $MSE = 39.84$, $p < .001$), and so was the **5.5** difference in percentage points for the **TM > CM** contrast, with a 95% SCI of 0.7 to 10.3 ($F(1,19) = 8.88$, $MSE = 34.08$, $p = .008$).

Latency data of match trials

The mean correct yes-response latencies as a function of Word Type and Repetition, separately for each trial block, are presented in Table 10 of Appendix C. Figure 16 shows the mean correct yes-response latencies for the AM-words, TM-words, and CM-words for the primary test, repetition 1, and repetition 2. These mean latencies were collapsed over trial block and participant group.

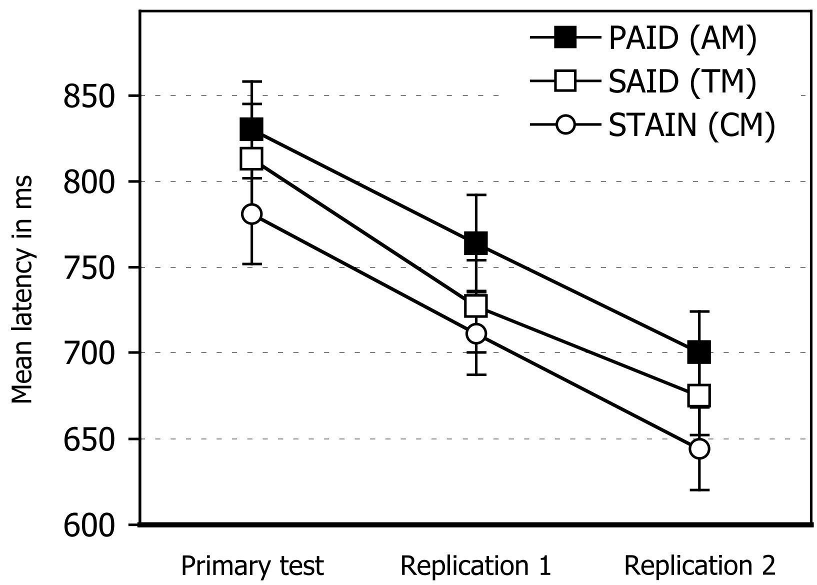

Figure 16. Mean correct yes-response latencies as a function of Word Type (AM-words vs. TM-words vs. CM-words) for primary test, replication 1 and replication 2 in Experiment 4. Error bars represent the standard error of the mean.

Again, the ratio of English friends and enemies is expected to influence match-trial performance and, as can be seen in Figure 16, this was in fact the case. Specifically, by comparing Figures 15 and 16 it can be verified that the patterns of yes-response latencies and false-negative error rates were nearly identical. For the primary test, response latencies were longer for AM-words (e.g., PAID), than for TM-words (e.g., SAID). Also, response latencies for CM-words (e.g., STAIN) were shorter than for TM-words.

Omnibus analysis of variance. The nine mean correct yes-response latencies obtained for combinations of Word Type and replication were subjected to statistical analysis. Table 26 presents the results of a 3 (Word Type: AM vs. TM vs. CM) by 3 (Repetition: primary test vs. repetition 1 vs. repetition 2) repeated-measures ANOVA. As can be confirmed by looking at Table 26, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position) and in particular of an effect of Repetition. Specifically, participants produced longer yes-response latencies in the primary test (M = 808 ms) than in repetition 1 (M = 734 ms) and repetition 2 (M = 673 ms). There was no evidence for a substantial interaction effect of Word Type and Repetition.

Table 26.
Analysis of variance on correct yes-response latencies for Experiment 4.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F\H0)*	η_p^2
• Repetition	183286.63	.76	1.52	120331.96	34.70	< .001	.646
Repetition × Participant	100349.06	.76	28.94	3467.45			
• Block Position	19278.68		1	19278.68	7.10	.015	.272
Block Position × Participant	51566.16		19	2714.01			
• Block Position × Repetition	8938.85	1.0	2.00	4469.43	3.93	.028	.171
Block Position × Repetition × Participant	43264.82	1.0	38.00	1138.55			
• Sequence	71880.05		1	71880.05	.68	.421	.036
Participant(Group)	1906945.92		18	105941.44			
• Sequence × Repetition	1238.38	.80	1.59	778.80	.23	.749	.012
Sequence × Repetition × Participant(Group)	99110.68	.80	28.62	3462.73			
• Word Type	83319.60	.84	1.67	49899.23	20.18	< .001	.515
Word Type × Participant	78444.84	.84	31.73	2472.62			
• Word Type × Repetition	2691.87	1.0	4.00	672.97	.44	.779	.023
Word Type × Repetition × Participant	116101.02	1.0	76.00	1527.65			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Planned contrasts. Returning to the results presented in Figure 16, there was a statistically significant main effect of Word Type, which, as can be seen in Table 26, accounted for a considerable percentage of variance. The overall Word Type effect was further inspected, exclusively for the primary test, with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was again set to .017. For the **AM > TM** contrast there was a difference of **17** ms, with a 95% SCI of -19 to 52, that was not statistically significant ($F(1,19) = 1.52$, $MSE = 1823.80$, $p = .233$). The **49** ms difference for the **AM > CM** contrast was statistically significant, with a 95% SCI of 11 to 86 ($F(1,19) = 11.73$, $MSE = 2034.86$, $p = .003$), but the **39** ms difference for the **TM > CM** contrast was not, with a 95% SCI of -10 to 74 ($F(1,19) = 4.10$, MSE = 2526.66, p = .057).

Error data of mismatch trials

A participant produced an error (i.e., a false-positive) in a mismatch trial when he or she pressed the "yes" button when presented with a printed word and a spoken rime that were actually incongruent with each other (e.g., BLOOD - /Yd/, BLOOD - /ud/, MOOD - /Yd/, MOOD - /}d/, and MOON - /en/). The mean percentages of false-positives as a function of Trial Type, Word Type, and Repetition, separately for each trial block, are presented in Table 11 of Appendix C. Figure 17 shows the mean percentages of false-positives for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials) for the primary test, repetition 1, and repetition 2. These mean percentages of false-positives were collapsed over trial block and participant group.

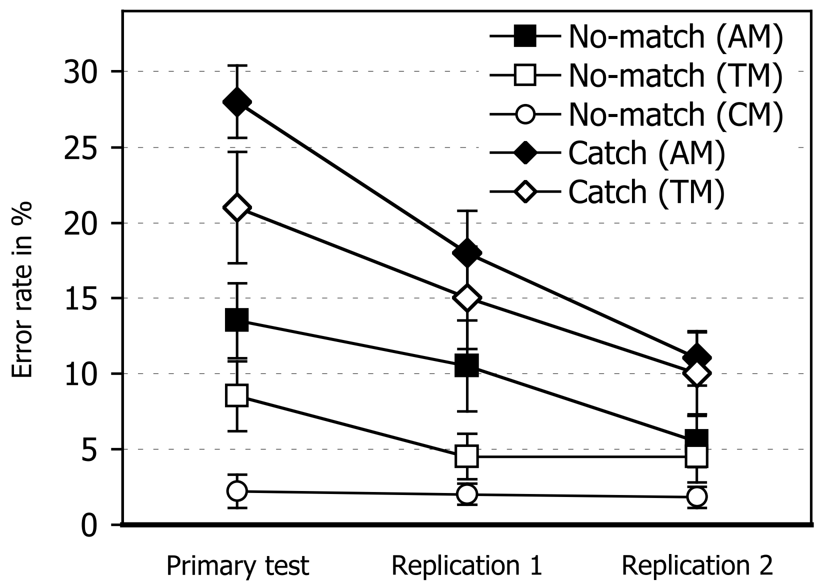

Figure 17. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test, replication 1 and replication 2 in Experiment 4. Error bars represent the standard error of the mean.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. Specifically, for no-match trials, words with atypical mappings (e.g., BLOOD) should induce more false-positive errors than words with typical mappings (e.g., MOOD) and consistent mappings (e.g., MOON). We also expect more errors for words with typical mappings than for words with consistent mappings. Looking at the data of the no-match trials in Figure 17, the ratio of friends and enemies was indeed associated, according to expectation, with the number of false-positives. For the primary test, participants made more errors on AM-words such as BLOOD than on TM-words such as MOOD. Also, the number of errors for CM-words such as MOON was, in contrast with Experiment 3, lower than for the words with typical mappings.

Omnibus analysis of variance. The mean percentages of false-positives were subjected to statistical analysis. Table 27 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 3 (Repetition: primary test vs. repetition 1 vs. repetition 2) repeated-measures ANOVA. The table also provides the results of non-parametric tests. As can be confirmed by inspection of Table 27, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position) and in particular of an effect of Repetition. Specifically, participants produced more errors in the primary test (M = 12.6%) than in repetition 1 (M = 8.7%) and repetition 2 (M = 5.7%).

Returning to the data of the mismatch trials, there was a statistically significant main effect of Word Type (AM vs. TM), which, as can be confirmed by inspection of Table 27, accounted for a considerable percentage of variance. There was no evidence for a substantial Word Type by Repetition interaction effect, or for a substantial three-way Trial Type by Word Type by Repetition interaction effect.

Planned contrasts. The overall Word Type effect for the no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons (which included CM-words), that kept familywise Type I errors at 5%. Hence, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals. The pairwise comparisons involved three planned contrasts that evaluated whether, exclusively for the primary test, (false-positive) error rates for words like BLOOD were higher than for words like MOOD (AM > TM) and for words like MOON (AM > CM), and higher for words like MOOD than for words like MOON (TM > CM). For the **AM > TM** contrast there was a difference of **5.0** percentage points, with a 95% SCI of -2.5 to 12.5, that was not statistically significant ($F(1,19) = 3.06$, $MSE = 81.58$, $p = .096$). The **11.2** difference in percentage points for the **AM > CM** contrast was statistically significant, with a 95% SCI of 4.5 to 18.0 ($F(1,19) = 18.91$, $MSE = 66.94$, $p < .001$). Finally, the **TM > CM** contrast showed a difference of **6.2** percentage points, with a 95% SCI of -2.4 to 13.2, that was not statistically significant ($F(1,19) = 5.51$, MSE = 70.89, p = .030).

Table 27.
Analysis of variance on percentages of false-positives for Experiment 4.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Repetition	470.16	1.0	2.00	235.08	22.33	< .001	< .001	.540
Repetition × Participant	400.06	1.0	38.00	10.53				
• Block Position	213.41		1	213.41	7.53	.013	.008	.284
Block Position × Participant	538.35		19	28.33				
• Block Position × Repetition	23.87	1.0	2.00	11.94	.59	.559	.494	.030
Block Position × Repetition × Participant	768.56	1.0	38.00	20.23				
• Sequence	60.00		1	60.00	.12	.732	.940	.007
Participant(Group)	8940.00		18	496.67				
• Sequence × Repetition	102.36	1.0	2.00	51.18	2.64	.085		.128
Sequence × Repetition × Participant(Group)	697.82	1.0	36.00	19.38				
• TT	5226.67		1	5226.67	25.00	< .001	< .001	.568
TT × Participant	3973.33		19	209.12				
• WT	881.67		1	881.67	7.55	.013	.019	.284
WT × Participant	2218.33		19	116.75				
• TT × WT	1.67		1	1.67	.01	.921	.096	.001
TT × WT × Participant	3131.67		19	164.83				
• TT × Repetition	643.33	.89	1.78	361.98	4.69	.019	.058	.198
TT × Repetition × Participant	2606.67	.89	33.77	77.19				
• WT × Repetition	263.33	.75	1.51	174.87	1.73	.199	.174	.084
WT × Repetition × Participant	2886.67	.75	28.61	100.89				
• TT × WT × Repetition	63.33	1.0	2.00	31.67	.53	.591	.566	.027
TT × WT × Repetition × Participant	2253.33	1.0	38.00	59.30				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (U|H0)* for Mann-Whitney test, *p* (χ^2*|H0)* for Friedman test and *p (□|H0)* for sign test.

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, recall that mismatch-trial performance should be worse for catch trials than for no-match trials (catch > no-match). Further, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping (catch (AM-words) > catch (TM-words)). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect (see introduction of Chapter 4), which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 17 shows, Trial Type was indeed associated with different numbers of false-positives, although the effect turned out smaller than in Experiment 3. Overall, participants made more errors on catch trials (*M* = 17.2%) than on no-match trials (*M* = 7.8%), in which the **catch > no-match** contrast gave an overall difference of **9.3** percentage points (*MSE* = 34.85, 95% CI 5.4 to 13.2). Table 27 indicates that this statistically significant main effect accounted for a considerable percentage of variance. However, as Figure 17 shows, and supported by a statistically significant

Trial Type by Repetition interaction effect (which accounted for a moderate percentage of variance, see Table 27), the Trial Type effect was larger in the primary test (**13.5** percentage points, 95% CI 6.7 to 20.3) than in repetition 1 (**9.0** percentage points, 95% CI 4.9 to 13.1) and repetition 2 (**5.5** percentage points, 95% CI 1.9 to 9.1).

Figure 17 also suggests that, overall, catch-trial performance was, as opposed to Experiment 3, not markedly worse for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a difference of **3.7** percentage points with a 95% CI of -1.4 to 8.7, that was not statistically significant ($F(1,19) = 2.30$, $MSE = 58.42$, $p = .146$). The same contrast was performed separately for the primary test, repetition 1, and repetition 2. For the primary test, this contrast gave a difference of **7.0** percentage points with a 95% CI of -0.9 to 14.9, for repetition 1 it was **3.0** percentage points with a 95% CI of -3.3 to 9.3, and for repetition 2 it was **1.0** percentage points with a 95% CI of -5.2 to 7.2.

Finally, Figure 17 shows that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was, contrary to expectation and in contrast to the effect found in Experiment 3, not larger for AM-words (9.2 percentage points) than for TM-words (9.5 percentage points), with a difference (in opposite direction) of **-0.3** percentage points ($MSE = 109.88$, 95% CI -7.3 to 6.6). Table 27 indicates that this interaction effect was not statistically significant and accounted for a nearly zero percentage of variance. The same contrast was performed separately for the primary test, repetition 1, and repetition 2. For the primary test, this contrast gave a difference of **2.0** percentage points with a 95% CI of -7.6 to 11.6, for repetition 1 it was **-3.0** percentage points with a 95% CI of -13.3 to 7.3, and for repetition 2 it was **0.0** percentage points with a 95% CI of -7.1 to 7.1.

Latency data of mismatch trials

The mean correct no-response latencies as a function of Trial Type, Word Type, and Repetition, separately for each trial block, are presented in Table 12 of Appendix C. Figure 18 shows the mean correct no-response latencies for the AM-words and TM-words (catch trials) and for the AM-words, TM-words, and CM-words (no-match trials), for the primary test, repetition 1, and repetition 2. These mean correct no-response latencies were collapsed over trial block and participant group.

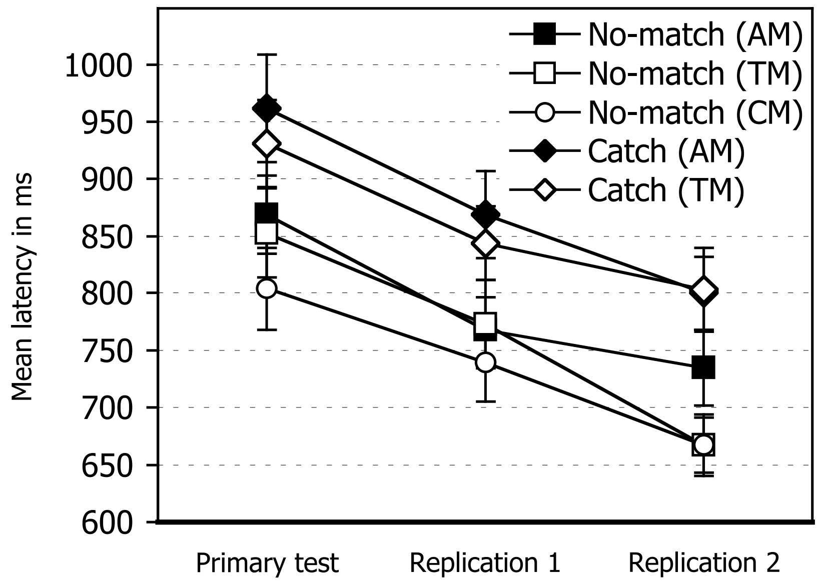

Figure 18. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for primary test, replication 1 and replication 2 in Experiment 4. Error bars represent the standard error of the mean.

Again, the ratio of English friends and enemies is expected to influence no-match-trial performance. As can be seen in Figure 18, this was indeed the case. For the primary test, response latencies were longer for AM-words (e.g., BLOOD), than for TM-words (e.g., MOOD). Also, response latencies for CM-words (e.g., MOON) were, in contrast to Experiment 3, shorter than for TM-words.

Omnibus analysis of variance. The mean no-response latencies were subjected to statistical analysis. Table 28 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 3 (Repetition: primary test vs. repetition 1 vs. repetition 2) repeated-measures ANOVA. As can be confirmed by looking at Table 28, preliminary analyses on the procedural variable indicated that there was no evidence for a substantial Sequence effect or an effect of temporal position of trial block (Block Position), but there was evidence for an effect of Repetition. Specifically, participants produced longer yes-response latencies in the primary test (M = 866 ms) than in repetition 1 (M = 786 ms) and repetition 2 (M = 721 ms).

Returning to the data of the no-match trials, there was a statistically significant main effect of Word Type (AM vs. TM), which, as can be confirmed by inspection of Table 28, accounted for a moderate percentage of variance. There was no evidence for a substantial Word Type by Replication interaction effect, or for a substantial three-way Trial Type by Word Type by Replication interaction effect.

Table 28.
Analysis of variance on correct no-response latencies for Experiment 4.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Repetition	210801.03	.70	1.39	151620.36	42.11	< .001	.689
Repetition × Participant	95124.30	.70	26.42	3601.00			
• Block Position	10138.41		1	10138.41	2.04	.170	.097
Block Position × Participant	94587.76		19	4978.30			
• Block Position × Repetition	2103.62	1.0	2.00	1051.81	.42	.661	.022
Block Position × Repetition × Participant	95555.72	1.0	38.00	2514.62			
• Sequence	21869.50		1	21869.50	.09	.767	.005
Participant(Group)	4357615.41		18	242089.75			
• Sequence × Repetition	12591.12	.74	1.48	8500.88	1.34	.272	.069
Sequence × Repetition × Participant(Group)	169411.57	.74	26.66	6354.33			
• TT	491324.50		1	491324.50	68.60	< .001	.783
TT × Participant	136091.25		19	7162.70			
• WT	28754.70		1	28754.70	4.41	.049	.188
WT × Participant	123979.71		19	6525.25			
• TT × WT	1339.54		1	1339.54	.13	.723	.007
TT × WT × Participant	196550.21		19	10344.75			
• TT × Repetition	2975.56	1.0	2.00	1487.78	.28	.759	.014
TT × Repetition × Participant	203092.94	1.0	38.00	5344.55			
• WT × Repetition	5137.71	1.0	2.00	2568.85	.71	.497	.036
WT × Repetition × Participant	137065.13	1.0	38.00	3606.98			
• TT × WT × Repetition	29318.73	.96	1.92	15273.87	1.94	.159	.093
TT × WT × Repetition × Participant	286887.76	.96	36.47	7866.16			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Planned contrasts. The overall Word Type effect for the no-match trials was, exclusively for the primary test, further inspected with three (Bonferroni-adjusted) pairwise comparisons (which included CM-words). The alpha level was again set to .017. For the **AM > TM** contrast there was a difference of **17** ms, with a 95% SCI of -53 to 87, that was not statistically significant (F(1,19) = 0.40, MSE = 7156.98, p = .536). The **66** ms difference (95% SCI 21 to 110) for the **AM > CM** contrast was statistically significant (F(1,19) = 15.25, MSE = 2816.53, p < .001), but the **49** ms difference for the **TM > CM** contrast (95% SCI -5 to 102) was not (F(1,19) = 5.71, MSE = 4150.73, p = .027).

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, recall that mismatch-trial performance should be worse for catch trials than for no-match trials (contrast catch > no-match). Further, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping (catch (AM-words) > catch (TM-words)). Moreover, the Trial Type main effect was expected to be embedded in a

Trial Type by Word Type interaction effect, which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 18 shows, Trial Type indeed had an effect on no-response latencies. As in Experiment 3, participants produced longer response latencies on catch trials (M = 868 ms) than on no-match trials (M = 778 ms), in which the **catch > no-match** contrast gave an overall difference of **90** ms (MSE = 1194.86, 95% CI 68 to 113). Table 28 indicates that this statistically significant main effect accounted for a considerable percentage of variance. In contrast to the error data, there was no evidence for a substantial Trial Type by Replication interaction effect in the latency data. In the primary test, the contrast gave a difference of **86** ms (95% CI 51 to 121), in repetition 1 it was **85** ms (95% CI 43 to 128), and in repetition 2 it was **100** ms (95% CI 71 to 130).

Figure 18 also shows that, catch-trial performance was indeed worse for AM-words than for TM-words, but only for the primary test and repetition 1. The overall **catch (AM-words) > catch (TM-words)** contrast gave a difference of **17** ms, with a 95% CI of -19 to 53, that was, in contrast to experiment 3, not statistically significant ($F(1,19) = 0.99$, MSE = 2984.33, p = .3324). The same contrast was performed separately for the primary test, repetition 1, and repetition 2. For the primary test, this contrast gave a difference of **30** ms with a 95% CI of -26 to 86, for repetition 1 it was **25** ms with a 95% CI of -28 to 77, and for repetition 2 it was **-3** ms with a 95% CI of -58 to 52.

Finally, Figure 18 shows that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect was, contrary to expectation and also in contrast to the effect found in Experiment 3, not larger for AM-words (86 ms) than for TM-words (95 ms), with a difference (in opposite direction) of **-9** ms (MSE = 6885.99, 95% CI -64 to 45). Table 28 indicates that this interaction effect was not statistically significant and accounted for a moderate percentage of variance. The same contrast was performed separately for the primary test, repetition 1, and repetition 2. For the primary test, this contrast gave a difference of **13** ms with a 95% CI of -73 to 100, for repetition 1 it was **-30** ms with a 95% CI of -58 to 117, and for repetition 2 it was **-71** percentage points with a 95% CI of -156 to 13.

Discussion

The primary test of Experiment 4 was an abbreviated version of Experiment 3. In many respects, the results of the primary test resembled the basic findings of Experiment 3. For the match and no-match trials, the patterns of error rates and response latencies of Experiments 3 and 4 showed reasonable similarity, although in Experiment 4 statistical support for an effect of English enemies on word perception was somewhat reduced. This was not entirely unexpected, because Experiment 4 only used a quarter of the number of participants that was tested in Experiment 3, resulting in considerably lower statistical power and precision. Experiment 4 also reproduced a sizeable effect of Trial Type. Relative to no-match trials, performance on catch trials was associated with elevated false-positive error rates and longer no-response latencies. However, the observed Trial Type effect in Experiment 4 was considerably smaller than in Experiment 3. Furthermore, in contrast to Experiment 3, no evidence was obtained in Experiment 4 that catch-trial performance was worse for AM-words than for TM-words, or that the Trial Type effect was larger for AM-words than for TM-words. Presumably, the modification of the experimental design that was utilised in Experiment 4 is to be held responsible for the different results patterns. To conclude, the design of Experiment 4 appears suitable for yielding an effect of English enemy neighbors on match-trial and no-match-trial performance and is capable of demonstrating an overall effect of Trial Type, and may thereby provide support for the idea that, in Dutch-English bilinguals, processing of an inconsistent English word indeed involves auxiliary coding of inappropriate phonology resulting from manifold spelling-to-sound associations.

Experiment 4 also examined the impact of presenting the entire experimental procedure three times to the participants, which, in fact, is the course of action taken in Experiments 5 and 7. In general, as might be expected, performance on all three trial types improved over repetition. Nevertheless, the effects of Word Type in match trials and no-match trials did not change markedly across repetitions. In contrast, the Trial Type effect appeared to diminish across repetitions, but only so in error rates. Hence, effects of Trial Type and Word Type can be observed across repetitions, and therefore we proceed to employ this procedure in Experiments 5 and 7 for the purpose of establishing a within-participants SOA manipulation.

EXPERIMENT 5

Mismatch-Trial Performance of Dutch-English Bilinguals and Native English Speakers: SOA Effects

Experiment 5 used the basic form (i.e., the primary test) of Experiment 4 to explore the effects of varying stimulus-onset asynchrony (SOA) of the visually presented printed word and auditory presented spoken rime. Furthermore, Experiment 5 tested two language groups, monolingual English speakers and bilingual Dutch speakers with English as their second language. Thus, in addition to investigating the SOA effect, Experiment 5 compared Dutch-English bilinguals with native English speakers on performing the print-to-speech correspondence task with respect to effects of Word Type, Trial Type, and SOA.

In the previous experiments (Experiments 2-4) employing the print-to-speech correspondence task, printed words and spoken rimes were always presented simultaneously over the visual and auditory modalities. Because of this temporal coincidence, it can be conceived that the coding generated by the spoken rime may act as a *context* for processing of the printed word. In other words, in this specific version of the task, the process of word perception may be affected by processing of the spoken rime. Consequently, for a match trial, a congruent spoken rime (e.g., /Ed/, derived from SAID) may possibly exert a positive influence on processing of a printed word (e.g., SAID), by aiding the appropriate orthographic-phonologic coding.

However, a simultaneously presented spoken rime may also act to obstruct perception of a printed word. This may be the case for all trial types if, in general, simultaneous bimodal information processing requires additional attention. What is more, interference in word processing for no-match trials may not only occur because of increased attentional demands but also by imposing an extraneous, unrelated candidate phonological structure. In like manner, for catch trials a spoken rime (e.g., /}d/, derived from BLOOD) may exert a negative influence on the processing of a printed word (e.g., MOOD) by imposing an extraneous candidate phonological structure that actually emerges in the initial conditions of word perception, thereby potentially fostering inappropriate phonological codings.

The SOA manipulation of Experiment 5 was aimed at exploring whether there is a change in match-trial, no-match-trial, and catch-trial performance if the spoken rime is presented not simultaneously with the printed word, but after a substantial time gap, thus moving the context of the spoken rime away from the relatively early phases of visual word processing. This aim relates to the third main research question specified in the Introduction section. If processing of the spoken rime is suspended, match and no-match-trial performance may improve because, under such conditions, processing of the printed word proceeds fairly unobstructed by processing of the spoken rime, yielding a stable, appropriate phonological coding that arises quickly. For the same reasons, suspension of presentation of the spoken rime also may improve catch-trial performance. However, besides this possible gain yielded by loosened attentional demands, catch-trial performance may improve because processing of the printed word takes place in relative absence of the context that comes forth by processing of the spoken rime. This improvement may occur if the appropriate phonological coding that arises in word perception gets increasingly stable in the course of word processing, thus reducing the impact of an imposing extraneous candidate phonological structure. Nevertheless, if we assume that word perception is a continuous, metastable process, we may observe that a spoken rime derived from an enemy neighbor still acts to restore the diminished, inappropriate local orthographic-phonologic coding to such a level that it undermines the ability to perceive a mismatch. Finally, for match trials, suspending presentation of the spoken rime may not improve but rather hurt performance, because the appropriate orthographic-phonologic coding can no longer benefit from congruence of print and speech.

A second way of presenting the spoken rime distant from print is not by suspension of presentation of the spoken rime but by suspension of presentation of the printed word. Thus, in this instance, the spoken rime is presented first and the printed word follows after a substantial time gap. This SOA condition is similar to the previous one in that processing of the printed word proceeds fairly unobstructed, for example due to relaxed attentional demands. The main difference, however, is that if the spoken rime is presented first the resulting coding still acts as a context for processing of the printed word, the same way as in the condition in which the printed word and spoken rime are presented simultaneously. Therefore, for a no-match trial, word perception should proceed fairly unobstructed. This is not expected in the case of catch trials where a spoken rime may exert a negative influence on processing of a printed word by imposing an extraneous candidate phonological structure. Finally, for a match trial, a congruent spoken rime may exert a positive influence on processing of the subsequent printed word, by aiding the appropriate orthographic-phonologic coding. Furthermore, perception of the printed word may proceed fairly unobstructed, again for example due to loosened attentional demands.

In sum, Experiment 5 contrasted three SOAs (see Table 29). In SOA 1, speech was initiated first and print followed after an interval of approximately 500 ms. In SOA 2 (which uses an SOA identical to Experiment 4) print and speech were initiated simultaneously. In SOA 3, print was initiated first and speech followed after an interval of approximately 500 ms. Thus, the difference between the three SOAs was that, in contrast to SOAs 1 and 2, early processing of the printed word in SOA 3 proceeds without contextual interference from simultaneous or earlier processing of the spoken rime. Further, SOA 2 was different from the other two SOAs in that print and speech were presented simultaneously whereas in the other SOAs they were not. Again, simultaneous presentation of print and speech may result in increased attentional demands, thereby hindering performance.

Table 29.
Alternating stimulus onset asynchrony (SOA) for Experiments 5 and 7.

	Time onset Print	Time onset Sound
Sound earlier than Print (SOA1)	t = 0 ms	t = -500 ms
Sound simultaneously with Print (SOA2)	t = 0 ms	t = 0 ms
Sound later than Print (SOA3)	t = 0 ms	t = 500 ms

For all types of trials, performance may be poorest in SOA 2 where simultaneous, bimodal processing of information may require increased attentional demands. However, for no-match trials, we did not expect an influence of contextual interference from processing of spoken rime, because print and speech were unrelated to each other. Such an influence was however expected for catch trials, since the spoken rimes were derived from enemy neighbors, and inappropriate phonology resulting from spelling-to-sound knowledge of the same enemy neighbors is assumed to emerge in the processing of inconsistent words. Also, for match trials an influence of contextual interference may be possible, because the appropriate phonological coding and the coding generated from processing the spoken rime are congruent with each other. Thus, for no-match trials the contrast SOA 1 < SOA 2 > SOA 3 may be expected. For catch trials this contrast was expected too, but also the contrast SOA 1 > SOA 3. Finally, for match trials the contrast SOA 1 < SOA 2 < SOA 3 may be expected if contextual interference from processing of the spoken rime does occur and the contrast SOA 1 < SOA 2 > SOA 3 if it does not occur.

Method

Participants and materials

A group of 36 Dutch-English bilinguals (Dutch participants) and a group of 24 native speakers of English (USA participants) served as participants. They were presented with the same materials as used in Experiment 4.

Experimental design

In Experiment 5 the same basic design was used as in Experiment 4. Again, eight groups of trials were created that represented specific combinations of Trial Type and Word Type. These eight combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). Table 30 presents the layout of the experimental design. As in Experiment 4, each participant was presented with each of the two trial blocks. Hence, for each participant data was obtained for each available combination of trial type and word type, using each word only once. Furthermore, the participants were observed in all three SOA conditions, in which they were tested three times on the same basic form of Experiment 4. Thus, SOA condition was treated as a within-subjects variable. The temporal order of the two trial blocks was counterbalanced across two different participant groups according to a single Latin square (participant group 1: Sequence A-B; participant group 2: Sequence B-A). Furthermore, within the two participant groups, the temporal order of SOA conditions was also Latin square counterbalanced (see Table 30) across six different participant subgroups (SOA sequence 1-2-3, SOA sequence 2-3-1, SOA sequence 3-1-2, SOA sequence 1-3-2, SOA sequence 3-2-1, and SOA sequence 2-1-3).

Table 30.
SOA block ordered by block position according to a standard Latin square (Sequences 1, 2, and 3) and a non-standard Latin square (Sequences 4, 5, and 6) for Experiments 5 and 7. The temporal order of SOA block is Latin-square counterbalanced across six different participant groups (Participant Group 1: Sequence 1-2-3; Participant Group 2: Sequence 2-3-1; Participant Group 3: Sequence 3-1-2; Participant Group 4: Sequence 1-3-2; Participant Group 5: Sequence 3-2-1; Participant Group 6: Sequence 2-1-3).

	Temporal Position of SOA		
Sequence	P1	P2	P3
S1	SOA1	SOA2	SOA3
S2	SOA2	SOA3	SOA1
S3	SOA3	SOA1	SOA2
S4	SOA1	SOA3	SOA2
S5	SOA3	SOA2	SOA1
S6	SOA2	SOA1	SOA3

Participants were randomly assigned to the different sequences. Again, the counterbalancing procedure was intended to disentangle the effect of temporal position of the procedural variables from the effects of the independent variables. This potential source of variance can be isolated and removed from the estimate of error variance, which may improve the efficiency of the design. Again, this was accomplished by adding participant subgroup (involving six SOA sequences) as a between-subjects variable in an ANOVA, and testing the effects against the resulting treatments×participants(group) error term.

Results

As in Experiments 2-4, participants were presented with eight different groups of trials representing specific combinations of Trial Type and Word Type: match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). Each participant responded three times (i.e., in SOA 1, SOA 2, and SOA 3) to 60 match trials which consisted of 20 AM-words (e.g., PAID), 20 TM-words (e.g., SAID), and 20 CM-words (e.g., STAIN). In addition, each participant responded three times (i.e., in SOA 1, SOA 2, and SOA 3) to 40 no-match trials and 20 catch trials. The no-match trials consisted of 10 AM-words (e.g., BLOOD), 10 TM-words (e.g., MOOD), and 20 CM-words (e.g., MOON), and the catch trials consisted of 10 AM-words (e.g., BLOOD) and 10 TM-words (e.g., MOOD). For each participant in each SOA condition, the correct response latencies within these eight groups were averaged and percentage of errors was calculated for each of the eight groups. Hence, the data that entered the statistical analyses consisted for each participant of nine latency means and nine percentages of false-negatives for match trials (e.g., for PAID, SAID, and STAIN), nine latency means and nine percentages of false-positives for no-match trials (e.g., for BLOOD, MOOD, and MOON), and six latency means and six percentages of false-positives for catch trials (e.g., for BLOOD and MOOD).

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. This resulted for the USA and Dutch participants in a rejection of a total of respectively 12.9% and 16.9% for the catch trials, respectively 6.2% and 4.5% for the no-match trials, and respectively 2.7% and 5.0% for the match trials. Further, for the USA and Dutch participants, less than 0.0% and 0.3%, respectively, of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. In all SOA conditions of the print-to-speech correspondence task, a trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. In the SOAs 1 and 2, this experimental procedure resulted in a cut-off that rejected all latencies greater than 2000 ms and in SOA 3 it resulted in a cut-off that rejected all latencies greater than 2500 ms. We did not consider further truncation, because the procedure resulted for the USA and Dutch participants in rejection of 0.4% (SOA 1: 0.1%, SOA 2: 0.5%, SOA 3: 0.6%) and 0.3% (SOA 1: 0.3%, SOA 2: 0.3%, SOA 3: 0.2%), respectively, of the correct response latencies and, as recommended by Ulrich and Miller (1994), this percentage should not be exceeded.

Data of match trials

A participant produced an error (i.e., a false-negative) in a match trial when he or she pressed the “no” button when presented with a printed word and a spoken rime that were actually congruent with each other (e.g., PAID - /ed/, SAID - /Ed/, and STAIN - /en/). The mean percentages of false-negatives as a function of Word Type and SOA condition, separately for each trial block, are presented in Table 13 of Appendix C. The mean latencies are presented in Table 14 of Appendix C. Figure 19 shows the mean percentages of false-negatives for the AM-words, TM-words, and CM-words for the three SOA conditions, separately for USA participants and Dutch participants. These mean percentages of false-negatives were collapsed over trial block and participant (sub)group.

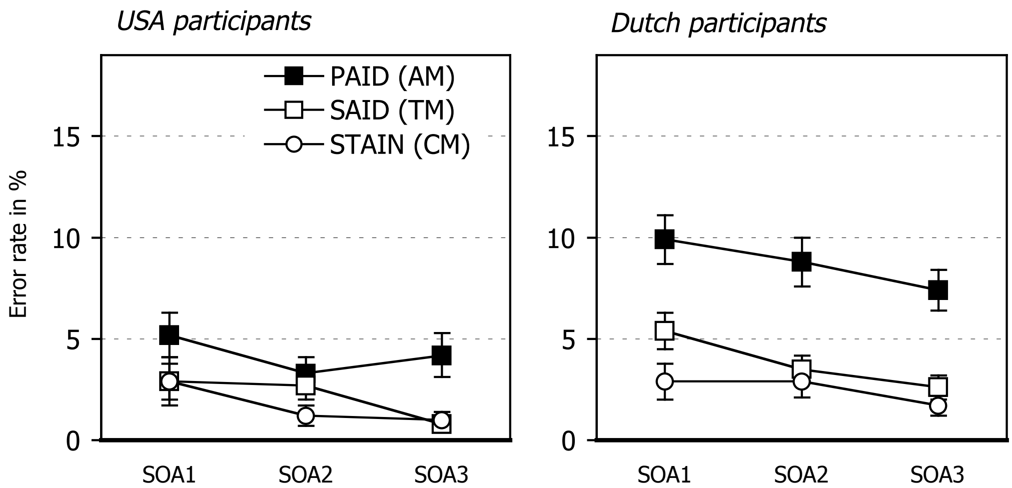

Figure 19. Mean percentages of false-negatives as a function of Word Type (AM-words vs. TM-words vs. CM-words) and SOA (SOA1 vs. SOA2 vs. SOA3) for USA participants (left panel) and Dutch participants (right panel) in Experiment 5. Error bars represent the standard error of the mean.

As in Experiments 3 and 4, match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 19, both for the USA participants and the Dutch participants the ratio of friends and enemies was indeed associated with the number of false-negatives. Replicating Experiments 3 and 4, USA and Dutch participants made overall more errors on AM-words such as PAID than on TM-words such as SAID. Furthermore, participants made fewer errors on CM-words such as STAIN than on words with typical mappings.

Omnibus analysis of variance. The nine mean percentages of false-negatives obtained for combinations of Word Type and SOA were subjected to statistical analysis. Tables 31 (USA participants) and 32 (Dutch participants) present the results of a 3 (Word Type: AM vs. TM vs. CM) by 3 (SOA condition: SOA 1 vs. SOA 2 vs.

SOA 3) repeated-measures ANOVA. The tables also provide the results of non-parametric tests.

As can be confirmed by inspecting Tables 31 and 32, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, or for a substantial effect of temporal position of SOA block (SOA Session). Specifically, USA and Dutch participants produced comparable numbers of errors across the three sessions (USA participants: M_1 = 3.6%, M_2 = 2.2%, M_3 = 2.4%; Dutch participants: M_1 = 5.9%, M_2 = 4.7%, M_3 = 4.4%). Although temporal position of SOA block did not account for much variance, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Planned contrasts. Returning to the results presented in Figure 19, for both language groups there was a statistically significant main effect of Word Type, which, as can be seen in Tables 31 and 32, accounted for a considerable percentage of variance. This overall effect was further inspected with the same three (Bonferroni-adjusted) pairwise comparisons performed in Experiments 3 and 4. The alpha level was set to .017. Starting with the data of the *USA participants*, for the **AM > TM** contrast there was a statistically significant difference of **2.1** percentage points, with a 95% SCI of 0.4 to 3.8 (F(1,18) = 10.15, MSE = 5.13, p = .005).

Table 31.
Analysis of variance on percentages of false-negatives for Experiment 5 (USA participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Session	29.85	1.0	2.00	14.93	2.54	.093	.368	.124
Session × Participant(Group)	211.53	1.0	36.00	5.88				
• Sequence	45.49		5	9.10	.24	.940	.617	.062
Participant(Group)	684.03		18	38.00				
• SOA	108.33	1.0	2.00	54.17	3.07	.059	.304	.146
SOA × Participant(Group)	634.72	1.0	36.00	17.63				
• Word Type	258.33	1.0	2.00	129.17	11.96	< .001	.015	.399
Word Type × Participant(Group)	388.89	1.0	36.00	10.80				
• Word Type × SOA	47.92	1.0	4.00	11.98	.97	.428		.051
Word Type × SOA × Participant(Group)	886.11	1.0	72.00	12.31				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (χ|H0)* for Kruskal-Wallis test and *p (χ²|H0)* for Friedman test.

Table 32.
Analysis of variance on percentages of false-negatives for Experiment 5 (Dutch participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Session	48.75	1.0	2.00	24.37	3.10	.053	.004	.094
Session × Participant(Group)	472.27	1.0	60.00	7.87				
• Sequence	136.11		5	27.22	.27	.927	.940	.043
Participant(Group)	3041.67		30	101.39				
• SOA	256.02	1.0	2.00	128.01	5.42	.007	.017	.153
SOA × Participant(Group)	1416.67	1.0	60.00	23.61				
• Word Type	2264.35	.75	1.50	1507.05	38.57	< .001	< .001	.563
Word Type × Participant(Group)	1761.11	.75	45.08	39.07				
• Word Type × SOA	40.74	1.0	4.00	10.19	.620	.649		.020
Word Type × SOA × Participant(Group)	1972.22	1.0	120.00	16.44				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p* $(\chi|H0)$ for Kruskal-Wallis test and *p* $(\chi^2|H0)$ for Friedman test.

The **2.5** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 1.1 to 3.9 ($F(1,18) = 23.14$, $MSE = 3.24$, $p < .001$), but the **0.4** difference in percentage points for the **TM > CM** contrast was not statistically significant, with a 95% SCI of -0.8 to 1.6 ($F(1,18) = 0.86$, $MSE = 2.43$, $p = .367$). Turning to the data of the *Dutch participants*, for the **AM > TM** contrast there was a statistically significant difference of **4.8** percentage points, with a 95% SCI of 2.7 to 6.9 ($F(1,30) = 33.38$, $MSE = 12.50$, $p < .001$). The **6.2** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 3.9 to 8.4 ($F(1,30) = 47.02$, $MSE = 14.51$, $p < .001$), and so was the **1.3** difference in percentage points for the **TM > CM** contrast, with a 95% SCI of 0.4 to 2.3 ($F(1,30) = 13.88$, $MSE = 2.34$, $p < .001$).

The same overall Word Type analyses were performed on correct yes-response latencies. For both language groups there was a statistically significant main effect of Word Type, which accounted for a considerable percentage of variance (USA participants: $\varepsilon = 1.0$, $F(2,36) = 16.92$, $MSE = 1212.22$, $p < .001$, $\eta^2 = .280$; Dutch participants: $\varepsilon = 1.0$, $F(2,60) = 41.23$, $MSE = 2215.25$, $p < .001$, $\eta^2 = .579$). Again, there was no evidence for a substantial Word Type by SOA interaction effect (USA participants: $\varepsilon = 1.0$, $F(4,72) = 0.40$, $MSE = 1230.04$, $p = .805$, $\eta^2 = .022$; Dutch participants: $\varepsilon = 1.0$, $F(4,120) = 0.48$, $MSE = 1384.81$, $p = .749$, $\eta^2 = .016$). These overall effects were further inspected with three (Bonferroni-adjusted) pairwise comparisons. Starting with the data of the *USA participants*, for the **AM > TM** contrast there was a difference of **13** ms, with a 95% SCI of -2 to 29, that was not statistically significant ($F(1,18) = 5.35$, $MSE = 397.47$, $p = .033$). The **34** ms difference for the **AM > CM** contrast was statistically significant, with a 95% SCI of 18 to 49 ($F(1,18) = 31.92$, $MSE = 424.13$, $p < .001$), and so was the **20** ms difference for the **TM > CM** contrast, with a 95% SCI of 5 to 35 (F(1,18) =12.87, MSE = 383.09, p = .002). Turning to the data of the *Dutch participants*, for the **AM > TM** contrast there was a statistically significant difference of **21** ms, with a 95% SCI of 6 to 35 (F(1,30) = 12.99, MSE = 585.50, p = .001). The **57** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 42 to 73 (F(1,30) = 88.14, MSE = 673.19, p < .001), and so was the **37** ms difference for the **TM > CM** contrast, with a 95% SCI of 18 to 55 (F(1,30) = 25.54, MSE = 957.38, p < .001).

SOA effect on error rates for match trials

Turning now to the principal analyses that compared the three SOA conditions, Figure 19 also shows the effect of SOA on the number of false-negative errors. We did not evaluate the SOA effect on response latencies because the timing of the different events in a trial was not identical across SOA conditions. As can be seen in Figure 19, both for the USA participants and the Dutch participants, error rates were higher for SOA 1 than for SOA 2, of which the error rates were higher than for SOA 3. This pattern of error rates, however, was not as expected. For match trials we expected the contrast SOA 1 < SOA 2 < SOA 3 if contextual interference from processing of the spoken rime does occur and the contrast SOA 1 < SOA 2 > SOA 3 if it does not occur. Tables 31 and 32 show that the main effect of SOA was not statistically significant for the USA participants but it was for the Dutch participants. As can be further seen in these tables, the SOA main effect accounted for a moderate percentage of variance. Furthermore, for both language groups, there was no evidence for a substantial Word Type by SOA interaction effect.

The overall effect of SOA was further inspected with orthogonal polynomial contrasts that evaluated the linear effect (e.g., SOA 1 > SOA 3, $\square_{\textbf{linear}}$) and the quadratic effect (e.g., SOA 1 < SOA 2 > SOA 2, $\square_{\textbf{quadratic}}$) across the three levels of SOA. For the USA participants the linear effect (SOA 1 > SOA 3) gave an estimated value for $\square_{\textbf{linear}}$ of **1.7** percentage points, with a 95% CI of -0.1 to 3.4, that was not statistically significant (F(1,18) = 3.89, MSE = 8.56, p = .064, $\square^2$ = .178). Further, no evidence was obtained for a quadratic effect. The estimated value of $\square_{\textbf{quadratic}}$ was **0.8** percentage points, with a 95% CI of -1.0 to 2.7, that was not statistically significant (F(1,18) = 0.87, MSE = 3.19, p = .363, $\square^2$ = .046). For the Dutch participants a statistically significant linear effect (SOA 1 > SOA 3) was obtained that gave an estimated value for $\square_{\textbf{linear}}$ of **2.2** percentage points, with a 95% CI of 0.9 to 3.5 (F(1,30) = 11.64, MSE = 7.32, p = .002, $\square^2$ = .280). Again, there was no evidence for a quadratic effect. The estimated value of $\square_{\textbf{quadratic}}$ was **0.1** percentage points, with a 95% CI of -2.3 to 2.6, that was not statistically significant (F(1,30) = 0.01, MSE = 8.42, p = .907, $\square^2$ = .000).

Data of mismatch trials

A participant produced an error (i.e., a false-positive) in a mismatch trial when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other (e.g., BLOOD - /Yd/, BLOOD - /ud/, MOOD - /Yd/, MOOD - /}d/, and MOON - /en/). The mean percentages of false-positives as a function of Trial Type, Word Type, and SOA condition, separately for each trial block, are presented in Table 15 of Appendix C. The mean latencies are presented in Table 16 of Appendix C. Figure 20 (see also Figure 21) shows the mean percentages of false-positives for the AM-words and TM-words both for the no-match trials (including CM-words) and the catch trials, and as a function of SOA. These mean percentages of false-positives were collapsed over trial block and participant (sub)group.

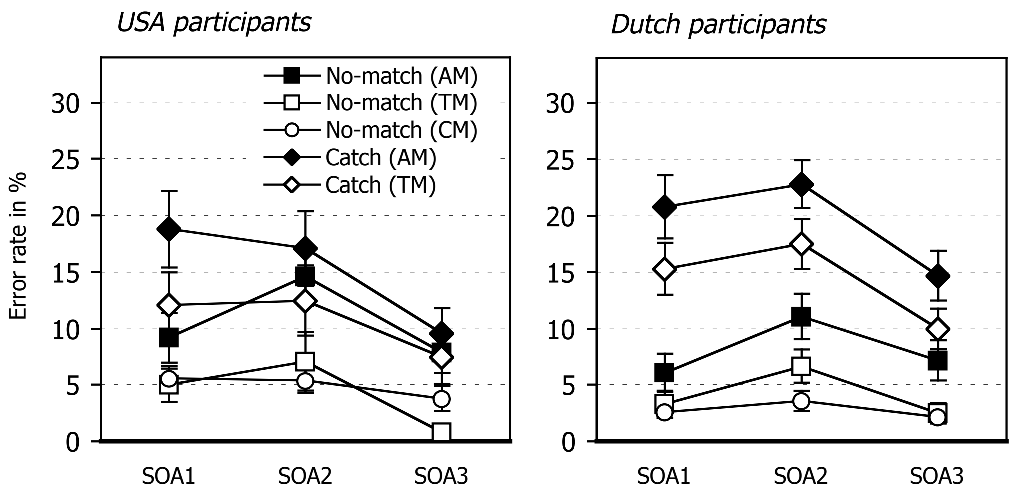

Figure 20. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials), Word Type (AM-words vs. TM-words vs. CM-words) and SOA (SOA1 vs. SOA2 vs. SOA3) for USA participants (left panel) and Dutch participants (right panel) in Experiment 5. Error bars represent the standard error of the mean.

As in Experiments 3 and 4, no-match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 21, both for the USA participants and the Dutch participants the ratio of friends and enemies was indeed associated with the number of false-positives. Replicating Experiments 3 and 4, USA and Dutch participants made more errors on AM-words such as BLOOD than on TM-words such as MOOD and CM-words such as MOON. However, the number of errors for CM-words was not markedly lower than for the words with typical mappings.

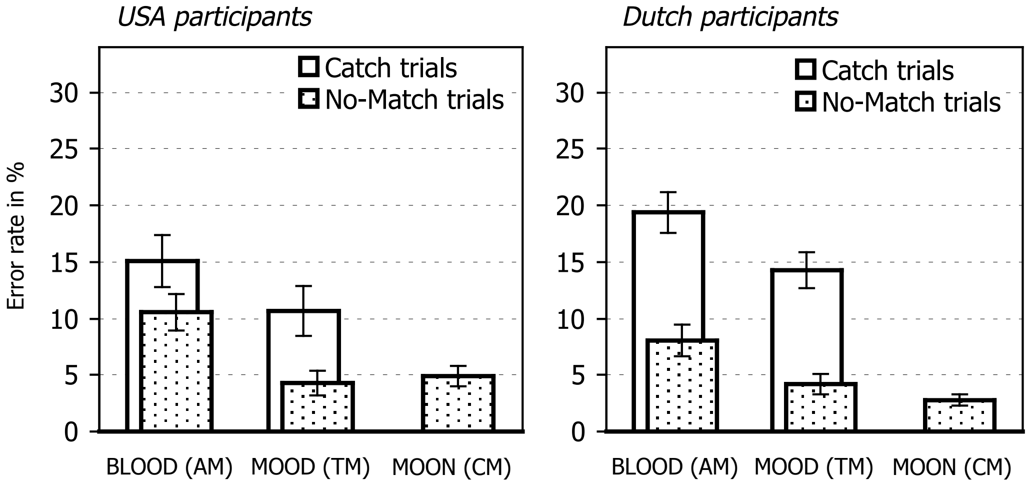

Figure 21. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) and collapsed over SOA for USA participants (left panel) and Dutch participants (right panel) in Experiment 5. Error bars represent the standard error of the mean.

Omnibus analysis of variance. The mean percentages of false-positives were subjected to statistical analysis. Tables 33 (USA participants) and 34 (Dutch participants) present the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 3 (SOA condition: SOA 1 vs. SOA 2 vs. SOA 3) repeated-measures ANOVA. The tables also provide the results of non-parametric tests. As can be confirmed by inspecting Tables 33 and 34, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of SOA block (SOA Session) which accounted for a substantial percentage of variance. Specifically, USA and Dutch participants produced fewer errors across sessions (USA participants: M_1 = 11.3%, M_2 = 8.3%, M_3 =5.7%; Dutch participants: M_1 = 10.6%, M_2 = 8.8%, M_3 = 6.3%). Further, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Table 33.
Analysis of variance on percentages of false-positives for Experiment 5 (USA participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Session	380.68	1.0	2.00	190.34	17.37	< .001	.017	.491
Session × Participant(Group)	394.46	1.0	36.00	10.96				
• Sequence	3697.57		5	739.51	1.47	.247	.067	.290
Participant(Group)	9035.42		18	501.97				
• SOA	2104.86	1.0	2.00	1052.43	12.09	< .001	.030	.402
SOA × Participant(Group)	3133.33	1.0	36.00	87.04				
• TT	2167.01		1	2167.01	9.75	.006	.134	.351
TT × Participant(Group)	4002.08		18	222.34				
• WT	2058.68		1	2058.68	25.09	< .001	< .001	.582
WT × Participant(Group)	1477.08		18	82.06				
• TT × WT	58.68		1	58.68	.70	.414	.664	.037
TT × WT × Participant(Group)	1510.42		18	83.91				
• TT × SOA	292.36	1.0	2.00	146.18	1.87	.169	.172	.094
TT × SOA × Participant(Group)	2816.67	1.0	36.00	78.24				
• WT × SOA	25.69	1.0	2.00	12.85	.27	.762	.913	.015
WT × SOA × Participant(Group)	1691.67	1.0	36.00	46.99				
• TT × WT × SOA	179.86	.96	1.92	93.59	.90	.413	.316	.047
TT × WT × SOA × Participant(Group)	3608.33	.96	34.59	104.31				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p* (χ^2*|H0)* for Friedman test and *p (□|H0)* for sign test.

Returning to the data of the mismatch trials, for both language groups there was a statistically significant main effect of Word Type, which, as can be confirmed by inspection of Tables 33 and 34, accounted for a considerable percentage of variance. There was no evidence for a substantial Word Type by SOA interaction effect, or of a substantial three-way Trial Type by Word Type by SOA interaction effect.

Planned contrasts. The overall Word Type effect for the no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons, which included CM-words. Starting with the data of the *USA participants*, for the **AM > TM** contrast there was a statistically significant difference of **6.2** percentage points, with a 95% SCI of 2.3 to 10.2 ($F(1,18) = 17.41$, $MSE = 26.93$, $p < .001$).

Table 34.
Analysis of variance on percentages of false-positives for Experiment 5 (Dutch participants).

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	*p*[a]	η_p^2
• Session	336.61	1.0	2.00	168.31	12.46	< .001	< .001	.293
Session × Participant(Group)	810.78	1.0	60.00	13.51				
• Sequence	2531.71		5	506.34	.85	.523	.533	.125
Participant(Group)	17798.61		30	593.29				
• SOA	2511.57	1.0	2.00	1255.79	14.20	< .001	< .001	.321
SOA × Participant(Group)	5305.56	1.0	60.00	88.43				
• TT	12352.08		1	12352.08	92.67	< .001	< .001	.755
TT × Participant(Group)	3998.61		30	133.29				
• WT	2268.75		1	2268.75	19.55	< .001	.001	.395
WT × Participant(Group)	3481.94		30	116.07				
• TT × WT	39.12		1	39.12	.40	.531	.597	.013
TT × WT × Participant(Group)	2926.39		30	97.55				
• TT × SOA	629.17	.99	1.97	318.83	3.78	.029	.024	.112
TT × SOA × Participant(Group)	4988.89	.99	59.20	84.27				
• WT × SOA	9.72	1.0	2.00	4.86	.07	.933	.511	.002
WT × SOA × Participant(Group)	4205.56	1.0	60.00	70.09				
• TT × WT × SOA	36.57	1.0	2.00	18.29	.30	.739	.648	.010
TT × WT × SOA × Participant(Group)	3611.11	1.0	60.00	60.19				

Note. Type III Sum of Squares with Huynh-Veldt ε adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p* (χ*\|H0)* for Kruskal-Wallis test, *p* (χ^2*\|H0)* for Friedman test and *p* (□*\|H0)* for sign test.

The **5.6** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 2.0 to 9.3 ($F(1,18) = 16.64$, $MSE = 22.82$, $p < .001$), but the **-0.6** difference in percentage points (in opposite direction) for the **TM > CM** contrast was not, with a 95% SCI of -3.0 to 1.7 ($F(1,18) = 0.48$, $MSE = 9.70$, p = .496). Turning to the data of the *Dutch participants*, for the **AM > TM** contrast there was a statistically significant difference of **4.0** percentage points, with a 95% SCI of 1.2 to 6.7 ($F(1,30) = 13.54$, $MSE = 21.08$, $p < .001$). The **5.3** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 2.5 to 8.1 ($F(1,30) = 23.28$, $MSE = 21.92$, $p < .001$), but the **1.3** difference in percentage points for the **TM > CM** contrast was not, with a 95% SCI of -0.6 to 3.3 ($F(1,30) = 3.14$, $MSE = 10.35$, p = .087).

The same overall Word Type analyses were performed on correct no-response latencies (see Figure 22). For both language groups there was a statistically significant main effect of Word Type, which accounted for a moderate percentage of variance (USA participants: $F(1,18) = 4.91$, $MSE = 3784.18$, $p = .040$, $\eta^2 = .214$; Dutch participants: $F(1,30) = 7.78$, $MSE = 3907.84$, $p = .009$, $\eta^2 = .206$). Again, there was no evidence for a substantial Word Type by SOA interaction effect (USA participants: $\varepsilon = 1.0$, $F(2,36) = 1.27$, $MSE = 3992.07$, $p = .293$, $\eta^2 = .066$; Dutch participants: □= 1.0, $F(2,60) = 0.21$, $MSE = 4130.16$, $p = .815$, $□^2 = .007$). These overall effects were, separately for the no-match trials, further inspected with (Bonferroni-adjusted) pairwise comparisons. Starting with the data of the *USA participants*, for the **AM > TM** contrast there was a statistically significant difference of **42** ms, with a 95% SCI of 6 to 77 ($F(1,18) = 9.41$, $MSE = 2203.20$, $p = .007$). The **50** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 21 to 79 ($F(1,18) = 21.18$, $MSE = 1426.83$ $p < .001$), but the **9** ms difference for the **TM > CM** contrast was not, with a 95% SCI of -14 to 32 ($F(1,18) = 0.97$, $MSE = 916.76$, $p = .337$). Turning to the data of the *Dutch participants*, for the **AM > TM** contrast there was a statistically significant difference of **35** ms, with a 95% SCI of 12 to 58 ($F(1,30) = 14.95$, $MSE = 1483.15$, $p < .001$). The **52** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 31 to 72 ($F(1,30) = 40.70$, $MSE = 1174.30$, $p < .001$), but the **16** ms difference for the **TM > CM** contrast was not, with a 95% SCI of -4 to 36 ($F(1,30) = 4.34$, $MSE = 1119.32$, $p = .046$).

Turning now to the analyses that contrasted no-match-trial and catch-trial performance, as in Experiments 3 and 4, mismatch-trial performance should be worse for catch trials than for no-match trials (catch > no-match). Further, catch-trial performance using a word with an atypical mapping should be worse than catch-trial performance using a word with a typical mapping (catch (AM-words) > catch (TM-words)). Moreover, the Trial Type main effect was expected to be embedded in a Trial Type by Word Type interaction effect, which is captured by the contrast Trial Type effect (AM-words) > Trial Type effect (TM-words).

As Figure 21 shows, Trial Type was indeed associated with different numbers of false-positives. Replicating Experiment 4, USA and Dutch participants made overall more errors on catch trials ($M_{\text{USA}} = 12.9\%$; $M_{\text{Dutch}} = 16.9\%$) than on no-match trials ($M_{\text{USA}} = 7.4\%$; $M_{\text{Dutch}} = 6.2\%$), in which the **catch > no-match** contrast gave an overall difference of **5.5** percentage points ($MSE = 37.06$, 95% CI 1.8 to 9.2) for the USA participants and **10.7** percentage points ($MSE = 22.21$, 95% CI 8.4 to 13.0) for the Dutch participants. Tables 33 and 34 indicate that these statistically significant main effects accounted for a considerable percentage of variance. Figure 21 also shows that, overall, catch-trial performance was worse for AM-words than for TM-words, both for the USA participants and the Dutch participants. For the USA participants the **catch (AM-words) > catch (TM-words)** contrast gave a difference of **4.4** percentage points with a 95% CI of 1.2 to 7.7, that was statistically significant ($F(1,18) = 8.35$, $MSE = 28.40$, $p = .010$) and for the Dutch participants the contrast gave a difference of **5.5** percentage points with a 95% CI of 1.8 to 8.6, that was also statistically significant ($F(1,30) = 9.66$, $MSE = 50.12$, $p = .004$). Similar effects were observed in Experiment 4, although these were not statistically supported. Finally, Figure 21 shows that with regard to the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, the overall Trial Type effect for the USA participants was, contrary to expectation, not larger for AM-words (4.6 percentage points) than for TM-words (6.4 percentage points), with a difference (in opposite direction) of **-1.8** percentage points (*MSE* = 55.94, 95% CI -6.3 to 2.7). For the Dutch participants the overall Trial Type effects for AM-words and TM-words were 11.3 and 10.1 percentage points, respectively, and the contrast gave a difference of **1.2** percentage points (*MSE* = 65.03, 95% CI -2.7 to 5.1). Tables 33 and 34 indicate that, as in Experiment 4, these interaction effects were not statistically significant and accounted for a low percentage of variance.

Turning to the latency data, in Figure 22 it can be seen that Trial Type also had an effect on no-response latencies. Overall, USA and Dutch participants produced longer latencies on catch trials (M_{USA} = 976 ms; M_{Dutch} = 966 ms) than on no-match trials (M_{USA} = 923 ms; M_{Dutch} = 877 ms), in which the **catch > no-match** contrast gave an overall difference of **53** ms (*MSE* = 1050.80, 95% CI 33 to 73) for the USA participants and **89** ms (*MSE* = 1564.15, 95% CI 70 to 108) for the Dutch participants. Tables 33 and 34 indicate that these statistically significant main effects accounted for a considerable percentage of variance.

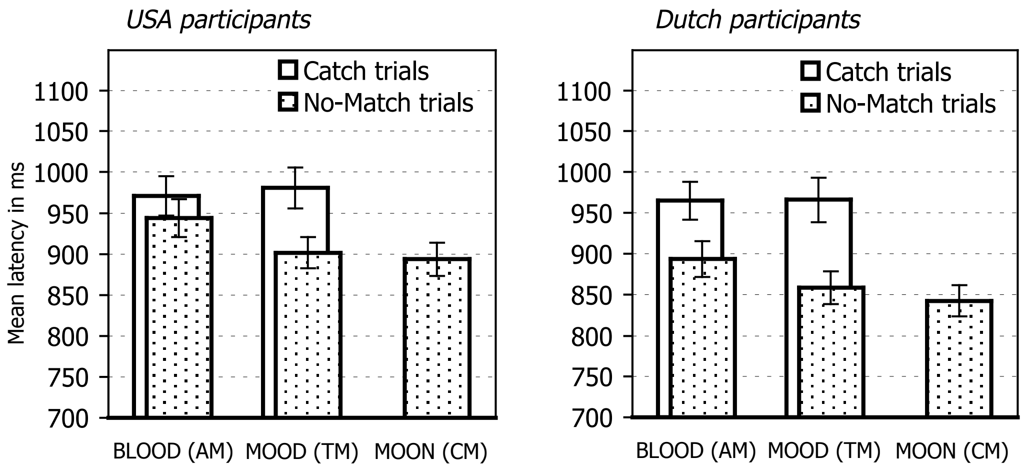

Figure 22. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) and collapsed over SOA for USA participants (left panel) and Dutch participants (right panel) in Experiment 5. Error bars represent the standard error of the mean.

Figure 22 further shows that, overall, catch-trial performance was, contrary to expectation, not worse for AM-words than for TM-words, both for the USA participants and the Dutch participants. For the USA participants the **catch (AM-words) > catch (TM-words)** contrast gave a difference (in opposite direction) of **-9** ms with a 95% CI of -43 to 24, that was not statistically significant ($F(1,18) = 0.34$, *MSE* = 3129.58, p = .565) and for the Dutch participants the contrast gave a difference (in opposite direction) of **-2** ms with a 95% CI of -29 to 26, that was also not statistically significant ($F(1,30) = 0.01$, *MSE* = 3353.22, p = .912). For the **Trial**

Type effect (AM-words) > Trial Type effect (TM-words) contrast, the overall Trial Type effect for the USA participants was, contrary to expectation, not larger for AM-words (27 ms) than for TM-words (78 ms), with a difference (in opposite direction) of **-51** ms (*MSE* = 8134.84, 95% CI -106 to 4). Likewise, for the Dutch participants, the overall Trial Type effects for AM-words and TM-words were 70 and 107 ms, respectively, and the contrast gave a difference (in opposite direction) of **-37** ms (*MSE* = 7068.60, 95% CI -77 to 4). Tables 33 and 34 indicate that these interaction effects were not statistically significant and accounted for a moderate percentage of variance. Finally, there was no evidence for a substantial three-way interaction effect of Trial Type, Word Type and SOA.

SOA effect on error rates for mismatch trials

Turning now to the principal analyses that compared the three SOA conditions, Figure 23 shows the effect of SOA on the number of false-positive errors. In these analyses, data of CM-words were excluded. Again, we did not evaluate the SOA effect on response latencies because the timing of the different events in a trial was not identical across SOA conditions. As can be seen in Figure 23, both for the USA participants and the Dutch participants, error rates were, as predicted, overall higher for SOA 2 than for SOAs 1 and 3. Further, error rates were lowest in SOA 3 and, as expected, in particular for catch trials. Tables 33 and 34 show that the main effect of SOA was statistically significant, both for the USA participants and for the Dutch participants. As can be seen in these tables, the SOA main effects accounted for a considerable percentage of variance. Furthermore, for the Dutch participants but not for the USA participants, there was evidence for a substantial Trial Type by SOA interaction effect. Inspection of Figure 23 indicates that for both language groups the Trial Type effect was largest in SOA 1 and, for the Dutch participants, smallest in SOA 3.

The overall effect of SOA was further inspected with orthogonal polynomial contrasts that evaluated, separately for the no-match trials and for catch trials, the linear and quadratic effects across the three levels of SOA. Starting with the data of the *no-match trials*, recall that for this type of trial we expected a quadratic effect (i.e., SOA 1 < SOA 2 > SOA 3) but not a linear effect (i.e., SOA 1 > SOA 3). Indeed, for both language groups this was the case. For the USA participants the linear effect gave an estimated value for $\boldsymbol{\psi}_{\textbf{linear}}$ of **2.7** percentage points, with a 95% CI of -0.5 to 6.0, that was not statistically significant ($F(1,18) = 3.07$, $MSE = 28.65$, $p = .097$, $\eta^2 = .146$). However, a substantial quadratic effect was apparent. The estimated value of $\boldsymbol{\psi}_{\textbf{quadratic}}$ was **10.2** percentage points, with a 95% CI of 3.6 to 16.9, that was statistically significant ($F(1,18) = 10.39$, $MSE = 40.10$, $p = .005$, $\eta^2 = .366$). Also for the Dutch participants there was no evidence for a linear effect.

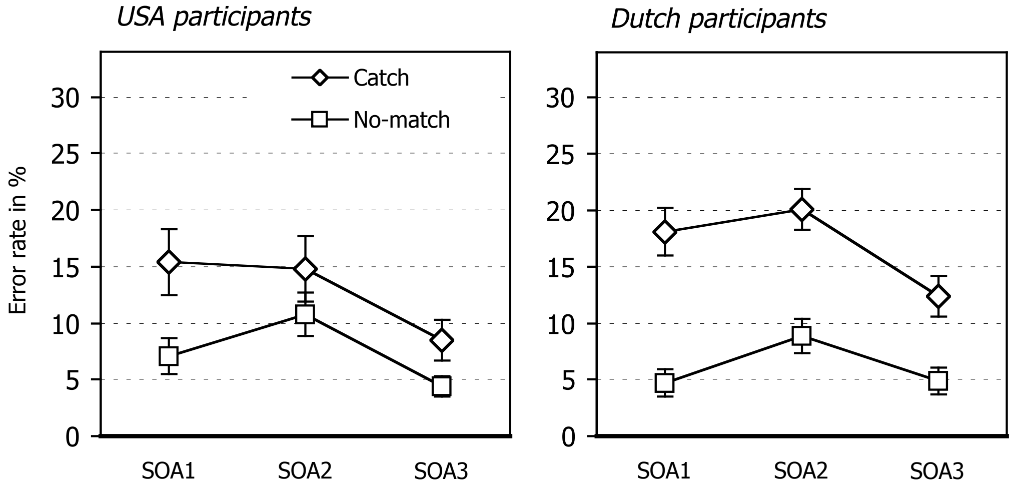

Figure 23. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and SOA (SOA1 vs. SOA2 vs. SOA3) for USA participants (left panel) and Dutch participants (right panel) in Experiment 5. Error bars represent the standard error of the mean.

The estimated value of $\boldsymbol{\psi}_{\textbf{linear}}$ was **0.1** percentage points (in opposite direction), with a 95% CI of -2.9 to 3.1 ($F(1,30) = 0.01$, $MSE = 39.10$, $p = .926$, $\eta^2 = .000$). There was however evidence for a substantial quadratic effect. The estimated value of $\boldsymbol{\psi}_{\textbf{quadratic}}$ was **8.2** percentage points, with a 95% CI of 3.9 to 12.5, that was statistically significant ($F(1,30) = 15.28$, $MSE = 26.37$, $p < .001$, $\eta^2 = .337$).

Turning to the data of the *catch trials*, recall that for this type of trial we expected the contrast SOA 1 < SOA 2 > SOA 3, as well as the contrast SOA 1 > SOA 3. In line with expectation, for the USA participants a substantial linear effect was observed with an estimated value for $\boldsymbol{\psi}_{\textbf{linear}}$ of **6.9** percentage points, with a 95% CI of 2.8 to 10.9, that was statistically significant ($F(1,18) = 12.81$, $MSE = 44.27$, $p = .002$, $\eta^2 = .416$). However, no evidence for a substantial quadratic effect was obtained. The estimated value of $\boldsymbol{\psi}_{\textbf{quadratic}}$ was **5.6** percentage points, with a 95% CI of -2.0 to 13.2, that was not statistically significant ($F(1,18) = 2.42$, $MSE = 52.26$, $p = .137$, $\eta^2 = .119$). In contrast, for the Dutch participants there was evidence for both a linear and a quadratic effect. The linear effect gave an estimated value for $\boldsymbol{\psi}_{\textbf{linear}}$ of **5.7** percentage points, with a 95% CI of 1.9 to 9.5, that was statistically significant ($F(1,30) = 9.39$, $MSE = 62.15$, $p = .005$, $\eta^2 = .238$). The estimated value of $\boldsymbol{\psi}_{\textbf{quadratic}}$ for the quadratic effect was **9.9** percentage points, with a 95% CI of 4.3 to 15.4, that was statistically significant ($F(1,30) = 13.27$, $MSE = 43.96$, $p = .001$, $\eta^2 = .307$).

Error analyses comparing USA and Dutch participants

Overall, Dutch participants produced more false-negative errors on match trials ($M = 5.0\%$) than the native, English speaking USA participants ($M = 2.7\%$). The statistically significant difference in group means was 2.3 percentage points (95% CI

0.9 to 3.7), $F(1,58) = 10.10$, $MSE = 7.49$, $p = .002$ (Mann-Whithey U, $p = .002$). There was also a statistically significant Word Type by language group interaction effect (□= .86, $F(1.71, 82.12) = 6.98$, $MSE = 26.18$, $p = .003$). The same way as in the word-naming experiment (Experiment 1) presented in Chapter 3, the two language groups were compared on the AM > TM contrast (e.g., PAID vs. SAID). We explored this effect because for the Dutch participants particularly high error rates were again observed for words like PAID. The analysis revealed that the difference for the AM > TM contrast was 4.8 percentage points for Dutch participants and for the USA participants it was 2.1 percentage points. The difference between these effects was 2.7 percentage points (95% CI 0.3 to 5.1) and statistically significant, $F(1,58) = 5.22$, $MSE = 10.30$, $p = .026$ (Mann-Whithey U, $p = .038$).

Further, Dutch participants produced, overall, more false-positive errors (M = 11.5%) on mismatch trials than the USA participants (M = 10.2%). The difference in group means was 1.3 percentage points (95% CI -2.3 to 5.0), and not statistically significant, $F(1,58) = 0.54$, $MSE = 47.51$, $p = .467$ (Mann-Whithey U, $p = .432$). Furthermore, contrasting no-match-trial and catch-trial performance, the Trial Type effect was considerably larger for Dutch participants (10.7 percentage points) than for USA participants (5.5 percentage points), with a difference of 5.2 percentage points (95% CI 1.0 to 9.4). In accordance, the Trial Type by language group interaction effect was statistically significant ($F(1,58) = 6.09$, $MSE = 64.16$, $p = .017$, Mann-Whithey U, $p = .015$).

With regard to the effect of SOA, there was no evidence for an overall SOA by language group interaction effect (□= 1.0, $F(2.00,96.00) = 0.73$, $MSE = 87.91$, $p = .484$). Furthermore, both for the no-match trials and the catch trials there was no evidence for an interaction effect between language group and the linear effect of SOA, nor between language group and the quadratic effect of SOA (no-match trials: $F_{\text{linear}}(1,48) = 1.66$, $MSE = 70.36$, $p = .204$, $F_{\text{quadratic}}(1, 48) = 0.31$, $MSE = 63.04$, $p = .581$; catch trials: $F_{\text{linear}}(1, 48) = 0.18$, $MSE = 110.89$, $p = .672$, $F_{\text{quadratic}}(1, 48) = 0.92$, $MSE = 94.14$, $p = .344$). Treating the analyses as a between-groups comparison of individual participants' effect (difference) scores, thereby using the regular within-groups sum of squares to estimate error variance, the differences in linear and quadratic effects between language groups were 2.8 (95% CI -2.0 to 7.7) and 2.0 percentage points (95% CI -5.2 to 9.2), respectively, for the no-match trials, and 1.2 (95% CI -5.7 to 8.0) and 4.2 percentage points (95% CI -6.1 to 14.6), respectively, for the catch trials.

Discussion

With regard to the Dutch-English bilinguals, Experiment 5 successfully replicated the main findings of Experiment 4. Once more, evidence was obtained that, in the print-to-speech correspondence task, perceiving a match or mismatch between an English printed word and a spoken rime is influenced by knowledge of English enemy neighbors. Again, this supports the hypothesis that bilingual processing of an English inconsistent word (e.g., MOOD) involves simultaneous coding of appropriate (e.g., /ud/) and inappropriate (e.g., /}d/) intermediate-grain size phonological structures that compete with each other. Moreover, it was once more demonstrated that, as indicated by a robust Trial Type effect, perceiving a mismatch between an English printed word and a spoken rime that is derived from an enemy neighbor is extraordinarily demanding. That is, relative to no-match trials (e.g., MOOD - /Yd/), performance on catch trials (e.g., MOOD - /}d/) was associated with elevated false-positive error rates and longer no-response latencies. Evidently, a spoken rime acts to restore a degraded, inappropriate phonological coding to such a degree that competition between appropriate and inappropriate codings is fully resumed. This reinforces the notion that processing of an inconsistent English word indeed involves auxiliary coding of an inappropriate phonological structure resulting from spelling-to-sound knowledge of enemy neighbors.

Furthermore, contrasting Dutch-English bilinguals and native English speakers, it was shown that performance of the two language groups on the print-to-speech correspondence task was quite similar. There were only two noticeable differences. One, Dutch-English bilinguals showed, relative to the native English speakers, rather poor match-trial performance, although predominantly on English AM-words (without Dutch neighbors) such as PAID. This was also observed in Experiment 1 comparing word-naming performance of similar Dutch-English bilinguals and native English speakers, and suggests that knowledge of spelling-to-sound mappings of this type of words is especially poor. Two, Dutch-English bilinguals showed a markedly larger effect of Trial Type. For the Dutch-English bilinguals, perceiving a mismatch in a catch trial was extraordinarily demanding and reveals that, in particular for this language group, spoken rimes presented in catch trials have a large potentiality to restore a degraded, inappropriate phonological coding. As suggested by the larger Trial Type effect, the resulting competition is relatively profound for Dutch-English bilinguals and probably reflects weaker knowledge of correct local orthographic-phonologic mappings for inconsistent English words.

In addition, Experiment 5 explored whether there is a change in performance on the print-to-speech correspondence task when the spoken rime is presented not simultaneously (SOA 2) but before (SOA 1) or after (SOA 3) presentation of the printed word. An examination of the SOA effects revealed that for both language groups, match-trial performance was better in SOA 3 (first print, then speech) than in SOA 2, the condition in which print and speech are presented simultaneously. This finding suggests that the context provided by processing of the spoken rime does not facilitate processing of the printed word. On the contrary, match-trial performance was, unexpectedly, poorest in SOA 1, the condition in which speech is presented first and print follows after a substantial time gap. Thus for match trials, it appears that a spoken rime that is congruent with the printed word does not aid but rather hinders processing of the printed word.

Further, for both language groups, no-match-trial performance was, as expected, poorest in SOA 2, the condition in which print and speech are presented simultaneously. There was no evidence for a difference in performance between SOAs 1 (first speech, then print) and 3 (first print, then speech). This finding suggests that presenting the spoken rime distant from print causes relatively unobstructed processing of the printed word. The resulting improvement in no-match-trial performance may be caused by loosened attentional demands, since in these conditions bimodal information processing proceeds in nonoverlapping time scales.

For catch trials, however, the observed SOA effect showed a different pattern. Although for both language groups catch-trial performance was, as expected, better in SOA 3 (first print, then speech) than in SOA 2 (print and speech presented simultaneously), catch-trial performance was also, as expected, better in SOA 3 than in SOA 1 (first speech, then print). Thus, suspending presentation of the spoken rime resulted in a diminished impact of an imposing extraneous candidate phonological structure. For example, a spoken rime such as /{d/ (derived from BLOOD) that was presented half a second after offset of presentation of a printed word such as MOOD resulted in better catch-trial performance (i.e., fewer errors) than when speech and print were presented concurrently. This finding may indicate that appropriate phonological codings become less susceptible to a reinstated inappropriate phonological coding as time progresses, indicating enhanced stability of the appropriate coding. It may however also indicate that, as time progresses, inappropriate phonological codings loose some of their potentiality to be reinstated. In either case, it appears that also under the conditions of SOA 3 competition between appropriate and inappropriate phonological codings is resumed, albeit to a relatively moderate degree.

In sum, observing poor catch-trial performance even in a condition where speech is presented after some substantial delay is consistent with the view that word perception is a continuous, metastable process. Apparently, the reading system never settles fully in the appropriate state, because even after a substantial time gap, the context that comes forth by processing of the spoken rime still has an impact on perception of the printed word. Thus, in a catch trial, a spoken rime may cause participants to perceive MOOD’s phonology to rhyme with that of BLOOD even when speech is presented half a second after off-set of the printed word. Nevertheless, under these conditions the impact seems to be reduced, which suggests that the essentials for resuming competition change in the course of time.

To conclude, in accordance with the phonological coherence hypothesis, processing of an inconsistent English word appears to involve auxiliary coding of an inappropriate phonological structure resulting from knowledge of English enemy neighbors. In the course of word processing this inappropriate coding is assumed to engage in competition with the appropriate coding. This competition is resolved quickly in that the inappropriate phonological structure that emerged in the initial conditions of word perception is effectively inhibited. However, the inappropriate phonological coding is not entirely annihilated. In fact, this coding seems to lie dormant in a state in which it can be readily reinstated by, for instance, a fostering context. Furthermore, in terms of the resonance framework, the inappropriate coding appears to continue as long as the appropriate phonological structure is in a resonant state with relevant, other knowledge structures.

5 When MOOD Rhymes with ROAD: Interlingual Phonological Coding and Language Mode

The evidence gathered in Chapter 4 supports the notion that the process of monolingual and bilingual word perception involves mandatory, intralingual phonological coding. This learning meets the first objective of this study. To recapitulate, Chapter 4 showed that processing of an inconsistent English word such as MOOD is affected by knowledge of enemy neighbors such as BLOOD. In particular, strong evidence was provided for the idea that, both for Dutch-English bilinguals and native English speakers, perception of an inconsistent word involves auxiliary coding of inappropriate phonology. This evidence came about largely by observing performance on catch trials, such as when MOOD is accompanied with the spoken rime /}d/ (derived from BLOOD). Performance on this type of trials was associated with high rates of false-positives. This was taken as an indication that the spoken rimes restored degraded, inappropriate phonological codings, thereby causing participants to perceive a word such as MOOD to rhyme with its enemy BLOOD.

Thus, in short, under specific conditions MOOD may be perceived to rhyme with BLOOD. But is it also true that MOOD may be perceived to rhyme with the Dutch word LOOD (i.e., rhyming with ROAD), resulting from manifold *cross-language* spelling-to-sound relations? This question emanates from the second objective of this study, which seeks to investigate whether the process of second-language English word perception involves mandatory, *interlingual* phonological coding due to spelling-to-sound knowledge of *Dutch* enemy neighbors. For example, the Dutch word LOOD is a Dutch enemy of the English words MOOD and BLOOD, because it has the same spelling body (-*OOD*) but is pronounced differently. This research goal is motivated by the hypothesis presented in the Introduction section, which states that bilingual word perception proceeds essentially language non-selectively (e.g., Van Heuven et al., 1998).

In Chapter 5, using the methodology of Chapter 4, it is investigated whether Dutch-English bilinguals engage spelling-to-sound knowledge of *both* their languages when reading English words. In particular, the second main research question specified in the Introduction section is addressed, asking whether the process of bilingual word perception involves simultaneous coding of cross-language phonology. To recapitulate, cross-language phonological coding concerns a collateral process of extraneous native Dutch phonology emerging simultaneously with appropriate (and inappropriate) non-native English phonology. Again, auxiliary coding of inappropriate Dutch phonology is assumed to result from profound reading experience with Dutch (enemy) neighbors.

Resorting again to the print-to-speech correspondence task, evidence for auxiliary coding of cross-language phonology may be observed in performance on catch trials, such as when MOOD is accompanied with a spoken rime derived from a Dutch enemy neighbor (e.g., /od/, derived from the Dutch enemy LOOD). Thus, if inappropriate Dutch phonology is actually part of the initial conditions of perception of MOOD, this spoken rime may restore the degraded, inappropriate coding to such a degree that participants may find it difficult or are actually unable to perceive a mismatch. That is, the spoken rime may put the degraded, inappropriate coding back into a competing state. Consequently, as was observed in Experiments 2-5, participants may react frequently with incorrect "yes" responses (i.e., false-positives). Furthermore, since for Dutch-English bilinguals spelling-to-sound knowledge of Dutch (enemy) words is assumed to be stronger (i.e., more self-consistent) than spelling-to-sound knowledge of English friends and enemies, it is expected that catch-trial performance will be even more defective when the spoken rime is not derived from an English but from a Dutch enemy neighbor.

EXPERIMENT 6

SIMULTANEOUS INTERLINGUAL CODING OF ENGLISH AND DUTCH PHONOLOGY

In Experiment 6 we used the print-to-speech correspondence task to investigate whether in Dutch-English bilinguals processing of an inconsistent English word involves simultaneous cross-language coding of English and Dutch phonology. The issue of auxiliary coding of inappropriate Dutch phonology is addressed by comparing performance on catch trials and no-match trials, separately for all three word types (AM, TM, and CM). Notice that in contrast to the previous experiments, Experiment 6 also allowed creation of catch trials for *consistent* words. The general predictions have the same basic form as in the previous experiments. For no-match trials we expected the contrasts AM > TM, AM > CM, and TM > CM, thus performance on a trial like BLOOD - /Yd/ (with /Yd/ derived from the unrelated word BRIDE) is worse than on a trial like MOOD - /Yd/, of which the performance is worse than on a trial like MOON - /en/ (with /en/ derived from the unrelated word VEIN). Further, for comparisons of *no-match trials* and *catch trials* (with spoken rimes now derived from *Dutch* enemy words) we expected the contrast catch > no-match, thus performance on trials like BLOOD - /od/ and MOOD - /od/ is worse than on trials like BLOOD - /Yd/ and MOOD - /Yd/. Furthermore, we expected the contrast catch (AM-words) > catch (TM-words), thus performance on trials like BLOOD - /od/ is worse than on trials like MOOD - /od/. Since Experiment 6 also included catch trials for consistent words, we further expected the related contrasts catch (AM-words) > catch (CM-words) and catch (TM-words) > catch (CM-words). Finally, we expected the following contrast: Trial Type effect (AM-words) > Trial Type effect (TM-words), thus the difference in performance on trials like BLOOD - /od/ and BLOOD - /Yd/ is larger than the difference in performance on trials like MOOD - /od/ and MOOD - /Yd/. In addition, we expected the following related contrasts: Trial Type effect (AM-words) > Trial Type effect (CM-words) and Trial Type effect (TM-words) > Trial Type effect (CM-words).

In addition to these general comparisons, Experiment 6 explored the effect of stimulus-list composition (i.e., language mode, Grosjean, 1997, 2001) on performing the print-to-speech correspondence task. This served to address the fourth main research question specified in the Introduction section. To recapitulate, studies of Dijkstra et al. (1998), De Groot et al. (2000), and Jared and Kroll (2001) have shown that stimulus-list composition affects the degree in which second-language words are processed language non-selectively (but see Van Hell & Dijkstra, 2002; Dijkstra & Van Hell, 2003). For example, it has been shown that when Dutch filler words are added to the stimulus list, lexical decisions of Dutch-English participants on interlingual homographs (e.g., BAD) take more time than on single-language control words (e.g., LOW). When the stimulus list does not contain any Dutch words, performance on words like BAD and LOW is more or less the same (e.g., see Dijkstra et al., 1998). The same holds for the use of spelling-to-sound knowledge from both languages in bilingual naming (Jared & Kroll, 2001). In order to evaluate the effect of stimulus-list composition in Experiment 6, two experimental conditions were compared. In the Dutch-fillers condition the stimulus materials were mixed with an additional 25% Dutch filler trials and in the (neutral) English-fillers condition they were mixed with an additional 25% English filler trials. It was expected that adding Dutch trials to the stimulus set would affect the participants' relative prominence of the non-target, here Dutch, language. Processing of Dutch words should increase the activity of the Dutch language, which may enhance (inappropriate) phonological coding of English words according to Dutch spelling-to-sound knowledge. This, in turn, may lead to increased false-positive errors and longer no-response latencies for catch trials.

However, because it is assumed that in the initial conditions of word perception all phonological structures are launched that have previously been associated with a spelling body, irrespective of the relative prominence of the non-target language, we are not sure whether to expect a very large impact of stimulus-list composition. The reason for this is explained in the Introduction section and boils down to the premise that we expect to detect all codings that originate from the initial conditions of word perception. Since it is assumed that the relative prominence of the non-target language does not affect the ballistic process of phonological coding, we may consider the possibility that adding Dutch filler trials to the stimulus set does not severely elevate the process of inappropriate Dutch phonological coding. Essentially, this reasoning corresponds with the viewpoint of Dijkstra and Van Hell (2003), who argue that word processing is always language non-selective.

Method

Participants and materials

A group of 60 Dutch-English bilinguals served as participants. They were presented with the 120 printed English words described in the Method section of Experiment 4 and a subset of the 120 spoken rimes described in the General Method section of Chapter 4. Specifically, for the match and no-match trials, the same spoken rimes were used as in Experiments 4 and 5. For the catch trials, however, a different set of spoken rimes was used (see Table 35). These spoken rimes were all derived from *Dutch* enemy neighbors. For example, to create a catch trial with the English word MOOD, a spoken rime was extracted from a recorded pronunciation of the Dutch word LOOD (which rhymes with the English word ROAD).

Table 35.
Examples of trial types for Experiments 6-8. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

Yes-trials			
Word Type	*Example*	Match Sound	Analogy
CM	*MOON*	/un/	"moon"
TM	*MOOD*	/ud/	"mood"
AM	*BLOOD*	/}d/	"blood"

No-trials			
No-Match Sound	Analogy	Catch Sound	Analogy
/en/	"vein"	/on/	"zoon"
/Yd/	"bride"	/ot/	"rood"
/Yd/	"bride"	/ot/	"rood"

Furthermore, the experimental materials were augmented with 40 filler trials (constituting 25% of the total number of trials). These filler trials consisted of either 40 Dutch (match and no-match) trials or 40 English (match and no-match) trials. In addition, a set of Dutch filler trials was created to serve as practise trials. Participants were informed that the experimental materials not only consisted of English words but also of Dutch words.

Experimental design

Experiment 6 used the same basic design as Experiments 4 and 5. In Experiment 6, nine groups of target trials were created that represented specific combinations of Trial Type and Word Type. These nine combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), catch trial (TM), and catch trial (CM). Table 36 presents the layout of the experimental design. As in Experiment 4, each participant was presented with each of the two trial blocks A and B. Hence, for each participant data was obtained for each available combination of trial type and word type, using every word once.

Table 36.
Experimental design for Experiment 6 (print-to-speech correspondence task). English words with Dutch neighbors are used for No-trials, and English words without Dutch neighbors for Yes-trials. Identical to Experiment 4, the word lists comprising each word type are split up in two sub word-lists (CM_1 and CM_2, TM_1 and TM_2, AM_1 and AM_2), according to the criterion whether or not a word can be used to create a Dutch catch trial. Combinations of word type and trial type (Match, No-Match, Dutch Catch) are systematically distributed over two trial blocks (A and B). The temporal order of trial block is Latin-square counterbalanced across two different participant groups (Participant Group 1: Sequence A-B; Participant Group 2: Sequence B-A). (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

No-trials

Trial Block	A	B
Sub Word-List	CM_1	CM_2
Trial Type	Dutch Catch	No-Match
Example Word	***MOON***	***SEEM***
Sub Word-List	TM_1	TM_2
Trial Type	Dutch Catch	No-Match
Example Word	***MOOD***	***SCAN***
Sub Word-List	AM_2	AM_1
Trial Type	No-Match	Dutch Catch
Example Word	***SWAN***	***BLOOD***

The temporal order of the two trial blocks was counterbalanced across two different participant groups according to a single Latin square (participant group 1: Sequence A-B; participant group 2: Sequence B-A). Participants were randomly assigned to the different sequences. Again, the counterbalancing procedure was intended to disentangle the effect of temporal position of the procedural variables from the effects of the independent variables.

For half of the participants (Dutch-fillers condition) the stimulus materials were mixed with (25%) Dutch filler trials and for the other half (English-fillers condition) they were mixed with (25%) English filler trials. Therefore, in the analyses, Filler Type was treated as a between-subjects variable.

Results

Participants were presented with nine different groups of target trials representing specific combinations of Trial Type and Word Type. Each participant responded to 60 match trials, which consisted of 20 AM-words (e.g., PAID), 20 TM-words (e.g., SAID), and 20 CM-words (e.g., STAIN). In addition, each participant responded to 30 no-match trials and 30 catch trials. Both the no-match trials and the catch trials consisted of 10 AM-words (e.g., BLOOD), 10 TM-words (e.g., MOOD), and 10 CM-words (e.g., MOON). For each participant, the correct response latencies within these nine groups were averaged and percentage of errors was calculated for each of the nine groups.

In Experiments 6-8 we refrained from analyses of match-trial data. We did so because in Experiments 6-8 match-trial performance was nearly identical to that observed in Experiments 2-5. Hence, the data that entered the statistical analyses consisted for each participant of three latency means and three percentages of false-positives for no-match trials, and three latency means and three percentages of false-positives for catch trials.

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. For the participants in the English-filler and Dutch-fillers conditions this resulted in a rejection of a total of 27.7% and 28.8%, respectively, for the catch trials and 7.0% and 7.7%, respectively for the no-match trials. Further, for both groups of participants, nearly 0% of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. In an experimental session a trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. In the analyses, this experimental procedure resulted in a cut-off that rejected all latencies greater than 2000. We did not consider further truncation, because for the participants in the English-filler and Dutch-fillers conditions the procedure resulted in rejection of 0.4% and 0.3%, respectively, of the correct response latencies and it is not recommended to exceed these percentages (Ulrich & Miller, 1994).

Error data

In a mismatch trial, a participant produced an error (i.e., a false-positive) when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other. The mean percentages of false-positives as a function of Trial Type, Word Type, and Filler Type, separately for each trial block, are presented in Table 17 of Appendix C. Figure 24 shows these data collapsed over trial block and participant (sub)group.

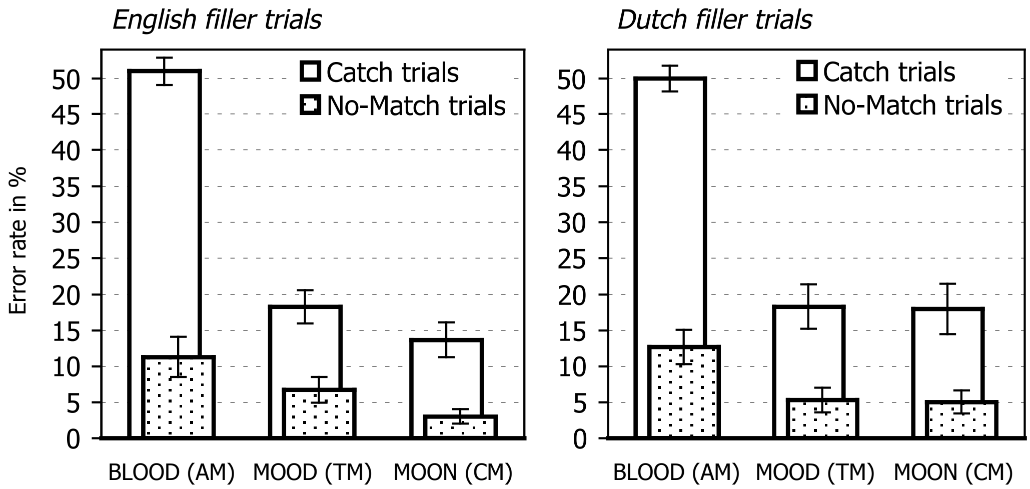

Figure 24. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 6. Error bars represent the standard error of the mean.

As in the previous experiments that evaluated the effect of intralingual consistency, no-match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 24, for both participant groups (English-fillers condition vs. Dutch-fillers condition) the ratio of friends and enemies was indeed associated with the number of false-positives. Replicating the findings of Experiments 3-5, participants made more false-positive errors on AM-words such as BLOOD than on TM-words such as MOOD and CM-words such as MOON. In addition, as was observed previously, the number of errors for CM-words was not markedly lower than for the words with typical mappings.

Omnibus analysis of variance. The mean percentages of false-positives were subjected to statistical analysis. Table 37 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 3 (Word Type: AM vs. TM vs. CM) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA in which Filler Type was treated as a between-subjects variable. The table also provides the results of non-parametric tests. As can be confirmed by looking at Table 37, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position), which accounted for a substantial percentage of variance. Again, participants produced fewer errors across temporal positions of trial blocks. Further, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Table 37.
Analysis of variance on percentages of false-positives for Experiment 6.

Source of variance	*SS*	□	*df*	*MS*	*F*	*p (F|H0)*	*p*[a]	□$_p$^{2}
• Block Position	1040.53		1	1040.53	14.45	< .001	< .001	.197
Block Position × Participant	4249.05		59	72.02				
• Sequence	1284.44		1	1284.44	3.71	.059	.149	.062
Participant(Group)	19395.56		56	346.35				
• TT	39271.11		1	39271.11	281.15	< .001	< .001	.834
TT × Participant(Group)	7822.22		56	139.68				
• WT	32973.89	.96	1.92	17138.16	223.07	< .001	< .001	.799
WT × Participant(Group)	8277.78	.96	107.74	76.83				
• TT × WT	13960.56	.84	1.67	8340.44	55.85	< .001	< .001	.499
TT × WT × Participant(Group)	13997.78	.84	93.74	149.33				
• Filler Type	71.11		1	71.11	.21	.652	.864	.004
Participant(Group)	19395.56		56	346.35				
• TT × Filler Type	4.44		1	4.44	.03	.859	.917	.001
TT × Participant(Group)	7822.22		56	139.68				
• WT × Filler Type	243.89	.96	1.92	126.76	1.65	.198		.029
WT × Participant(Group)	8277.78	.96	107.74	76.83				
• TT × WT × Filler Type	90.56	.84	1.67	54.10	.36	.659		.006
TT × WT × Participant(Group)	13997.78	.84	93.74	149.33				

Note. Type III Sum of Squares with Huynh-Veldt □-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (U|H0)* for Mann-Whitney test, *p* (□^{2}*|H0)* for Friedman test and *p* (□*|H0)* for sign test.

Returning to the data of the mismatch trials, there was a statistically significant main effect of Word Type, which, as can be confirmed by inspecting Table 37, accounted for a considerable percentage of variance. There was no evidence for a substantial Word Type by Filler Type interaction effect, or for a substantial three-way Trial Type by Word Type by Filler Type interaction effect.

Planned contrasts. The Word Type effect for no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons. Again, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals (95% SCI). In this analysis, the data of the two participant groups (Filler Type) were merged. For the **AM > TM** contrast there was a statistically significant difference of **6.0** percentage points, with a 95% SCI of 1.7 to 10.3 ($F(1,58) = 11.59$, $MSE = 93.22$, $p = .001$). The **8.0** difference in percentage points for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 3.6 to 12.4 ($F(1,58) = 20.21$, $MSE = 95.00$, $p < .001$), but the **2.0** difference in percentage points for the **TM > CM** contrast was not, with a 95% SCI of -1.0 to 5.0 (F(1,58) = 2.70, MSE = 44.43, p = .106).

Turning now to the data for the catch trials, the same pairwise comparisons were performed as for the no-match trials. For the **catch (AM-words) > catch (TM-words)** contrast there was a statistically significant difference of **32.2** percentage points, with a 95% SCI of 26.9 to 37.4 (F(1,58) = 230.57, MSE = 134.63, p < .001). The **34.7** difference in percentage points for the **catch (AM-words) > catch (CM-words)** contrast was also statistically significant, with a 95% SCI of 29.1 to 40.3 (F(1,58) = 233.82, MSE = 154.20, p < .001), but the **2.5** difference in percentage points for the **catch (TM-words) > catch (CM-words)** contrast was not, with a 95% SCI of -1.2 to 6.2 (F(1,58) = 2.85, MSE = 65.89, p = .097).

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, mismatch-trial performance should, again, be worse for catch trials than for no-match trials (catch > no-match). Furthermore, the Trial Type main effect was expected to be embedded in an Trial Type by Word Type interaction effect, which is further inspected by the contrasts Trial Type effect (AM-words) > Trial Type effect (TM-words), Trial Type effect (AM-words) > Trial Type effect (CM-words), and Trial Type effect (TM-words) > Trial Type effect (CM-words).

As Figure 24 shows, Trial Type was indeed associated with different numbers of false-positives. Compared to the primary test of Experiment 4, in which spoken rimes for catch trials were derived from English enemy words, error rates for catch trials in Experiment 6 reached stunningly high levels, up to 50% for AM-words and 18% for TM-words. To compare, in the primary test of Experiment 4 we observed no more than 28% and 21% of false positives for AM-words and TM-words, respectively. Overall, in Experiment 6 participants made more errors on catch trials (M = 28.2%) than on no-match trials (M = 7.3%), in which the **catch > no-match** contrast gave an overall difference of **20.9** percentage points (MSE = 45.08, 95% CI 18.4 to 23.3, collapsed over Filler Type). Table 37 indicates that this statistically significant main effect accounted for a considerable percentage of variance. For comparison with Experiment 4, when the data of the CM-words were discarded, the error percentages for catch trials and no-match trials were 34.4 and 9.0, respectively, in which the **catch > no-match** contrast gave an overall difference of **25.4** percentage points (MSE = 63.31, 95% CI 22.5 to 28.3, collapsed over Filler Type). For the primary test of Experiment 4, this overall difference was markedly smaller, namely 13.5 percentage points (95% CI 6.7 to 20.3).

Finally, Figure 24 shows that the Trial Type effect was larger for the AM-words than for the other word types. This differential effect was supported by a statistically significant Trial Type by Word Type interaction effect, which, as is shown in Table 37, accounted for a considerable percentage of variance. There was no further evidence for a substantial Trial Type by Filler Type interaction effect.

The Trial Type by Word Type interaction effect was further inspected with three (Bonferroni-adjusted) pairwise comparisons. Again, the alpha level was set to .017, and estimates of differences were provided by 95% simultaneous confidence intervals. In this analysis, the data of the two participant groups (Filler Type) were merged. The Trial Type effect was larger for AM-words than for TM-words, in which the **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast gave a statistically significant difference of **26.2** percentage points, with a 95% SCI of 18.8 to 33.6 ($F(1,58) = 75.94$, $MSE = 270.49$, $p < .001$). Furthermore, the Trial Type effects were larger for AM-words and TM-words than for CM-words. The **26.7** difference in percentage points for the **Trial Type effect (AM-words) > Trial Type effect (CM-words)** contrast was statistically significant, with a 95% SCI of 18.4 to 35.0 ($F(1,58) = 62.70$, $MSE = 340.23$, $p < .001$), but the **0.5** difference in percentage points for the **Trial Type effect (TM-words) > Trial Type effect (CM-words)** contrast was not, with a 95% SCI of -4.4 to 5.4 ($F(1,58) = 0.06$, MSE = 118.19, p = .802).

Latency data

The mean correct no-response latencies as a function of Trial Type, Word Type, and Filler Type, separately for each trial block, are presented in Table 18 of Appendix C. Figure 25 shows the mean correct no-response latencies for the AM-words, TM-words, and CM-words both for the no-match trials and the catch trials. These mean correct no-response latencies were collapsed over trial block and participant (sub)group.

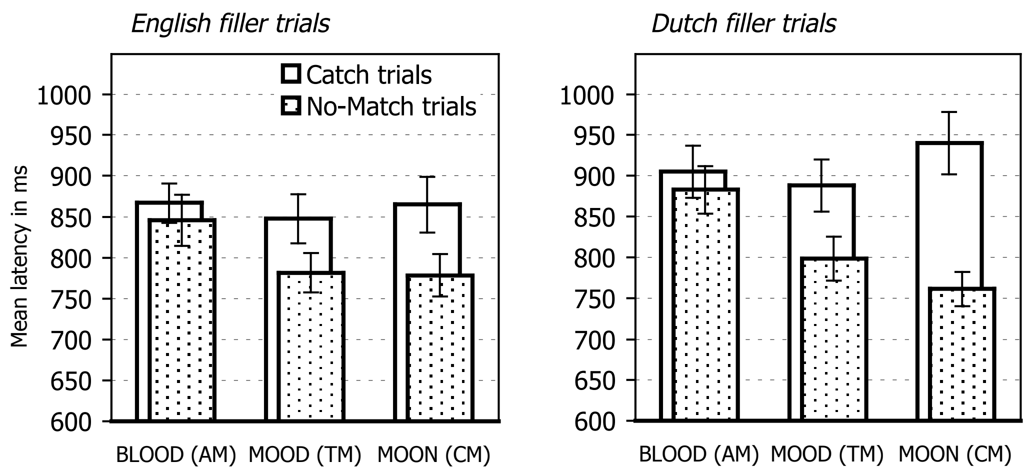

Figure 25. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 6. Error bars represent the standard error of the mean.

Once more, no-match-trial performance should be influenced by spelling-to-sound knowledge of English enemy neighbors. As can be seen in Figure 25, for both groups of participants (English-fillers condition vs. Dutch-fillers condition) this was indeed the case. Replicating the findings of Experiments 3-5, correct no-response latencies were longer for AM-words such as BLOOD than for TM-words such as MOOD and CM-words such as MOON. Again, response latencies for CM-words were not markedly shorter than for the words with typical mappings.

Omnibus analysis of variance. The mean correct no-response latencies were subjected to statistical analysis. Table 38 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 3 (Word Type: AM vs. TM vs. CM) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA in which Filler Type was treated as a between-subjects variable. As can be confirmed by looking at Table 38, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position), which accounted for a substantial percentage of variance.

Table 38.
Analysis of variance on correct no-response latencies for Experiment 6.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	95541.63		1	95541.63	13.64	< .001	.188
Block Position × Participant	413245.37		59	7004.16			
• Sequence	232664.18		1	232664.18	2.05	.158	.035
Participant(Group)	6358415.07		56	113543.13			
• TT	530073.88		1	530073.88	58.31	< .001	.510
TT × Participant(Group)	509085.82		56	9090.82			
• WT	145733.51	1.0	2.00	72866.75	15.20	< .001	.213
WT × Participant(Group)	536922.53	1.0	112.00	4793.95			
• TT × WT	181131.71	.97	1.94	93343.25	10.94	< .001	.163
TT × WT × Participant(Group)	927364.71	.97	108.67	8533.97			
• Filler Type	92096.01		1	92096.01	.81	.372	.014
Participant(Group)	6358415.07		56	113543.13			
• TT × Filler Type	33872.40		1	33872.40	3.73	.059	.062
TT × Participant(Group)	509085.82		56	9090.82			
• WT × Filler Type	1539.11	1.0	2.00	769.55	.161	.852	.003
WT × Participant(Group)	536922.53	1.0	112.00	4793.95			
• TT × WT × Filler Type	34105.22	.97	1.94	17575.56	2.06	.134	.035
TT × WT × Participant(Group)	927364.71	.97	108.67	8533.97			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Again, participants responded faster across temporal positions of trial blocks. Further, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance. Returning to the data of the mismatch trials, there was a statistically significant main effect of Word Type, which, as is shown in Table 38, accounted for a substantial percentage of variance. There was no evidence for a substantial Word Type by Filler Type interaction effect, or for a substantial three-way Trial Type by Word Type by Filler Type interaction effect.

Planned contrasts. The Word Type effect for the no-match trials was further inspected with three (Bonferroni-adjusted) pairwise comparisons. The alpha level was again set to .017. In this analysis, the data of the two participant groups (Filler Type) were merged. For the **AM > TM** contrast there was a statistically significant difference of **74** ms, with a 95% SCI of 38 to 110 ($F(1,58) = 25.88$, $MSE = 6310.53$, $p < .001$). The **94** ms difference for the **AM > CM** contrast was also statistically significant, with a 95% SCI of 56 to 131 ($F(1,58) = 37.21$, $MSE = 7063.92$, $p < .001$), but the **20** ms difference for the **TM > CM** contrast was not, with a 95% SCI of -6 to 45 ($F(1,58) = 3.70$, $MSE = 3181.30$, p = .059).

Turning now to the data of the catch trials, the same pairwise comparisons were performed as for the no-match trials. For the **catch (AM-words) > catch (TM-words)** contrast there was a difference of **18** ms that was not statistically significant, with a 95% SCI of -19 to 55 ($F(1,58) = 1.42$, $MSE = 6770.11$, $p = .239$). The **-16** ms difference (in opposite direction) for the **catch (AM-words) > catch (CM-words)** contrast was not statistically significant either, with a 95% SCI of -58 to 26 ($F(1,58) = 0.92$, $MSE = 8729.43$, $p = .342$), nor was the **-34** ms difference (in opposite direction) for the **catch (TM-words) > catch (CM-words)** contrast, with a 95% SCI of -74 to 5 ($F(1,58) = 4.54$, $MSE = 7746.64$, $p = .037$). In sum, there was evidence for a Word Type effect on latencies for no-match trials, but not for catch trials.

Turning now to the primary analyses that contrasted no-match-trial and catch-trial performance, Figure 25 shows that Trial Type again had an effect on no-response latencies. As was observed in the previous experiments, participants produced longer response latencies on catch trials ($M = 885$ ms) than on no-match trials ($M = 809$ ms), in which the **catch > no-match** contrast gave an overall difference of **77** ms ($MSE = 3120.19$, 95% CI 56 to 97, collapsed over Filler Type). Table 38 indicates that this statistically significant main effect accounted for a considerable percentage of variance. When the data of the CM-words were discarded, the mean no-response latencies for catch trials and no-match trials were 877 ms and 828 ms, respectively, in which the **catch > no-match** contrast gave an overall difference of **49** ms ($MSE = 3516.70$, 95% CI 28 to 71, collapsed over Filler Type). For the primary test of Experiment 4, this overall difference was somewhat larger, 86 ms (95% CI 51 to 121).

Finally, Figure 25 shows that, contrary to expectation, the Trial Type effect was smaller for the AM-words than for the other word types, which was substantiated by a statistically significant Trial Type by Word Type interaction effect that accounted for a moderate percentage of variance (Table 38). There was no further evidence for a substantial Trial Type by Filler Type interaction effect. The Trial Type by Word Type interaction effect was further examined with three (Bonferroni-adjusted) pairwise comparisons in which the alpha level was set to .017. In this analysis, the data of the two participant groups (Filler Type) were merged. The **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast gave a difference (in opposite direction) of **-56** ms that was not statistically significant, with a 95% SCI of -113 to 2 (F(1,58) = 5.74, MSE = 16317.11, p = .020). The **-110** ms difference (in opposite direction) for the **Trial Type effect (AM-words) > Trial Type effect (CM-words)** contrast was statistically significant, with a 95% SCI of -177 to -43 (F(1,58) = 16.30, MSE = 22251.86, p < .001), and so was the **-54** ms difference (in opposite direction) for the **Trial Type effect (TM-words) > Trial Type effect (CM-words)** contrast, with a 95% SCI of -104 to -4 (F(1,58) = 7.10, MSE = 12361.57, p = .010).

With respect to the odd differential effect of trial type, it should be noted that in Experiment 6 error rates were exceptionally high for catch trials involving AM-words. Thus, a large percentage of these trials were discarded, trials that represent the more difficult items and that should be associated with relatively long no-response latencies. It seems, therefore, reasonable to assume that the demise of these long latencies may account for the unexpectedly fast catch-trial performance for AM-words.

Analyses evaluating the effect of stimulus-list composition

As can be seen in Figure 25, the two groups of participants receiving different types of fillers produced nearly identical results in the print-to-speech correspondence task. Overall, participants in the Dutch-fillers condition produced more false-positive errors on mismatch trials (M = 18.2%) than participants in the English-fillers condition (M = 17.3%). The difference in group means was 0.9 percentage points (95% CI -3.1 to 4.9) and not statistically significant, F(1,58) = 0.20, MSE = 59.44, p = .657 (Mann-Whithey U, p = .864). In addition, no-response latencies were longer in the Dutch-fillers condition (M = 863 ms) than in the English-fillers condition (M = 831 ms). The difference in group means was 32 ms (95% CI -39 to 103) and not statistically significant, F(1,58) = 0.81, MSE = 18941.70, p = .372.

Furthermore, contrasting no-match-trial and catch-trial performance, the Trial Type effect on error rates and no-response latencies was larger in the Dutch-fillers condition (**21.1** percentage points and **96** ms) than in the English-fillers condition (**20.7** percentage points and **57** ms), with statistically non-significant group differences of 0.4 percentage points (MSE = 92.62, 95% CI -4.5 to 5.4) and 39 ms (MSE = 6257.79, 95% CI -2 to 80). In accordance, as is confirmed by inspecting Tables 37 and 38, the Filler Type by Trial Type interaction effects for error rates and response latencies were not statistically significant.

Discussion

Experiment 6 extends the findings of Experiments 2-5 by demonstrating the occurrence of interlingual spelling-to-sound consistency effects in Dutch-English bilinguals. Replicating Experiments 2-5, mismatch-trial performance was again found to be worse for inconsistent words with atypical mappings than for inconsistent words with typical mappings. Again, these findings substantiate the idea that bilingual processing of an English intralingual inconsistent word involves simultaneous coding of appropriate and inappropriate English intermediate-grain size phonological structures that compete with each other.

Most intriguing, however, is the observation that spelling-to-sound knowledge of *Dutch* enemy neighbors affected the ability to perceive a mismatch between an English printed word and a spoken rime. As in the previous experiments, the results obtained with the print-to-speech correspondence task were compelling: Rejecting a catch trial that consisted of a printed inconsistent word (e.g., BLOOD) and a spoken rime that was derived from a Dutch enemy of the word (e.g., /od/, as in the Dutch word LOOD) appeared to demand a remarkably large effort. Participants responded very frequently with false-positives, thus indicating that they, for instance, perceived BLOOD's phonology to rhyme with the rime of LOOD. To elaborate, for catch trials in Experiment 6, the false-positive error rate for TM-words was already as high as 18%, but for AM-words it reached the extraordinarily high level of 50%. In comparison, for no-match trials that contained spoken rimes derived from unrelated words, the false-positive error rates for TM-words reached a more moderate level of 6%, and for AM-words it was 12%. This finding strongly supports the hypothesis that the perception of an interlingually inconsistent word such as BLOOD involves auxiliary cross-language coding of inappropriate intermediate-grain size Dutch phonology.

A major shift from the previous experiments was that, in order to create catch trials, Experiment 6 used spoken rimes derived from Dutch enemy neighbors. Experiment 4, on the other hand, included the same printed words but used spoken rimes derived from English enemy neighbors. Note that Experiments 4, 5, and 6 employed the same basic design. To evaluate the effect of this shift, we may take the Trial Type effects (catch > no match) observed in Experiment 6 for AM-words (38.5 percentage points) and TM-words (12.3 percentage points) and compare these to the effects found in the primary test of Experiment 4 (AM-words: 14.5 percentage points; TM-words: 12.5 percentage points). If we make that comparison it becomes evident that the Trial Type effect for AM-words was considerable larger in Experiment 6 than in Experiment 4. More formally, performing a one-way (between-subjects) ANOVA on the Trial Type effect scores, revealed that the 24.0 percentage points difference (95% CI 14.7 to 33.3) between Experiment 6 (Dutch enemies) and Experiment 4

(English enemies) was statistically significant, $F(1,78) = 26.26$, $MSE = 328.97$, $p < .001$ (Mann-Whitney U test, $p < .001$).

To observe a relatively large Trial Type effect with this new type of spoken rimes was expected for Dutch-English bilinguals. For these participants, spelling-to-sound knowledge of Dutch words is assumed to be much stronger (i.e., more self-consistent) than spelling-to-sound knowledge of English words. It appears, therefore, that a strong-rule Dutch phonological coding that has been inhibited for an AM-word such as BLOOD (e.g., /od/, as in the Dutch word LOOD) is restored quite instantly by a fostering sound stimulus. Consequently, in a catch trial, a degraded, inappropriate Dutch phonological coding is pulled more readily into competition than an inappropriate English coding, causing relatively defective perception of a mismatch between print and sound.

In sum, consistent with the notion of a leading role of phonology in bilingual visual word perception, the Trial Type effects observed in Experiment 6 support a language non-selective view in which bilingual perception of an inconsistent word involves simultaneous cross-language phonological coding. Furthermore, as predicted by the relative self-consistency of spelling-to-sound associations, auxiliary coding of inappropriate Dutch phonology seems to occur even more firmly than auxiliary coding of inappropriate English phonology.

The results of Experiment 6 support the conclusions reached by Brysbaert et al. (1999) and Jared and Kroll (2001) that second-language reading not only engages spelling-to-sound knowledge of the target language, but also that of the non-target, native language. Moreover, it extends our current knowledge of the influence of cross-language enemy neighbors on word processing such as gathered by Jared and Kroll (2001). These investigators observed that the impact of enemy spelling-to-sound knowledge on second-language word naming was larger if it concerned the relatively weak second language (i.e., English) than if it concerned the relatively strong native language (i.e., French). Furthermore, even when the non-target, native language was aroused the impact remained relatively small. What this finding seems to point out is that auxiliary coding of inappropriate phonology due to knowledge of native-language enemy neighbors is relatively insubstantial. Such a conclusion, however, does not agree with the results of the present experiment, that in fact suggest that auxiliary coding of inappropriate, native Dutch phonology is quite strong. From the principle of self-consistency this finding indeed makes sense, because spelling-to-sound knowledge in the native language is much stronger than in the second language. Nevertheless, even if it is assumed that auxiliary coding of native-language phonology occurs more firmly than of second-language phonology, the fact remains that for second-language word-naming performance the impact of native, enemy spelling-to-sound knowledge was rather small in Jared and Kroll's (2001) study.

How can this contradictory state of affairs be understood? The key to the solution may be to assume that auxiliary coding of inappropriate Dutch phonology takes place in the initial conditions of word perception, a phase that may be exposed by the print-to-speech correspondence task and not by the word naming task. In these initial conditions, appropriate and inappropriate phonological structures emerge in proportion to their statistical dominance. However, incorrect codings are rapidly constrained by semantic feedback as the system moves towards an appropriate phonological structure. In fact, an incorrect Dutch phonological coding may be suppressed quite effectively when phonologic-semantic dynamics are tuned towards expectations of phonological codings from the target language. Consequently, because in the word-naming task dynamics relevant to words from the non-target language are inhibited, cross-language competition between local orthographic-phonologic resonances is resolved quickly, resulting in relatively unobstructed word-naming performance.

Finally, Experiment 6 incorporated a stimulus-list composition variable (i.e., Filler Type) to alter language mode. It was expected that adding Dutch filler trials to the stimulus set would increase the participants' relative prominence of the Dutch, non-target language. Yet, a comparison of the Dutch-fillers and English-fillers conditions gave no clear evidence that it caused enhanced (inappropriate) phonological coding according to Dutch spelling-to-sound knowledge. We may cautiously interpret this finding as suggesting a nearly zero effect of stimulus-list composition. If this conclusion is correct, the results of Experiment 6 are clearly not in accordance with observations of marked stimulus-list composition effects such as reported in Dijkstra et al. (1998). However, the discrepancy may be understood by noting that Experiment 6 focuses on the ballistic process of phonological coding, whereas in studies of, for instance, Dijkstra et al. (1998) the focus is on global-level linguistic coding (e.g., lexical decision). Possibly, the former type of coding is more autonomous with respect to the relative prominence of the non-target language than the latter.

EXPERIMENT 7

INTERLINGUAL PHONOLOGICAL CODING AND LANGUAGE MODE

To sum up, Dutch-English bilinguals may perceive an English word such as MOOD to rhyme with a Dutch word such as LOOD. The previous experiment was specifically designed to invoke this phenomenon. Taken together, the findings of Experiment 6 reveal the impact of spelling-to-sound knowledge of Dutch enemy neighbors on second-language word perception. There was, however, no indication that the impact of Dutch enemy neighbors was affected by language mode: A standard list-composition manipulation to increase the Dutch language's activity did not seem to affect the process of interlingual phonological coding. At first sight, this null-effect may point to the ballistic nature of auxiliary phonological coding in the initial conditions of word perception. However, it could also be the case that, in Experiment 6, performance on no-match and catch trials had reached a point where it could not deteriorate any further. In fact, participants in Experiment 6 certainly produced quite a lot of false-positive errors. It is therefore possible that when the Dutch language is activated the impact of Dutch enemy neighbours is increased after all, but we are unable to observe such an increase because of a floor effect.

Experiment 7 presents another effort to find out whether the relative prominence of the Dutch, non-target language affects the process of interlingual phonological coding. For that purpose two filler conditions were compared again, the Dutch-fillers and the English-fillers condition. In order to achieve an overall reduction in error rate, Experiment 7 included an SOA manipulation that followed the same scheme as used in Experiment 5 (see Chapter 4). Since this particular design entails considerable repetition of trials and also includes an SOA condition that appears to invoke fewer error responses (i.e., SOA 3, see Experiment 5), overall performance may improve significantly. This, in turn, may give room for a potentially detrimental impact of Dutch enemy neighbours to emerge.

Furthermore, the SOA manipulation of Experiment 7 enables us, as was aimed in Experiment 5, to explore whether a change in performance on the print-to-speech correspondence task occurs when the spoken rime is presented not simultaneously (SOA 2) but before (SOA 1) or after (SOA 3) presentation of the printed word. To recapitulate, the difference between SOA 3 and SOAs 1 and 2 was that early processing of the printed word in SOA 3 proceeds without contextual interference from simultaneous or earlier processing of the spoken rime. SOA 2 was different from the other two SOAs in that print and speech were presented simultaneously over the visual and auditory modalities whereas in the other SOAs they were not. Recall that in Experiment 5 no-match-trial performance was poorest in SOA 2. There was no evidence for a difference in performance between SOAs 1 and 3. Furthermore, for catch trials, performance in SOA 3 was better than in SOA 2 and SOA 1. Hence, delaying presentation of the spoken rime resulted in a diminished impact of an imposing extraneous candidate phonological structure. As time elapses, appropriate phonological codings become less susceptible, although not impervious, to reinstated inappropriate phonological coding.

With regard to the SOA manipulation, Experiment 7 again poses the question whether in SOA 3 (first print, then speech), the context that is provided by processing of the spoken rime still has an impact on the perception of the printed word. In other words, the relevant question is whether in a catch trial a spoken rime may cause participants to perceive MOOD’s phonology to rhyme with that of LOOD even when speech is presented half a second after the off-set of the printed word.

Method

Participants and materials

A group of 24 Dutch-English bilinguals served as participants. They were, except for one difference, presented with the same materials as used in Experiment 6. The difference was that, contrary to Experiment 6 but consistent with experiment 5, in Experiment 7 catch trials for consistent words were replaced with no-match trials.

Experimental design

Experiment 7 used the same basic design as Experiments 5 and 6. Eight groups of target trials were created that represented specific combinations of Trial Type and Word Type. These eight combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), and catch trial (TM). Each participant was presented with each of the two trial blocks A and B. Hence, for each participant data was obtained for each available combination of trial type and word type, using every word once. Furthermore, the participants were observed in all three SOA conditions in which they were tested three times with the same stimulus set (see also the method section of Experiment 5). The temporal order of the two trial blocks was counterbalanced across two different participant groups according to a single Latin square (participant group 1: Sequence A-B; participant group 2: Sequence B-A). In addition, within the two participant groups, the temporal order of SOA condition was Latin square counterbalanced across six different participant subgroups (SOA sequence 1-2-3, SOA sequence 2-3-1, SOA sequence 3-1-2, SOA sequence 1-3-2, SOA sequence 3-2-1, and SOA sequence 2-1-3; see also Table 30 in Experiment 5 of Chapter 4). Participants were randomly assigned to the different sequences. Again, the counterbalancing procedure was intended to disentangle the effect of temporal position of the procedural variables from the effects of the independent variables. This potential source of variance can be isolated and removed from the estimate of error variance, which may improve the efficiency of the design. Again, this was accomplished by adding participant subgroup (involving six SOA sequences) as a between-subjects variable in an ANOVA, and testing the effects against the resulting treatments×participants(group) error term.

Finally, as in Experiment 6, Filler Type was varied between participants. In the Dutch-fillers condition the stimulus materials were mixed with (25%) Dutch filler trials and in the English-fillers condition they were mixed with (25%) English filler trials.

Results

Participants were presented with eight different groups of target trials representing specific combinations of Trial Type and Word Type. Each participant responded three times (i.e., in SOA 1, SOA 2, and SOA 3) to 60 match trials which consisted of 20 AM-words, 20 TM-words, and 20 CM-words. In addition, each participant responded three times (i.e., in SOA 1, SOA 2, and SOA 3) to 40 no-match trials and 20 catch trials. The no-match trials consisted of 10 AM-words, 10 TM-words, and 20 CM-words, and the catch trials consisted of 10 AM-words and 10 TM-words. For each participant in each SOA condition, the correct response latencies within these eight groups were averaged and percentage of errors was calculated for each of the eight groups.

As in the Experiment 6, we refrained from analyses of the match-trial data. Hence, for each participant the data consisted of nine latency means and nine percentages of false-positives for no-match trials (e.g., for BLOOD, MOOD, and MOON), and six latency means and six percentages of false-positives for catch trials (e.g., for BLOOD and MOOD).

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. For the participants in the English-fillers and Dutch-fillers conditions this resulted in a rejection of a total of 32.4% and 30.3%, respectively, for the catch trials, and 5.5% and 5.3%, respectively, for the no-match trials. Furthermore, for both groups of participants, less than 0.1% of the trials were excluded because of apparatus failure or because the response latency was shorter than 200 ms. In all SOA conditions, a trial was cancelled if the participant failed to respond within 2000 ms after onset of the printed word. In the SOAs 1 and 2, this experimental procedure resulted in a cut-off that rejected all latencies greater than 2000 ms and in SOA 3 it resulted in a cut-off that rejected all latencies greater than 2500 ms. For the participants in the English-fillers and Dutch-fillers conditions the procedure resulted in rejection of 0.2% (SOA 1: 0.1%, SOA 2: 0.4%, SOA 3: 0.1%) and 0.3% (SOA 1: 0.1%, SOA 2: 0.4%, SOA 3: 0.3%), respectively, of the correct response latencies. Following recommendations of Ulrich and Miller (1994), we did not consider further data trimming.

Error data

In a mismatch trial, a participant produced an error (i.e., a false-positive) when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other. The mean percentages of false-positives as a function of Trial Type, Word Type, SOA condition, and Filler Type, separately for each trial block, are presented in Table 19 of Appendix C. Figure 26 (see also Figure 27) shows the mean percentages of false-positives for the AM-words and TM-words both for the no-match trials (including CM-words) and the catch trials, and as a function of SOA. These mean percentages of false-positives were collapsed over trial block and participant (sub)group.

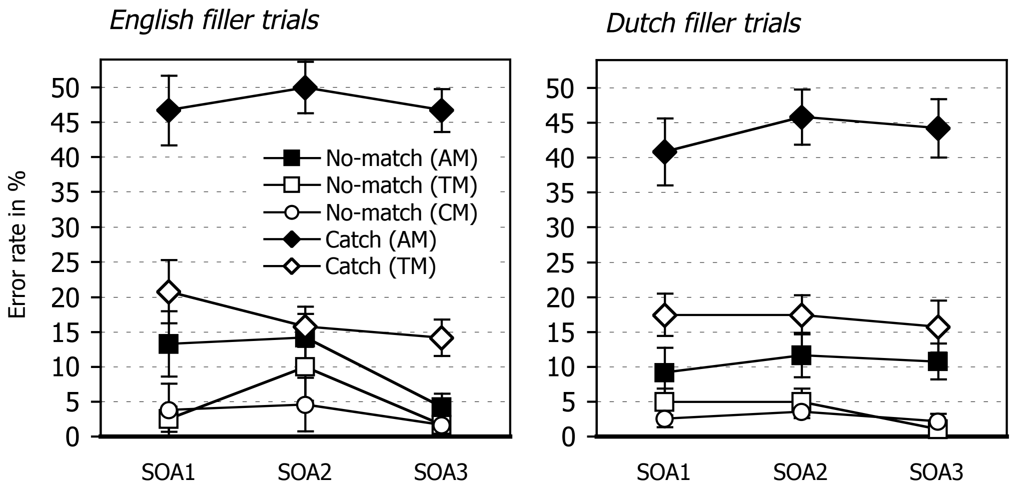

Figure 26. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials), Word Type (AM-words vs. TM-words vs. CM-words) and SOA (SOA1 vs. SOA2 vs. SOA3) for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 7. Error bars represent the standard error of the mean.

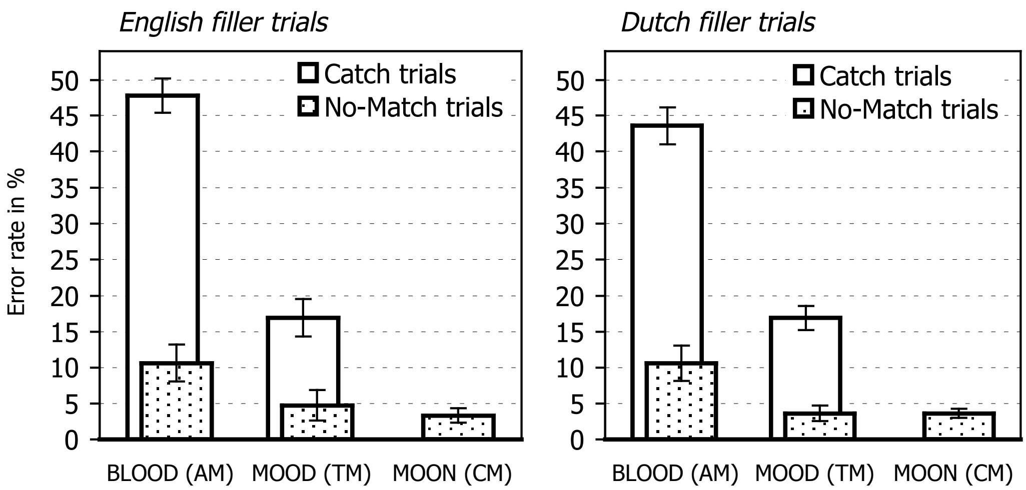

Figure 27. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) and collapsed over SOA for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 7. Error bars represent the standard error of the mean.

As can be seen in Figure 27, the number of false-positives again depended upon the ratio of friends and enemies. Replicating Experiments 3-6, overall participants made more errors on AM-words than on TM-words (**6.4** percentage points, 95% SCI 3.0 to 9.8) and CM-words (**7.1** percentage points, 95% SCI 3.1 to 11.1). The number of errors for CM-words was not markedly lower than for the words with typical mappings (**0.7** percentage points, 95% SCI -1.9 to 3.3).

Omnibus analysis of variance. The mean percentages of false-positives were further subjected to statistical analysis. Table 39 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 3 (SOA condition: SOA 1 vs. SOA 2 vs. SOA 3) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA in which Filler Type was treated as a between-subjects variable. The table also provides the results of non-parametric tests. As can be confirmed by looking at Table 39, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of SOA block (SOA Session) which accounted for a substantial percentage of variance. Specifically, participants produced fewer errors across sessions. Further, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

As Figure 27 shows, error rates for catch trials again reached high levels, although in a somewhat more moderate degree than in Experiment 6. Specifically, error rates for AM-words ranged from 44% to 48%, and for TM-words it ranged from 15% to 19%.

Planned contrasts. With regard to the analyses that contrasted overall no-match-trial and catch-trial performance, Figure 27 indicates that Trial Type was again associated with different numbers of false-positives. Replicating Experiment 6, overall participants made more errors on catch trials (M = 31.3%) than on no-match trials (M = 7.4%), in which the **catch > no-match** contrast gave an overall difference of **24.0** percentage points (MSE = 11.67, 95% CI 21.9 to 26.0, collapsed over SOA condition and Filler Type). Table 39 indicates that this statistically significant main effect accounted for a considerable percentage of variance. Furthermore, as revealed by a statistically significant Trial Type by Word Type interaction effect (Table 39), the Trial Type effect was larger for AM-words (35.1 percentage points) than for TM-words (12.8 percentage points). The corresponding **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, collapsed over SOA condition and Filler Type, gave a difference of **22.4** percentage points (95% CI 17.8 to 26.9).

Table 39.
Analysis of variance on percentages of false-positives for Experiment 7.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Session	666.94	1.0	2.00	333.47	18.53	< .001	< .001	.507
Session × Participant(Group)	647.79	1.0	36.00	17.99				
• Sequence	2647.57		5	529.51	.92	.499	.289	.278
Participant(Group)	6887.50		12	573.96				
• SOA	754.86	1.0	2.00	377.43	2.99	.069	.237	.200
SOA × Participant(Group)	3025.00	1.0	24.00	126.04				
• TT	41328.13		1	41328.13	463.13	< .001	< .001	.975
TT × Participant(Group)	1070.83		12	89.24				
• TT × WT	9000.35		1	9000.35	99.31	< .001	< .001	.892
TT × WT × Participant(Group)	1087.50		12	90.63				
• TT × SOA	168.75	1.0	2.00	84.38	1.04	.368	.195	.080
TT × SOA × Participant(Group)	1941.67	1.0	24.00	80.90				
• TT × WT × SOA	259.03	1.0	2.00	129.51	2.11	.143	.572	.149
TT × WT × SOA × Participant(Group)	1475.00	1.0	24.00	61.46				
• Filler Type	125.35		1	125.35	.22	.649	.728	.018
Participant(Group)	6887.50		12	573.96				
• TT × Filler Type	42.01		1	42.01	.47	.506	.618	.038
TT × Participant(Group)	1070.83		12	89.24				
• TT × SOA × Filler Type	146.53	1.0	2.00	73.26	.91	.418		.070
TT × SOA × Participant(Group)	1941.67	1.0	24.00	80.90				
• TT × WT × Filler Type	125.35		1	125.35	1.38	.262	.153	.103
TT × WT × Participant(Group)	1087.50		12	90.63				
• TT × WT × SOA × Filler Type	209.03	1.0	2.00	104.51	1.70	.204		.124
TT × WT × SOA × Participant(Group)	1475.00	1.0	24.00	61.46				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (□|H0)* for Kruskal-Wallis test, *p (□²|H0)* for Friedman test and *p (□|H0)* for sign test.

Figure 27 also shows that, overall, catch-trial performance was poorer for AM-words than for TM-words. Again replicating Experiment 6, the **catch (AM-words) > catch (TM-words)** contrast gave a difference of **28.7** percentage points with a 95% CI of 24.9 to 32.6, that was statistically significant ($F(1,18) = 41.28$, $MSE = 240.27$, $p < .001$).

Latency data

The mean correct no-response latencies as a function of Trial Type, Word Type, SOA condition, and Filler Type, separately for each trial block, are presented in Table 20 of Appendix C. Figure 28 shows the mean correct no-response latencies for the AM-words, TM-words, and CM-words both for the no-match trials and the catch trials for the English-fillers condition and the Dutch-fillers condition. These mean correct no-response latencies were collapsed over SOA, trial block and participant (sub)group.

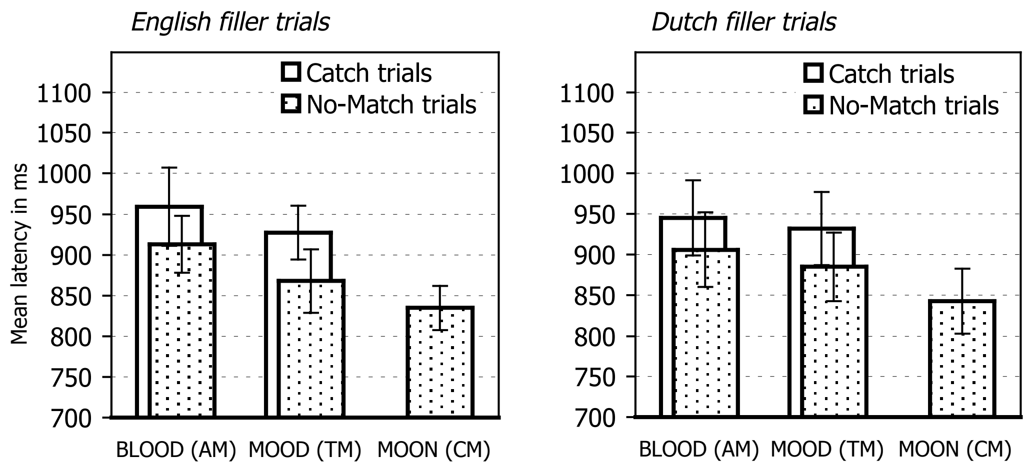

Figure 28. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) and collapsed over SOA for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 7. Error bars represent the standard error of the mean.

As can be seen in Figure 28, the ratio of friends and enemies was associated with correct no-response latencies. Response latencies were longer overall for AM-words than for TM-words (**33** ms, 95% SCI -21 to 88) and CM-words (**71** ms, 95% SCI 32 to 109), and response latencies for CM-words were shorter than those for TM-words (**37** ms, 95% SCI 3 to 72).

Omnibus analysis of variance. Table 40 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 2 (Word Type: AM vs. TM) by 3 (SOA condition: SOA 1 vs. SOA 2 vs. SOA 3) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA in which Filler Type was treated as a between-subjects variable. As can be confirmed by looking at Table 40, preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of trial block (Block Position) which accounted for a moderate percentage of variance. Again, participants responded faster across temporal positions of trial blocks. Furthermore, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, for the significance tests and the accompanying (simultaneous) confidence intervals, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Table 40.
Analysis of variance on percentages of false-positives for Experiment 7.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Session	65495.11	1.0	2.00	32747.56	8.39	.001	.318
Session × Participant(Group)	140539.17	1.0	36.00	3903.87			
• Sequence	509227.23		5	101845.45	.37	.857	.135
Participant(Group)	3271357.79		12	272613.15			
• TT	164021.28		1	164021.28	7.41	.019	.382
TT × Participant(Group)	265556.13		12	22129.68			
• TT × WT	1958.34		1	1958.34	.07	.792	.006
TT × WT × Participant(Group)	322718.13		12	26893.18			
• TT × SOA	132375.52	1.0	2.00	66187.76	7.42	.003	.382
TT × SOA × Participant(Group)	214148.25	1.0	24.00	8922.84			
• TT × WT × SOA	26698.84	1.0	2.00	13349.42	.49	.621	.039
TT × WT × SOA × Participant(Group)	659362.25	1.0	24.00	27473.43			
• Filler Type	4.25		1	4.25	.00	.997	.000
Participant(Group)	3271357.79		12	272613.15			
• TT × Filler Type	1906.53		1	1906.53	.09	.774	.007
TT × Participant(Group)	265556.13		12	22129.68			
• TT × SOA × Filler Type	26822.90	1.0	2.00	13411.45	1.50	.243	.111
TT × SOA × Participant(Group)	214148.25	1.0	24.00	8922.84			
• TT × WT × Filler Type	96.84		1	96.84	.00	.953	.000
TT × WT × Participant(Group)	322718.13		12	26893.18			
• TT × WT × SOA × Filler Type	449.22	1.0	2.00	224.61	.01	.992	.001
TT × WT × SOA × Participant(Group)	659362.25	1.0	24.00	27473.43			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Planned contrasts. Turning to the analyses that contrasted overall no-match-trial and catch-trial performance, Figure 28 indicates that Trial Type had an effect on no-response latencies. As was observed in the previous experiments, participants produced longer response latencies on catch trials (M = 941 ms) than on no-match trials (M = 893 ms), in which the **catch > no-match** contrast gave an overall difference of **48** ms (MSE = 2952.89, 95% CI 15 to 81, collapsed over SOA condition and Filler Type). Table 40 indicates that this statistically significant main effect accounted for a considerable percentage of variance.

The interaction effect of Trial Type and Word Type was not statistically significant (see Table 40). The Trial Type effect was 43 ms for AM-words and for TM-words it was 53 ms. The corresponding **Trial Type effect (AM-words) > Trial Type effect (TM-words)** contrast, collapsed over SOA condition and Filler Type, gave a difference (in opposite direction) of **-10** ms (95% CI -82 to 61). With regard to catch-trial performance, Figure 28 indicates that, overall, (catch-trial) latencies were longer for AM-words than for TM-words. The **catch (AM-words) > catch (TM-words)** contrast gave a difference of **23** ms with a 95% CI of -22 to 68, that was not statistically significant ($F(1,18) = 1.16$, $MSE = 5449.41$, $p = .295$).

Analyses evaluating the effect of stimulus-list composition

Experiment 7 was a revised version of Experiment 6 and designed in such a way that a potential influence of stimulus-list composition (i.e., Filler Type) should have a greater likelihood of being observed. In Experiment 7, the effect of the Filler Type variable was of primary interest. Yet, in spite of all modifications, participants in the Dutch-fillers and English-fillers conditions produced nearly identical results on the print-to-speech correspondence task (see Figure 27). This was, in fact, also observed in Experiment 6. In Experiment 7, participants in the Dutch-fillers condition produced, overall, fewer false-positive errors ($M = 18.7\%$) than participants in the English-fillers condition ($M = 20.0\%$). The difference in group means was 1.3 percentage points (95% CI -6.7 to 4.0) and not statistically significant, $F(1,22) = 0.26$, $MSE = 39.87$, $p = .614$ (Mann-Whithey U, $p = .728$). Likewise, no-response latencies in the Dutch-fillers condition ($M = 917$ ms) and in the English-filler condition ($M = 917$ ms) were virtually identical. The difference in group means was less than a millisecond (95% CI -112 to 113) and not statistically significant, $F(1,22) = 0.00$, $MSE = 17615.56$, $p = .997$. With regard to the Trial Type effect on error rates and no-response latencies, this was smaller in the Dutch-fillers condition (**23.2** percentage points and **43** ms) than in the English-filler condition (**24.7** percentage points and **53** ms), with statistically non-significant group differences of 1.5 percentage points ($MSE = 22.80$, 95% CI -5.6 to 2.5) and 10 ms ($MSE = 5076.37$, 95% CI -71 to 50). Accordingly, as is shown in Tables 39 and 40, the Filler Type by Trial Type interaction effects for error rates and response latencies were not statistically significant. Moreover, Tables 39 and 40 show that there were no statistically significant two- or three-way interaction effects concerning Filler Type, SOA condition and Trial Type.

SOA effect on error rates

Also of interest in experiment 7 was the effect of SOA on the number of false-positive errors. The effect of SOA condition was already examined in Experiment 5 that used English instead of Dutch enemy words. In the analyses that follow, data of CM-words were excluded. As in Experiment 5, we did not evaluate the SOA effect on response latencies because the timing of the different events in a trial was not identical across SOA conditions. In the analyses that follow, the data of the English-filler and Dutch-filler conditions were collapsed.

As was observed in Experiment 5, error rates were, overall, higher for SOA 2 than for SOAs 1 and 3. Furthermore, error rates were lowest in SOA 3. However, unlike what was found in Experiment 5, they appeared particularly low for no-match trials. Tables 39 and 40 however indicate that the main effect of SOA was not statistically significant. Furthermore, the SOA main effects for error rates and response latencies did not account for much variance. Finally, unlike Experiment 5, there was no evidence for a substantial Trial Type by SOA interaction effect.

The same orthogonal polynomial contrasts as in Experiment 5 were conducted to explore the linear (i.e., SOA 1 > SOA 3) and quadratic (i.e., SOA 1 < SOA 2 > SOA 3) effects across the three levels of SOA. Regarding the data of the *no-match trials*, recall that for this type of trial we expected a quadratic but not a linear effect (see Experiment 5). The pattern of results resembles that of Experiment 5 (see Figure 20). For the no-match-trial data of Experiment 7, the linear effect gave an estimated value for $\psi_{\mathbf{linear}}$ of **3.1** percentage points, with a 95% CI of -0.1 to 6.4, that was not statistically significant ($F(1,12) = 4.41$, $MSE = 26.56$, $p = .058$, $\eta^2 = .269$). Unlike in Experiment 5, however, the quadratic effect was not statistically significant either. The estimated value of $\psi_{\mathbf{quadratic}}$ was **8.5** percentage points, with a 95% CI of -0.4 to 17.5 $F(1,12) = 4.34$, $MSE = 67.19$, $p = .059$, $\eta^2 = .266$).

Turning to the data of the *catch trials*, recall that for this type of trial we expected both a linear and a quadratic effect. Indeed, in Experiment 5, both effects were observed. In the present experiment however there was no evidence for either effect. The linear effect gave an estimated value for $\psi_{\mathbf{linear}}$ of **1.3** percentage points, with a 95% CI of -2.7 to 5.2, that was not statistically significant ($F(1,12) = 0.49$, $MSE = 38.54$, $p = .499$, $\eta^2 = .039$). The estimated value of $\psi_{\mathbf{quadratic}}$ for the quadratic effect was **2.9** percentage points, with a 95% CI of -6.5 to 12.3, that was not statistically significant ($F(1,12) = 0.46$, $MSE = 74.65$, $p = .512$, $\eta^2 = .037$).

Discussion

In line with Experiment 6, the present findings show that spelling-to-sound knowledge of Dutch enemy neighbors has an influence on second-language word perception. For Dutch-English bilinguals, again, rejecting a catch trial that consisted of a printed inconsistent word and a spoken rime that was derived from a Dutch enemy appears rather difficult. This finding substantiates the language non-selective view of bilingual word perception, which assumes auxiliary cross-language coding of inappropriate Dutch phonology.

Experiment 7 further explored the effect of SOA condition on performing the print-to-speech correspondence task. Although no-match-trial performance was, as in Experiment 5, poorest in SOA 2 (i.e., where print and speech are presented simultaneously), there was no statistical support for an SOA effect, neither for no-match trials nor for catch trials. However, it should be noted that, as compared to

Experiment 5, the present experiment had considerably fewer participants at its disposal. Consequently, statistical power and precision was relatively low and, hence, the null effects should be interpreted with caution.

Nevertheless, observing poor catch-trial performance even in a condition where speech is presented after some substantial delay (i.e., SOA 3) supports the idea that the reading system never settles fully in the appropriate state, because even after a substantial time gap, the context that comes about by processing of the spoken rime is still effective. Thus, in a catch trial, a spoken rime may cause participants to perceive MOOD's phonology to rhyme with that of LOOD even when speech is presented as much as half a second after off-set of the printed word.

The primary goal of Experiment 7, however, was to attempt once again to show that the relative prominence of the Dutch, non-target language affects the process of interlingual phonological coding. In Experiment 6, no evidence for that was found, and it was suggested that this was due to the fact that participants produced so many errors that it diminished a possible impact of language mode. Therefore, in the present experiment, steps were taken to accomplish that participants produced fewer errors. It was hypothesised that an overall reduction in (false-positive) error rate might be achieved by way of adding an SOA manipulation. If such a reduction would indeed occur, Dutch enemy neighbours might still be found to have a relatively strong influence on task performance when the participants' language mode is tuned towards the Dutch, non-target language. Indeed, overall rates of false-positive errors in Experiment 7 appeared somewhat lower than in Experiment 6. More formally, the reduction across experiments was 2.4 percentage points (95% CI -1.1 tot 5.8), but it was not statistically significant ($F(1,82) = 1.85$, $MSE = 51.90$, p =.177; Mann-Whithey U, p =.139). It appears, therefore, that the procedural changes that were meant to improve performance applied to Experiment 7 were of limited success. More specifically, they did not result in a substantial difference between the Dutch-fillers and English-fillers conditions. As in Experiment 6, a nearly zero effect of stimulus-list composition occurred. In other words, it appears that adding Dutch filler trials to the stimulus set did not cause enhanced (inappropriate) phonological coding according to Dutch spelling-to-sound knowledge. But note once again that in Experiment 7 statistical power and precision was relatively limited. Possibly, this is due to the fact that the experimental design treated Filler Type as a between-participants variable. This inefficient design feature may have caused a relatively high error variance.

The next experiment provides one final attempt to find out whether the relative prominence of the Dutch, non-target language affects the process of interlingual phonological coding. In Experiment 8, which is basically a replication of Experiment 6 (and hence does not include an SOA variable), Filler Type is planned as within-participants variable. Thus, participants are presented twice with the stimulus materials used in Experiment 6, once in the Dutch-fillers condition and once in the

English-fillers condition. This design modification may give rise to reduced error variance and hence improved efficiency. A further modification was included with the purpose to improve overall task performance. This modification involved an adjustment in task instruction. Unlike in the previous experiments, participants were not required to make a quick response, but only to perform *as accurately* as possible. As was explained previously, if performance improves substantially, a potentially detrimental effect of Dutch enemy neighbours might be revealed. In that case, an effect of stimulus-list composition might emerge.

EXPERIMENT 8

STIMULUS-LIST COMPOSITION VARIED WITHIN PARTICIPANTS

Method

Participants and materials

A group of 8 Dutch-English bilinguals served as participants. They were presented with the same materials as used in Experiment 6. Originally, the number of participants was planned to add up to 16. This number could, unfortunately, not be achieved because of administrative restrictions.

Experimental design

Experiment 8 used the same basic design as Experiment 6. Thus, nine groups of target trials were created that represented specific combinations of Trial Type and Word Type. These nine combinations were: Match trial (AM), match trial (TM), match trial (CM), no-match trial (AM), no-match trial (TM), no-match trial (CM), catch trial (AM), catch trial (TM), and catch trial (CM). As in the previous experiments, the temporal order of the two trial blocks was counterbalanced across two different participant groups according to a single Latin square.

Whereas in Experiment 6 Filler Type was varied between participants, in Experiment 8 each participant performed the print-to-sound correspondence task both in the Dutch-fillers condition and in the English-fillers condition. In the Dutch-fillers condition the stimulus materials were mixed with (25%) Dutch filler trials and in the English-fillers condition they were mixed with (25%) English filler trials. In the analyses, Filler Type was treated as a *within*-subjects variable. Finally, the temporal order of the two Filler Type conditions was counterbalanced across two different participant groups according to a single Latin square. Participants were randomly assigned to the different sequences.

Results

Participants were presented with nine different groups of target trials. Each participant responded two times (i.e., in the Dutch-fillers and English-fillers conditions) to 60 match trials, which consisted of 20 AM-words (e.g., PAID), 20 TM-words (e.g., SAID), and 20 CM-words (e.g., STAIN). In addition, each participant responded two times to 30 no-match trials and 30 catch trials. Both the no-match trials and the catch trials consisted of 10 AM-words (e.g., BLOOD), 10 TM-words (e.g., MOOD), and 10 CM-words (e.g., MOON). For each participant, the correct response latencies within these nine groups were averaged and percentage of errors was calculated for each of the nine groups. For each participant, the data that entered the statistical analyses consisted of six latency means and six percentages of false-positives for no-match trials, and six latency means and six percentages of false-positives for catch trials.

Data filtering

Response latencies of trials on which the participant responded incorrectly were excluded from the latency analyses. This resulted in a rejection of a total of 25.6% for the catch trials and 9.4% for the no-match trials. None of the trials was excluded because of apparatus failure or because the response latency was shorter than 200 ms. A trial was cancelled if the participant failed to respond within 6 seconds after onset of the printed word. In the analyses, this experimental procedure resulted in a cut-off that rejected all latencies greater than 6 seconds. There were no latencies that exceeded this limit. Latencies were further truncated to 2500 ms, which resulted in rejection of 1.0% of the correct response latencies. This 2500 ms cut-off was suggested by a visual inspection of a scatter plot.

Error data

In a mismatch trial, a participant produced an error (i.e., a false-positive) when he or she pressed the “yes” button when presented with a printed word and a spoken rime that were actually incongruent with each other. Figure 29 shows the mean percentages of false-positives for the AM-words, TM-words, and CM-words both for the no-match trials and the catch trials, and as a function of Filler Type (English-fillers condition vs. Dutch-fillers condition). These mean percentages of false-positives were collapsed over trial block and participant (sub)group.

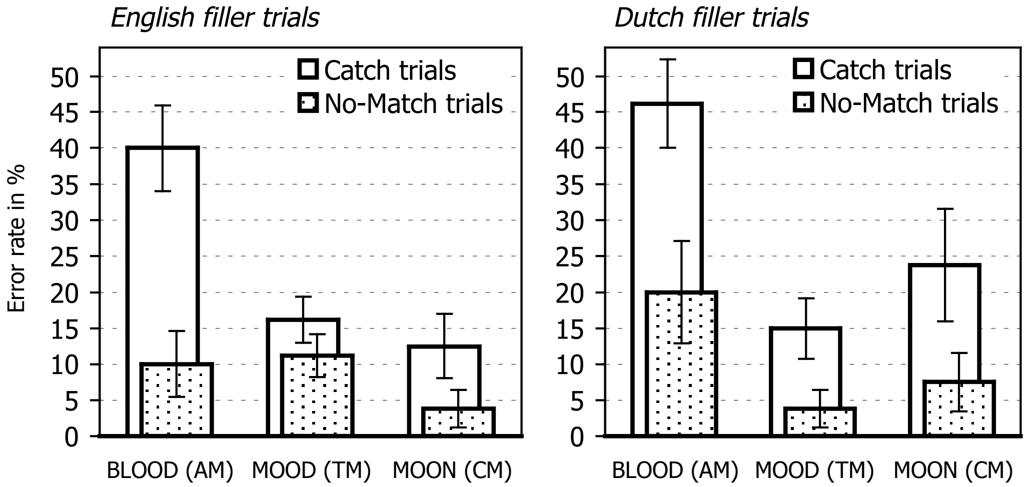

Figure 29. Mean percentages of false-positives as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 8. Error bars represent the standard error of the mean.

As can be seen in Figure 29, the number of false-positives depended upon the ratio of friends and enemies, both for no-match trials and catch trials. Overall, participants made more errors on AM-words than on TM-words and CM-words. The number of errors for CM-words was not markedly lower than for the words with typical mappings. Also apparent in Figure 29 is an effect of Trial Type, which appears largest for AM-words. These general findings match those that were observed in Experiments 3-7.

Omnibus analysis of variance. Table 41 provides the results of a 2 (Trial Type: catch trial vs. no-match trial) by 3 (Word Type: AM vs. TM vs. CM) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA. The table includes the results of non-parametric tests. Preliminary analyses on the procedural variables indicated that there was no evidence for a substantial (Filler Type) Sequence effect, but there was evidence for an effect of temporal position of Filler Type, which accounted for a substantial percentage of variance. Again, participants produced fewer errors across sessions. Furthermore, adding participant subgroup (Filler Type sequence) as a between-subjects variable turned out to improve statistical power and precision in a notable way. Therefore, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

Table 41.
Analysis of variance on percentages of false-positives for Experiment 8.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	p^a	η_p^2
• Block Position	2.81		1	2.81	.07	.800	.727	.018
Block Position × Participant	153.39		4	38.35				
• Filler Type Position	200.93		1	200.93	21.22	.004	.016	.780
Filler Type Position × Participant	56.81		6	9.47				
• Sequence Block Position	16.67		1	16.67	.02	.888	.663	.006
Participant(Group)	2983.33		4	745.83				
• Sequence Filler Type Position	2016.67		1	2016.67	4.03	.091	.147	.402
Participant(Group)	3000.00		6	500.00				
• TT	6337.50		1	6337.50	53.68	< .001	.008	.899
TT × Participant(Group)	708.33		6	118.06				
• WT	6418.75	1.0	2.00	3209.38	17.61	< .001	.003	.746
WT × Participant(Group)	2187.50	1.0	12.00	182.29				
• TT × WT	1768.75	.94	1.87	943.59	3.53	.067	.088	.371
TT × WT × Participant(Group)	3004.17	.94	11.25	267.11				
• Filler Type	337.50		1	337.50	5.93	.051	.999	.497
Filler Type × Participant(Group)	341.67		6	56.94				
• Filler Type × Sequence Filler Type Position	1204.17		1	1204.17	21.15	.004	.020	.779
Filler Type × Participant(Group)	341.67		6	56.94				
• TT × Filler Type	66.67		1	66.67	.44	.533	.999	.068
TT × Filler Type × Participant(Group)	916.67		6	152.78				
• WT × Filler Type	793.75	1.0	2.00	396.88	7.37	.008	.045	.551
WT × Filler Type × Participant(Group)	645.83	1.0	12.00	53.82				
• TT × WT × Filler Type	152.08	1.0	2.00	76.04	.87	.443	.381	.127
TT × WT × Filler Type × Participant(Group)	1045.83	1.0	12.00	87.15				

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.
[a] *P*-value of nonparametric test, *p (U|H0)* for Mann-Whitney test, *p (χ^2|H0)* for Friedman test and *p (□|H0)* for sign test.

Latency data

Figure 30 shows the mean correct no-response latencies for the AM-words, TM-words, and CM-words both for the no-match trials and the catch trials, and as a function of Filler Type. These mean correct no-response latencies were collapsed over trial block and participant (sub)group.

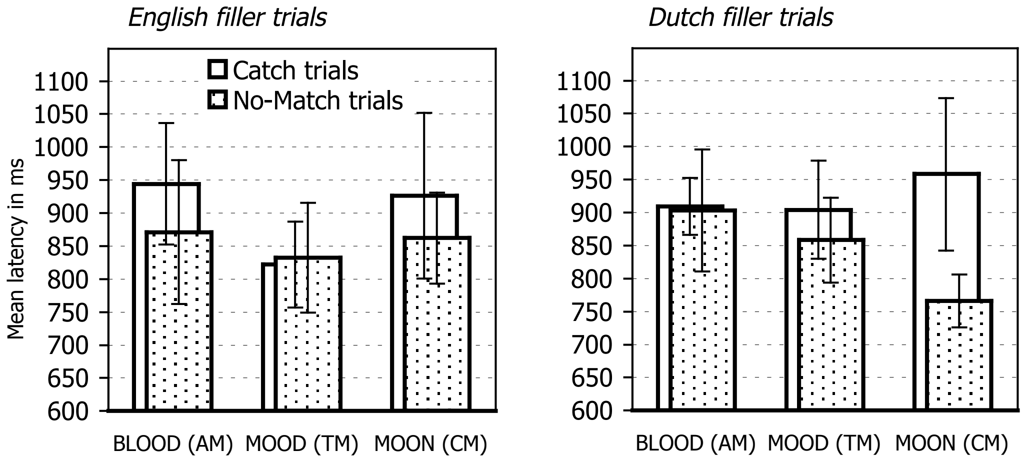

Figure 30. Mean correct no-response latencies as a function of Trial Type (catch trials vs. no-match trials) and Word Type (AM-words vs. TM-words vs. CM-words) for the English-fillers condition (left panel) and the Dutch-fillers condition (right panel) in Experiment 8. Error bars represent the standard error of the mean.

The pattern of response latencies mirrors that observed in Experiments 6 and 7. As can be seen in Figure 30, for no-match trials the ratio of friends and enemies was associated with correct no-response latencies, according to an expected rank order of Word Type. The rank order for catch trials, however, was not as expected. This was also the case in Experiment 6 and may be related to the high rates of false-positive errors in these conditions.

Omnibus analysis of variance. Table 42 presents the results of a 2 (Trial Type: catch trial vs. no-match trial) by 3 (Word Type: AM vs. TM vs. CM) by 2 (Filler Type: English-fillers condition vs. Dutch-fillers condition) repeated-measures ANOVA. Preliminary analyses on the procedural variables indicated that there was no evidence for a substantial Sequence effect, but there was evidence for an effect of temporal position of Filler Type, which accounted for a considerable percentage of variance. Again, participants responded faster across temporal positions of trial blocks. Furthermore, adding participant subgroup as a between-subjects variable turned out to improve statistical power and precision in a notable way. Hence, the treatments×participants(group) interaction sum of squares was used to estimate error variance.

As in Experiment 7, the primary goal of Experiment 8 was to evaluate the effect of Filler Type on mismatch-trial performance. In an effort to improve task performance, participants received instructions that stressed accuracy over speed. In addition, Filler Type was included as a within-participants variable in order to reduce error variance. This should make the design more efficient.

Table 42.
Analysis of variance on correct no-response latencies for Experiment 8.

Source of variance	*SS*	ε	*df*	*MS*	*F*	*p (F|H0)*	η_p^2
• Block Position	15006.25		1	15006.25	4.82	.093	.547
Block Position × Participant	12446.50		4	3111.63			
• Filler Type Position	25043.06		1	25043.06	13.89	.010	.698
Filler Type Position × Participant	10814.88		6	1802.48			
• Sequence Block Position	408465.04		1	408465.04	.97	.380	.196
Participant(Group)	1680480.42		4	420120.10			
• Sequence Filler Type Position	5490.38		1	5490.38	.01	.914	.003
Participant(Group)	1680480.42		4	420120.10			
• TT	84372.04		1	84372.04	6.45	.044	.518
TT × Participant(Group)	78531.79		6	13088.63			
• WT	41795.40	1.0	2.00	20897.70	1.77	.212	.228
WT × Participant(Group)	141862.67	1.0	12.00	11821.89			
• TT × WT	10497.02	1.0	2.00	5248.51	.23	.795	.038
TT × WT × Participant(Group)	268986.08	1.0	12.00	22415.51			
• Filler Type	12330.67		1	12330.67	1.22	.311	.169
Filler Type × Participant(Group)	60476.29		6	10079.38			
• Filler Type × Sequence Filler Type Position	171873.38		1	171873.38	17.05	.006	.740
Filler Type × Participant(Group)	60476.29		6	10079.38			
• TT × Filler Type	27812.04		1	27812.04	3.82	.098	.389
TT × Filler Type × Participant(Group)	43642.96		6	7273.83			
• WT × Filler Type	3879.02	1.0	2.00	1939.51	.27	.770	.043
WT × Filler Type × Participant(Group)	87125.08	1.0	12.00	7260.42			
• TT × WT × Filler Type	63581.65	1.0	2.00	31790.82	1.77	.595	.083
TT × WT × Filler Type × Participant(Group)	215010.17	1.0	12.00	17917.51			

Note. Type III Sum of Squares with Huynh-Veldt ε-adjusted degrees of freedom.

Analyses evaluating the effect of stimulus-list composition

Table 41 indicates that the main effect of Filler Type accounted for a non-trivial percentage of variance. Specifically, false-positive error rates were larger in the Dutch-fillers condition than in the English-fillers condition. The difference was 3.8 percentage points (95% CI 0.0 to 7.5) but not statistically significant, although the value of *p* came very near the nominal 5% level of significance. In addition to that, Table 41 reveals a statistically significant Filler Type by Sequence interaction effect that qualifies the Filler Type main effect. Simple effects were conducted (using the regular treatments×participants interaction sum of squares to estimate error variance), which revealed an asymmetric cross-over effect. For participants who first performed in the Dutch-fillers condition followed by the English-fillers condition, error rates were highest in the Dutch-fillers condition. The difference was 10.8 percentage points (95% CI 1.8 to 19.9) and statistically significant ($F(1,3) = 14.49$, $MSE = 16.20$, $p = .032$). In contrast, for participants who performed in these conditions in reversed order, error rates were highest in the English-fillers condition. This difference was 3.3 percentage points (95% CI -0.4 to 7.1) but not statistically significant ($F(1,3) = 8.00$, $MSE = 2.79$, $p = .066$).

Table 41 also shows that neither the interaction effect of Filler Type and Trial Type, nor that of Filler Type, Trial Type, and Word Type was statistically significant. In addition, none of these effects showed statistically significant interactions with Sequence. The Filler Type by Word Type interaction effect, however, was statistically significant. Simple effects showed that, overall, the Filler Type effect for TM-words (-4.4 percentage points, 95% SCI -13.8 tot 5.0, $F(1,6) = 2.33$, $MSE = 32.81$, $p = .178$) was in opposite direction from that for the AM-words (8.1 percentage points, 95% SCI -1.3 to 17.5, $F(1,6) = 8.05$, $MSE = 32.81$, $p = .030$) and CM-words (7.5 percentage points, 95% SCI 0.8 to 14.2, $F(1,6) = 13.50$, $MSE = 16.67$, $p = .010$).

The Filler Type effect on response latencies is evaluated in Table 42. The table shows that the main effect of Filler Type accounted for a moderate percentage of variance, which however was not statistically significant. Specifically, correct no-response latencies were 23 ms longer (95% CI -28 to 73) in the Dutch-fillers condition than in the English-fillers condition. However, a qualifying Filler Type by Sequence interaction effect was apparent that was statistically significant. Simple effects (using the regular treatments×participants interaction sum of squares to estimate error variance) revealed an asymmetric cross-over effect. For participants who first performed in the Dutch-fillers condition and then in the English-fillers condition, response latencies were longest in the Dutch-fillers condition. The difference was 107 ms (95% CI 32 to 183) and statistically significant ($F(1,3) = 20.53$, $MSE = 1120.78$, $p = .020$). In contrast, for participants who performed in these conditions in the opposite order, response latencies were longest in the English-fillers condition. This difference was 62 ms (95% CI -45 to 169) but not statistically significant ($F(1,3) = 3.42$, $MSE = 2243.63$, $p = .162$).

Further inspection of Table 42 indicates that the Filler Type effect may be different for no-match trials and catch trials. Although the Filler Type by Trial Type interaction effect was not statistically significant, simple effects revealed a statistically significant effect of Filler Type for catch trials (57 ms, 95% SCI 10 to 104, $F(1,6) = 12.77$, $MSE = 1012.48$, $p = .012$) and a statistically non-significant effect of Filler Type for no-match trials (-12 ms, 95% SCI -114 to 91, $F(1,6) = 0.11$, $MSE = 4767.56$, $p = .749$). Furthermore, Table 42 shows that neither the interaction effect of Filler Type and Word Type, nor that of Filler Type, Trial Type, and Word Type was statistically significant. In addition, none of these effects showed statistically significant interactions with Sequence.

Discussion

The basic observations of Experiments 3-7, involving effects of Word Type and Trial Type, recurred in Experiment 8. Yet, the primary object of this final experiment was to see whether at this point evidence could be found for an effect of stimulus-list composition (i.e., a Filler Type effect). Experiment 8 differed from Experiment 6 in modified task instructions that stressed accuracy over speed. This was intended to improve task performance and hence should increase the chance of demonstrating a negative impact of stimulus-list composition. Furthermore, the Filler Type variable was appropriately changed so that repeated measures were obtained, thereby possibly improving the efficiency of the design.

Overall, false-positive error rates in Experiment 8 were indeed somewhat lower than in Experiment 6, with a difference of 1.4 percentage points (95% CI -4.4 tot 7.1) that was however not statistically significant ($F(1,66) = 0.24$, $MSE = 58.60$, $p = .630$; Mann-Whithey U, $p = .393$). In other words, the design modifications did not bring about much of an improvement in performance. However, changing Filler Type to a within-participants variable appeared to reduce error variance rather well. Tables 37, 38, 41, and 42 show that the Filler Type × participants(group) interaction sum of squares in Experiment 8 was substantially lower than the estimate of error variance in Experiment 6 (less than 17% of the value of Experiment 6), even though the latter experiment contained a considerably larger number of participants.

Whereas Experiments 6 and 7 yielded null-effects of stimulus-list composition, Experiment 8 actually found some evidence of an effect of this variable. Comparison of the Dutch-fillers and English-fillers conditions showed a Filler Type effect on error rates and latencies. That is, participants performed poorest in the Dutch-fillers condition. Furthermore, for errors the effect appeared larger than in Experiments 6 and 7. Statistical support, however, was not obtained. Nevertheless, the effect appeared fairly large for the group of participants who performed first in the Dutch-fillers condition and then in the English-fillers condition. It would seem that the presence of trials that involve Dutch words causes enhanced phonological coding according to Dutch spelling-to-sound knowledge—given the prerequisite that an experimental session starts off with Dutch filler trials, promoting a Dutch language mode at the outset. If, however, an English language mode is promoted first, Dutch filler trials apparently do not deteriorate task performance beyond the opposite, positive effect of temporal position of Filler Type. This may suggest that the bilinguals' language mode is under limited strategic control and does not switch easily from non-native to native language mode.

In particular for specific word types, an influence of stimulus-list composition was in fact quite apparent. Specifically, for no-match trials, the observed Filler Type effect was not the same across the three word types. For AM-words and CM-words, evidence was obtained for a Filler Type effect, but not for TM-words.

Furthermore, the Filler Type effect appeared to depend on the type of trial. For catch trials more than for no-match trials, the presence of Dutch filler trials seemed to result in relatively poor task performance. If indeed in catch trials Dutch fillers cause enhanced phonological coding according to Dutch spelling-to-sound knowledge, then spoken rimes (derived from Dutch enemy words) should have more capacity to restore degraded, inappropriate phonological codings in these trials. This is not expected for no-match trials since in this type of trials such codings are not fostered. The reason is that in no-match trials spoken rimes are derived from unrelated words that do not correspond to any of the inappropriate phonological codings.

In sum, it was once again assumed that adding Dutch filler trials to the stimulus set would increase the participants' relative prominence of the Dutch, non-target language. Under the specific conditions of Experiment 8 this resulted in tenuous evidence that it caused slightly enhanced (inappropriate) phonological coding according to Dutch spelling-to-sound knowledge. Contrary to the findings of Experiments 6 and 7, the results of Experiment 8 thus appear to converge with the stimulus-list composition effects as, for example, reported in Dijkstra et al. (1998). Nevertheless, the presently observed effects of Filler Type seem rather small, which is more in line with the conclusions of Dijkstra and Van Hell (2003), who reviewed evidence for the claim that bilingual visual word perception always proceeds language non-selectively.

Further experimentation should pinpoint the specific conditions under which a reliable influence of stimulus-list composition in the print-to-speech correspondence task may be obtained. For the moment, the relevant findings of Chapter 5 indicate that the influence of stimulus-list composition is rather small, substantially smaller than that observed in reading tasks such as lexical decision. The difference between such tasks and the present one may be caused by the fact that in Experiments 6-8 the focus is on processes of phonological coding in the initial conditions of word perception and in other studies it is on more global-level linguistic coding, which may be more dependent on the relative prominence of the non-target language.

6 General Discussion

The present study investigated the nature of phonological coding in the perception of written words. Our first step was to introduce the concept of manifold relations, a simple notion that proved a constructive background for almost every theoretical aspect of this study. In brief, cognitive systems deal with inconsistent experience by creating manifold associations between surface forms and multiple functions. As a result, the act of processing a surface form initially causes the cognitive system to consider all previously learned functions. This idea was exercised in a series of experiments that took spelling-to-sound relations as the basic object of investigation. In an orthography, manifold relations arise when spellings have more than one possible pronunciation, such as in case of the spelling body $-OOD$, which has different pronunciations in the neighbouring words MOOD and BLOOD. The resultant inconsistency of spelling-to-sound entails ambiguity of pronunciation, the upshot of which may be seen in reading performance. To explain the classic spelling-to-sound consistency effect in word reading, it is generally assumed that spelling-to-sound knowledge of enemy neighbors causes coding of extraneous inappropriate phonology, which underscores the mandatory nature of phonological coding. Consistent with the phonological coherence hypothesis (Van Orden & Goldinger, 1994) and, more generally, the strong phonological theory (Frost, 1998), an effect of manifold spelling-to-sound relations is an indication that visual word perception involves obligatory simultaneous coding of appropriate and inappropriate phonology that compete with each other. Framed in dynamic systems theory, ambiguity of spelling-to-sound engenders multistable perception, meaning that there are multiple stable attractor solutions corresponding to the competing pronunciations.

Of primary interest in this study is that manifold spelling-to-sound relations also span *across* writing systems: Two languages may assign entirely different pronunciations to an identical spelling. Returning to the example of the spelling body $-OOD$, the pronunciation in a Dutch word (e.g., LOOD) is quite different from the way it is pronounced in cross-language neighbors such as MOOD and BLOOD. Recognizing the significance of interlingual manifold relations, we exploited a unique quality of Dutch-English bilinguals that may pose a formidable challenge for the idea that phonological coding is a central and primary constituent of visual word perception. If this view provides an accurate picture of word perception, we should find phonological coding be as important in bilingual word perception as it is in monolingual word perception (Brysbaert et al., 1999). The aforementioned quality of Dutch-English bilinguals is that, when reading English words, they can map spelling to sound according to English but also to Dutch associations. In other words, generally, for spelling bodies that occur in more than one language, bilinguals can develop manifold, interlingual spelling-to-sound associations.

These considerations give rise to an intriguing hypothesis, which forms the heart of this thesis. Essentially, it combines two ideas, namely, the fundamental role of phonology in printed word perception and, with respect to bilingual reading, language non-selective processing. With regard to the population of Dutch-English bilinguals, this hypothesis states that in the process of English word perception, inappropriate phonology from the native, Dutch language emerges simultaneously with correct English phonology (Jared & Kroll, 2001; Van Wijnendaele & Brysbaert, 2002). Thus, for a Dutch-English bilingual, visual processing of MOOD may elicit extraneous inappropriate phonology according to Dutch spelling-to-sound associations, which yields a phonological structure (i.e., /od/) that, incidentally, rhymes with the English word ROAD. Hence, in this study, creating the experimental conditions wherein MOOD is perceived to rhyme with the Dutch word LOOD demonstrates that extraneous, simultaneous cross-language phonology emerges in bilingual word perception—an inference that strongly reinforces the generality of the strong phonological theory.

MAJOR FINDINGS AND CONCLUSIONS

Our investigation hinges on a cohesive set of eight psycholinguistic experiments. Each experiment orchestrated a large number of behavioural observations involving skilled readers processing a single printed word. The basic goal was to learn more about general principles of phonological coding in monolingual and bilingual visual word perception. A total of 376 research participants contributed to this goal. They were all individually tested in concise blocks of trials wherein they responded to various sets of carefully chosen linguistic stimuli. Experimentation took place in two different countries: The Netherlands and the United States of America. Altogether, our Dutch and American participants faithfully generated the vast amount of nearly 90,000 responses, which granted us a massive and fairly exclusive collection of observations. The results of these experiments together demonstrate a clear and persistent influence of manifold spelling-to-sound relations on word processing. This finding reveals several principles of phonological coding in visual word perception, which, in turn, set a number of potential constraints for contemporary accounts of monolingual and bilingual word perception in reading. The major findings are the following.

Multistability in Intralingual Phonological Coding

English words reside among several neighbors; some of which are friends and others are enemies. Our first objective was to assess the impact of English enemy neighbors on the visual perception of English words. Chapter 3 described a standard word-naming experiment in which monolingual native English speakers and Dutch-English bilinguals performed the task of reading aloud English printed words. This experiment examined the effect of manifold intralingual spelling-to-sound relations on word-naming performance. The relevant contrasts reported in Experiment 1 are shown in the forest plots of Figures 31 and 32 (the dotted bars). It was found that spelling-to-sound knowledge of strong English enemy neighbors not only hinders monolingual but also bilingual naming performance, which replicates findings of previous research (e.g., Jared, 1997, 2002; Jared & Kroll, 2001; Jared et al., 1990). This classic spelling-to-sound consistency effect was interpreted suggesting that phonology is fundamental not only to monolingual but also to bilingual visual word perception (Drieghe & Brysbaert, 2002; Van Wijnendaele & Brysbaert, 2002), assuming that the effect results from competition between appropriate and inappropriate phonological codings.

In Chapter 4, explicit evidence was presented for the fundamental idea that competition results from inappropriate phonological coding. In Experiments 2 and 3, Dutch-English bilinguals performed the print-to-speech correspondence task. Figures 31-34 provide forest plots that summarize the relevant contrasts for Experiments 2-8. It was found that perceiving a match or mismatch between an English printed word (e.g., BLOOD) and an unrelated spoken rime was affected by knowledge of strong English enemy neighbors (e.g., MOOD). Essentially, this finding represents a spelling-to-sound consistency effect such as observed in word-naming performance, and, in the present case, it indicates that processing an English inconsistent word by Dutch-English bilinguals involves simultaneous coding of competing appropriate and inappropriate phonological structures. In terms of dynamic systems theory, this is due to *multistability*, which means that the same stimulus may support multiple percepts (e.g., Hock et al., 2003; Ploeger et al., 2002; Van Orden et al., 1997).

Specific evidence for the assumption that word perception involves coding of inappropriate phonology came from a particular class of mismatch trials where an inconsistent printed English word (e.g., MOOD) was accompanied by a spoken rime not derived from an unrelated word (e.g., BRIDE), but from an *enemy neighbor* (e.g., BLOOD, not visually presented in the trial). Responding to these “catch trials” appeared an extraordinary difficult test: A surprisingly large number of catch trials produced a false-positive error (see Figure 33), indicating that the printed word’s phonology was perceived to rhyme with that of its enemy. In other words, to evoke in an experiment that MOOD rhymes with BLOOD strongly suggests that perception of an inconsistent word involves extraneous coding of intralingual enemy phonology. In effect, this is an unequivocal demonstration of the mandatory nature of phonological coding.

This conclusion was further supported by the main results of Experiments 4 and 5. In addition, Experiment 5 compared performance of native English speakers and Dutch-English bilinguals on the print-to-speech correspondence task. It was found that the two language groups performed quite similar: Both groups showed clear evidence for coding of intralingual enemy phonology. However, the class of catch trials appeared particularly demanding for the Dutch-English bilinguals. This finding was explained by noting that for these participants knowledge of correct local orthographic-phonologic mappings for inconsistent English words may be relatively weak, therefore permitting a larger impact of coding of extraneous enemy phonology.

Interlingual Phonological Coding in Bilingual Word Perception

The impact of enemy neighbors on word perception was further investigated in Chapter 5. In this chapter, however, we allowed for more remote vicinity. That is, cross-language *Dutch* enemy neighbors were introduced, with the purpose of investigating the impact of manifold *interlingual* spelling-to-sound relations. Accordingly, in Experiments 6-8 the question was addressed whether, in English word perception, Dutch-English bilinguals not only engage knowledge of English but also of *Dutch* spelling-to-sound relations. Put differently, it sought support for the hypothesis that bilingual word perception involves coding of inappropriate, cross-language phonology (cf. Jared & Kroll, 2001).

The Dutch-English bilinguals of Experiments 6-8 again performed the print-to-speech correspondence task. As was found in Experiments 2-5, perceiving a match or mismatch between an English printed word and an unrelated spoken rime was affected by knowledge of strong (intralingual) English enemy neighbors (see Figures 31 and 32, for an overview). Once again, particular evidence was obtained from the class of catch trials. The catch trials of Chapter 5 consisted, as in Chapter 4, of an inconsistent printed English word (e.g., MOOD) simultaneously presented with a spoken rime that was derived from an enemy neighbor. Whereas in Chapter 4 the spoken rime was derived from an English enemy word (e.g., /}d/ as in BLOOD), in Chapter 5 it was derived from a *Dutch* enemy (e.g., /od/ as in the Dutch word LOOD). As was observed for the catch trials of Chapter 4, a spectacular large number of false-positive responses was shown: Dutch-English bilinguals frequently pressed the “yes” button when a word like MOOD was presented jointly with a Dutch pronunciation of the spelling body (see Figure 33). Thus, under these conditions, the printed word’s phonology appears every so often to be perceived to rhyme with that of a cross-language enemy. In the present case, to induce experimentally that MOOD rhymes with the Dutch word LOOD indicates that perception of an inconsistent English word includes coding of *cross-language* enemy phonology. This is consistent with the view that phonological coding in printed word perception proceeds essentially language non-selectively (cf. Van Wijnendaele & Brysbaert, 2002). Again, framed in terms of dynamic systems theory, this may be based on multistable, interlingual spelling-to-sound dynamics.

Interlingual Phonology and Language Mode

When processing words in a second language, bilingual readers have to deal with more than one language system. A key experimental factor that may influence the degree in which bilingual word processing proceeds language non-selectively relates to the composition of the stimulus-list, that is, whether words from one or both of the languages are included in the experiment. According to the language mode hypothesis of Grosjean (1997, 2001), this should affect the participants' language mode, which is assumed to alter the relative prominence of the non-target language system (but see Dijkstra & Van Hell, 2003). This reasoning is in agreement with several psycholinguistic studies that showed that language non-selective word processing occurs primarily when the bilingual reader's non-target language system is engaged (e.g., Dijkstra et al., 1998; see also De Groot et al., 2000).

Experiments 6-8 all included a stimulus-list composition variable. In view of the aforementioned psycholinguistic studies, it was expected that adding Dutch filler trials to the stimulus set would increase the participants' relative prominence of the Dutch, non-target language. Still, neither in Experiment 6 nor in Experiment 7 it affected performance on the print-to-speech correspondence task (see Figures 35 and 36). In these experiments, specifically, the presence of Dutch filler trials did not appear to cause enhanced coding of inappropriate cross-language phonology. However, Experiment 8, which included modified task instructions and a more efficient design, did produce some evidence for it. In sum, the effects of stimulus-list composition observed in Experiments 6-8 seem rather trivial, which indicates that stimulus-list composition has a smaller effect on performance on the print-to-speech correspondence task than on performance on regular reading tasks such as word naming and lexical decision.

As was explained previously, this difference across tasks can be accounted for by considering that performance on the print-to-speech correspondence task reflects processes of phonological coding in the initial conditions of word perception, of which the ballistic nature circumvents an external influence such as of language mode. It follows that the small impact of adding Dutch filler trials to the stimulus sets in Experiments 6-8 is an indication that phonological coding in bilingual word perception proceeds in an unyieldingly language non-selective fashion (cf. Brysbaert et al., 1999; Dijkstra & Van Hell, 2003). That is to say, when Dutch-English bilinguals read an interlingual inconsistent word such as MOOD, they always initiate both English and Dutch phonological structures according to language-specific spelling-to-sound relations. Take notice that the critical outcome is that language non-selective processing appears to occur even when the stimulus-list composition does *not* favor a Dutch language mode. Once more, recall that this finding diverges from that of earlier studies (e.g., Dijkstra et al., 1998), which yielded substantial evidence of language non-selective processing especially when participants were in the appropriate language mode.

The apparent discrepancy between our findings and those obtained in previous research (e.g., Dijkstra et al., 1998) suggests an extension of the standard language non-selective access hypothesis. Apparently, what we have learned is that the degree in which bilingual word processing attracts word knowledge of both the languages varies according to the time course of the form-function dynamics, with language non-selective processing occurring principally during early and mandatory (local-level) phonological coding. Again, language non-selective processing is expected to occur during the early phases of bilingual word perception, the initial conditions wherein codings of all phonological structures are launched that have previously been associated with a particular spelling body. For the bilingual reader, these structures consist of cross-language phonological codings arising from language-specific mappings between spelling and sound. Thus, according to our findings, phonological coding in bilingual word perception involves steady language non-selective processing, which seems relatively unaltered by the language mode the bilingual reader is in. Alternatively, the relative prominence of the bilingual's two languages may influence the later phases of bilingual word perception. Global-level processing of second-language words, which is required for performing laboratory tasks such as lexical decisions on interlingual homographs, seems to proceed in a relatively monolingual fashion, with the non-target language exerting little influence. However, when the bilingual reader is in a situation with two languages, the non-target language moves in thereby causing measurable interference in task performance. Therefore, as opposed to Dijkstra and Van Hell (2003), our standpoint assumes a weak version of the language mode hypothesis. This is consistent with a more general perspective on word processing, in which phonological assembly is mandatory, but the use of lexical knowledge may be subject to strategic control (Drieghe & Brysbaert, 2002; see also Frost, 1998).

Within the resonance approach (Van Orden & Goldinger, 1994), it is possible that a monolingual language mode elicits (global-level) phonologic-semantic feedback of the target language, while phonologic-semantic feedback relevant to words from the non-target language is inhibited. Consequently, impending inappropriate linguistic codings with respect to the non-target language are not supported by expectations generated by semantic feedback and thus have no potential to cause obstructions in task performance. A bilingual language mode, on the other hand, produces expectations for words from either language. As a result, feedback from the semantic level supports inappropriate, non-target linguistic codings, the result of which is causing Dutch-English bilinguals to respond slowly on interlingual homographs such as BAKER and BROOD.

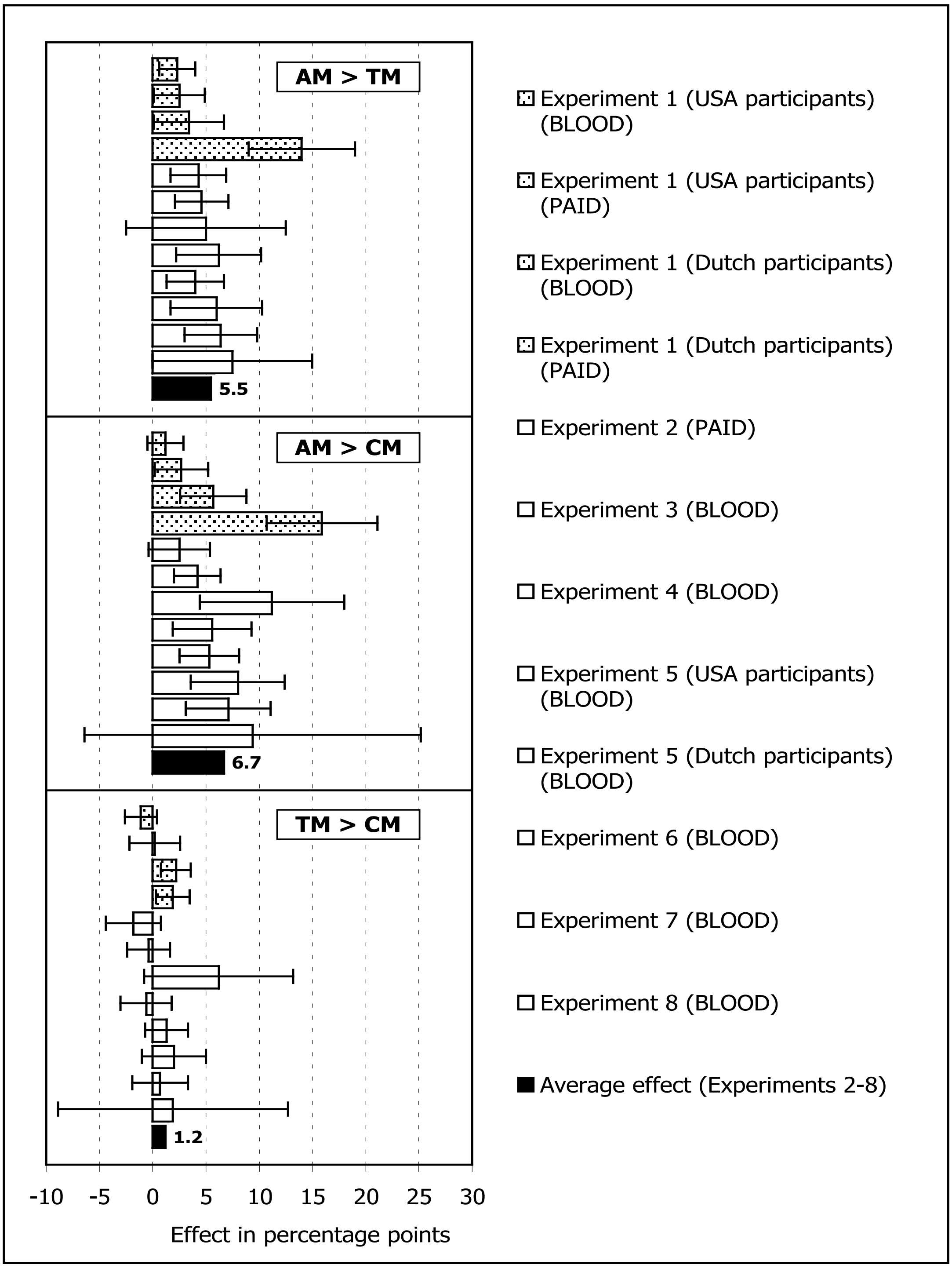

Figure 31. Overview of experimental effects for Experiments 1-8 regarding three planned contrasts that evaluated whether error rates for words like BLOOD were higher than for words like MOOD (AM > TM: upper panel) and for words like MOON (AM > CM: middle panel), and higher for words like MOOD than for words like MOON (TM > CM: lower panel). Contrasts are estimated with 95% (simultaneous) confidence intervals. In each panel of the figure, the rank order of the bars mirrors that of the elements in the legend. The extensions (PAID) and (BLOOD) indicate which wordlist is used for no-trials. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

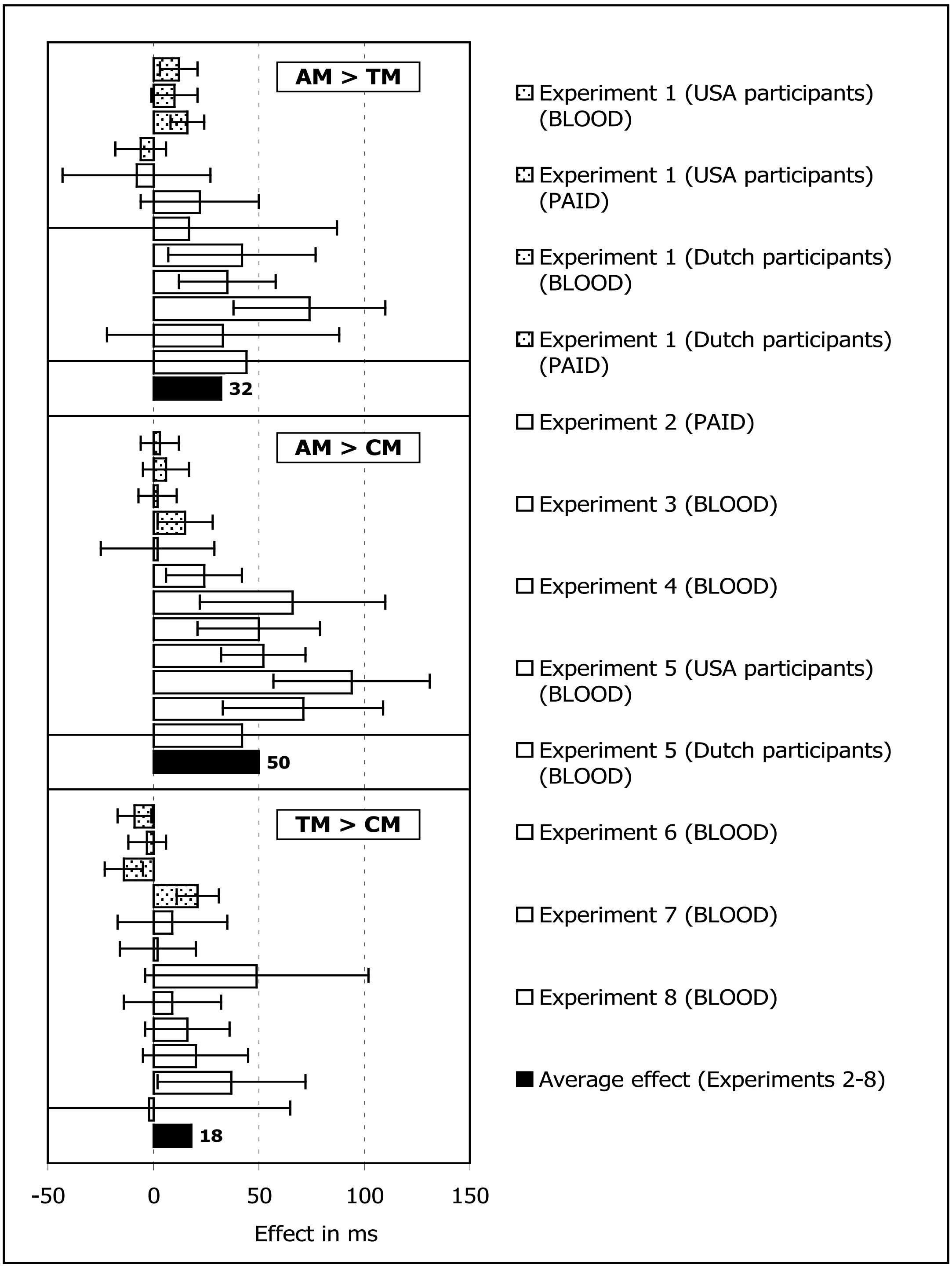

Figure 32. Overview of experimental effects for Experiments 1-8 regarding three planned contrasts that evaluated whether response latencies for words like BLOOD were longer than for words like MOOD (AM > TM: upper panel) and for words like MOON (AM > CM: middle panel), and longer for words like MOOD than for words like MOON (TM > CM: lower panel). Contrasts are estimated with 95% (simultaneous) confidence intervals. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

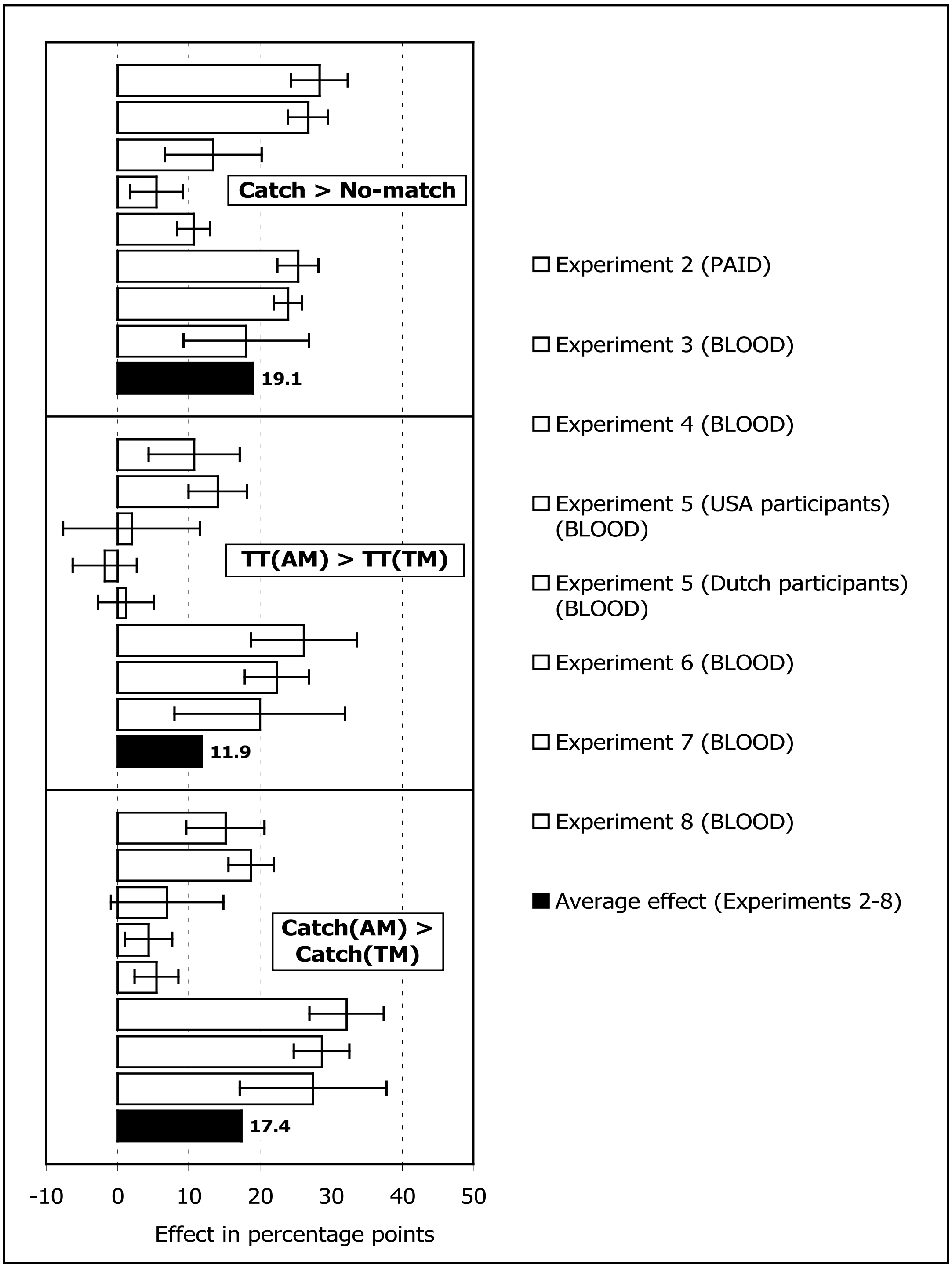

Figure 33. Overview of experimental effects for Experiments 2-8 regarding three planned contrasts that evaluated whether (1) error rates for catch trials were higher than for no-match trials (catch > no-match: upper panel), (2) this Trial Type (TT) effect was larger for AM-words than for TM-words (TT(AM) > TT(TM): middle panel), and (3) in catch trials error rates for words like BLOOD were higher than for words like MOOD (catch (AM) > catch (TM) : lower panel). Contrasts are estimated with 95% (simultaneous) confidence intervals. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

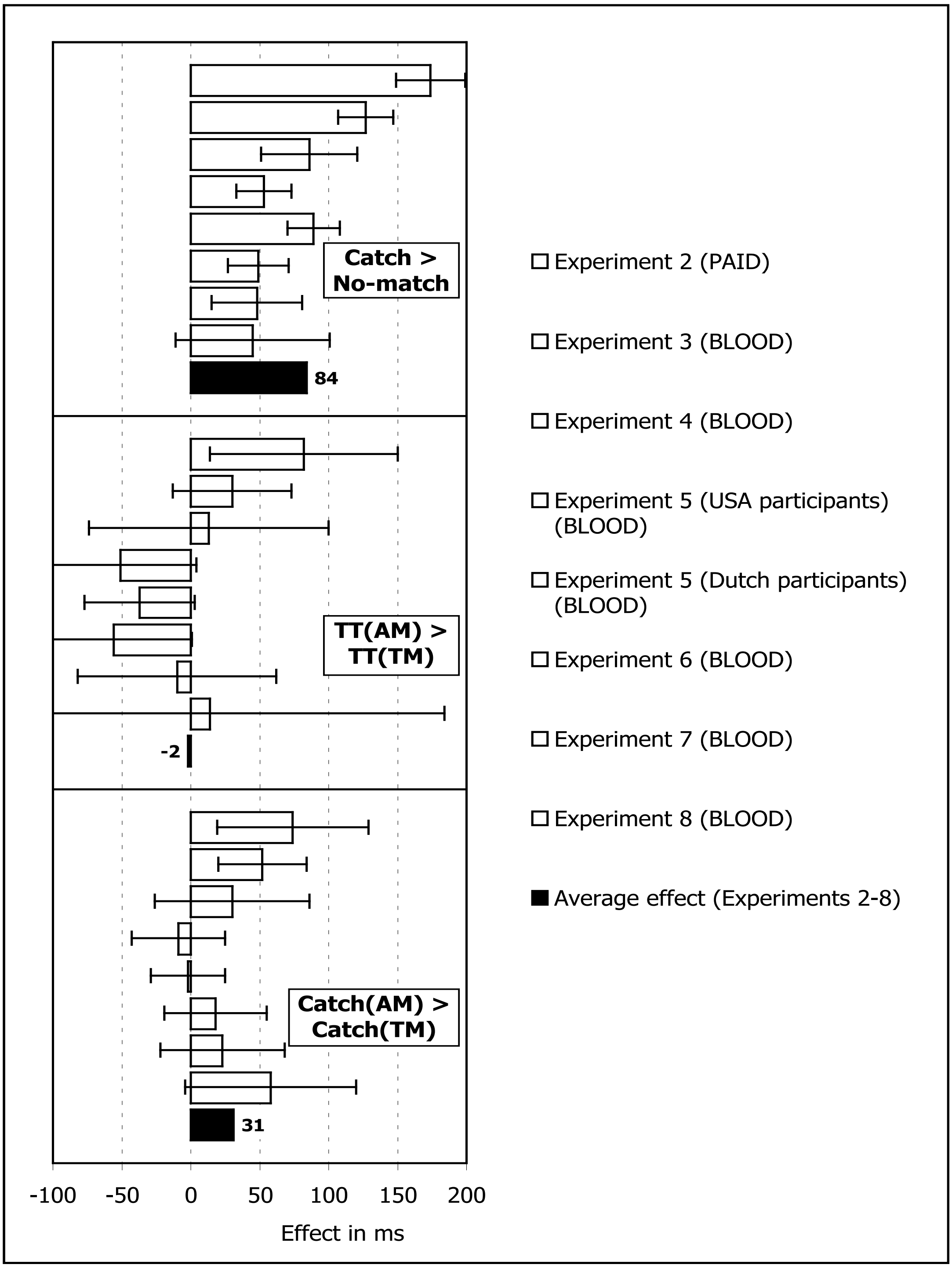

Figure 34. Overview of experimental effects for Experiments 2-8 regarding three planned contrasts that evaluated whether (1) response latencies for catch trials were longer than for no-match trials (catch > no-match: upper panel), (2) this Trial Type (TT) effect was larger for AM-words than for TM-words (TT(AM) > TT(TM): middle panel), and (3) in catch trials response latencies for words like BLOOD were longer than for words like MOOD (catch (AM) > catch (TM) : lower panel). Contrasts are estimated with 95% (simultaneous) confidence intervals. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

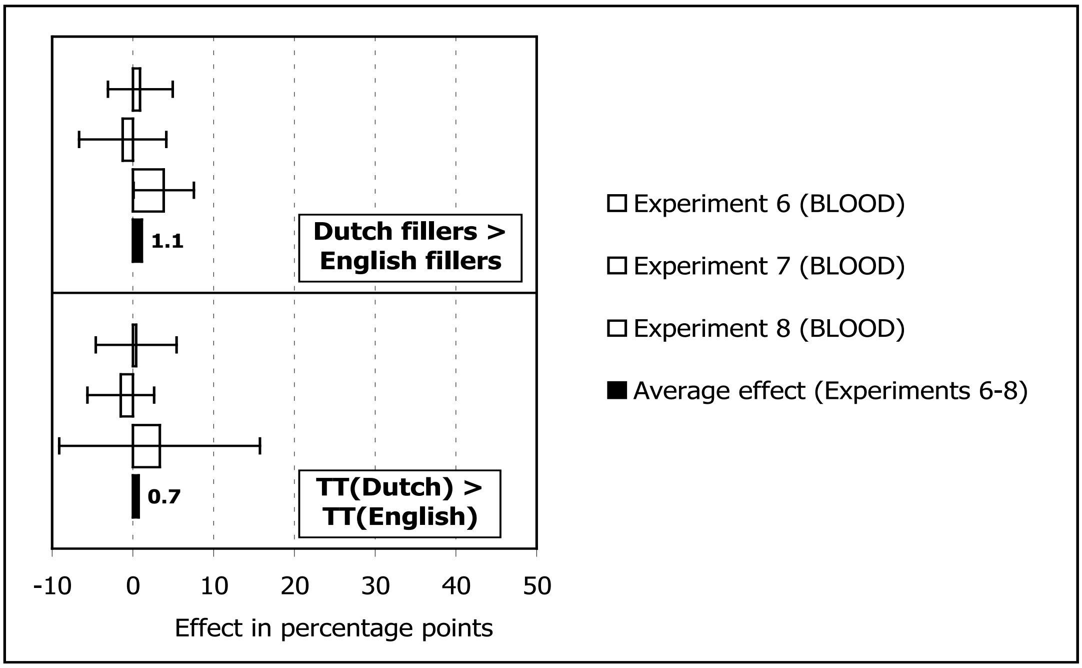

Figure 35. Overview of experimental effects for Experiments 6-8 regarding three planned contrasts that evaluated whether (1) error rates in the Dutch-fillers condition were higher than in the English-fillers condition (upper panel) and (2) the Trial Type (TT) effect was larger in the Dutch-fillers condition than in the English-fillers condition (lower panel). Contrasts are estimated with 95% (simultaneous) confidence intervals. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

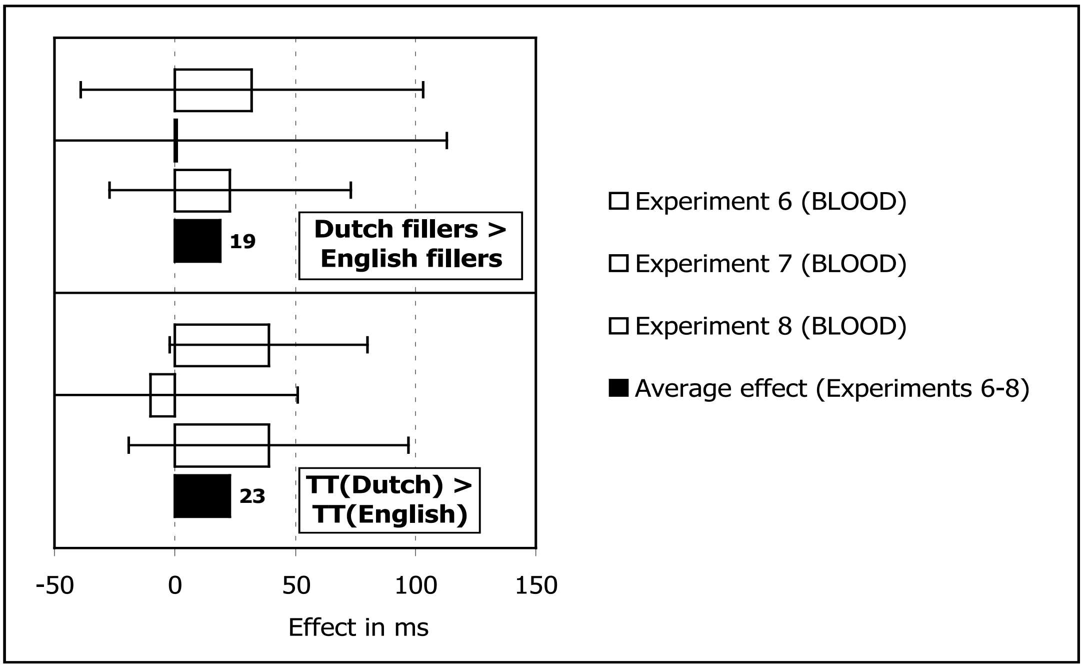

Figure 36. Overview of experimental effects for Experiments 6-8 regarding three planned contrasts that evaluated whether (1) response latencies in the Dutch-fillers condition were longer than in the English-fillers condition (upper panel) and (2) the Trial Type (TT) effect was larger in the Dutch-fillers condition than in the English-fillers condition (lower panel).

Contrasts are estimated with 95% (simultaneous) confidence intervals. (CM = consistent mappings; TM = typical mappings; AM = atypical mappings.)

Phonological Coding is a Metastable Dynamic Process

In the standard version of the print-to-speech correspondence task, spoken rimes and printed words are presented simultaneously. However, we also brought systematic variation in the temporal order of these auditory and visual stimuli. A major observation in Chapters 4 and 5 was that, for catch trials, suspending presentation of the spoken rime resulted in improved task performance (Experiments 5 and 7). Critically, however, under these conditions, performance on catch-trials was still found to be worse than on no-match trials. Thus, a spoken rime derived from an enemy word (e.g., /}d/ that is presented somewhat later than a printed word such as MOOD causes better performance than when speech and print are presented more or less concurrently. Two possible explanations were provided for this effect, which appear not to be mutually exclusive. It could be the case, for one, that an appropriate coding becomes more stable as time progresses, which protects it against interference from inappropriate phonological codings. However, it could also be the case that an inappropriate coding itself becomes increasingly unstable over time. This may cause that, in a catch trial, a spoken rime no longer succeeds in reinstating an inappropriate phonological coding. In either case, it appears that even under these conditions competition perseveres between appropriate and inappropriate phonological codings, albeit to a relatively moderate degree.

As explained in the Introduction section, this view exploits the idea of *metastability*, which asserts that the reading system never fully settles in the appropriate state but always remains slightly unstable. Metastability brings about that the system can consider alternative possibilities on a *continuous* basis, although only one alternative may be expressed in perception and action. Therefore, a trailing linguistic context may have an impact on word perception even if processed well past the moment the system reaches a state that, for instance, supports a word-identification response. This is consistent with a framework that entails a fluid metaphor of word perception, which conceives perception as a continuous flow of system states reflecting constraints emergent between printed forms and linguistic functions (Gottlob et al., 1999; Van Orden et al., 1996, 1997, 1999).

A study conducted by Van Orden et al. (1999) provides a rather interesting demonstration of the impact of a trailing context on word perception. This study used the phrase-evaluation task, which is a variant of the classic semantic categorization task. In the classic version of this task (e.g., Van Orden, 1987), which was described in the Introduction section, participants are first presented with a category name (e.g., flower) and then asked to decide whether a target word (e.g., ROWS) is an exemplar of the presented semantic category. In such an experiment, it may be observed that, for example, the homophone ROWS is mistakenly identified as “rose”, because the preceding context “flower” constraints perception of ROWS in a way that strongly favors the identity of the sound-alike exemplar “rose”. To provide a general account of these findings, Van Orden et al. (1999) argued that reading might involve a continuous interaction between environmental and cognitive constraints. For example, in case of processing a homophone such as ROWS, matching and mismatching constraints in any part of the system are combined in the ratio of constraints (i.e., the control parameter) that favors or opposes the exemplar identity. Introducing a context that favors the “rose” identity could yield a sudden change in the identity of ROWS. Van Orden et al. reasoned that if this outcome is controlled by a ratio of context and stimulus constraints, then the temporal order of presentation of context and stimulus should not be crucial. The radical prediction they put forth was that a trailing context might determine homophone identity *after* word identification should have supplied a stable “rows” identity.

This prediction was tested by means of a phrase-evaluation experiment. In this experiment, participants saw a stimulus word for 200 ms followed by a pattern mask, *before* they were presented with a comparison phrase. Their task was to judge whether the stimulus word (e.g., BREAK) and the trailing comparison phrase (e.g., “part of a car”) were closely related in meaning (which, in the current example, is obviously not the case). As in the original categorization task, the phrase-evaluation experiment included homophones (e.g., BREAK) and control words (e.g., BRAVE). Recall, a homophone such as BREAK is inherently ambiguous because its phonology is shared with another word (i.e., ”brake”). This results in a relatively weak capacity to maintain the proper identity. It was expected that the constraints due to the trailing comparison phrase (e.g., “part of a car”) would elicit the false identity of “brake”, resulting in erroneous phrase evaluations in which, for example, BREAK is mistaken for “part of a car”. This is exactly what Van Orden et al. (1999) found. In the phrase-evaluation task, participants produced substantially more phrase-evaluation errors to homophones than to controls. Crucially, it turns out that the trailing context leads to false identities of homophones such as BREAK long *after* word identification should have occurred! This finding is quite intriguing, because, as one would expect, skilled readers of English almost never make a mistake in reading a common word such as BREAK.

There appears to be a close resemblance between the trailing context effects found by Van Orden et al. (1999) and our Trial Type effects obtained in the print-to-speech correspondence task. That is, similar to the phrase-evaluation task, our Experiment 5 includes trials in which participants are presented with a stimulus word (e.g., MOOD) followed by what essentially is a trailing comparison sound (i.e., /}d/, derived from BLOOD). Spelling-to-sound inconsistent words such as MOOD contain ambiguous mappings between spelling body and several candidate phonological bodies. In a catch trial, the trailing comparison sound provides strong constraints to elicit erroneous coding of enemy phonology, which appears to occur even past the moment when correct phonological coding should have reached stability. In sum, the dramatic effect of a trailing linguistic context on perception of printed words, such as found in our print-to-speech correspondence task and in the phrase-evaluation experiment of Van Orden et al. (1999), may be a signature of metastability. By preserving some degree of instability, the reading system preserves its receptiveness to potentially alternative states, and thus gains in flexibility.

Metastability may present an essential characteristic of the reading system's processing dynamics. Due to metastability, linguistic codings are protected against complete stabilisation. This renders the reading system a high degree of flexibility, which allows it to switch easily from one state to another. What's more, the trailing context effects we discussed show that such a shift can be triggered rather swiftly. Back to the example of BREAK. With respect to this homophonic word we can assume strong intrinsic stimulus constraints (i.e., an attractor) in favor of the proper identity "break". Hence, the ratio of all possible constraints (including the opposing ones) would set the eventual balance to correct identification. However, when BREAK receives company from a trailing evaluation phrase such as "part of car", the additional constraints toward misperception place, with a sudden jump, the balance in favor of the identity "brake". Regarding the print-to-speech correspondence task, strong intrinsic stimulus constraints of an inconsistent word such as MOOD set the balance to appropriate phonology. Opposing constraints, in contrast, such as presentation of a trailing (enemy) evaluation sound, may rapidly set the balance to inappropriate phonology. In sum, rapid switching between competing states of form-function dynamics may be endorsed by multistability. Ambiguities, as in spelling-to-sound inconsistent and homophonic words, require a processing architecture that supports multistable states and fast transitions among them. Metastability offers an avenue for these fast transitions, by preserving some degree of instability for dominant linguistic codings.

As a final point, Van Orden et al. (1999) comment that an observed shift in identity of homophonic words like BREAK "… seems to indicate two qualitatively different processes, as though a shift had occurred from a data-driven (bottom-up) process to a conceptually driven (top-down) process." (Van Orden et al., 1999, page 53). However, as they further point out, from a dynamic systems perspective, processing a word entails a continuous interaction of matching and mismatching contextual and stimulus constraints. In a system that exhibits discontinuous behavior, the outcome of this interaction may eventually be reconfigured inordinately by continuous changes in a single control parameter. This occurs at the critical values of the control parameter (i.e., the bifurcation point), and gives rise to the occurrence of sudden jumps in the identity of homophonic words like BREAK. Therefore, as Van Orden et al. (1999) argue, the apparent shift from a bottom-up to a top-down process reflecting homophone errors may just as reasonably be interpreted as a bifurcation phenomenon. Conceivably, this reasoning may also apply to fast transitions between phonological states. Since it may be true that ambiguous spelling-to-sound mappings in inconsistent words give rise to multistable perception, it can be argued that spelling-to-sound inconsistent words are positioned near a bifurcation point. However, the experimental manipulations used in our study and that of Van Orden et al. (1999) did not use *continuous* control variables, which, in catastrophe theory, is a prerequisite for demonstrating the occurrences of bifurcations. Therefore, in order to learn more about the complex dynamics of language processing, further research should clarify the precise nature of fast transitions of linguistic states observed for ambiguous linguistic materials such as homographs, homophones, and multiple-grain size phonological clusters. For experimental workers, this might entail devising contextual variables that allow for continuous (i.e., gradual) experimental manipulation. On the face of it, this may appear a rather complicated task, since word characteristics are difficult to place on a continuous scale. Alternatively, there are several other, quite obvious ways to accomplish continuous variation of the independent variable. For example, the relative time course of linguistic stimuli can be changed gradually, as well as several physical aspects of the stimuli (i.e., degrading, adding noise).

PHONOLOGICAL CODING IN PRINTED WORD PERCEPTION

In a recent paper, Dijkstra et al. (1999) called attention to the problem of the neglected role of phonology in bilingual reading research. Along with the work of Brysbaert et al. (1999), Jared and Kroll (2001), and Van Wijnendaele and Brysbaert (2002), our study can be considered as an attempt to compensate for this lack of interest. In addition, the present investigation contributes to a larger body of knowledge on the cognitive dynamics of word perception. Our finding that manifold spelling-to-sound relations have a similar impact on first-language and second-language word reading argues for a general standpoint in which phonological assembly is viewed as fundamental to both monolingual and bilingual word perception. We also supplied further evidence that phonological coding in second-language word reading is affected not only by within-language but also cross-language spelling-to-sound knowledge. More than ever, such a finding underlines the universal and mandatory nature of phonological coding.

Phonology is Fundamental to Reading

The view that phonological coding essentially is a mandatory process comes with a strong theoretical foundation and is substantiated by a bulk of converging empirical evidence. The strong phonological theory proposed by Frost (1998) asserts that phonological processing is the *default* operation of the reading system (cf. Van Orden, 1987; Van Orden et al., 1990). The basic claim of this general reading theory is that "… all writing systems are phonological in nature and their primary aim is to convey phonological structures, that is, words, regardless of the graphemic structure adopted by each system." (Frost, 1998, page 89). In other words, the phonological root of all writing systems implies that reading is based on the mechanisms by which speech is perceived and produced (Lukatela et al., 2001; Liberman, 1995; Mattingly, 1985). Phonological coding in visual word perception, therefore, is a primary function of the reading system, and it is inevitably launched following the visual presentation of a printed word. In addition, as explicated in the phonological coherence hypothesis (Van Orden & Goldinger, 1994), the leading role of phonology follows directly from the high correlation between orthographic and phonologic structure, which is more systematic than that between orthographic and semantic structure, or between phonologic and semantic structure. Consequently, orthographic-phonologic activation dynamics cohere earliest, providing immediate further constraints on word perception.

Phonological Structure is Bottom-up Assembled and Top-down Shaped

Integral to the strong phonological theory is that phonological coding is a mandatory process that assembles phonological structure via knowledge of multiple-grain size spelling-to-sound correspondences. Furthermore, the strong phonological theory stipulates the computational principles of how exactly phonological structure is derived from print. In a strong phonological model, phonology is always assembled bottom-up through a single basic mechanism (e.g., a parallel distributed network of elementary processing units). However, this process of phonological coding does not necessarily result in a fully detailed phonological description. For example, in so-called 'deep orthographies', such as English or Hebrew, the profusion of ambiguous and idiosyncratic spelling-to-sound mappings often bring difficulties in the assembly process (i.e., the orthographic depth hypothesis, e.g., Frost & Katz, 1992; Frost, Katz, & Bentin, 1987), which cause impoverished and underspecified phonological structures (e.g., Berent & Perfetti, 1995; Frost, 1995; Gronau & Frost, 1997). However, irregularities of spelling-to-sound also occur in more shallow orthographies, and even devoid of stringent ambiguity, phonological assembly may still generate imprecise, noisy codings. Therefore, in order to yield a sufficiently detailed phonological structure, the bottom-up computation may require further shaping by top-down lexical knowledge (cf. Van Orden, 1987). This process employs the reader's word knowledge and eventually takes care of ambiguities or of gaps in what has been formed by the assembly process. In other words, lexical involvement does not contravene assembled phonology (Carello, Lukatela, & Turvey, 1994; see also Carello et al., 1992, for a thorough discussion).

Thus, in a strong phonological model, the process of phonological coding involves an interaction of bottom-up assembly and top-down shaping, with the relative contributions depending on various stimulus constraints and task demands (Frost, 1998; see also Carello et al., 1992). Such a model is very similar in nature to the frameworks proposed by Lukatela and colleagues (Lukatela et al., 1989; Lukatela, Lukatela, Carello, & Turvey, 1999) and Van Orden and Goldinger (1994), which lean on principles of adaptive resonance theory (e.g., Grossberg, 1995). In short, these resonance frameworks for visual word perception describe phonological assembly as a primary process, evolving in a bidirectional connective matrix that links orthographic and phonologic processing units. A clean-up process is assumed, which acts to resolve the phonological noise that may arise during initial word processing (e.g., Van Orden, 1987). This process, resonant equilibration, evolves as stable feedback loops stemming from a bidirectional connective matrix linking phonologic and semantic processing units (or lexical units, e.g., Frost, 1998; Lukatela et al., 1999).

Principles of Phonological Coding

Consider, just once more, the English word BLOOD. When its printed form meets the eye, the reading system is triggered to create immediately and irreversibly all phonological structures that have previously been associated with the spelling body *–OOD*. For a skilled reader of English, the resultant phonological ambiguity appears barely disruptive, since word reading usually proceeds fast and accurately. Thus, when reading an inconsistent word, enemy phonology is quickly spotted and stifled. However, it is not completely wiped out. On the contrary, it is preserved, and at the same time suppressed. This explains why it is possible that, even when appropriate phonology has overcome all opposition, it may all of a sudden be overturned by latent enemy phonology.

At this point it may be worth reviewing a number of candidate principles that constrain the process of phonological coding. In what follows, we give a brief overview of six principles that derive from the theoretical frameworks and empirical evidence discussed thus far. These principles could serve as future evaluation points to assess the way phonology is implemented in current models of monolingual and bilingual visual word perception.

(1) *Phonological coding is mandatory*. The obligatory nature of phonological coding appears a primary characteristic of the reading system. It delineates a processing architecture that, following input of a printed word, necessarily and immediately launches a process of phonological coding.
(2) *Phonological coding is an early process*. In general, highly self-consistent (correlated) mappings speed up form-function integration, and since self-

consistency for spelling-to-sound mappings is higher than for other mappings, phonological codings stabilize early relative to other linguistic codings.

(3) *Phonological coding is initially ballistic*. The automatic launch of phonological assembly seems to cause inappropriate phonological coding in the initial conditions of word perception. Given the premises that, one, reading experience with inconsistent words gives rise to manifold associations, and, two, phonological coding is launched as a rule, processing an inconsistent word should result in the simultaneous coding of all phonological structures that have previously been associated with its spelling body. Furthermore, once this process is triggered, there is no return. In other words, phonological coding acts as a ballistic process, creating phonological structure on a purely automatic basis without opportunity for strategic control, or concern for goals that emerge in later processing.

(4) *Phonological coding is multistable*. There is an obvious hitch to the automatic launch of phonological assembly. Coding of inappropriate phonology causes phonological ambiguity, which must be resolved for coherent form-function integration. In the reading system, ambiguous phonology leads to multistability, which acts to preserve the array of multiple percepts. However, in word processing, ambiguities are quickly resolved. That is to say, the hampering competition that goes on between appropriate and inappropriate codings persists only momentarily, and may be of no consequence for proper reading performance. Alternatively, the fact that coding of inappropriate phonology is tolerated can be seen as a indication of flexible processing.

(5) *Instability of phonological codes is preserved*. Even though ambiguity is resolved rather quickly, multistability prevents that inappropriate phonological codings disappear altogether. Conversely, appropriate phonological structure is restrained from fully settling down. Such qualities, entertaining alternative possibilities and preventing dominant codings to reach complete stabilization, is apparent for a system that incorporates metastability. A reading system that is always slightly unstable defines a processing architecture that, in response to environmental constraints, permits rapid transitions between system states. Such a quality may be of general importance, for in written language processing, phonology and meaning are dynamically constructed and may require instant modification as discourse constraints change over time.

(6) *Phonological coding allows for lexical involvement*. In printed word perception, phonology is principally derived from a bottom-up process of phonological assembly. However, the product of phonological coding may be noisy, and in a highly interactive processing system comprising of orthographic, phonologic and semantic codes, top-down feedback from semantics may act as a clean-up mechanism.

The Relation of Reading to Speech

In conclusion, the strong phonological theory of Frost (1998) offers a coherent account of the primary role of phonology in visual word perception. At the core of this account rests the assumption that phonological assembly is a mandatory process. The present findings extend the available empirical evidence on phonological coding in monolingual and bilingual word reading. Ultimately, our findings yield unequivocal evidence of the mandatory nature of phonological coding by demonstrating that bilingual word processing may actually initiate and preserve enemy phonology arising from knowledge of cross-language spelling-to-sound knowledge.

There is, however, a more fundamental argument why phonological coding is so important in reading. As has already been brought up, this concerns the intimate relation of reading to speech (e.g., Liberman, 1995; Mattingly, 1985). Human communication is first and foremost dependent on the spoken word, and one major cognitive activity associated with it is phonological processing. For example, it is widely acknowledged that in speech comprehension, phonological structures are the basic constituents on which linguistic and other cognitive processes operate (see Carello et al., 1992). Obviously, by the time reading acquisition sets off, humans normally have at their disposal a fully developed language system. Starting out from this simple fact, Carello et al. (1992) argue that the processes connected to those by which speech is produced and perceived should place the major constraint on word perception (see also Gough, 1972). The authors bolster their point with the following instructive excerpt from Liberman (1991): “The seemingly sensible strategy for the reader is to use the optical shapes to access phonological structures early in the reading process. Once the reader has done that, he has put the hard part of reading behind him, for everything else will be done automatically by language processes that he commands by virtue of his humanity.” (Liberman, 1991, pp. 242-243).

Implications for Computational Models of Visual Word Perception

Reflecting upon two decades of reading research, Carello et al. (1992) asked with bewilderment why some reading theorists remained “… stubbornly nonphonological, both in denying the plausibility of taking advantage of extant phonological processes and overlooking the phonological basis of writing systems.” (p. 212). Indeed, in our estimation, this stubbornness should be deemed rather unfortunate. It is not unlikely that it has caused many research programmes to drift away from more worthwhile courses. The obsessive preoccupation with non-linguistic reading processes, for example, may have withheld researchers in asking the questions that are nowadays recognized as addressing central issues in reading research. In any case, the former practise of discarding the hypothesis that phonology plays a central role in printed word perception may have been a costly mistake. It is beyond question that the advance in our understanding of the cause, development, and treatment of reading disorders, both in theoretical and practical sense, has suffered greatly because of this ignorance.

Fortunately, these days there appears to be much more consensus about the role of phonology. In many recent publications, the question is no longer asked *whether* phonology is involved in word perception, but when and how. It looks as if more and more researchers are acknowledging the importance of phonology in visual word perception. This is noticeably reflected in the architecture and processing dynamics of contemporary computational models of visual word perception. In the past 15 years, several models have been developed that successfully simulate the process of phonological coding in word reading.

A well-known example of a computational model that incorporates spelling-to-sound conversion is the parallel distributed developmental model of Seidenberg and McClelland (1989). In this model (and in its successor, see Plaut, McClelland, Seidenberg, & Patterson, 1996), orthographic and phonological information is represented in terms of distributed patterns of activity over separate groups of elementary processing units connected in a network. Phonological coding is accomplished through cooperative and competitive interactions among these processing units via a matrix of weighted connections. The Seidenberg and McClelland (1989) model has shown to present an important milestone for theory development in reading research. In the earlier years, the empirical literature on visual word perception was dominated by the dual route theory of Coltheart (1978). This theory of word naming postulates that two independent procedures are required for converting print to speech. In the first procedure, word pronunciation is directly addressed in a mental lexicon and in the second it is assembled through the application of grapheme-to-phoneme correspondence rules. The two routes towards pronunciation are specially tailored to handle different kinds of words. One route (i.e., the lexical or addressed route) is appropriate for pronouncing both high-frequency regular words (e.g., MINT) and irregular words (e.g., PINT), whereas the other (i.e., the nonlexical or assembled route) is suitable only for pronouncing low-frequency regular words or pseudowords. Thus, the nonlexical or assembled route is incapable of handling irregular words and the lexical or addressed route is incapable of handling pseudowords. However, the Seidenberg and McClelland (1989) model demonstrated that a *single* system, a parallel distributed connectionist network, can be successfully trained to assemble phonology *both* for (low-frequency) regular words (e.g., MINT) and for irregular ones (e.g., PINT).

This brings us to a second computational model, the dual route cascaded (DRC) model proposed by Coltheart and his colleagues (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, & Ziegler, 2001), which essentially implements the classic dual route theory (Coltheart, 1978). In this recent model, which operates in the localist framework, a phonological representation is formed either indirectly through a relatively slow rule-based system of grapheme-to-phoneme correspondence rules (i.e., assembled phonology via the nonlexical route) or directly via a relatively fast lexical look-up (i.e., addressed phonology via the lexical route).

Although the DRC model assigns a central function to phonological computation, the way it is currently parameterized is at odds with the fast and mandatory nature of phonological coding. In order to accommodate this key aspect of phonological coding, the DRC model would require a small modification. Lukatela, Eaton, Lee, Carello, and Turvey (2002) propose that such a modification would not need calling into question the basic architecture of the model. A simple change in the DRC model's parameter settings would be sufficient. This would involve imposing a delay on the start of processing on the *lexical route* instead of on the nonlexical route, and assigning higher weights to activation on the nonlexical route. Such a modification would have the effect that assembled phonology via the nonlexical route *precedes* addressed phonology via the lexical route. However, this simple change in the parameter settings of the model would actually present a rather drastic break with the classic dual route theory, because it effectively rejects the delayed phonology hypothesis (see also Lukatela & Turvey, 1994a, 1994b; Van Orden et al., 1990, for discussions).

A number of other interesting computational models of word processing have been proposed, with the purpose to account for performance in different experimental tasks (e.g., lexical decision). For example, the multiple read-out model (MROM) put forward by Grainger and Jacobs (1996) provides remarkable insight into the dynamics of orthographic processing. The model is based on the assumption that word perception involves the computation of several types of codes (i.e., orthographic, phonological, and semantic codes) in parallel. Because of modelling considerations, MROM does not incorporate phonological processes, although the authors acknowledge that phonological codes can affect performance in perceptual identification and lexical decision tasks. However, it appears that MROM does not reserve a central role for phonology. If in the future MROM will be extended to include computation of phonological structure, and if the model is intended to accommodate the major empirical findings on phonological assembly in word perception, it should be designed to reflect the mandatory nature of phonological coding.

The same consideration holds for a new model for bilingual word processing, the BIA+ model (Dijkstra & Van Heuven, 2002) that is currently implemented in the localist connectionist model SOPHIA (Van Heuven & Dijkstra, in preparation). One notable difference with the original model is that the BIA+ model now includes phonological representations, with the purpose to address the claim that phonology plays an important role in printed word perception (Dijkstra et al., 1999). However, the model's processing dynamics in the first stages of word recognition basically involve activation of lexical orthographic candidates, which, in turn, activate their corresponding phonological and semantic representations. Consequently, during word reading, phonological representations are activated *later* than orthographic representations, thus essentially complying with a key aspect of dual route theory, delayed phonology. In other words, representation and processing in the BIA+ model do not conform to the phonological coherence hypothesis (Van Orden & Goldinger, 1994), which states that phonological coding is an *early* and *primary* source of constraint on word perception.

Final Word

We bring a close to this chapter by noting that our conclusions regarding the mandatory nature of phonological coding in word reading have more general implications for research on cognitive processes. For example, a large part of studies investigating information processing or the workings of human memory involve laboratory experiments that use printed words as stimuli. The basic action of participants is to read these words and respond to them. Thus, in a sense, many of these experiments are really word-reading experiments. In only a few occasions, however, researchers are mindful that their laboratory task may also involve phonological coding and that their experimental findings are potentially influenced, or worse, contaminated by it. In the spirit of Carello et al. (1992) we therefore ask: Can laboratory experiments remain stubbornly nonphonological? Our point is that researchers who use some sort of word-reading task in their experiments (e.g., lexical decision, perceptual identification) should be strongly aware that processing of word stimuli may include fast and mandatory phonological coding, which can have a significant impact on task performance. To prevent systematic bias or to isolate this source of variance, researchers should carefully balance their word stimuli on relevant phonological dimensions.

References

Abraham, F. D., Abraham, R. H., & Shaw, C. D. (1991). *A visual introduction to dynamical systems theory for psychology*. Santa Cruz, CA: Aerial Press.

Altenberg, E. P., & Cairns, H. S. (1983). The effects of phonotactic constraints on lexical processing in bilingual and monolingual subjects. *Journal of Verbal Learning and Verbal Behavior*, *22*, 174-188.

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (Eds.) (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: BMJ Books.

American Psychological Association (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, *10*, 565-575.

Anderson, J. A. (1990). *Cognitive psychology and its implications*. W. A. Freeman and Company: New York.

Baayen, H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66*, 423-437.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 386-393.

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language selective access. *Journal of Memory and Language*, *26*, 658-672.

Berent, I., & Perfetti, C.A. (1995). A ROSE is a REEZ - the 2-cycles model of phonology assembly in reading English. *Psychological Review*, *102*, 146-184.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, *25*, 447-457.

Borgwaldt, S. R. (2003). *From onset to entropy*. Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Bosman, A. M. T. (1994). *Reading and spelling in children and adults: evidence for a single route model*. Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Bosman, A. M. T., & de Groot, A. M. B. (1995). Evidence for assembled phonology in beginning and fluent readers as assessed with the first-letter-naming task. *Journal of Experimental Child Psychology*, *59,* 234-259.

Bosman, A. M. T., & de Groot, A. M. B. (1996). Phonologic mediation is fundamental to reading: evidence from beginning readers. *The Quarterly Journal of Experimental Psychology*, *49A*, 715-744.

Bosman, A.M.T., van Leerdam, M., & de Gelder, B. (2000). The /O/ in OVER is different from the /O/ in OTTER: Phonological effects in Dutch children with and without dyslexia. *Developmental Psychology, 36*, 817-825.

Bosman, A. M. T., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben & M. Fayol, (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 173-194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bradley, L., & Bryant, P. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature*, *271*, 746-747.

Brown, G. D. A., & Watson, F. L. (1994). Spelling-to-sound effects in single-word reading. *British Journal of Psychology, 85*, 181-202.

Brysbaert, M. (in press). Bilingual visual word recognition: Evidence from masked phonological priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: State-of-the-art*. Hove, UK: Psychology Press.

Brysbaert, M. (2001). Prelexical phonological coding of visual words in Dutch: Automatic after all. *Memory & Cognition, 29*, 765-773.

Brysbaert, M. (1998). Word recognition in bilinguals: Evidence against the existence of two separate lexicons. *Psychologica Belgica*, *38*, 163-175.

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 137-148.

Burnage, G. (1990). CELEX: A guide for users. Nijmegen, The Netherlands: SSN.

Carello, C, Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost & L. Kratz (Eds.), *Orthography, phonology, morphology, and meaning*. (pp. 211-226) Amsterdam: North-Holland.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*, 378-399.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11,* 63-65.

Clark, H. H. (1973). The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.

Coltheart, M. (1981) The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505.

Coltheart, M., Davelaar, E., Jonasson, J.R., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention an performance XI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589-608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204-256.

Coltheart, V., Patterson, K., & Leahy, J. (1994). When a ROWS is a ROSE: Phonological effects in written word comprehension. *Quarterly Journal of Experimental Psychology, 47A,* 915-955.

Cotton, J. W. (1989). Interpreting data from two-period crossover design (also termed the replicated 2×2 Latin square design). *Psychological Bulletin, 106*, 503-515.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 532-575.

Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 618-632.

De Groot, A. M. B., & Kroll, J. F. (Eds.). (1997) *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, N.J.: Lawrence Erlbaum Associates.

De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for nonselective access to bilingual memory. *The Quarterly Journal of Experimental Psychology, 53A,* 397-428.

Dennis, I., & Newstead, S. E. (1981). Is phonological recoding under strategic control? *Memory & Cognition, 9,* 472-477.

Dijkstra, A., Grainger, J., & Van Heuven, W. J. B. (1999). Recognizing cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496-518.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42,* 445-464.

Dijkstra, T., & Van Hell, J.G. (2003). Testing the language mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism, 6*, 2-16.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism, Language and Cognition*, *5*, 175-197.

Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA-model and bilingual word recognition. In J. Grainger and A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.

Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*, 177-196.

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism, Language and Cognition*, *1*, 51-66.

Dearborn, W. F. (1906). *The psychology of reading: An experimental study of the reading pauses and movements of the eye*. Doctoral dissertation. Columbia University, Faculty of Philosophy, United States of America.

Dodge, R. (1900) Visual perception during eye movement. *Psychological Review, 7,* 454-465.

Drieghe, & Brysbaert, (2002). Strategic effects in associative priming with words, homophones, and pseudohomophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 951-961.

Estes, W. K. (1997). On the communication of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin and Review*, *4*, 330-341.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 288-298.

Frost, R. (1995). Phonological computation and missing vowels: Mapping lexical involvement in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 398-408.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*, 71-99.

Frost, R., & Katz, L. (Eds.). (1992). *Orthography, phonology, morphology, and meaning*. Amsterdam: North-Holland.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104-115.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than *p* values: Estimation rather than hypothesis testing. *British Medical Journal, 292*, 746-750.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 305-315.

Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods and Instruments, 12*, 395-427.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674-691.

Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say “blay!” Attentional dynamics of delayed naming. *Journal of Memory and Language, 37,* 190-216.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1122-1139.

Gottlob, L. R., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 561-574.

Gough, P. B. (1972). One second of reading. In J. K. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 331-358). Cambridge, MA: MIT Press.

Gough, P. B., & Cosky, M. J. (1977). One second of reading again. In N. J. Castellan, D. B. Pisoni, & G. R. Potts, *Cognitive theory 2* (pp. 271-288) Hillsdale, NJ: Erlbaum.

Grainger, J. (1993). Visual word recognition in bilinguals. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 11-25). Amsterdam/Philadelphia: John Benjamins.

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology, 39A,* 295-319.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*, 623-647.

Gronau, N., & Frost, R. (1997). Prelexical phonologic computation in a deep orthography: Evidence from backward masking in Hebrew. *Psychonomic Bulletin and Review*, *4*, 107-112.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225-254). Mahwah, NJ: Erlbaum.

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1-22). Oxford, England: Basil Blackwell.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*, 1-51.

Grossberg, S. (1995). The attentive brain. *American Scientist, 83,* 438-449.

Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93*, 46-74.

Harris, R. (1992). *Cognitive processing in bilinguals*. New York: North Holland.

Hays, W. L. (1994). *Statistics*. (4th ed.). New York: Holt, Rinehart & Winston.

Hock, H. S., Schöner, G., & Giese, M. (2003). The dynamical foundations of motion pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception & Psychophysics, 65*, 429-457.

Hooper, D. A., & Paap, K. P. (1997). The use of assembled phonology during performance of a letter search task and its dependence on the presence and proportion of word stimuli. *Journal of Memory and Language, 37,* 167-187.

Huey, E. B. (1968). *The psychology and pedagogy of reading*. Cambridge, Mass: The M.I.T. Press (Original work published in 1908).

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. The *Behavioral and Brain Sciences*, *8*, 689-740.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8*, 3-7.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition - sampling the state of the art. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1311-1334.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36,* 505-529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, *46*, 723-750.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language, 44,* 2-31.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*, 687-715.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General*, *120*, 358-394.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: a parallel distributed processing account. *Journal of Memory and Language*, *32*, 474-516.

Kawamoto, A. H., Farrar, W. T., & Kello, C. T. (1994). When two meanings are better than one: Modelling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1233-1247.

Kawamoto, A. H., & Zemblidge, J. (1992). Pronunciation of homographs. *Journal of Memory & Language*, *31*, 349-374.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or the body in question: Further dissection of the process of sub-lexical spelling-sound translation. In M. Coltheart (Ed.). *Attention and performance XII: The psychology of reading* (pp. 449-470). London: Erlbaum.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 44,* 592-617.

Kirk, R. (1995). *Experimental design: Procedures fore the behavioral sciences* (3rd.). Pacific Grove, CA: Brooks/Cole.

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 61*, 213-218.

Kroll, J. F., & De Groot, A. M. B. (Eds.). (in press). *Handbook of bilingualism: Psycholinguistic approaches.* Cambridge: Oxford University Press.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Lam, A. S. L., Perfetti, C. A., & Bell, L. (1991). Automatic phonetic transfer in bidialectal reading. *Applied Psycholinguistics, 12,* 299-311.

Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 285-294.

Liberman, A. M. (1995). The relation of speech to reading and writing. In B. de Gelder & J. Morais (Eds.), *Speech and Reading* (pp. 17-32). Erlbaum UK Taylor and Francis, Hove, UK.

Liberman, A. M. (1991). Observations from the sidelines. *Reading and Writing: An Interdisciplinary Journal, 3*, 429-433.

Loftus, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology, 36*, 102-105.

Loftus, G. R. (1993). Editorial comment. *Memory & Cognition, 21*, 1-3.

Loftus, G. R. (1995). Data analysis as insight. *Behavior Research Methods, Instruments, and Computers, 27*, 57-59.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5,* 161-171.

Loftus, G. R. (2002). Analysis, interpretation, and visual presentation of experimental data. In H. Pashler (Ed.), *Stevens' handbook of experimental psychology* (Vol. 4, pp. 339-390). New York: John Wiley and Sons.

Loftus, G. R., Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin and Review, 1*, 476-490.

Lukatela, G. Feldman, I. B., Turvey, M. T., Carello, C., & Katz, L. (1989). Context effects in bi-alphabetical word perception. *Journal of Memory & Language*, *28*, 237-254.

Lukatela, G., Turvey, M. T., Feldman, I. B., Carello C., & Katz, L. (1989). Alphabet priming in bi-alphabetical word perception. *Journal of Memory & Language*, *28*, 214-236.

Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General*, *123*, 107-128.

Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General*, *123*, 331-335.

Macnamara, J., & Kushnir, S. (1971). Linguistic independence of bilinguals: The Input Switch. *Journal of Verbal Learning and Verbal Behavior*, *10*, 480-487.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203-220.

Mattingly, I. G. (1992). Linguistic awareness and orthographic form. In R. Frost and L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 11-26). Amsterdam: North-Holland.

Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Pacific Grove, CA: Brooks/Cole.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

McCuster, L. X., Hillinger, M. L., & Bias, R. C. (1981). Phonological recoding and reading. *Psychological Bulletin, 88,* 217-245.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *The Quarterly Journal of Experimental Psychology*, *43A*, 907-912.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452-467.

Myers, J. L. (1979). *Fundamentals of experimental design* (3rd ed.). Boston: Allyn & Bacon.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior*, *22*, 526-534.

Nickerson, R. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241-301.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement, 76* (3, part 2).

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & Bell, L. C. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and masked priming. *Journal of Memory and Language, 30,* 473-485.

Perfetti, C. A., Bell, L. C., & Delaney, S. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence form backward masking. *Journal of Memory and Language, 27,* 59-70.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in visual Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 101-118.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost and L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 227-248). Amsterdam: North-Holland.

Peter, M., & Turvey, M. T. (1994). Phonological codes are early sources of constraint in visual semantic categorization. *Perception & Psychophysics, xx*, xx-xx.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56-116

Ploeger, A., Van der Maas, H. L. J., & Hartelman, P. A. I. (2002). Stochastic catastrophe analysis of switches in the perception of apparent motion. *Psychonomic Bulletin and Review, 9*, 26-42.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785-794.

Port, R. F., & Van Gelder, T. (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, Mass: The M.I.T. Press

Raaijmakers, J. G. W. (2003). A further look at the "Language-as-Fixed-Effect Fallacy". *Canadian Journal of Experimental Psychology, 57*, 141-151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The Language-as-Fixed-Effect Fallacy": Common misconceptions and alternative solutions. *Journal of Memory and language, 41*, 416-426.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 29-53.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, N. J.: Prentice-Hall.

Reese, H. W. (1997). Counterbalancing and other uses of repeated-measures Latin-square designs: Analyses and interpretations. *Journal of Experimental Child Psychology, 64*, 137-158.

Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology, 57*, 221-237.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, *10*, 645-657.

Schmidt, F. L. (1996) Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1*, 115-129

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and language, 25*, 605-618.

Seaman, M. A., & Serlin, R. C. (1998). Equivalence confidence intervals for two-group comparisons of means. *Psychological Methods, 3*, 403-411.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383-404.

Sim, J., & Reid, N. (1999). Statistical inference by confidence intervals: Issues of interpretation and utilization. *Physical Therapy, 79,* 186-195.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition, 6*, 410-415.

Stone, G. O. (1994). Combining connectionist and symbolic properties in a single process. In S. D. Lima, R. L. Corrigan, and G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 417-444). Amsterdam/Philidelphia: John Benjamins Publishing Company.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 744-774.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, xx*, 25-32.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms.* New York: Erlbaum.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124,* 107-136.

Tryon, W.W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods, 6*, 371-386.

Tzelgov, J., Henik, A., Sneg, R., & Baruch, O. (1996). Unintentional reading via the phonological route: The Stroop effect with cross-script homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 336-339.

Ulrich, R., & Miller, J. (1994) Effects on truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123*, 34-80.

Van Hell, J.G. (1998). *Cross-language processing and bilingual memory organization.* Unpublished doctoral dissertation, University of Amsterdam, The Netherlands.

Van Hell, J.G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review, 9*, 780-789.

Van Heuven, W. J. B. (2000). *Visual word recognition in monolingual and bilingual readers: Experiments and computational modelling*. Unpublished doctoral dissertation, University of Nijmegen, The Netherlands.

Van Heuven, W. J. B., & Dijkstra, T. (in preparation). The Semantic, Orthographic, and Phonological Interactive Activation Model [working title].

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458-483.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15,* 181-198.

Van Orden, G. C., Atchison, C. S., & Podgornik, M. N. (1996). *When a ROWS is not a ROSE: Null effects and the absence of cognitive structures*. Unpublished manuscript.

Van Orden, G. C., Stone, G. O., Garlington, K. L., Markson, L. R., Pinnt, G. S., Simonfy, C. M., Brichetto, T. (1992). "Assembled phonology and reading: A case study in how theoretical perspective shapes empirical investigation. In R. Frost and L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 249-292). Amsterdam: North-Holland.

Van Orden, G. C., Jansen op de Haar, M. A., & Bosman, A. M. T. (1997). Complex dynamic systems also predict dissociations but they do not reduce to autonomous components. *Cognitive Neuropsychology, 14,* 131-165.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1269-1291.

Van Orden, G. C., Holden, J. G., Podgornik, M. N., & Atchison, C. S. (1999). What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology, 11*, 45-79.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 371-386.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, *97*, 488-522.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science*, *25*, 111-172.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier estimation. *The Quarterly Journal of Experimental Psychology*, *47A*, 631-650.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 616-627.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Wijk, A. (1966). *Rules of pronunciation for the English language*. Oxford: Oxford University Press.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd.). New York: McGraw-Hill.

Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language*, *34*, 567-593.

Ziegler, J. C., Van Orden, G. C., & Jacobs, A. M. (1997). Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 845-860.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What's the pronunciation of _OUGH and the spelling for /u/? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, and Computers, 29,* 600-618.

Appendices

Appendix A

Printed word stimuli for Experiments 1-8. WT = Word Type; NOL = number of letters; CLF = CELEX log frequency; K&F = Kucera & Francis frequency; LK&F = log Kucera & Francis frequency; NOF = number of friends; LSFF = log summed frequency of friends; LBF = log bigram frequency; FAM = familiarity; CONC = concreteness; IMAG = imageability.

No Dutch Neighbors

Word	WT	NOL	CLF	K&F	LK&F	NOF	LSFF	LBF	FAM	CONC	IMAG
peace	CM	5	1.96	201	2.30	1	2.30	5.99	571	309	446
force	CM	5	2.23	232	2.37	1	2.36	6.20	552	331	437
world	CM	5	2.87	816	2.91	1	2.90	6.14	607	532	560
course	CM	6	2.79	465	2.67	1	2.67	6.22	564	389	391
point	CM	5	2.57	402	2.60	2	2.64	5.92	538	464	481
bird	CM	4	1.64	32	1.51	3	2.35	5.69	592	602	614
girl	CM	4	2.46	225	2.35	3	2.35	5.55	645	607	634
with	CM	4	3.85	7289	3.86	3	3.02	6.24	608	268	287
fix	CM	3	1.28	14	1.15	3	2.39	5.70			
prince	CM	6	1.54	33	1.52	4	2.82	5.98	506	542	606
wait	CM	4	2.08	94	1.97	5	2.04	6.00	577	317	357
cape	CM	4	1.21	21	1.32	7	2.19	5.97	521	581	566
teach	CM	5	1.66	41	1.61	7	3.04	5.97	552	353	429
ground	CM	6	2.29	193	2.29	9	3.05	5.98	574	558	513
grass	CM	5	1.94	56	1.75	10	2.78	5.70	587	599	602
bluff	CM	5	0.79	8	0.90	10	1.81	5.14			
lung	CM	4	1.00	17	1.23	12	2.29	5.57	546	569	576
spy	CM	3	0.93	9	0.95	15	3.45	5.43			
stain	CM	5	0.87	6	0.78	16	2.84	5.98	534	535	533
stick	CM	5	1.74	41	1.61	19	2.60	5.80	528	604	517
said	TM	4	3.46	1961	3.29	1	3.00	5.78	600	306	313
tall	TM	4	1.84	59	1.77	13	3.38	6.04	585	439	514
trash	TM	5	0.60	3	0.48	18	2.11	5.71	541	588	599
catch	TM	5	1.86	43	1.63	8	2.10	5.96			
thread	TM	6	1.05	15	1.18	9	2.95	6.50	522	607	568
leak	TM	4	0.91	2	0.30	11	2.28	5.96	514	472	545
dear	TM	4	2.08	56	1.75	15	3.20	6.15	536	326	374
heard	TM	5	2.46	247	2.39	1	2.38	6.06			
feast	TM	5	1.11	3	0.48	5	2.73	5.94	457	542	610
cheat	TM	5	0.64	3	0.48	12	2.60	6.49	549	329	457
weight	TM	6	1.99	101	2.00	3	2.40	5.57	578	462	521
grey	TM	4	1.87	14	1.15	3	3.01	6.03			
load	TM	4	1.39	46	1.66	4	2.42	5.73	562	439	461
north	TM	5	2.19	209	2.32	2	2.44	6.27			
both	TM	4	2.80	730	2.86	1	2.86	6.10	598	322	298
through	TM	7	2.99	972	2.99	1	2.99	6.38	587	274	320
mouth	TM	5	2.15	106	2.03	3	2.54	6.16	572	568	613
brown	TM	5	2.00	183	2.26	8	3.12	6.02			
pull	TM	4	1.84	60	1.78	3	2.47	5.94	565	360	446
brush	TM	5	1.28	46	1.66	10	1.98	5.36	579	589	570
paid	AM	4	2.10	149	2.17	6	2.61	5.80	577	386	429
shall	AM	5	2.36	267	2.43	1	2.43	6.10			237

wash	AM	4	1.62	39	1.59	1	1.57	6.08	632	447	522
watch	AM	5	2.05	85	1.93	1	1.91	5.93	576	487	525
plead	AM	5	0.79	5	0.70	4	2.26	5.71			393
steak	AM	5	0.93	10	1.00	2	2.01	5.80	558	646	647
bear	AM	4	1.87	58	1.76	4	2.04	6.25	526	585	572
beard	AM	5	1.37	26	1.41	1	1.41	6.14	538	580	630
breast	AM	6	1.67	11	1.04	1	1.04	5.91	555	580	597
great	AM	5	2.82	670	2.83	2	1.81	6.00	588	311	390
height	AM	6	1.55	37	1.57	1	1.54	5.62	590	376	472
key	AM	3	1.87	92	1.96	1	1.96	5.29	603	612	618
broad	AM	5	1.64	89	1.95	1	1.92	5.88	523	399	463
worth	AM	5	2.02	96	1.98	1	1.97	6.24	542	257	275
cloth	AM	5	1.68	46	1.66	4	1.68	5.74	561	580	547
though	AM	6	2.82	442	2.65	2	2.66	6.51			
youth	AM	5	1.81	82	1.91	1	1.91	6.23	551	439	507
known	AM	5	2.42	266	2.42	6	3.11	5.76	545	226	310
dull	AM	4	1.54	28	1.45	6	1.78	5.91			373
bush	AM	4	1.65	14	1.15	2	1.71	5.82	532	585	549

With Dutch Neighbors

Word	WT	NOL	CLF	K&F	LK&F	NOF	LSFF	LBF	FAM	CONC	IMAG
she	CM	3	3.64	2858	3.46	5	3.70	6.74	599	406	474
then	CM	4	3.27	1377	3.14	11	3.48	6.92	610	190	204
bet	CM	3	1.55	20	1.30	12	3.35	6.27	527	403	453
theft	CM	5	0.90	10	1.00	4	2.69	6.77	499	361	436
such	CM	4	2.99	1303	3.11	2	3.29	5.85	521	248	219
hold	CM	4	2.20	170	2.23	9	3.20	5.84	596	416	416
hook	CM	4	1.51	5	0.70	8	3.05	5.90	497	525	541
rich	CM	4	2.10	79	1.90	2	3.03	5.69	575	377	467
cup	CM	3	1.80	45	1.65	4	3.02	5.39	595	539	558
them	CM	4	3.37	1789	3.25	4	3.02	6.91	542	344	368
spent	CM	5	2.16	105	2.02	11	3.01	6.44			
pact	CM	4	1.29	5	0.70	5	2.88	5.82	361	372	364
charge	CM	6	1.88	126	2.10	3	2.69	6.03			
moon	CM	4	1.74	66	1.82	6	2.47	5.96	585	581	585
walk	CM	4	2.10	103	2.01	3	2.41	5.92	625	452	505
seem	CM	4	2.34	229	2.36	2	2.36	6.13	549	226	249
sane	CM	4	0.91	8	0.90	7	2.25	6.08	508	290	364
shift	CM	5	1.56	42	1.62	7	2.18	5.90			
nude	CM	4	0.82	21	1.32	3	1.61	5.55			
jerk	CM	4	0.85	2	0.30	3	1.57	5.93	452	441	479
camp	TM	4	1.87	79	1.90	10	2.11	6.03	541	571	588
scan	TM	4	0.43	5	0.70	15	3.65	5.73			
bar	TM	3	1.84	85	1.93	11	2.94	6.25	592	565	596
yard	TM	4	1.58	45	1.65	5	2.43	6.05	522	553	568
barn	TM	4	1.05	31	1.49	3	1.66	6.10	466	614	589
gasp	TM	4	0.73	3	0.48	3	1.34	5.91	457	409	491
sat	TM	3	2.39	150	2.18	14	3.39	5.99			
where	TM	5	3.02	939	2.97	3	3.29	6.76	608	256	255
twist	TM	5	1.15	18	1.26	7	2.31	5.72	510	423	529
give	TM	4	2.67	392	2.59	2	2.75	5.95	595	326	383
two	TM	3	3.13	1510	3.18	4	3.60	5.71	596	383	445
come	TM	4	2.93	632	2.80	2	3.21	6.33	608	355	322
son	TM	3	2.22	173	2.24	3	2.42	6.27			560
none	TM	4	2.10	108	2.03	3	3.15	6.18	569	288	425
mood	TM	4	1.68	37	1.57	4	3.04	5.95	541	234	394

work	TM	4	2.91	772	2.89	1	2.88	6.22	603	402	458
nose	TM	4	1.88	60	1.78	7	3.03	6.03	584	628	605
ghost	TM	5	1.32	11	1.04	4	3.05	5.75	505	379	552
hour	TM	4	2.21	179	2.25	6	3.08	6.27	608	375	408
love	TM	4	2.57	237	2.37	3	2.39	6.04	619	311	569
swamp	AM	5	0.71	5	0.70	1	0.70	5.15	438	570	600
swan	AM	4	0.74	3	0.48	3	0.78	5.66			
war	AM	3	2.54	492	2.69	1	2.67	6.35	582	477	551
ward	AM	4	1.41	27	1.43	1	1.40	6.21			
warn	AM	4	1.08	11	1.04	1	1.04	6.19	491	315	359
wasp	AM	4	0.40	2	0.30	1	0.30	6.09			
what	AM	4	3.42	1909	3.28	2	3.00	6.41	608	293	249
here	AM	4	2.88	749	2.87	3	2.91	6.22	599	285	278
christ	AM	6	1.74	97	1.99	1	1.99	5.71			
five	AM	4	2.45	385	2.59	7	2.63	6.05	553	365	529
go	AM	2	2.99	633	2.80	3	3.22	5.92	618	337	364
home	AM	4	2.70	566	2.75	4	2.76	6.17	626	539	599
won	AM	3	1.85	70	1.85	3	1.49	6.20	575	324	418
stone	AM	5	1.98	62	1.79	12	2.44	5.82	564	614	585
blood	AM	5	2.18	133	2.12	2	2.15	5.33	571	613	620
fork	AM	4	1.15	15	1.18	3	1.52	6.30	584	592	598
lose	AM	4	1.91	58	1.76	2	2.49	5.89	534	299	373
lost	AM	4	2.34	174	2.24	3	2.61	6.13			
four	AM	4	2.51	416	2.62	2	2.63	6.36	553	365	491
move	AM	4	2.27	171	2.23	2	2.35	6.08	572	390	413

Appendix B

Printed word stimuli with corresponding transcriptions of sound stimuli (see Ziegler et al., 1996, for key to phonetic symbols). Source words for no-match sounds are in parentheses.

No Dutch Neighbors

Consistent Mappings

Word	Match Sound	No-Match Sound
peace	is	os (dose)
force	ors	Irs (fierce)
world	Rld	arld
course	ors	ars (sparse)
point	Ont	Ent (meant)
bird	Rd	ord (ward)
girl	Rl	arl
with	IT	uT (truth)
fix	Iks	@ks (wax)
prince	Ins	Ens (hence)
wait	et	it (beat)
cape	ep	op (slope)
teach	iC	IC (which)
ground	Wnd	End (friend)
grass	@s	Is (miss)
bluff	}f	@f (staff)
lung	}G	@G (gang)
spy	Y	u (too)
stain	en	}n (sun)
stick	Ik	ok (joke)

Typical Mappings

Word	Match Sound	Catch Sound	No-Match Sound
said	Ed	ed	id (plead)
tall	cl	@l	Il (fill)
trash	@S	cS	}S (brush)
catch	@C	cC	IC (witch)
thread	Ed	id	od (code)
leak	ik	ek	ok (soak)
dear	Ir	Er	R (her)
heard	Rd	Ird	ard (yard)
feast	ist	Est	ost (toast)
cheat	it	Et	Ut (foot)
weight	et	Yt	}t (but)
grey	e	i	o (go)
load	od	cd	id (read)
north	orT	RT	irT
both	oT	cT	ET (death)
through	u	o	e (may)
mouth	WT	uT	IT (with)
brown	Wn	on	cn (fawn)
pull	Ul	}l	El (bell)
brush	}S	US	Ish (wish)

Atypical Mappings

paid	ed	Ed	id (plead)
shall	@l	cl	Il (fill)
wash	cS	@S	}S (brush)
watch	cC	@C	IC (witch)
plead	id	Ed	od (code)
steak	ek	ik	ok (soak)
bear	Er	Ir	R (her)
beard	Ird	Rd	ard (yard)
breast	Est	ist	ost (toast)
great	Et	it	Ut (foot)
height	Yt	et	}t (but)
key	i	e	o (go)
broad	cd	od	id (read)
worth	RT	orT	irT
cloth	cT	oT	ET (death)
though	o	u	e (may)
youth	uT	WT	IT (with)
known	on	Wn	cn (fawn)
dull	}l	Ul	El (bell)
bush	US	}S	Ish (wish)

Appendix B *(continued)*

With Dutch Neighbors

Consistent Mappings

Word	Match Sound	Dutch Catch Sound	No-Match Sound
she	i	} (ze)	o (go)
then	En	En (den)	In (pin)
bet	Et	Et (smet)	at (plot)
theft	Eft	Eft (heft)	Ift (lift)
such	}C	}g (kuch)	iC (beach)
hold	old	old (gold)	Yld (child)
hook	Uk	ok (rook)	ek (steak)
rich	IC	Ig (zich)	}C (touch)
cup	}p	}p (hup)	ip (heap)
them	Em	Em (rem)	}m (from)
spent	Ent	Ent (krent)	Wnt (count)
pact	@kt	akt (exact)	Ikt (strict)
charge	arJ	arJ} (marge)	orJ (gorge)
moon	un	on (boon)	en (vein)
walk	ck	alk (kalk)	Ilk (milk)
seem	im	em (zeem)	om (foam)
sane	en	an} (liane)	}n (son)
shift	Ift	Ift (gift)	@ft (draft)
nude	ud	ud} (etude)	Yd (bride)
jerk	Rk	Erk (kerk)	ark (bark)

Typical Mappings

Word	Match Sound	Dutch Catch Sound	English Catch Sound	No-Match Sound
camp	@mp	amp (kramp)	amp	Imp (shrimp)
scan	@n	an (pan)	an	on (bone)
bar	ar	ar (kar)	or	Ir (clear)
yard	ard	ard (hard)	ord	Rd (bird)
barn	arn	arn (barn)	orn	Rn (stern)
gasp	@sp	asp (rasp)	asp	Isp (crisp)
sat	@t	at (nat)	at	Et (great)
where	Er	Ir} (ere)	Ir	or (door)
twist	Ist	Ist (kist)	Yst	}st (bust)
give	Iv	ive (prive)	Yv	@v (have)
two	u	o (zo)	o	i (flea)
come	}m	om} (slome)	om	um (room)
son	}n	On (bon)	an	@n (than)
none	}n	on} (schone)	on	in (bean)
mood	Ud	od (rood)	}d	Yd (bride)
work	Rk	Ork (vork)	ork	ark (bark)
nose	oz	os} (pose)	uz	ez (blaze)
ghost	cst	Ost (kost)	cst	ist (priest)
hour	Wr	ur (tour)	or	Ir (here)
love	}v	ov} (dove)	uv	Yv (five)

Appendix B *(continued)*

Atypical Mappings

Word	Match Sound	Dutch Catch Sound	English Catch Sound	No-Match Sound
swamp	amp	amp (kramp)	@mp	Imp (shrimp)
swan	an	an (pan)	@n	on (bone)
war	or	ar (kar)	ar	Ir (clear)
ward	ord	ard (hard)	ard	Rd (bird)
warn	orn	arn (barn)	arn	Rn (stern)
wasp	asp	asp (rasp)	@sp	Isp (crisp)
what	at	at (nat)	@t	Et (great)
here	Ir	Ir} (ere)	Er	or (door)
christ	Yst	Ist (kist)	Ist	}st (bust)
five	Yv	ive (prive)	Iv	@v (have)
go	o	o (zo)	u	i (flea)
home	om	om} (slome)	}m	um (room)
won	an	On (bon)	}n	@n (than)
stone	on	on} (schone)	}n	in (bean)
blood	}d	od (rood)	Ud	Yd (bride)
fork	ork	Ork (vork)	Rk	ark (bark)
lose	uz	os} (pose)	oz	ez (blaze)
lost	cst	Ost (kost)	cst	ist (priest)
four	or	ur (tour)	Wr	Ir (here)
move	uv	ov} (dove)	}v	Yv (five)

Appendix C

Tabulations of error rates and response latencies as a function of procedural and experimental variables for Experiments 2-8.

Table 1
Mean percentages of false negatives (standard errors in parentheses) as a function of Word Type, Trial Block, and Sequence of Trial Block in Experiment 2.

Trial Block A

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	5.5 (1.5)	5.0 (1.8)	2.0 (0.9)
B-A-D-C	7.5 (1.9)	7.0 (1.5)	2.0 (1.4)
C-D-A-B	3.5 (1.3)	3.0 (1.1)	0.0 (0.0)
D-C-B-A	4.0 (1.5)	4.5 (1.5)	2.0 (1.2)

Trial Block B

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	13.5 (2.9)	5.0 (1.8)	1.0 (0.7)
B-A-D-C	9.0 (1.8)	2.0 (1.2)	3.0 (1.1)
C-D-A-B	6.5 (1.7)	2.0 (1.2)	2.5 (1.2)
D-C-B-A	7.0 (1.9)	3.0 (1.3)	1.5 (0.8)

Trial Block C

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	6.0 (1.3)	5.0 (1.4)	0.5 (0.5)
B-A-D-C	4.5 (1.4)	4.0 (1.8)	1.5 (0.8)
C-D-A-B	6.5 (2.1)	5.0 (1.4)	2.0 (0.9)
D-C-B-A	3.5 (1.5)	4.5 (1.5)	1.0 (0.7)

Trial Block D

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	10.5 (1.7)	3.0 (1.3)	1.0 (0.7)
B-A-D-C	10.0 (1.9)	2.5 (1.2)	1.0 (1.0)
C-D-A-B	12.5 (2.3)	3.5 (1.3)	1.0 (0.7)
D-C-B-A	10.0 (2.5)	4.0 (1.1)	1.5 (0.8)

Appendix C *(continued)*

Table 2
Mean correct yes-response latencies (standard errors in parentheses) as a function of Word Type, Trial Block, and Sequence of Trial Block in Experiment 2.

Trial Block A

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	812 (23)	791 (31)	763 (25)
B-A-D-C	782 (38)	763 (36)	705 (33)
C-D-A-B	719 (23)	670 (24)	684 (24)
D-C-B-A	678 (25)	674 (22)	676 (22)

Trial Block B

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	819 (30)	786 (31)	736 (31)
B-A-D-C	860 (40)	838 (44)	770 (38)
C-D-A-B	735 (27)	698 (26)	659 (24)
D-C-B-A	746 (24)	704 (23)	679 (19)

Trial Block C

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	770 (32)	730 (28)	729 (26)
B-A-D-C	727 (24)	691 (29)	686 (28)
C-D-A-B	826 (43)	774 (36)	770 (38)
D-C-B-A	715 (25)	700 (17)	693 (23)

Trial Block D

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	762 (33)	738 (27)	681 (24)
B-A-D-C	760 (21)	699 (23)	674 (24)
C-D-A-B	787 (28)	773 (35)	721 (32)
D-C-B-A	825 (25)	786 (25)	746 (21)

Appendix C *(continued)*

Table 3
Mean percentages of false positives (standard errors in parentheses) as a function of Trial Type, Word Type, Trial Block, and Sequence of Trial Block in Experiment 2.

Trial Block A

Sequence	AM (List 2)		TM (List 1)		CM (List 1)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	4.5 (1.5)	30.5 (3.4)	-	6.0 (1.8)
B-A-D-C	-	2.5 (1.2)	20.0 (2.8)	-	5.5 (2.1)
C-D-A-B	-	2.5 (1.0)	11.5 (2.9)	-	3.0 (1.5)
D-C-B-A	-	2.5 (1.2)	12.0 (3.0)	-	1.0 (0.7)

Trial Block B

Sequence	AM (List 1)		TM (List 2)		CM (List 2)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	59.5 (4.9)	-	-	4.0 (1.3)	3.0 (1.6)
B-A-D-C	57.0 (3.8)	-	-	4.5 (1.1)	5.5 (1.8)
C-D-A-B	43.5 (4.5)	-	-	2.0 (1.2)	1.0 (0.7)
D-C-B-A	41.0 (5.0)	-	-	1.0 (0.7)	1.5 (0.8)

Trial Block C

Sequence	AM (List 2)		TM (List 1)		CM (List 1)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	25.5 (2.7)	-	-	2.5 (1.2)	1.5 (1.1)
B-A-D-C	18.0 (2.4)	-	-	3.5 (1.8)	3.5 (1.5)
C-D-A-B	27.5 (3.3)	-	-	2.0 (0.9)	4.0 (1.5)
D-C-B-A	24.0 (2.9)	-	-	3.5 (1.3)	3.0 (1.3)

Trial Block D

Sequence	AM (List 1)		TM (List 2)		CM (List 2)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	7.0 (1.9)	19.0 (2.7)	-	3.5 (1.3)
B-A-D-C	-	10.0 (2.6)	17.5 (2.4)	-	2.5 (1.2)
C-D-A-B	-	10.0 (2.4)	18.0 (2.6)	-	2.0 (1.2)
D-C-B-A	-	15.5 (2.2)	24.5 (3.6)	-	2.5 (1.2)

Appendix C *(continued)*

Table 4
Mean correct no-response latencies (standard errors in parentheses) as a function of Trial Type, Word Type, Trial Block, and Sequence of Trial Block in Experiment 2.

Trial Block A

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	780 (25)	951 (36)	-	820 (27)
B-A-D-C	-	747 (31)	841 (41)	-	738 (33)
C-D-A-B	-	697 (25)	813 (32)	-	743 (31)
D-C-B-A	-	676 (24)	772 (28)	-	712 (28)

Trial Block B

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	951 (30)	-	-	745 (24)	772 (22)
B-A-D-C	1115 (46)	-	-	822 (41)	827 (35)
C-D-A-B	942 (43)	-	-	711 (20)	711 (27)
D-C-B-A	879 (30)	-	-	708 (19)	721 (22)

Trial Block C

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	838 (33)	-	-	737 (23)	734 (25)
B-A-D-C	794 (35)	-	-	701 (27)	695 (23)
C-D-A-B	954 (43)	-	-	840 (41)	822 (39)
D-C-B-A	766 (21)	-	-	720 (25)	701 (23)

Trial Block D

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	802 (24)	824 (28)	-	749 (26)
B-A-D-C	-	759 (24)	765 (25)	-	711 (28)
C-D-A-B	-	837 (38)	857 (35)	-	762 (32)
D-C-B-A	-	851 (31)	927 (35)	-	806 (24)

Appendix C *(continued)*

Table 5
Mean percentages of false negatives (standard errors in parentheses) as a function of Word Type, Trial Block, and Sequence of Trial Block in Experiment 3.

Trial Block A

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	4.0 (1.3)	3.3 (2.1)	6.0 (1.9)
B-A-D-C	5.3 (2.4)	1.3 (0.9)	4.0 (1.9)
C-D-A-B	2.0 (1.4)	1.3 (0.9)	1.3 (0.9)
D-C-B-A	2.0 (1.1)	1.3 (0.9)	0.7 (0.7)

Trial Block B

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	8.7 (3.6)	1.3 (0.9)	2.0 (1.1)
B-A-D-C	12.0 (2.4)	0.7 (0.7)	2.7 (1.5)
C-D-A-B	6.7 (2.3)	1.3 (0.9)	2.7 (1.5)
D-C-B-A	8.0 (2.6)	0.7 (0.7)	1.3 (1.3)

Trial Block C

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	1.3 (1.3)	2.7 (1.5)	2.7 (1.5)
B-A-D-C	3.3 (1.6)	1.3 (0.9)	2.7 (1.2)
C-D-A-B	4.0 (2.1)	2.0 (1.1)	2.0 (1.1)
D-C-B-A	1.3 (0.9)	2.7 (1.5)	2.0 (1.1)

Trial Block D

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	5.3 (1.7)	0.7 (0.7)	1.3 (0.9)
B-A-D-C	7.3 (1.8)	0.0 (0.0)	1.3 (0.9)
C-D-A-B	11.3 (3.5)	2.0 (1.1)	2.7 (1.2)
D-C-B-A	10.7 (2.3)	2.0 (1.1)	4.0 (1.6)

Appendix C *(continued)*

Table 6
Mean correct yes-response latencies (standard errors in parentheses) as a function of Word Type, Trial Block, and Sequence of Trial Block in Experiment 3.

Trial Block A

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	809 (39)	789 (38)	807 (47)
B-A-D-C	727 (41)	737 (36)	722 (29)
C-D-A-B	657 (32)	641 (24)	653 (28)
D-C-B-A	669 (28)	696 (34)	668 (31)

Trial Block B

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	816 (32)	747 (26)	758 (22)
B-A-D-C	866 (44)	760 (37)	811 (34)
C-D-A-B	692 (41)	626 (24)	634 (25)
D-C-B-A	721 (27)	689 (25)	713 (30)

Trial Block C

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B-C-D	709 (26)	696 (23)	705 (28)
B-A-D-C	653 (22)	654 (18)	653 (20)
C-D-A-B	769 (42)	735 (33)	751 (38)
D-C-B-A	710 (25)	722 (26)	721 (31)

Trial Block D

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B-C-D	711 (21)	660 (22)	700 (23)
B-A-D-C	713 (29)	659 (27)	690 (21)
C-D-A-B	800 (47)	685 (33)	699 (40)
D-C-B-A	824 (33)	739 (29)	770 (34)

Appendix C *(continued)*

Table 7
Mean percentages of false positives (standard errors in parentheses) as a function of Trial Type, Word Type, Trial Block, and Sequence of Trial Block in Experiment 3.

Trial Block A

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	17.3 (3.0)	28.0 (4.0)	-	5.3 (1.7)
B-A-D-C	-	11.3 (1.9)	19.3 (5.2)	-	4.7 (1.3)
C-D-A-B	-	11.3 (1.7)	15.3 (4.0)	-	8.0 (2.4)
D-C-B-A	-	10.7 (2.7)	14.7 (3.8)	-	6.0 (1.9)

Trial Block B

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	44.0 (5.7)	-	-	8.0 (2.2)	7.3 (1.5)
B-A-D-C	46.7 (5.2)	-	-	10.7 (2.1)	9.3 (2.1)
C-D-A-B	33.3 (5.7)	-	-	3.3 (1.3)	7.3 (2.1)
D-C-B-A	39.3 (4.3)	-	-	2.7 (1.2)	3.3 (1.6)

Trial Block C

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	34.0 (5.0)	-	-	1.3 (0.9)	8.7 (2.2)
B-A-D-C	24.0 (4.0)	-	-	3.3 (1.3)	7.3 (1.8)
C-D-A-B	39.3 (6.4)	-	-	2.0 (1.1)	5.3 (1.3)
D-C-B-A	44.7 (4.2)	-	-	1.3 (0.9)	10.0 (2.0)

Trial Block D

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	1.3 (0.9)	17.3 (3.4)	-	6.7 (2.5)
B-A-D-C	-	2.0 (1.4)	18.0 (4.0)	-	6.0 (2.1)
C-D-A-B	-	3.3 (1.6)	24.7 (5.2)	-	6.7 (2.5)
D-C-B-A	-	7.3 (1.5)	42.0 (3.9)	-	10.0 (2.2)

Appendix C *(continued)*

Table 8
Mean correct no-response latencies (standard errors in parentheses) as a function of Trial Type, Word Type, Trial Block, and Sequence of Trial Block in Experiment 3.

Trial Block A

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	850 (49)	972 (47)	-	820 (39)
B-A-D-C	-	784 (30)	881 (42)	-	777 (35)
C-D-A-B	-	671 (29)	755 (35)	-	672 (33)
D-C-B-A	-	733 (40)	820 (33)	-	726 (36)

Trial Block B

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	971 (32)	-	-	802 (38)	754 (25)
B-A-D-C	1007 (53)	-	-	887 (42)	843 (39)
C-D-A-B	786 (37)	-	-	695 (37)	663 (32)
D-C-B-A	920 (56)	-	-	776 (31)	770 (37)

Trial Block C

	AM (List 2)		TM (List 1)		CM (List 1)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	921 (41)	-	-	702 (20)	735 (29)
B-A-D-C	875 (41)	-	-	688 (19)	727 (29)
C-D-A-B	1025 (56)	-	-	749 (38)	780 (45)
D-C-B-A	996 (71)	-	-	733 (36)	764 (41)

Trial Block D

	AM (List 1)		TM (List 2)		CM (List 2)
Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
A-B-C-D	-	687 (24)	822 (30)	-	718 (31)
B-A-D-C	-	692 (21)	848 (32)	-	717 (22)
C-D-A-B	-	693 (30)	842 (39)	-	710 (28)
D-C-B-A	-	814 (51)	1008 (48)	-	823 (50)

Appendix C *(continued)*

Table 9
Mean percentages of false negatives (standard errors in parentheses) as a function of Replication, Word Type, Trial Block, and Sequence of Trial Block in Experiment 4.

Trial Block A1

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	11.0 (2.8)	8.0 (2.9)	5.0 (2.2)
B-A	9.0 (3.8)	6.0 (3.1)	0.0 (0.0)

Trial Block B1

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	14.0 (2.2)	8.0 (2.0)	3.0 (1.5)
B-A	18.0 (5.5)	9.0 (3.1)	1.0 (1.0)

Trial Block A2

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	6.0 (1.6)	8.0 (3.3)	1.0 (1.0)
B-A	4.0 (2.2)	4.0 (1.6)	1.0 (1.0)

Trial Block B2

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	8.0 (2.9)	4.0 (1.6)	0.0 (0.0)
B-A	16.0 (4.0)	5.0 (2.2)	2.0 (2.0)

Trial Block A3

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	9.0 (3.1)	4.0 (1.6)	0.0 (0.0)
B-A	4.0 (2.2)	4.0 (1.6)	1.0 (1.0)

Trial Block B3

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	9.0 (3.1)	1.0 (1.0)	0.0 (0.0)
B-A	8.0 (2.5)	4.0 (2.2)	1.0 (1.0)

Appendix C *(continued)*

Table 10
Mean correct yes-response latencies (standard errors in parentheses) as a function of Replication, Word Type, Trial Block, and Sequence of Trial Block in Experiment 4.

Trial Block A1

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	897 (50)	902 (57)	785 (40)
B-A	764 (34)	778 (43)	762 (48)

Trial Block B1

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	814 (40)	796 (39)	780 (41)
B-A	849 (60)	775 (54)	795 (54)

Trial Block A2

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	775 (45)	747 (40)	755 (39)
B-A	717 (43)	694 (40)	683 (26)

Trial Block B2

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	814 (37)	750 (39)	710 (37)
B-A	753 (42)	718 (40)	696 (39)

Trial Block A3

Sequence	AM (List 2)	TM (List 1)	CM (List 1)
A-B	708 (43)	705 (35)	666 (47)
B-A	663 (34)	647 (35)	644 (31)

Trial Block B3

Sequence	AM (List 1)	TM (List 2)	CM (List 2)
A-B	728 (35)	674 (28)	635 (31)
B-A	696 (38)	672 (38)	629 (33)

Appendix C *(continued)*

Table 11
Mean percentages of false positives (standard errors in parentheses) as a function of Replication, Trial Type, Word Type, and Trial Block in Experiment 4.

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
Replication	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
Primary Test	-	13.5 (2.5)	21.0 (3.7)	-	2.5 (1.4)
Replication 1	-	10.5 (3.0)	15.0 (3.4)	-	2.5 (1.2)
Replication 2	-	5.5 (1.7)	10.0 (2.7)	-	1.5 (1.1)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
Replication	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
Primary Test	28.0 (2.4)	-	-	8.5 (2.3)	2.0 (1.2)
Replication 1	18.0 (2.8)	-	-	4.5 (1.5)	1.5 (0.8)
Replication 2	11.0 (1.8)	-	-	4.5 (1.7)	2.0 (0.9)

Table 12
Mean correct no-response latencies (standard errors in parentheses) as a function of Replication, Trial Type, Word Type, and Trial Block in Experiment 4.

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
Replication	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
Primary Test	-	869 (34)	931 (38)	-	789 (39)
Replication 1	-	768 (28)	844 (32)	-	718 (33)
Replication 2	-	735 (33)	803 (37)	-	658 (25)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
Replication	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
Primary Test	962 (47)	-	-	853 (39)	817 (39)
Replication 1	869 (38)	-	-	773 (39)	760 (38)
Replication 2	799 (32)	-	-	667 (27)	678 (26)

Appendix C *(continued)*

Table 13
Mean percentages of false negatives (standard errors in parentheses) as a function of SOA, Language, Word Type, and Trial Block in Experiment 5.

Trial Block A

Language and SOA	AM (List 1)	TM (List 2)	CM (List 2)
English Speakers			
SOA 1	5.8 (1.5)	2.9 (1.3)	3.8 (1.2)
SOA 2	2.1 (0.8)	1.7 (1.0)	2.1 (0.8)
SOA 3	3.8 (1.5)	0.8 (0.6)	1.7 (0.8)
Dutch Speakers			
SOA 1	7.8 (1.5)	4.4 (0.9)	5.0 (1.5)
SOA 2	6.1 (1.3)	4.4 (0.9)	3.3 (1.1)
SOA 3	4.2 (1.3)	3.6 (1.1)	0.8 (0.6)

Trial Block B

Language and SOA	AM (List 2)	TM (List 1)	CM (List 1)
English Speakers			
SOA 1	4.6 (1.5)	2.9 (1.5)	2.1 (1.0)
SOA 2	4.6 (1.3)	3.8 (1.2)	0.4 (0.4)
SOA 3	4.6 (1.3)	0.8 (0.6)	0.4 (0.4)
Dutch Speakers			
SOA 1	11.9 (1.7)	6.4 (1.6)	0.8 (0.5)
SOA 2	11.4 (1.7)	2.5 (0.9)	2.5 (1.1)
SOA 3	10.6 (1.5)	1.7 (0.6)	2.5 (0.7)

Appendix C *(continued)*

Table 14
Mean correct yes-response latencies (standard errors in parentheses) as a function of SOA, Language, Word Type, and Trial Block in Experiment 5.

Trial Block A

Language and SOA	AM (List 1)	TM (List 2)	CM (List 2)
English Speakers			
SOA 1	610 (21)	586 (22)	574 (16)
SOA 2	745 (29)	709 (28)	704 (23)
SOA 3	1239 (33)	1202 (33)	1200 (29)
Dutch Speakers			
SOA 1	693 (23)	694 (24)	679 (22)
SOA 2	755 (28)	744 (29)	711 (25)
SOA 3	1110 (27)	1108 (28)	1073 (25)

Trial Block B

Language and SOA	AM (List 2)	TM (List 1)	CM (List 1)
English Speakers			
SOA 1	580 (19)	593 (22)	556 (19)
SOA 2	738 (25)	731 (30)	709 (22)
SOA 3	1219 (33)	1231 (32)	1188 (30)
Dutch Speakers			
SOA 1	719 (29)	688 (33)	633 (22)
SOA 2	765 (25)	721 (23)	684 (21)
SOA 3	1149 (26)	1107 (24)	1066 (24)

Appendix C *(continued)*

Table 15
Mean percentages of false positives (standard errors in parentheses) as a function of SOA, Language, Sequence of SOA, Trial Type, Word Type, and Trial Block in Experiment 5.

Sound earlier than Print (SOA1)

English speakers

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	-	20.0 (8.2)	20.0 (9.1)	-	7.5 (4.8)
1-3-2	-	10.0 (4.1)	7.5 (7.5)	-	5.0 (5.0)
2-1-3	-	10.0 (5.8)	27.5 (7.5)	-	0.0 (0.0)
2-3-1	-	7.5 (4.8)	5.0 (2.9)	-	5.0 (5.0)
3-1-2	-	0.0 (0.0)	2.5 (2.5)	-	2.5 (2.5)
3-2-1	-	7.5 (4.8)	10.0 (4.1)	-	2.5 (2.5)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	40.0 (4.1)	-	-	10.0 (5.8)	7.5 (4.8)
1-3-2	12.5 (7.5)	-	-	5.0 (2.9)	12.5 (4.8)
2-1-3	25.0 (6.5)	-	-	5.0 (2.9)	12.5 (2.5)
2-3-1	2.5 (2.5)	-	-	7.5 (4.8)	5.0 (5.0)
3-1-2	7.5 (4.8)	-	-	0.0 (0.0)	7.5 (4.8)
3-2-1	25.0 (8.7)	-	-	2.5 (2.5)	0.0 (0.0)

Sound earlier than Print (SOA1)

Dutch speakers

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	-	5.0 (3.4)	21.7 (3.1)	-	1.7 (1.7)
1-3-2	-	3.3 (2.1)	11.7 (3.1)	-	0.0 (0.0)
2-1-3	-	10.0 (6.8)	23.3 (8.0)	-	1.7 (1.7)
2-3-1	-	0.0 (0.0)	6.7 (3.3)	-	1.7 (1.7)
3-1-2	-	10.0 (5.2)	18.3 (6.0)	-	5.0 (2.2)
3-2-1	-	8.3 (4.0)	10.0 (6.3)	-	5.0 (3.4)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	31.7 (4.8)	-	-	0.0 (0.0)	5.0 (2.2)
1-3-2	25.0 (10.2)	-	-	5.0 (5.0)	5.0 (3.4)
2-1-3	16.7 (4.9)	-	-	8.3 (3.1)	1.7 (1.7)
2-3-1	10.0 (3.7)	-	-	3.3 (2.1)	3.3 (2.1)
3-1-2	20.0 (5.2)	-	-	3.3 (3.3)	1.7 (1.7)
3-2-1	21.7 (8.7)	-	-	0.0 (0.0)	0.0 (0.0)

***Table 15** (continued)*

Sound simultaneously with Print (SOA2)

English speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	30.0 (7.1)	15.0 (11.9)	-	2.5 (2.5)
1-3-2	-	12.5 (9.5)	10.0 (7.1)	-	0.0 (0.0)
2-1-3	-	15.0 (5.0)	25.0 (8.7)	-	5.0 (2.9)
2-3-1	-	12.5 (4.8)	12.5 (4.8)	-	0.0 (0.0)
3-1-2	-	7.5 (2.5)	2.5 (2.5)	-	0.0 (0.0)
3-2-1	-	10.0 (4.1)	10.0 (7.1)	-	0.0 (0.0)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	22.5 (10.3)	-	-	2.5 (2.8)	5.0 (5.0)
1-3-2	12.5 (4.8)	-	-	2.5 (2.5)	10.0 (4.1)
2-1-3	35.0 (6.5)	-	-	7.5 (2.5)	20.0 (4.1)
2-3-1	12.5 (9.5)	-	-	20.0 (13.5)	7.5 (4.8)
3-1-2	5.0 (2.9)	-	-	2.5 (2.5)	7.5 (2.5)
3-2-1	15.0 (6.5)	-	-	7.5 (4.8)	7.5 (4.8)

Sound simultaneously with Print (SOA2)

Dutch speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	10.0 (5.2)	15.0 (3.4)	-	1.7 (1.7)
1-3-2	-	6.7 (3.3)	6.7 (3.3)	-	1.7 (1.7)
2-1-3	-	13.3 (6.7)	18.3 (4.8)	-	8.3 (4.8)
2-3-1	-	8.3 (3.1)	23.3 (6.1)	-	8.3 (3.1)
3-1-2	-	8.3 (4.0)	21.7 (5.4)	-	5.0 (2.2)
3-2-1	-	20.0 (6.3)	20.0 (6.8)	-	1.7 (1.7)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	23.3 (3.3)	-	-	5.0 (2.2)	1.7 (1.7)
1-3-2	13.3 (3.3)	-	-	10.0 (5.2)	0.0 (0.0)
2-1-3	28.3 (6.0)	-	-	6.7 (3.3)	5.0 (3.4)
2-3-1	23.3 (4.9)	-	-	1.7 (1.7)	5.0 (2.2)
3-1-2	21.7 (8.3)	-	-	3.3 (3.3)	1.7 (1.7)
3-2-1	26.7 (2.1)	-	-	13.3 (4.9)	3.3 (2.1)

***Table 15** (continued)*

Sound later than Print (SOA3)

English speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	2.5 (2.5)	7.5 (4.8)	-	0.0 (0.0)
1-3-2	-	5.0 (2.9)	10.0 (10.0)	-	2.5 (2.5)
2-1-3	-	2.5 (2.5)	5.0 (2.9)	-	5.0 (2.9)
2-3-1	-	15.0 (6.5)	10.0 (10.0)	-	2.5 (2.5)
3-1-2	-	7.5 (2.5)	2.5 (2.5)	-	0.0 (0.0)
3-2-1	-	15.0 (5.0)	10.0 (4.1)	-	0.0 (0.0)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	2.5 (2.5)	-	-	0.0 (0.0)	2.5 (2.5)
1-3-2	15.0 (6.5)	-	-	2.5 (2.5)	5.0 (2.9)
2-1-3	12.5 (4.8)	-	-	0.0 (0.0)	12.5 (6.3)
2-3-1	0.0 (0.0)	-	-	0.0 (0.0)	2.5 (2.5)
3-1-2	7.5 (2.5)	-	-	2.5 (2.5)	7.5 (4.8)
3-2-1	20.0 (7.1)	-	-	0.0 (0.0)	5.0 (2.9)

Sound later than Print (SOA3)

Dutch speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	5.0 (2.2)	10.0 (3.7)	-	0.0 (0.0)
1-3-2	-	8.3 (4.8)	3.3 (2.1)	-	0.0 (0.0)
2-1-3	-	5.0 (3.4)	8.3 (4.8)	-	3.3 (2.1)
2-3-1	-	1.7 (1.7)	5.0 (2.2)	-	1.7 (1.7)
3-1-2	-	5.0 (2.2)	13.3 (6.7)	-	1.7 (1.7)
3-2-1	-	18.3 (7.0)	20.0 (3.7)	-	1.7 (1.7)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	16.7 (3.3)	-	-	3.3 (2.1)	0.0 (0.0)
1-3-2	15.0 (5.0)	-	-	3.3 (2.1)	3.3 (2.1)
2-1-3	5.0 (2.2)	-	-	0.0 (0.0)	1.7 (1.7)
2-3-1	11.7 (6.5)	-	-	0.0 (0.0)	3.3 (3.3)
3-1-2	16.7 (7.1)	-	-	3.3 (3.3)	8.3 (3.1)
3-2-1	23.3 (6.1)	-	-	5.0 (3.4)	1.7 (1.7)

Appendix C *(continued)*

Table 16
Mean correct no-response latencies (standard errors in parentheses) as a function of SOA, Language, Sequence of SOA, Trial Type, Word Type, and Trial Block in Experiment 5.

Sound earlier than Print (SOA1)

English speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	631 (64)	716 (75)	-	573 (68)
1-3-2	-	698 (47)	769 (77)	-	670 (36)
2-1-3	-	716 (32)	820 (61)	-	650 (18)
2-3-1	-	626 (49)	657 (63)	-	602 (48)
3-1-2	-	696 (48)	799 (64)	-	692 (48)
3-2-1	-	693 (123)	766 (102)	-	546 (45)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	740 (89)	-	-	662 (59)	653 (69)
1-3-2	795 (87)	-	-	707 (42)	709 (60)
2-1-3	735 (17)	-	-	667 (31)	638 (37)
2-3-1	610 (54)	-	-	529 (21)	548 (24)
3-1-2	877 (24)	-	-	702 (53)	763 (88)
3-2-1	661 (71)	-	-	645 (96)	608 (94)

Sound earlier than Print (SOA1)

Dutch speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	694 (70)	791 (55)	-	640 (36)
1-3-2	-	748 (76)	855 (57)	-	713 (65)
2-1-3	-	764 (78)	858 (103)	-	687 (85)
2-3-1	-	781 (46)	877 (93)	-	736 (47)
3-1-2	-	715 (38)	759 (40)	-	636 (43)
3-2-1	-	707 (55)	719 (27)	-	629 (33)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	776 (75)	-	-	719 (81)	673 (63)
1-3-2	908 (95)	-	-	713 (95)	735 (112)
2-1-3	827 (66)	-	-	685 (58)	668 (69)
2-3-1	789 (52)	-	-	733 (34)	714 (44)
3-1-2	835 (47)	-	-	675 (25)	651 (16)
3-2-1	743 (40)	-	-	701 (60)	662 (49)

***Table 16** (continued)*

Sound simultaneously with Print (SOA2)

English speakers

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	-	760 (58)	813 (78)	-	694 (49)
1-3-2	-	891 (22)	864 (59)	-	771 (32)
2-1-3	-	969 (77)	1014 (70)	-	879 (56)
2-3-1	-	804 (53)	790 (80)	-	747 (48)
3-1-2	-	921 (49)	864 (21)	-	819 (47)
3-2-1	-	788 (104)	809 (49)	-	680 (61)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	768 (59)	-	-	716 (45)	701 (26)
1-3-2	900 (36)	-	-	862 (41)	885 (33)
2-1-3	997 (30)	-	-	880 (45)	875 (35)
2-3-1	853 (71)	-	-	751 (34)	748 (42)
3-1-2	881 (86)	-	-	839 (13)	897 (41)
3-2-1	801 (112)	-	-	749 (76)	767 (77)

Sound simultaneously with Print (SOA2)

Dutch speakers

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	-	723 (49)	771 (35)	-	653 (37)
1-3-2	-	687 (68)	775 (69)	-	666 (79)
2-1-3	-	819 (64)	913 (129)	-	756 (90)
2-3-1	-	942 (63)	1004 (66)	-	897 (71)
3-1-2	-	740 (32)	807 (74)	-	697 (45)
3-2-1	-	771 (34)	893 (75)	-	734 (34)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
SOA Sequence	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
1-2-3	814 (73)	-	-	678 (24)	691 (30)
1-3-2	741 (71)	-	-	664 (60)	606 (37)
2-1-3	901 (69)	-	-	782 (66)	754 (63)
2-3-1	1008 (67)	-	-	870 (33)	835 (53)
3-1-2	874 (48)	-	-	700 (29)	681 (23)
3-2-1	852 (71)	-	-	757 (58)	775 (57)

***Table 16** (continued)*

Sound later than Print (SOA3)

English speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	1117 (58)	1092 (62)	-	1092 (35)
1-3-2	-	1392 (46)	1441 (55)	-	1319 (52)
2-1-3	-	1257 (45)	1353 (79)	-	1267 (39)
2-3-1	-	1266 (61)	1276 (46)	-	1179 (27)
3-1-2	-	1448 (52)	1486 (75)	-	1347 (69)
3-2-1	-	1318 (42)	1328 (74)	-	1232 (40)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	1143 (56)	-	-	1101 (51)	1094 (27)
1-3-2	1392 (55)	-	-	1373 (27)	1396 (62)
2-1-3	1341 (41)	-	-	1228 (44)	1328 (29)
2-3-1	1254 (17)	-	-	1192 (22)	1214 (21)
3-1-2	1524 (102)	-	-	1402 (85)	1403 (70)
3-2-1	1211 (54)	-	-	1237 (70)	1184 (48)

Sound later than Print (SOA3)

Dutch speakers

Trial Block A

SOA Sequence	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
1-2-3	-	1097 (37)	1137 (46)	-	1051 (31)
1-3-2	-	1155 (85)	1228 (90)	-	1094 (73)
2-1-3	-	1119 (98)	1147 (117)	-	1076 (95)
2-3-1	-	1231 (34)	1291 (78)	-	1146 (48)
3-1-2	-	1230 (55)	1304 (38)	-	1169 (25)
3-2-1	-	1175 (30)	1263 (57)	-	1165 (45)

Trial Block B

SOA Sequence	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
1-2-3	1113 (39)	-	-	1071 (32)	1101 (44)
1-3-2	1179 (115)	-	-	1093 (87)	1041 (59)
2-1-3	1157 (87)	-	-	1062 (54)	1126 (63)
2-3-1	1233 (37)	-	-	1179 (35)	1159 (54)
3-1-2	1337 (16)	-	-	1184 (44)	1138 (22)
3-2-1	1276 (62)	-	-	1200 (74)	1186 (46)

Appendix C *(continued*

Table 17
Mean percentages of false positives (standard errors in parentheses) as a function of Filler Type, Trial Type, Word Type, and Trial Block in Experiment 6.

Trial Block A

Filler Type	AM (List 1)		TM (List 2)		CM (List 2)	
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial
None	-	13.1 (2.3)	15.6 (1.5)	-	12.8 (1.9)	-
English	-	11.3 (2.8)	18.3 (2.3)	-	13.7 (2.4)	-
Dutch	-	12.7 (2.4)	18.3 (3.1)	-	18.0 (3.5)	-

Trial Block B

Filler Type	AM (List 2)		TM (List 1)		CM (List 1)	
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial
None	47.4 (2.5)	-	-	7.7 (1.6)	-	2.1 (0.7)
English	51.0 (1.9)	-	-	6.7 (1.8)	-	3.0 (1.0)
Dutch	50.0 (1.8)	-	-	5.3 (1.7)	-	5.0 (1.6)

Table 18
Mean correct no-response latencies (standard errors in parentheses) as a function of Filler Type, Trial Type, Word Type, and Trial Block in Experiment 6.

Trial Block A

Filler Type	AM (List 1)		TM (List 2)		CM (List 2)	
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial
None	-	840 (22)	839 (22)	-	866 (25)	-
English	-	846 (31)	848 (30)	-	865 (34)	-
Dutch	-	883 (29)	888 (32)	-	940 (38)	-

Trial Block B

Filler Type	AM (List 2)		TM (List 1)		CM (List 1)	
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial
None	897 (27)	-	-	789 (24)	-	796 (24)
English	867 (24)	-	-	782 (24)	-	779 (26)
Dutch	905 (32)	-	-	799 (27)	-	762 (21)

Appendix C *(continued*

Table 19
Mean percentages of false positives (standard errors in parentheses) as a function of SOA, Filler Type, Trial Type, Word Type, and Trial Block in Experiment 7.

Sound earlier than Print (SOA 1)

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
Filler Type	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	-	13.3 (4.7)	20.8 (4.5)	-	0.8 (0.8)
Dutch	-	9.2 (3.6)	17.5 (3.0)	-	3.3 (1.9)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
Filler Type	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	46.7 (5.0)	-	-	2.5 (1.8)	6.7 (3.6)
Dutch	40.8 (4.8)	-	-	5.0 (1.9)	5.0 (2.3)

Sound simultaneously with Print (SOA 2)

Trial Block A

	AM (List 1)		TM (List 2)		CM (List 2)
Filler Type	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	-	14.2 (3.4)	15.8 (2.9)	-	4.2 (2.6)
Dutch	-	11.7 (3.2)	17.5 (2.8)	-	4.2 (1.9)

Trial Block B

	AM (List 2)		TM (List 1)		CM (List 1)
Filler Type	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	50.0 (4.8)	-	-	10.0 (4.9)	5.0 (2.3)
Dutch	45.8 (4.0)	-	-	5.0 (1.9)	3.3 (1.4)

Table 19 *(continued)*

Sound later than Print (SOA 3)

Trial Block A

Filler Type	AM (List 1)		TM (List 2)		CM (List 2)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	-	4.2 (1.9)	14.2 (2.6)	-	0.8 (0.8)
Dutch	-	10.8 (2.6)	15.8 (3.4)	-	3.3 (1.4)

Trial Block B

Filler Type	AM (List 2)		TM (List 1)		CM (List 1)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	46.7 (3.1)	-	-	1.7 (1.1)	2.5 (1.8)
Dutch	44.2 (4.2)	-	-	0.8 (0.8)	2.5 (1.3)

Table 20
Mean correct no-response latencies (standard errors in parentheses) as a function of SOA, Filler Type, Trial Type, Word Type, and Trial Block in Experiment 7.

Sound earlier than Print (SOA 1)

Trial Block A

Filler Type	AM (List 1)		TM (List 2)		CM (List 2)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	-	705 (32)	696 (50)	-	657 (31)
Dutch	-	724 (42)	672 (32)	-	633 (32)

Trial Block B

Filler Type	AM (List 2)		TM (List 1)		CM (List 1)
	Catch Trial	No-Match Trial	Catch Trial	No-Match Trial	No-Match Trial
English	709 (79)	-	-	710 (47)	685 (38)
Dutch	714 (56)	-	-	701 (36)	693 (40)

***Table 20** (continued)*

Sound simultaneously with Print (SOA 2)

Trial Block A

Filler Type	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
English	-	827 (45)	871 (34)	-	742 (31)
Dutch	-	824 (61)	854 (62)	-	740 (51)

Trial Block B

Filler Type	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
English	933 (80)	-	-	782 (41)	756 (30)
Dutch	873 (62)	-	-	817 (61)	758 (55)

Sound later than Print (SOA 3)

Trial Block A

Filler Type	AM (List 1) Catch Trial	AM (List 1) No-Match Trial	TM (List 2) Catch Trial	TM (List 2) No-Match Trial	CM (List 2) No-Match Trial
English	-	1207 (47)	1214 (49)	-	1058 (26)
Dutch	-	1171 (44)	1270 (57)	-	1119 (56)

Trial Block B

Filler Type	AM (List 2) Catch Trial	AM (List 2) No-Match Trial	TM (List 1) Catch Trial	TM (List 1) No-Match Trial	CM (List 1) No-Match Trial
English	1236 (41)	-	-	1111 (39)	1113 (39)
Dutch	1248 (78)	-	-	1137 (49)	1116 (42)

Samenvatting (Dutch Summary)

Dynamiek van fonologisch coderen in tweetalige visuele woordperceptie

De rol van fonologie in het leesproces

Dit proefschrift gaat over de complexe informatiestroom in het brein die ontstaat zodra we een geschreven woord zien. Visuele woordperceptie, zoals dit mentale proces wordt genoemd, is een belangrijk onderdeel van het lezen en waarschijnlijk een van de meest onderzochte processen in de experimentele psychologie. In het hier gerapporteerde onderzoek wordt ingezoomd op wat kan worden beschouwd als de basis van woordverwerking: *fonologisch coderen*. Dit betreft een geautomatiseerd proces waarbij het leessysteem de visuele vorm van een woord transformeert naar een klankstructuur. De rol van dit proces in visuele woordperceptie blijkt veel belangrijker dan tot voor kort werd gedacht. Volgens recente inzichten moet fonologisch coderen niet als bijproduct worden gezien, maar juist als een primaire functie van het leessysteem. Deze gedachte komt formeel tot uitdrukking in de fonologische coherentie hypothese en, meer algemeen, in de sterke fonologische theorie van visuele woordherkenning. Kort gezegd wordt hierin gesteld dat het aanmaken van een klankstructuur zich snel, automatisch en vroeg in het woordperceptieproces voltrekt. Met andere woorden, zodra een geschreven woord het netvlies raakt lanceert ons leessysteem onmiddellijk en onverbiddelijk het proces van fonologisch coderen.

Ambiguïteit in spelling-naar-klank relaties

Het automatische en dwingende karakter van fonologisch coderen komt fraai tot uiting bij het lezen van zogenaamde spelling-naar-klank inconsistente woorden. Dit zijn woorden waarbij de relatie tussen spelling en uitspraak niet eenduidig is. De Engelse orthografie is berucht om dit soort woorden. Het kan in het Engels namelijk gebeuren dat dezelfde spellingcluster in verschillende woorden op een andere manier wordt uitgesproken. Neem bijvoorbeeld de spellingcluster *–OOD*. Deze klinkt in het Engelse woord MOOD totaal anders dan in het buurwoord BLOOD. Het kenmerkende van een inconsistent woord is dus de *meervoudige* spelling-naar-klank relatie, waarmee het woord dubbelzinnig in uitspraak is. Onderzoek heeft aangetoond dat de verwerking van inconsistente woorden relatief stroef verloopt. Dit staat bekend als het *consistentie-effect*. Gevonden is dat, vergeleken met eenduidige Engelse woorden zoals MOON, het lezers veel tijd kost om inconsistente woorden op te lezen en bovendien worden er meer fouten bij gemaakt. Het consistentie-effect kan worden verklaard door aan te nemen dat, als gevolg van de onvermijdelijkheid van fonologisch coderen, een ambigue spellingcluster zoals *–OOD* wordt omgezet in alle klankstructuren waar het ooit in verschillende woorden mee is geassocieerd. Dus het lezen van een woord als BLOOD brengt met zich mee dat additioneel de klankstructuur van de slotlettergreep van MOOD wordt gegenereerd. BLOOD is een lastig te verwerken woord omdat de simultaan gegenereerde correcte en incorrecte fonologische structuren met elkaar in competitie gaan, en deze competitie moet worden opgelost.

Tussentalige spelling-naar-klank relaties

Ten grondslag aan deze studie is het gegeven dat veel tweetaligen leesvaardig zijn in de tweede taal en dat meervoudige spelling-naar-klank relaties zich ook *tussen* twee orthografieën kunnen voordoen: dezelfde spellingcluster kan in twee talen verschillend worden uitgesproken. Neem opnieuw de spellingcluster *–OOD*, deze klinkt in het Nederlandse woord LOOD alweer heel anders dan in de Engelse buurwoorden MOOD en BLOOD. De tussentalige ambiguïteit van spelling naar klank kan verrassende implicaties hebben, zowel voor het leesproces in de tweede taal als voor woordperceptie in het algemeen. Als fonologisch coderen inderdaad een automatisch proces is dan kan worden verondersteld dat dit ook voor tweetalige visuele woordperceptie geldt. Voor iemand met het Nederlands als moedertaal die Engelse woorden leest zou dat als consequentie hebben dat bij een tussentalig ambigu woord als MOOD additioneel de klankstructuur van de slotlettergreep van een Nederlands woord als LOOD wordt gegenereerd. Een dergelijk proces, *simultaan tussentalig fonologisch coderen*, mag als een opmerkelijk verschijnsel worden beschouwd. Het zou niet alleen impliceren dat bij tweetalige woordperceptie fonologie evenzeer een centrale rol speelt, maar ook dat de aard van fonologisch coderen nog dwingender is dan we tot nu toe beseften. Het additioneel genereren van fonologische structuren die zijn gebaseerd op spelling-naar-klank relaties uit de irrelevante taal strookt in algemene zin met een overheersende opvatting over tweetalige woordverwerking. Deze wordt verwoord door de zogenaamde 'language non-selective access hypothesis'. Volgens deze hypothese beschikken tweetaligen over een enkel geïntegreerd taalsysteem en zijn zij niet in staat om bij woordverwerking kennis van de irrelevante taal uit te schakelen. Dus in principe zou kunnen worden gesteld dat de notie van simultaan tussentalig fonologisch coderen een synthese vormt van enerzijds het denkbeeld van de fundamentele rol van fonologie in woordperceptie en anderzijds die van niet-taalselectieve woordverwerking.

Kort overzicht van dit proefschrift

In deze studie wordt de dynamiek onderzocht van het proces van fonologisch coderen bij het verwerken van Engelse woorden door Nederlands-Engels tweetaligen en door personen die het Engels als moedertaal hebben. Het eerste hoofdstuk biedt een literatuuroverzicht van het recente empirische onderzoek naar visuele woordperceptie en de daarmee samenhangende theoretische ontwikkelingen. Hierin worden twee onderzoeksgebieden bijeengebracht, namelijk die van de rol van fonologie in visuele woordperceptie en die van tweetalige woordverwerking. Het effect van meervoudige spelling-naar-klank relaties op het leesproces wordt besproken vanuit het perspectief van dynamische systemen theorie, waarbij connectionistische inzichten en resonantie-theorie een centrale rol spelen. In deze bespreking wordt het idee van het automatische en dwingende karakter van fonologisch coderen doorgetrokken naar het domein van tweetalige visuele woordperceptie. Hierdoor wordt het mogelijk de notie van niet-taalselectieve woordverwerking vanuit een andere invalshoek te bekijken. Dit blijkt van belang wanneer de invloed van de relatieve prominentie van de twee talen op het leesproces nader wordt beschouwd.

Het uitgangspunt van deze studie is dus de hypothese dat fonologie een primaire rol speelt in visuele woordperceptie. De vraag of het proces van fonologisch coderen ook in tweetalige woordperceptie een automatisch en dwingend karakter heeft wordt in Hoofdstukken 3, 4 en 5 experimenteel onderzocht. Hoofdstuk 3 rapporteert een woordbenoemingsexperiment (Experiment 1) waarin de taakprestaties van zowel Nederlandse als Amerikaanse proefpersonen bekeken worden. In Hoofdstuk 4 wordt een nieuw laboratoriumexperiment geïntroduceerd, de spelling-naar-klank correspondentietaak, waarmee in meer detail het proces van fonologisch coderen kan worden bestudeerd. Dit hoofdstuk rapporteert een serie van vier samenhangende experimenten, waarin de invloed van meervoudige spelling-naar-klank relaties binnen de Engelse orthografie centraal staat. In Experimenten 2, 3 en 4 namen alleen Nederlandse proefpersonen deel, in Experiment 5 zowel Nederlandse als Amerikaanse.

Hoofdstuk 5 beschrijft de Experimenten 6, 7 en 8. In dit hoofdstuk neemt de opvatting dat tweetalige woordperceptie zich niet selectief voltrekt een belangrijke plaats in, waarbij de vraag centraal staat of Nederlands-Engels tweetaligen spelling-naar-klank kennis van beide talen inzetten. Dit betreft, zoals gezegd, de kwestie van simultaan tussentalig fonologisch coderen in tweetalige woordperceptie, als gevolg van meervoudige (tussentalige) spelling-naar-klank relaties. Hoofdstuk 6, tenslotte, vat de belangrijkste resultaten van de experimenten samen en bepreekt de theoretische implicaties ervan. Daarbij wordt ingegaan op het principe van multistabiliteit in dynamische systemen, het aanverwante idee dat fonologisch coderen een metastabiel proces is en worden recente modellen van visuele woordperceptie geëvalueerd met betrekking tot de plaats die zij aan fonologie toekennen.

Veronderstellingen, voorspellingen en typerende bevindingen

De experimenten die in dit project zijn uitgevoerd maken vanaf Hoofdstuk 4 gebruik van de *spelling-naar-klank correspondentietaak*. In deze laboratoriumtaak krijgen proefpersonen op een computerscherm een Engels woord zoals MOOD te zien en tegelijkertijd horen ze over een koptelefoon de slotlettergreep *(–OOD*) uitgesproken. Deze slotlettergreep kan correct, maar ook incorrect worden uitgesproken. In het laatste geval hoort een proefpersoon in feite de slotlettergreep van een ander woord uitgesproken, bijvoorbeeld de spellingcluster *–IDE* van het Engelse woord BRIDE. De taak van de proefpersonen is om snel te beslissen of de visuele stimulus en de gesproken stimulus met elkaar in overeenstemming zijn. Dit doen ze door een ja- of nee-knop in te drukken.

Vrienden en vijanden uit de buurt. In de experimenten maakten we gebruik van twee algemene woordcategorieën. De woorden die de proefpersonen te zien kregen waren ofwel consistent (b.v. MOON), ofwel, als gevolg van meervoudige spelling-naar-klank relaties, inconsistent (b.v. MOOD). Een woord zoals MOON is spelling-naar-klank consistent omdat alle buurwoorden “vrienden” zijn. Dit wil zeggen dat alle Engelse woorden die eindigen op de spellingcluster *–OON* op elkaar rijmen. Een spelling-naar-klank inconsistent woord zoals MOOD, daarentegen, heeft niet alleen “vrienden” maar ook “vijanden”. De bevriende buren van MOOD zijn Engelse woorden waarin de spellingcluster *–OOD* op dezelfde wijze wordt uitgesproken (b.v. FOOD, SNOOD), en de vijandelijke buren van MOOD zijn Engelse woorden waarin deze spellingcluster op een andere wijze wordt uitgesproken (b.v. BLOOD). De buurwoorden MOOD en BLOOD zijn dus allebei spelling-naar-klank inconsistent (en elkaars vijand). Het verschil tussen deze woorden is echter dat het aantal vrienden en vijanden ongelijk is. MOOD schaart zich onder een grote groep vrienden en kent nauwelijks vijanden, terwijl BLOOD juist veel vijanden heeft en amper een vriend. Dit heeft tot gevolg dat voor MOOD het buurwoord BLOOD een zwakke vijand is en andersom dat voor BLOOD het buurwoord MOOD een sterke vijand is. Omdat het woord MOOD meer vrienden dan vijanden heeft wordt deze *typisch* inconsistent genoemd, en daar bij BLOOD het omgekeerde het geval is wordt deze *atypisch* inconsistent genoemd.

Asymmetrische spelling-naar-klank relaties en fonologisch coderen. De meervoudige spelling-naar-klank associaties die door leeservaring ontstaan kunnen bij een inconsistent woord dus asymmetrisch zijn. Bij een typisch inconsistent woord zoals MOOD leidt kennis van vele vrienden zoals FOOD en SNOOD tot een sterke associatie tussen *–OOD* en de correcte klank, en leidt kennis van een futiele vijand zoals BLOOD tot een zwakke additionele (incorrecte) associatie. Omgekeerd leidt bij een atypisch inconsistent woord zoals BLOOD het beperkte aantal vrienden tot een correcte maar zwakke associatie tussen spelling en klank, terwijl kennis van de vele vijanden tot gevolg heeft dat er een additionele incorrecte maar sterke associatie tussen ontstaat. Voor het leesproces heeft deze asymmetrie als consequentie dat, vanwege de sterke vijandelijke spelling-naar-klank associatie bij een atypisch inconsistent woord als BLOOD, het veel moeilijker is om de competitie tussen de correcte en incorrecte fonologische structuren op te lossen dan bij een typisch inconsistent woord.

Experimenteren met MOOD, BLOOD en LOOD. In de hier gerapporteerde experimenten lieten we onze proefpersonen drie typen woorden zien, namelijk atypisch inconsistente woorden (b.v. BLOOD), typisch inconsistente woorden (b.v. MOOD) en consistente woorden (b.v. MOON). Het essentiële verschil tussen de drie woordtypen is dus de mate van vijandigheid van de buren, waarbij BLOOD veel vijanden heeft, MOOD weinig en MOON in het geheel geen. Dit zou tot uiting moeten komen in de prestaties van de proefpersonen op de spelling-naar-klank correspondentietaak. Als we, zoals gezegd, uitgaan van het het idee dat a) fonologisch coderen een automatisch en onvermijdelijk proces is, en dat b) hieruit voortvloeit dat in woordverwerking een spellingcluster die spelling-naar-klank ambigu is wordt omgezet in alle daarmee geassocieerde klankstructuren, dan zou de hevigheid van de competitie die ontstaat bij een woord met veel vijanden leiden tot langere beslissingstijden en veel vergissingen.

Zijn deze verwachtingen uitgekomen? Zoals beschreven moesten de proefpersonen snel beslissen of een visueel aangeboden Engels woord in overeenstemming was met de over de koptelefoon uitgesproken slotlettergreep van dat woord. Nogmaals, in deze taak kunnen twee situaties zich voordoen: de slotlettergreep kan correct worden uitgesproken, of incorrect. In het eerste geval dienen de proefpersonen de ja-knop in te drukken en in het tweede geval de nee-knop. Wat we nu in bijvoorbeeld Experiment 3 vonden was dat wanneer de slotlettergreep van een woord *correct* werd uitgesproken Nederlandse proefpersonen bij atypisch inconsistente woorden zoals BLOOD langzamer de ja-knop indrukten dan bij de andere twee woordtypen. De gemiddelde responstijd was ongeveer 800 ms voor woorden als BLOOD tegen ongeveer 750 ms voor woorden als MOOD en MOON. Bovendien vergisten de proefpersonen zich relatief vaak bij de atypisch inconsistente woorden; in ongeveer 7% van beslissingen werd bij woorden als BLOOD abusievelijk de nee-knop ingedrukt terwijl bij woorden als MOOD en MOON dit bij slechts ongeveer 2% van de beslissingen gebeurde.

In de situatie dat de slotlettergreep van het geschreven woord niet correct maar *incorrect* werd uitgesproken hoorden proefpersonen de uitspraak van de slotlettergreep van een ander, ongerelateerd Engels woord. Ook in deze situatie bleek de tijd die nodig was om de juiste knop in te drukken, in dit geval de nee-knop, het langst bij woorden als BLOOD (ongeveer 800 ms tegen ongeveer 780 ms bij de twee andere woordtypen), en werd bij dit woordtype tevens relatief vaak ten onjuiste op de ja-knop gedrukt. Dit gebeurde in ongeveer 8% van de beslissingen terwijl bij de andere twee woordtypen de proportie fouten beperkt bleef tot ongeveer 4%. Deze resultaten onderbouwen de algemene veronderstelling dat het leesproces wordt beïnvloed door kennis van vijandelijke woordburen. Het laat nadrukkelijk zien dat wanneer meervoudige spelling-naar-klank associaties zijn gebaseerd op ervaring met veel vijanden, er bij het lezen van een inconsistent woord niet alleen de correcte, maar tevens een incorrecte, rivaliserende klankstructuur wordt gegenereerd.

Het meest in het oog springende resultaat verkregen we echter door van een list gebruik te maken. In een specifiek aantal gevallen waarbij de slotlettergreep van een inconsistent woord (b.v. BLOOD) incorrect werd uitgesproken werd niet, zoals daarvoor, de slotlettergreep van een ongerelateerd woord gebruikt (b.v. *–IDE* van BRIDE), maar die van een *vijandelijk woord* (b.v. *–OOD* van MOOD). In deze conditie lieten we de proefpersonen dus een woord als BLOOD zien en werd over de koptelefoon de slotlettergreep van het vijandelijke buurwoord MOOD uitgesproken. Het omgekeerde gebeurde ook. Hierbij werd een woord als MOOD getoond en kregen de proefpersonen de slotlettergreep van het woord BLOOD te horen. Omdat ook in deze situaties de slotlettergreep incorrect wordt uitgesproken is het indrukken van de nee-knop de juiste reactie. Maar wat deden onze Nederlandse proefpersonen? De resultaten waren verrassend. Werden in de vorige conditie op woorden als BLOOD en MOOD nog respectievelijk 7% en 2% fouten gemaakt, in deze nieuwe conditie waren de foutpercentages maar liefst 42% en 24%. Vooral het grote aantal keer dat de proefpersonen bij atypisch inconsistente woorden de ja-knop indrukten is opzienbarend. De resultaten wijzen erop dat proefpersonen in deze taak een woord als BLOOD kortstondig waarnemen als rijmend met de vijand MOOD. Dat dit experimenteel kan worden bewerkstelligd verschaft een specifieke aanwijzing dat het waarnemen van een inconsistent woord gepaard gaat met het onbedoeld genereren van een incorrecte klankstructuur. Dit onderstreept nogmaals het automatische en dwingende karakter van het proces van fonologisch coderen.

Onze Nederlands-Engels tweetalige proefpersonen kregen dus in een specifiek aantal gevallen een spelling-naar-klank inconsistent woord te zien waarbij de slotlettergreep van een Engelse vijand werd uitgesproken. Maar wat zou er gebeuren als we niet een Engelse maar een *Nederlandse* vijand gebruikten? Dat probeerden we uit in de Experimenten 6, 7 en 8. Daarin lieten we Nederlandse proefpersonen weer woorden zien zoals BLOOD, MOOD en MOON en werd opnieuw de slotlettergreep incorrect uitgesproken. In de ene conditie gebeurde dat weer door de slotlettergreep van een ongerelateerd woord te gebruiken (b.v. *–IDE* van BRIDE), en in de andere conditie gebeurde dat door die van een Nederlandse vijand te gebruiken (b.v. *–OOD* van het Nederlandse woord LOOD). In een dergelijk geval lieten we dus de proefpersoon een woord als BLOOD zien en werd over de koptelefoon de slotlettergreep van het Nederlandse woord LOOD uitgesproken. Net als in de eerder beschreven situatie waarin de proefpersonen de slotlettergreep van een ongerelateerd Engels woord of vijandelijk Engels buurwoord uitgesproken hoorden is ook hier weer het indrukken van de nee-knop de juiste reactie. Dit bleek echter helemaal niet zo gemakkelijk. De proefpersonen drukten zeer frequent de ja-knop in, waarmee ze feitelijk aangaven dat in hun waarneming een Engels woord als BLOOD rijmt met het Nederlandse LOOD. Voor woorden als BLOOD gebeurde dit maar liefst in ongeveer 50% van de beslissingen en voor woorden als MOOD in ongeveer 18%. Deze resultaten ondersteunen het idee dat de perceptie van een inconsistent Engels woord zoals BLOOD gepaard gaat met het onbedoeld genereren van een incorrecte klankstructuur die gebaseerd is op kennis van Nederlandse spelling-naar-klank relaties. Dit sluit aan bij de opvatting dat fonologisch coderen in visuele woordperceptie zich in principe niet taalselectief voltrekt. Tot slot, onze bevinding dat het effect van meervoudige spelling-naar-klank relaties op het lezen van woorden uit de eerste taal vergelijkbaar is met die op het lezen van woorden uit de tweede taal pleit voor een algemeen standpunt waarin het proces van fonologisch coderen wordt gezien als fundamenteel voor zowel eentalige als tweetalige woordperceptie. De observatie dat fonologisch coderen bij lezen in de tweede taal wordt beïnvloed door kennis van spelling-naar-klank relaties uit de eerste taal benadrukt het universele en dwingende karakter van fonologisch coderen.

Curriculum Vitae

Martin van Leerdam was born on December 12, 1969 in Curaçao. He received his early education at the Rijksscholengemeenschap in Schagen, The Netherlands. In 1989 he started his study at the University of Amsterdam, where in 1995 he graduated *cum laude* in cognitive psychology. For his Master's thesis, an experimental study of letter perception, he received an academic award from the Department of Psychology of the University of Amsterdam. From November 1996 to October 2002 he was a Ph.D student at the Psychonomy Section of the same department, where he conducted Monte Carlo computer simulations of procedures for response-time outliers, took an interest in the use of interval-estimation in statistical data analysis, and carried out experimental research on bilingual visual word perception. His academic interests further include the field of applied criminology; since December 2002 he works as a researcher at the FIOD-ECD, the Dutch Fiscal Intelligence and Investigation Service.