Language-specific listening:
the case of phonetic sequences

Cover design: Linda van den Akker, Inge Doehring
Cover illustration: Nicola Dorn, Inge Doehring
Printed and bound by Ponsen & Looijen bv, Wageningen

Language-specific listening: the case of phonetic sequences

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op donderdag 28 juni 2001,
des namiddags om 1.30 uur precies

door
Andrea Christine Weber
geboren op 24 augustus 1972 te München, Duitsland

Promotor:	Prof. dr. A. Cutler

Manuscriptcommissie:	Prof. dr. W. Vonk
	Dr. T. Dijkstra
	Prof. dr. W. Strange (City University of New York)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany

Acknowledgments

Finally, the moment has come that I anticipated so happily throughout the Ph.D. project: The work is done and it is time to write the acknowledgments. I owe a lot of thanks to many people who contributed to this thesis in different ways.

First of all, I would like to thank Anne Cutler, my promoter. Her trust in me to complete such a project and her endless 24-hour enthusiasm for science gave me the motivation to keep going. Her door was always open whenever I needed her advise, and especially in the last stages of writing, her immediate corrections were a great help. Thank you Anne!

Next, I would like to thank James McQueen and Natasha Warner for taking over big parts of my supervision. They both provided instant help whenever I came with one of my many questions. I also want to thank James for explaining again my data to me when I was sometimes not so sure about them anymore, for introducing me into the secrets of Shortlist, and also for being there when "China's red lantern" scared me. After about a year, Natasha joined our group and also became a supervisor. Her great knowledge of phonetics and phonology left me sometimes wondering how one can possibly memorize all those details. I also want to thank her for her patient fight against all my passive constructions and many more of my robust mistakes in English.

Delphine Dahan and Anne Pier Salverda I want to thank for introducing me into the methodology of eye tracking. Of course all my Ph.D. colleagues contributed essentially to the success of my thesis, too. They made the stay at the institute very pleasant by sharing with me the daily *Freud und Leid* of a Ph.D. student. Kerstin Mauth was my office mate during the first two years. We not only started a huge collection of teas, but we also "collected" dialectal differences between Bavaria and Rheinland ("Mist, mein *feil* is weg..." "Deine Datei ist weg?" "Nein, mein *feil*!" "Ach, dein Pfeil..."). Arie van der Lugt and Dirk Janssen I also want to thank for introducing me to the Max Planck world and my Unix machine with all the programs that come with it.

The support of the technical group was wonderful. Especially, I want to thank John Nagengast and Johan Weustink for the many times they equipped me to test my participants abroad. They also saved my trip to the US with their support over email and over the phone for weeks, after my equipment was stolen in New York. Just before Christmas they got me up and running again. That was a great Christmas present.

In 1999 Winifred Strange invited me to work with her in New York, where I spent a fantastic half year. Her warm welcome meant a great deal. She gave me the chance to continue my own research and to also learn about vowel perception. In addition, I want to thank her for the opportunity to collect my data in Florida.

To collect data usually meant travelling for me. To get a hold of all my participants and to run the experiments so smoothly would not have been possible without the help of a number of people. Especially, I want to thank Eileen Griffiths, Kanae Nishi, Edeltraud Kienberger and Anja Suchan.

I want to thank Andrea Krott and Simone Sprenger in advance who agreed to be my paranimfen and will stand by me during the defense.

Thanks also to Mirjam Broersma, who translated the Dutch summary that you find at the end of the book.

My parents I want to thank for all their support and for their unquestioning belief and trust in me.

And of course Thomas!

Table of Contents

CHAPTER 1

CHAPTER 2

Chapter 2a

CHAPTER 3

Chapter 3a

Chapter 3b

CHAPTER 4

CHAPTER 5

Introduction

CHAPTER 1

Prelexical processing of speech

Understanding spoken language is a fascinating human cognitive skill. The complex processes that are necessary to get from a highly variable speech signal to the meaning of an utterance are handled by listeners with astonishing ease. Whereas there is an infinite number of possible sentences that a listener might hear, there is presumably only a finite number of words. Models of spoken-language understanding therefore assume that individual words (rather than whole sentences) are stored in the mental lexicon of a listener. A lexical entry can represent individual words, but also groups of word forms, and each lexical entry contains different types of information, ranging from orthographic and phonological information, through morphological, syntactic and semantic information, to pragmatic information. A central component in the understanding of spoken language is therefore to recognize words and to access their lexical entries. The processing of the speech signal that takes place in order to achieve word recognition is called prelexical processing. This thesis will be mainly concerned with the prelexical processing of speech.

One problem of prelexical processing that the listener has to solve is the problem of variability. A given word can vary in its acoustic realization for many reasons. Biological differences in the vocal tract of individual speakers can influence acoustic realizations, as can the age of a speaker, gender, dialect, speech rate, and speech style (e.g., formal or colloquial). Different environments influence the acoustics, too. There is also variation in the speech signal due to coarticulation, because sound segments are constantly influenced in their realization by neighboring segments. For example, the vowel formants in the vowel /a/ differ depending on whether /a/ is preceded by a bilabial stop consonant as in /ba/, an alveolar stop as in /da/, or a velar stop as in /ga/.

Even though the acoustic structure of speech is so variable, it is nevertheless very important for lexical access, because words are after all distinguished from one another via acoustic-phonetic information. To solve the problem of variation most theories assume that some sort of phonetic units are extracted from the acoustic-phonetic information, classifying the speech signal into 'units of perception'. These units of perception allow the mapping of the auditory signal onto stored lexical knowledge. Although most theories agree on the necessity of such mental representations, there is no agreement about their nature. Researchers have argued, inter alia, for acoustic-phonetic features (Eimas & Corbit, 1973), articulatory gestures (Liberman & Mattingley, 1985), phonemes (Fowler, 1984), and syllables (Mehler, Dommergues, Frauenfelder, & Seguí, 1981) as units of perception.

Besides the variation mentioned above there is also a higher level of variation in spoken language, which is called phonologically conditioned variation. One example of phonologically conditioned variation is assimilation, in which a feature of a sound spreads to a neighboring sound. The English prefix *in*, for example, occurs in many English words such as *intolerable*, *incapable*, and *improper*. The nasal in the prefix varies because it takes over the place feature of subsequent stop consonants. In *intolerable* the nasal is realized as alveolar [n], in *incapable* as velar [ŋ], in *improper* as bilabial [m]. Because the place of articulation of the nasal and the subsequent stop consequently match, the place of articulation of the nasal may give already a strong cue about the place of articulation of the subsequent stop. Numerous studies have shown indeed that listeners are sensitive to phonologically conditioned variation in their native language (see for example Costa, Cutler, & Sebastián-Gallés, 1998; Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Otake, Yoneyama, Cutler, & van der Lugt, 1996; van Donselaar, Kuijpers, & Cutler, 1999). They use their knowledge of phonological regularities, and can even profit from phonologically conditioned variation in spoken-word recognition.

Another problem of prelexical processing that the listener has to solve is the problem of segmentation. Whereas in many written languages, spaces unambiguously indicate word boundaries, in spoken language most of the time such clear boundary markers are missing. Speech is rather continuous, but nevertheless listeners have to segment the speech stream into individual words in order to convey the meaning of an utterance. Most theories assume that the segmentation problem is solved by a so-called competition process. Multiple candidate words that match the input are activated by the speech signal and compete with each other for recognition. The input *wild beast*, for example, will activate *wild* and *beast*, but among other alternatives also *why*, *while*, and *bee*. The candidates that can best account for the whole input win the competition (and victory in the competition is the moment of word recognition). But in addition, phonological restrictions and rules can provide information about likely word boundaries. The phonology of a language, for example, also restricts how phonemes may be combined within syllables. Whereas the phoneme sequence /ml/ never occurs within syllables in English, the phoneme sequence /sl/ can occur in syllable-initial position, as in the English word *sleigh*. Thus, sequences that can only occur across syllables in English, like /ml/, mark a syllable boundary and may therefore also cue a possible word boundary. Just as with the variation problem, listeners have been found to use their knowledge of phonological restrictions and regularities to help them solve the segmentation problem and to parse the speech stream into words (see for example Cutler & Norris, 1988; McQueen, 1998; Otake, Hatano, Cutler, & Mehler, 1993; Suomi, McQueen, & Cutler, 1997).

The use of phonological information for word recognition is a subconscious process of which listeners are usually unaware. It is also a process listeners cannot suppress. Most phonological information is language-specific. Every language has its own set of phonemes and phonological rules and restrictions on how to combine them to language-specific phonetic sequences. Listeners learn these rules and regularities when they acquire their native language. However, humans are able to learn to understand and to communicate in more than one language. Because the phonologies of different languages are never the same, different languages often contradict each other in the restrictions they make about speech. There is ample evidence that listeners make use of phonological rules and regularities of their native language for word recognition in that language. However, what role does the native phonology play in spoken-word recognition in a second, non-native language? Can listeners suppress the influence of native phonology when they are listening to a non-native language, with a phonological structure that differs from the listeners' native language?

First, assimilation rules can differ between languages. In German, for example, there is a fricative assimilation rule, due to which the velar fricative [x] occurs after back vowels (e.g., German *lacht*, 'laughs', is realized as [laxt]), and the palatal fricative [ç] occurs after front vowels (e.g., German *Licht*, 'light', is realized as [lɪçt]), but Dutch has no such assimilation rule. For native German listeners, spoken Dutch may therefore violate a native fricative assimilation rule (e.g., Dutch *licht*, 'light', is realized as [lɪxt]). Are listeners sensitive to the violation of native assimilation rules in a non-native language?

Second, phonological cues to word boundaries, such as constraints on syllable structure, can differ between languages. Whereas for example the sequence /sl–/ is a syllable onset in English (e.g., *sleigh*), the sequence does not occur within syllables in German. Thus, it marks a syllable boundary in German but not in English. Are listeners sensitive to the violation of native phonotactic constraints in a non-native language, just as they might be sensitive to the violation of native assimilation rules? And do listeners rely on native phonological cues for likely word boundaries in the segmentation of a non-native language even though these cues may be harmful for the recognition process, or do they solely make use of non-native cues to locate word boundaries?

Third, if someone is listening to a non-native language, not only different candidate words in that language will match the speech signal and will therefore get activated, native words can also be phonologically similar to the non-native speech input. For a native Dutch listener who has good knowledge of English, the English word *desk* /dɛsk/ is not only phonologically similar in onset to the English word *debt* /dɛt/ for example, but also the Dutch word *deksel*, 'lid', /dɛksəl/ is similar in onset. Do listeners activate native candidate words during the recognition process of non-native words or can they suppress the activation of native candidates that are irrelevant? The three questions outlined above are the focus of this thesis and will be addressed by comparing the performance of native listeners and non-native listeners during prelexical processing.

Throughout this thesis I will use the terms native and non-native word recognition rather than monolingual and bilingual word recognition. Although the term bilingual in general refers to a person who can speak and understand two languages, there are many degrees of bilingualism. There is no terminological consensus in the literature on whether only a speaker whose proficiency in two languages is comparable to monolingual native speakers is a bilingual or whether also a speaker with a somewhat lower proficiency in one of the languages is a bilingual. Even speakers with minimal knowledge of a second language are sometimes referred to as bilinguals. The proficiency of the participants in a second language varied considerably in this thesis, and none of the participants grew up with two languages; rather, they acquired their knowledge of a second language through secondary education. The term non-native listener seemed to describe all of the participants more neutrally.

Models of native word recognition

Early models of word recognition were developed on the basis of data obtained in reading tasks (e.g., Forster, 1976; Morton, 1969), but were often treated as if they could account also for spoken-language processing. Only later was it realized that the data on visual-word recognition did not necessarily apply to listening, because of the temporal nature of the speech signal. (Speech is distributed in time, whereas writing is distributed in space.) Since the 1970s, however, a number of models have been developed specifically for spoken-word recognition. These models differ one from another in particular in two points. First, they vary in the assumptions they make about the nature of the representations that make contact with the lexicon. Second, the models diverge with respect to the way information flows between the different components of the processing system. Different components are responsible for different processing stages and are ordered from relatively low-level acoustic-phonetic to higher stages in the processing system involving the lexicon. Interactive models not only allow information to flow from lower to higher components but also allow top-down information flow, whereas autonomous models assume that flow of information in one direction, from the bottom up, is sufficient for spoken-word recognition. Below, the three most influential models of spoken-word recognition, the Cohort model, TRACE, and Shortlist are described.

The Cohort model

In the Cohort model (Marslen-Wilson, 1987, 1990; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978), spoken-word recognition is divided into three sub-levels. At the first level, the word-recognition system makes contact with acoustic-phonetic representations of the speech input. During this stage a set of candidate words (the *cohort*) is activated in a strictly bottom-up data-driven manner. The cohort consists of all words that match the beginning of the speech input. Thus /p/ activates, among others, *plastic*, *power*, *prey*, and *pool*. As more information arrives, processing elements that do not match the input drop out of this initial cohort of potential words. If the next sound is an /l/, for example, the cohort is reduced to words beginning /pl/; an incoming /a/ reduces still further and so on until one candidate word remains (e.g., *plastic*). In the later version of the model (Marslen-Wilson, 1987), the all-or-none rejection of candidate words due to early mismatch was changed into downgrading a candidate word's activation rather than removing it entirely from the cohort in order to allow for recognition of mispronounced words. The model later assumed that items are represented featurally, rather than phonemically. Thus, not only items that share initial phonemes, but also items that share initial features are in the cohort. Because a word can diverge from all other existing words before the actual end of the word (at its uniqueness point), the Cohort model assumes that word recognition can be achieved before word offset. For example, because no words other than *February* begin with /fɛb/, those three sounds suffice to reduce the cohort to one word. After recognition of one word, the onset for the next cohort can be anticipated to start at the end of that word (anticipation strategy). On the second level a selection process chooses a single item from the word-initial cohort. Unlike the first level, this selection process is sensitive to different knowledge-driven constraints, including syntactic/ semantic context. Finally, at the highest level the selected lexical items are integrated into the available syntactic/semantic discourse. The fact that the model gives highest priority to word-initial information causes a problem since it cannot fully explain how listeners can recover from errors when the wrong word-initial cohort is activated, because information later in the word is not considered. Because the Cohort model has never successfully been computationally implemented no simulations can be run using it.

TRACE

TRACE (Elman & McClelland, 1988; McClelland 1979, 1987, 1991; McClelland & Elman, 1986) is a connectionist model, based on McClelland and Rumelhart's interactive activation (IA) model of visual-word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). The claim for interactivity in spoken-word recognition is made most strongly in TRACE. There are three levels of processing in the model: the feature level, the phoneme level, and the word level. On each level are processing elements (nodes) with resting activation values. While processing speech, activation values of relevant nodes increase, and when a threshold is reached, activation spreads to connected nodes. Bi-directional connections exist between nodes within each level and at adjacent levels. Each node is represented separately in each one of successive time slices, to mimic the temporal nature of speech. Activation spreads upward from the feature level, while within the feature level inhibition is exerted toward features incompatible with those activated by the input. As higher-level nodes are excited bottom-up, their increasing activation leads them to exert influences from the top down, which in turn increases activation from the bottom up, and so on. Recognition occurs when a node's activation reaches a certain threshold. How long it takes for a word to be recognized depends on its frequency and the number and frequency of similar words in the lexicon. Not only words which match the onset of the input become activated, but also words that match any other portion of the input. For example, the input *flight* /flaɪt/ will activate not only all words starting with /f/, but also all words starting with /l/ and /aɪ/ and /t/. A serious weakness of TRACE is that the size of the lexicon that can be used for simulations is in practice severely restricted, because the full lexicon is involved in the competition process and the entire network is multiplied for every time slice. TRACE cannot handle a realistic estimate of the size of the adult lexicon. This is, however, due to restrictions of the computer hardware, and is not a weakness of the model itself. The model is also unable to identify the nature of mispronunciations. Although a mispronounced word activates the wrong set of phonemes, the system will be unable to tell at the lexical level which phoneme was mispronounced because top-down feedback will correct the errorful information at the phoneme level.

Shortlist

Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997) is a strictly autonomous connectionist model of spoken-word recognition. Word recognition in Shortlist occurs in two distinct stages. First, a set of candidate words is accessed. Bottom-up information alone determines which candidates have a high enough degree of fit with the input to be considered as members of the shortlist. As in TRACE, the candidates can span different portions of the input and are activated by the speech at any moment. The model can therefore recognize words no matter when they begin. However, in contrast to TRACE, only the short-listed candidates are wired into a small interactive activation network, containing only as many connections as are needed for the particular set of words being processed. The candidates compete there for recognition. Only those words that provide an optimal parse of the input win the competition. Given the input *bus ticket* /bʌstɪkət/, for example, the candidates *bus*, *bust*, *stick*, and *ticket*, among others, will be activated. Candidates like *bust* and *stick* will lose the competition, because they inhibit each other (/st/ cannot be part of both words) and they cannot provide a complete parse of the input (e.g., *stick* leaves /bʌ/ and /ət/ unaccounted for). Because the competition stage is limited to a small candidate set, the model can be implemented with a vocabulary of over 25,000 words. Several experimental findings suggest that besides competition between activated words, more explicit segmentation strategies are used by listeners, too. Listeners, for instance, disfavor the parsing of an input that leaves a residue which cannot be a possible word. (A possible word must consist of at least one syllable.) This has been called the Possible Word Constraint (PWC; Norris et al., 1997). For example, it is harder to detect *apple* in the nonsense sequence *fapple* than in *vuffapple*, because it is only in the later that a possible word (rather than a single consonant) is left after segmenting *apple* out of the input (Norris et al., 1997). The Shortlist model has been refined over the past few years in order to accommodate these findings. The competition process is now modulated by the presence of multiple cues to the location of word boundaries in the input. These cues are then used by Shortlist to calculate the bottom-up support for candidate words; for instance, candidate words that violate the PWC with respect to the tentative boundaries are now penalized.

Models of non-native word recognition

Many theories have been proposed to account for how listeners recognize spoken words when they listen to a language which is not their native language. However, only recently have researchers started to express their theories in explicit functional models. As in native word recognition, models for the visual domain preceded models for the auditory domain. Whereas, for example, the BIA (Bilingual Interactive Activation) model is an implemented model for non-native visual-word recognition, there is to date no functioning model for auditory non-native word recognition. The only attempt so far to model non-native spoken-word recognition is BIMOLA (Bilingual Model of Lexical Access), but this model is still in development. Below, the BIA model and BIMOLA are described.

The BIA (Bilingual Interactive Activation) model

The BIA model (Dijkstra & van Heuven, 1998; Dijkstra, van Heuven, & Grainger, 1998) is an implemented model for non-native visual-word recognition and is based on the visual IA (Interactive Activation) model of McClelland and Rumelhart (1981).[1] The BIA model consists of four layers of nodes: letter features, letters, words, and languages. There are connections between nodes at each level as well as between nodes of different levels. On the word level, words from different languages are represented in an integrated lexicon. The BIA model is nonselective in the sense that initially word candidates from both languages are activated. Then, lateral inhibition between words and top-down feedback from words to letters suppress non-target candidate words. In addition, the language level nodes modulate the lexical activity of the two languages. Word nodes activate their language nodes, and language nodes send top-down inhibition to word nodes from the other language. The language nodes thereby collect activation of all words from one lexicon and suppress all words in the other lexicon. Dijkstra and van Heuven (1998) argue that the relative activity of the language nodes can be set dependent on the situation the listener is in. For example, in an experimental situation, listeners can preactivate the relevant language node of the experiment because of the language in which the instructions are given, and thereby give the lexical entries from that language a boost. Similarly, listeners may preactivate the language node which corresponds to the language their partner in a conversation is speaking. Because Dijkstra et al. (1998) found that different experimental tasks can induce a change in response patterns, they suggested that the BIA model should allow variation in parameter settings pertaining to decision criteria which depend, for example, on general task demands.

BIMOLA (Bilingual Model of Lexical Access)

BIMOLA (Grosjean, 1988, 1997; Léwy & Grosjean, in preparation) is the only attempt so far to model spoken-word recognition in a non-native language, but the model is still in development. Grosjean wants the model to account for both situations in which only one language is relevant, although the listener knows two languages (monolingual mode), and situations in

---

[1] *The IA model provides a theoretical framework that has been used in a number of models, including TRACE.*

which language mixing takes place (bilingual language mode). BIMOLA assumes two language networks that are both independent and interconnected at the same time (Grosjean, 1997). This assumption is, however, based on evidence coming from the production of speech rather than from perception. Grosjean argues that language networks are independent, because a speaker who is highly proficient in two languages is perfectly able to speak in one language only. Second, he proposes that language networks are interconnected, because a speaker who is highly proficient in two languages can switch between the two languages quite readily when speaking with interlocutors who also know the two languages well. Like TRACE, BIMOLA distinguishes a feature level, a phoneme level, and a word level. The feature level in BIMOLA is common to both languages and uses binary, ternary, and multivalued features to define a metric space of phonemes. Both phoneme and word level each have two independent subsets for the two languages. These subsets for the two languages are then enclosed in one larger subset. BIMOLA assumes not only inhibition in the phoneme and word levels but also excitation. Furthermore, instead of having language nodes, the model incorporates top-down preactivation via which external information about the listener's language mode and higher linguistic information activate words of the appropriate lexicon. Because BIMOLA is still in development, it remains to be established whether this model could simulate results of human listeners performing spoken-word recognition in a non-native language.

Differences between the models of word recognition

For native spoken-word recognition two interactive models (the Cohort model and TRACE) and one autonomous model (Shortlist) were described. Both the Cohort model and TRACE assume that items are represented featurally, rather than phonemically. However, whereas the Cohort model gives highest priority to word-initial information, in TRACE also words which match any other portion of the input are activated. There is also variation in the two models on how much top-down information flow is allowed. Whereas the Cohort model allows top-down information flow only on one level, TRACE allows top-down information flow on all three levels. The autonomous Shortlist model is a strictly bottom-up model, which assumes that items are represented phonemically. As in TRACE, in Shortlist candidate words are activated by the speech at any moment. The only two models on non-native word recognition (BIA and BIMOLA) are both interactive models. Whereas the BIA model is developed for non-native visual-word recognition, BIMOLA is developed for auditory non-native word recognition. BIA assumes letter features as contact representations whereas BIMOLA assumes sound features. The BIA model allows top-down information flow from the language nodes level to the word level, as well as from the word level to the letter level. BIMOLA allows top-down information flow from what is called higher linguistic information to the word level, as well as from the word level to the phoneme level. A major advantage of TRACE, Shortlist, and the BIA model is that these models have been computationally implemented. All three models have been used to simulate successfully a large body of experimental data. Note that the list of word recognition models described above is not exhaustive. For example, a model similar to the Cohort model was developed by Cole and Jakimik (1978, 1980). Another model of spoken-word recognition is the Neighborhood Activation Model (NAM), developed by Luce and Pisoni (1998).

Methodologies

In the last three decades, various experimental methods have been developed to investigate different aspects of how human listeners process both native and non-native speech. The methods differ for instance with respect to the amount of meta-linguistic processing they require, and whether they put the listener under time pressure or not. For example, some paradigms require judgments from the listener about the speech signal (off-line tasks). Other paradigms provide more insight into the on-line processing of speech by forcing speeded reactions of the listener to the speech signal (reaction time experiments). In some paradigms the listener has to perform only one task, while other paradigms involve more than one task (a dual task), for example, listening to speech and also detecting predetermined sounds. The following three paradigms are particularly relevant for this thesis: phoneme monitoring, word spotting, and eye tracking.

Phoneme monitoring

In 1969, Foss introduced the phoneme-monitoring task to the field of psycholinguistics (Foss, 1969). Phoneme monitoring is a dual-task paradigm, in which participants are asked to listen to speech and to detect predetermined target sounds. Listeners monitoring for /t/, for example, have to press a button as quickly as possible when they detect a /t/ in any stimulus, and reaction times are measured. Reaction times are assumed to reflect variations in speech processing, which means that longer reaction times are associated with greater processing load. The present thesis (Chapter 2) uses the generalized phoneme monitoring procedure (Frauenfelder & Seguí, 1989; Seguí & Frauenfelder, 1986), in which the target sound can occur anywhere in the stimulus, rather than at a prespecified position. Phoneme monitoring has featured in many studies of prelexical processing, and can easily be applied across languages, because listeners do not have to be highly proficient in the non-native language in order to perform phoneme monitoring in that language. For an overview of the paradigm see Connine and Titone (1996).

Word spotting

In 1988, Cutler and Norris introduced word spotting as an experimental paradigm (Cutler & Norris, 1988). In a word-spotting experiment, participants are asked to detect any embedded real words in spoken nonsense contexts. Reaction times and miss rates are the dependent measures. Embedded words can in principle occur anywhere in the nonsense sequence, in initial, internal, or final position. Listeners do not know what the embedded words are in advance. For instance, given *vuffaple*, listeners should detect *apple* and press a button as quickly as possible once they have spotted it; then they have to say *apple* aloud. The task was designed to study the segmentation of continuous speech, because it requires listeners to segment words out of nonsense contexts. Word spotting can only be applied across languages when the listeners are highly proficient in the non-native language, since they can only spot words they know and recognize quickly in that language. For an overview of the paradigm see McQueen (1996). The experiments in Chapter 3 of this thesis use word spotting.

Eye tracking

Despite an early study by Cooper (1974), detailed exploration of the eye-tracking paradigm for spoken-word recognition has only recently begun (see for example Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In an eye-tracking experiment, participants receive spoken instructions to click on pictures of objects on a screen using a computer mouse (e.g., "click on the apple"), while their eye movements are monitored. The picture that is being mentioned in the instructions is presented along with distractor pictures. It is assumed that the probability of fixating a picture reflects the activation of the name that is associated with the picture. Locations and latencies of eye movements provide information about lexical access in spoken-word recognition. In contrast to both phoneme monitoring and word spotting, eye tracking does not require listeners to make an overt decision about what they have heard. For an overview of the paradigm see Tanenhaus and Spivey-Knowlton (1996). In the present thesis, eye tracking was used in the experiment in Chapter 4.

Structure of the thesis

Chapter 2 reports a number of Dutch and German phoneme-monitoring experiments that investigate how violation of different assimilation rules and violation of phonotactic constraints affects the processing of spoken language by both native and non-native listeners. Chapter 3 describes word-spotting experiments in which both English and German listeners were presented with English speech stimuli. The experiments address the problem of word segmentation in a non-native language, and the use of native and non-native phonotactic cues to word boundaries. Chapter 4 reports an eye-tracking experiment in which Dutch listeners were presented with English spoken instructions to click on pictures of objects on a screen. The experiment investigates the activation of native candidate words during non-native word recognition. Chapter 5 highlights and ties together the main results of Chapters 2, 3, and 4. In addition, the results of simulations of some non-native experimental data with Shortlist are described, and resulting implications for models of non-native spoken-word recognition are discussed.

Help or hindrance: How violation of different assimilation rules affects spoken-language processing[2]

CHAPTER 2A

Abstract

Four phoneme-detection studies tested the conclusion from recent research (see Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996; Quené, van Rossum, & van Wijck, 1998) that spoken-language processing is inhibited by violation of obligatory assimilation processes in the listeners' native language. In Experiment 1, native listeners of German detected a target fricative in monosyllabic Dutch nonwords, half of which violated progressive German fricative place assimilation. In contrast to the earlier findings, listeners detected the fricative more quickly when assimilation was violated than when no violation occurred. This difference was not due to purely acoustic factors, since in Experiment 2 native Dutch listeners, presented with the same materials, showed no such effect. In Experiment 3, German listeners again detected the fricative more quickly when violation occurred in both mono– and bisyllabic native nonwords, further ruling out explanations based on non-native input or on syllable structure. Finally, Experiment 4 tested whether the direction in which the rule operates (progressive or regressive) controls the direction of the effect on phoneme-detection responses. When regressive German place assimilation for nasals was violated, German listeners detected stops more slowly, exactly as had been observed in previous studies of regressive assimilation. It is argued that a combination of low expectations in progressive assimilation and novel popout causes facilitation of processing, whereas not fulfilling high expectations in regressive assimilation causes inhibition.

---

[2] *A slightly adapted version of this chapter appeared in Language and Speech (Weber, 2001).*

Introduction

The language input listeners have to process is far from consistent, not least because incoming continuous speech is subject to many phonological adjustment processes. Variation in length of phonemes, vowel reduction, elision of vowels, consonant and vowel epenthesis, reduction of consonant clusters, varying position of word stress, and assimilation all occur constantly in spoken language. Despite all the variability, native listeners have little trouble in understanding spoken language. They can accommodate to and even profit from rule-bound variation.

The present study focuses on the role in spoken-language processing of one type of phonological rule, namely assimilation. Assimilation is the process by which an inherent feature in a sound segment is altered under the coarticulatory influence of neighboring segments. The direction of assimilation can be regressive (i.e., a later segment affects an earlier one) or progressive (i.e., an earlier segment affects a later one) and is always an adjustment of the sound segment to its context. Rules of assimilation can be either optional or obligatory. If a rule is optional, both realizations, the adjusted and the unadjusted, are legal. If a rule is obligatory, there is only one legal standard realization. The English phrase *ten bikes* offers a site for optional regressive assimilation: The nasal can be realized in colloquial speech with a bilabial segment as [tɛm baɪks] and in more careful speech with an alveolar nasal as [tɛn baɪks]. The place feature of the bilabial stop can be spread to the preceding nasal. An example of optional progressive assimilation is found in the two possible realizations of the German word *leben*, 'live', either as [le:bm̩] or as [le:bn̩] in more careful speech. In this case, the place feature of the bilabial stop can be spread to the following nasal. In contrast, regressive place assimilation for nasals is obligatory in Japanese. The Japanese morpheme *san*, 'three', occurs in many compound words: *sangatsu*, 'March', *sanban*, 'third', *sanju*, 'thirty'. In the first of these, the final nasal of the first syllable is realized as velar [ŋ], in the second as bilabial [m], in the third as dental-alveolar [n]. Place of articulation of the nasal differs as a function of the place of articulation of the following segment.

Recently a number of studies have investigated assimilation in Dutch, English, and Japanese, via phoneme detection or word recognition tasks (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van

Donselaar, in preparation; Otake et al., 1996; Quené et al., 1998). Those studies have shown a highly consistent pattern of results even though they were testing different languages and different assimilation rules and were using different experimental tasks. The results showed that spoken-language processing is neither facilitated nor interfered with by optional assimilation, but is inhibited by violation of obligatory assimilation.

For English, Koster (1987) investigated optional regressive place assimilation. Regressive place assimilation in English occurs optionally between morphemes in connected speech. The alveolar stop /t/ in *sweet girl* is either maintained or takes over the velar feature of the subsequent stop. The listeners' task was to press a button as soon as they heard the velar stop /g/ in [gɜl] (*girl*).[3] Koster did not find any evidence for exploitation of place cues in assimilated words: Listeners were equally fast in detecting /g/ in [gɜl] after [swit] or [swik]. Thus, if assimilation is optional, speed of detection is unaffected by whether or not an immediately preceding consonant is assimilated to the target segment.

Gaskell and Marslen-Wilson (1996) investigated effects of place assimilation in English on the recognition of spoken words. They used a cross-modal repetition priming paradigm for their first experiments. English participants listened to sentences that were truncated after a prime word, the last sound of which was either assimilated to the following removed word or not. If, for example, the prime word *wicked* was originally followed by the word *prank*, *wicked* was realized in one case as [wɪkɪb] and in a second as [wɪkɪd]. The word *prank* was cut off after recording. After hearing the truncated sentence, participants had to make a lexical decision on the now visually presented target word *wicked*. Reactions to the visual targets were equally fast after assimilated or unassimilated auditory prime words. In a second experiment, listeners heard the complete sentence where the assimilated form [wɪkɪb] was either followed by *prank* where the feature change in [wɪkɪb] is phonologically viable or by *game* where the feature change is not viable. Visual-word recognition was affected if the assimilated form was followed by the phonologically non-viable context: The lexical decision on the target word *wicked* was delayed. Apparently, recognition of spoken words was not affected by optional assimilation but was impaired by inappropriately applied assimilation.

---

[3] *Throughout Chapter 2, all examples and allophonic variants will be transcribed phonetically. Phonemes will be transcribed in slashes.*

In a follow-up experiment, Gaskell and Marslen-Wilson found similar effects with the phoneme monitoring task (Gaskell & Marslen-Wilson, 1998). They investigated whether processing of the second segment interacted with the presence of the immediately preceding segment. Phoneme detection of the second segment was considerably slower when it occurred in a non-viable context (e.g., /k/ in *fun camp* realized as *[fʌm kæmp]) than when it occurred in an unchanged context (e.g., /k/ in [fʌn kæmp]), but there was no advantage for /k/ following viably assimilated segments ([fʌŋ kæmp]) over unchanged segments. Thus, inappropriate application of assimilation rules significantly slowed processing while lawful assimilation failed to facilitate it.

A similar result obtains for Dutch. Dutch allows optional voicing assimilation across obstruent sequences. Thus, in Dutch the word *kaas*, 'cheese', before *boer*, 'monger', may be realized with a voiced fricative as [kaz.buːr] instead of [kas.buːr]. In a phoneme monitoring task, it was shown that voice assimilation did not facilitate recognition of the subsequent consonant /b/ (Kuijpers & van Donselaar, in preparation). Dutch listeners detected the target segment equally fast in Dutch words whether the preceding segment was lawfully assimilated or unassimilated in an optional assimilation case. On the other hand, when the target is preceded by misapplication of assimilation, detection was significantly slowed. Dutch listeners found it harder to detect the target /p/ in *kaasplank*, 'cheese board', if the fricative was voiced *[kaz.plɑŋk], than /p/ in [kas.plɑŋk] with an unvoiced fricative. The first form is not an assimilation environment, so that voicing in that position is inappropriate and consequently interfered with processing.

In Japanese, assimilation of place for a nasal and a following stop consonant is obligatory. The nasal must be homorganic with the following consonant in words like *tombo*, 'dragonfly', where the moraic nasal is realized as bilabial [m] before the bilabial stop /b/ and in *kondo*, 'this time', where the moraic nasal is alveolar before the stop consonant /d/. In the study of Otake et al. (1996), Japanese listeners responded equally rapidly and accurately to moraic nasals irrespective of their place of articulation. When asked to respond to the following stop, however, the same listeners were sensitive to the violation of the obligatory place assimilation. Their RTs in a phoneme monitoring task using real Japanese words were significantly slower in rule-violating items (heterorganic nasal and following stop consonant) than in lawfully assimilated items (homorganic nasal and following stop consonant).

Phonological adjustment processes, however, differ between languages. Listeners may use the phonological knowledge of their native language when listening to a non-native language. For instance, the place assimilation rule for nasals tested by Otake et al. (1996) in Japanese is optional in certain environments in Dutch. In order to find out how the knowledge of one's native language affects the perception of spoken non-native languages, Otake et al. presented Dutch listeners with the same Japanese materials (which, for the Dutch listeners, were nonwords). These listeners, for whom assimilation of nasal-stop sequences is optional, showed no difference in their detection times for stop consonant targets preceded by nasals matched versus unmatched in place of articulation. For Dutch listeners no violation of their native phonology was involved.

Processing of a non-native language might, however, be influenced by violations of native language phonotactic constraints (even though the sequences are permissible in the language in which they were produced). Obligatory phonological rules of the native language of a listener may be violated when listening to a non-native language because these phonological rules do not apply in the non-native language. Although it can be argued that violation of assimilation rules is not usually encountered in native spoken language, it can occur in a non-native language. This situation is encountered by people learning a foreign language, or hearing someone speak their own language with a foreign accent. Listening to a non-native language that incorporates rules not valid for the native language appears to be no problem (see Otake et al., 1996), but what if the non-native language violates rules that do hold in the native language? Experiment 1 of the present study sought to examine whether phoneme detection during processing of a non-native language is sensitive to the violation of a native assimilation rule.

At the same time two other factors were changed relative to previous studies. The previous studies tested assimilation at least across a syllable boundary (see Otake et al., 1996), some even across a word boundary (see Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation). The reason for this may be that most studies have tested optional assimilation processes, and optional assimilation does not occur (in the languages tested) within a syllable. However, recent research in speech perception suggests that sub-lexical units such as the syllable can be crucial in speech segmentation and recognition (see Cutler, 1995 for a review). Therefore, syllable structure may have had a crucial influence on previous findings. Thus far, how assimilation violation is processed in monosyllables has not been tested. Experiment 1 therefore uses monosyllables. A direct comparison between the processing of assimilation violation in monosyllables and bisyllables follows in Experiment 3.

Another constant factor thus far was the direction in which the assimilation rule operates. Previous experiments have only tested regressive assimilation. In regressive assimilation, a segment has an effect on the preceding segment, whereas in progressive assimilation a segment has an effect on the following segment. Both regressive and progressive assimilations form phonotactically legal segment strings, and violations of both types of assimilation result in phonotactically illegal sequences. But regressive and progressive assimilation contexts differ in the kind of expectations that they can induce in listeners.

In any phonological sequence, the set of possible later segments is always restricted, to greater or lesser extent, by the sequential phonotactics of the language. For any two-segment string, for example, listeners can therefore develop expectations about what the second segment will be on the basis of the information they hear in the first segment. The use of such information as soon as it becomes available, can help the recognition process of spoken language. When listeners develop expectations about what the next segment will be, the incoming segments can then be evaluated against these expectations. A regressive assimilation rule imposes strong constraints on these expectations, by limiting the set of possible continuations in a specific way. Under Japanese regressive place assimilation, for example, if the segment following a bilabial nasal consonant is a stop, it must also be bilabial. A violation of regressive assimilation thus results in a violation of these expectations; a segment that was not a member of the small set of expected continuations is heard. Progressive assimilation, on the other hand, does not act to impose particular limits on the set of possible continuations; instead, it acts to specifically exclude certain continuations. Under German progressive fricative assimilation, for example, the velar fricative is explicitly ruled out after front vowels (see below). Violation of progressive assimilation therefore results in a different kind of violation of the expectations set up by the first sound in a two-segment sequence; a segment that is a member of a small set of impossible continuations is heard.

As already noted, inhibition effects have been shown in previous experiments when listeners' expectations about an upcoming segment were defeated in regressive assimilation violations. Given that progressive assimilation constrains the set of possible continuations in a different way to regressive assimilation, one can ask whether the defeat of the expectations caused by progressive assimilation will also result in an inhibition effect. The present experiments addressed this issue. Experiments 1, 2, and 3 tested a progressive assimilation rule. Experiment 4 tested a regressive assimilation rule.

In the first experiment of the present study, listeners were presented with non-native language input, in which a progressive native assimilation rule was violated within syllables. Two closely related languages were used: Dutch and German. The distribution of the palatal fricative [ç] and the velar fricative [x] in standard German provided the phonological background. Many phonologists, including Trubetzkoy (1939) and Wurzel (1980) have discussed these two allophonic variants, the distribution of which is predictable. The two fricatives stand in complementary distribution in German: The velar fricative [x] occurs after back vowels, the palatal fricative [ç] after front vowels, glides, sonorant consonants, word-initially, and in the diminutive suffix *−chen* (see for example Hall, 1989). Thus, the place of articulation of a vowel specifies the place of articulation for the following fricative (progressive assimilation). The suffix *-chen* is an exception, as preceding vowel quality does not affect the fricative in this case, but the fricative is conditioned by a morpheme boundary. This leads to a few apparent minimal pairs, such as [kuːxən] (*Kuchen*, 'cake') and [kuːçən] (*Kuhchen*, 'small cow'). Whereas German *lacht*, 'laughs', is realized as [laxt] with a velar fricative due to the preceding back vowel, German *Licht*, 'light', is realized as [lɪçt] with a palatal fricative due to the preceding front vowel. It would violate German fricative assimilation to realize the German word *Licht* as *[lɪxt] with a velar fricative and would result in the illegal sequence *[ɪx].

This distribution does not apply for standard Dutch, since the Dutch phoneme repertoire contains only the velar form of the fricative (Booij, 1995). Gussenhoven (1992) has pointed out that in some varieties of Dutch a post-velar or uvular fricative [χ] rather than the velar form occurs. For German, Kohler (1990) found variation between [x] and [χ] after some back vowels. However, the velar fricative [x] is possible in both standard German and Dutch. Dutch *lacht*, 'laughs', and Dutch *licht*, 'light', are both realized with a velar fricative in postvocalic word position as [lɑxt] and [lɪxt], regardless of the place of articulation of the preceding vowel. These different distributions make it possible to use German and Dutch to ask whether the native phonological structure influences processing of a non-native language. When German listeners attend to Dutch, they hear repeated violations of German fricative assimilation. Do German listeners show an inhibition effect (as observed in the other assimilation studies) when they are listening to Dutch items that violate obligatory German progressive fricative assimilation within syllables?

The generalized phoneme monitoring procedure (Frauenfelder & Seguí, 1989; Seguí & Frauenfelder, 1986), in which a predetermined target segment can occur anywhere in the stimulus, was chosen as the experimental task. Phoneme monitoring involves two tasks for the participants, listening to speech and detecting a predetermined target segment (for an overview see Connine & Titone, 1996). The measured response times are assumed to reflect variations in speech processing, where longer RTs are associated with greater processing load.

Two types of phonotactically legal Dutch monosyllables were examined. All were nonwords both in German and in Dutch to avoid any potential lexical effects from cognates across the closely related languages. One type of monosyllable contained a front vowel followed by the velar fricative [x] in penultimate position (e.g., [bɛxt]). The other type contained a back vowel followed by the velar fricative [x] in penultimate position (e.g., [bɑxt]). The nonwords with back vowels were possible sequences in both standard Dutch and German. The nonwords with front vowels violated a German phonotactic constraint, but were legal in Dutch. In the first experiment, German listeners were presented with the Dutch speech stimuli. Their task was to detect the target fricative [x] in the Dutch nonwords. An inhibition effect for violation of the German phonotactic constraint would show that listeners make use of their native phonological structure for phoneme recognition while listening to a non-native language.

Experiment 1: Germans monitoring for [x] in Dutch

Method

Participants. Twenty-four students of the University of Regensburg in Germany were paid to take part in the experiment. They were all native speakers of German and had no knowledge of Dutch.

Materials. A list of 28 monosyllabic items, nonwords in Dutch and German, was selected, and with the help of the CELEX database checked for existing words (Baayen, Piepenbrock, & van Rijn, 1993). All items ended with the velar fricative [x] followed by the stop /t/, having the syllable structure CVxt or CCVxt (such as [hɔxt] and [frɪxt]). This syllable structure is common in both Dutch and German. No phonotactic constraints of either language, except the fricative assimilation in question, were violated in these nonwords. Only phonemes that occur in both languages were used, with one exception: The Dutch nonword *wocht* [ʋɔxt] is realized with a labiodental approximant, while in German it would be realized with a labiodental fricative (Booij, 1995; Wiese, 1996). The labiodental approximant does not occur in German. This small difference was reckoned unlikely to have significant influence on processing. Fourteen of the chosen nonwords contained the front vowels /ɛ/ or /ɪ/, while 14 other nonwords contained the back vowels /a/ or /ɔ/. Only short vowels were used since short vowels predominate in closed syllables in German. Because all items had to be nonwords in both languages that violated no phonotactic constraints other than fricative assimilation, it was not possible to find 14 matched pairs of nonwords that differed only in the vowel. This was taken into account in the statistical analyses. The nonwords are listed in Appendix 2−1, p. 153.

In addition, 308 mono− and bisyllabic filler nonwords, also legal nonwords in both Dutch and German, were selected. Eighty-four of the fillers contained the fricative [x] in a variety of positions in the nonwords. All fillers contained one of the four vowels /ɛ/, /ɪ/, /a/, or /ɔ/. From the complete set of 336 items, four different pseudo-random orders were constructed, with the restriction that for at least two items before a target item, only fillers without the target fricative [x] were used. Fourteen similar practice items were created and presented at the beginning of the experiment. Three pauses were put in the experiment, one after the practice list and two more in the experiment itself, after every 112 items.

Procedure. All materials were recorded onto a DAT tape in a sound-proof booth by a female native speaker of Dutch. The experimental stimulus nonwords were recorded two or three times and the best pronunciation was selected for use in the experiment. Although only phonemes that occur in both languages were used, there are of course phonetic differences especially in vowel quality between the two languages. Two native Dutch speakers listened to the materials and confirmed that they sounded Dutch. Speech stimuli were down-sampled during transfer to a computer to 16 kHz.

Each item was labeled using the Xwaves speech editor. Additionally point labels were put in the experimental items at the beginning of the fricative [x]. Each nonword was then transferred as an individual speech file to the hard-disk of a personal computer. Stimulus presentation, timing and data collection were performed using the NESU (Nijmegen Experiment Set-Up) experiment control software.

German participants were tested one at a time in a sound-proof booth. They were told that they were to listen for the target fricative [x] in a series of Dutch nonwords, and they were instructed to press the button in front of them with their preferred hand as fast as possible if they detected [x] in any of the nonwords. Written instructions were given, telling the participants to respond to the sound represented in orthography as *ch* as in the word *Nacht*, 'night'. In German orthography, both the velar and the palatal fricative are realized as *ch*. To make the task clear, additional oral instructions were given using German example words with the velar fricative only. Response times were measured from the onset of each target nonword. Each participant heard the practice list first, followed, after a short pause, by all experimental stimuli in one of the four pseudo-randomized orders. The experiment lasted approximately 18 minutes.

Results

Prior to statistical analysis, RTs (response times), which were originally measured from the onset of the items, were adjusted so as to measure from the onset of the target fricative [x]. Missed responses were treated as errors. All RTs lay within the range of 100 to 1500 ms. Mean RTs and mean error rates are given in Table 2−1.

Table 2−1. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants in Experiment 1 to the penultimate velar fricative [x] after back or front vowels in monosyllabic Dutch nonwords.*

Measure	Back vowel [bɑxt]	Front vowel *[bɛxt]
RTs	498	470
Errors	0.8%	1.1%

Instead of responding more slowly, German participants detected the fricative [x] 28 ms more quickly when the progressive German fricative assimilation rule was violated than when no violation occurred. Computed RTs were submitted to Analyses of Variance (ANOVAs), with both participants (F_1) and items (F_2) as the repeated measure. The pattern in the RTs was significant by participants ($F_1(1, 23) = 4.82$, $p < .04$). By items, however, the effect did not reach significance ($F_2(1, 26) = 3.01$, $p = .09$).

The reason why the items analysis failed to be significant was found in one of the illegal items. RTs to this particular item were on average 93 ms slower than RTs to the other 13 items in that context. Also this particular item showed the highest standard deviation. There was, however, a particularly slow item like this in the legal context as well. RTs to this item were on average 145 ms slower than RTs to the other 13 items in that context. Again, this particular item showed the highest standard deviation for its context. When both items were excluded, the mean RTs were 463 ms to illegal items and 488 ms to legal items. An analysis resulted in a significant effect of context for participants and items ($F_1(1, 23) = 6.69$, $p < .02$; $F_2(1, 24) = 5.74$, $p < .03$). The low percentage of errors indicates that participants had no problems detecting the target in the two types of nonwords. An error analysis revealed no significant main effect.

Whereas in all earlier studies violation of assimilation resulted in slower RTs (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996), RTs in Experiment 1 were faster to items containing such violations. Why did this facilitation effect occur?

The first and most obvious explanation was that some acoustic factors caused facilitation of processing rather than inhibition. Therefore, an analysis of the duration of the target sound was performed. It is possible that RTs might be influenced by differences in the length of the presented target fricative, as length provides a simple measure of acoustic difference between targets across contexts. For the remaining 26 items (after excluding the two items with particularly slow RTs) the fricative [x] was on average 14 ms shorter after front vowels, with an average length of 177 ms, than after back vowels, with the average length of 191 ms ($t(12) = 1.86$, $p > .07$). A correlation analysis with the RT data showed that there was no tendency for length to be correlated with RTs after both back and front vowels (after back vowels: $r(13) = .30$, $p > .3$; after front vowels: $r(13) = .23$, $p > .4$).

Because in Dutch the fricative does not assimilate to the preceding vowel, the vowel might assimilate to the fricative instead. The velar fricative could have caused the preceding front vowel to have been produced lower and /or further back than elsewhere. This would be apparent in the first and the second formant of the front vowel. A comparison of F1 and F2 in target items such as [pɪxt] and [pɛxt] and fillers with the same vowels but no velar fricative such as [bɪft] and [blɛmp] showed no difference between the vowels before [x] and elsewhere. The formants were measured in the last third of the vowel and an inspection of the means of F1 and F2 for both front vowels in target items and fillers suggested that there was no lowering or backing in the target items compared to the fillers. Thus, vowel quality could not have been a cue for the upcoming fricative.

Another way to test whether the results of Experiment 1 were due to acoustic confounds is to present the materials to participants for whom they violate no rules: Dutch listeners. If the results of Experiment 1 are due to violation of phonological constraints, they should not replicate for participants who lack the constraint. If, however, Dutch listeners showed a difference in their RTs, there might be unintended acoustic differences between the two sets of nonwords. Items had been chosen for Experiment 1 that were nonwords in both languages, German and Dutch. Once again the use of nonwords excluded lexical effects for both Dutch and for German listeners.

Experiment 2: Dutch monitoring for [x] in Dutch

Method

Participants. Twenty-four native speakers of Dutch, students at the University of Nijmegen in the Netherlands, took part in the experiment. They were paid for their participation.

Materials. The same Dutch materials and the same lists as described in Experiment 1 were used.

Procedure. The same procedure as in Experiment 1 was used. The only difference was that for the Dutch participants all the materials were legal Dutch nonwords that contained no phonological violation. The participants were tested one, two, or three at a time in separate sound-proof booths. They were told that they would listen to Dutch nonwords. Instructions were given again in writing and orally using Dutch example words such as *nacht*, 'night', and *geld*, 'money' (the velar fricative can be realized in Dutch orthography both as *ch* and as *g*).

Results

Mean RTs (from onset of the target fricative) and mean error rates are given in Table 2–2. Missed responses and one RT being slower than 1500 ms were treated as errors.

For the Dutch participants, listening to their native language, no phonological violation occurred in the materials. Accordingly, they showed no difference in their RTs between the two types of monosyllabic nonwords. Whether a front vowel or a back vowel preceded the target fricative [x] made a difference of only 8 ms in the mean RTs of these participants. The effect was, as expected, significant neither by participants nor by items (F_1 & $F_2 < 1$). Again, the low percentage of errors indicates that the participants had no problems performing the task. An error analysis revealed no significant main effect.

Table 2‒2. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of Dutch participants in Experiment 2 to the penultimate velar fricative [x] after back or front vowels in monosyllabic Dutch nonwords.*

Measure	Back vowel [bɑxt]	Front vowel [bɛxt]
RTs	531	539
Errors	2.1%	1.4%

Because the RTs of German listeners in Experiment 1 were faster than those of Dutch listeners in Experiment 2, a post hoc analysis was performed to check for the presence of interaction effects of language and context. A two factor mixed ANOVA was used, with language of the listener as the between-participants factor and context, with the two levels front and back vowel, as the repeated measures factor. Neither the main effect of language nor that of context reached significance by both participants and items. The interaction of language and context did not quite reach significance either by participants ($F_1(1, 46) = 3.71, p = .06$) or by items ($F_2(1, 26) = 3.30, p = .08$). A t‒test showed that German listeners' reactions to the velar fricative [x] after a front vowel were significantly faster than Dutch listeners' reactions ($t_1(46) = 2.29$, $p < .03$; $t_2(13) = 5.82$, $p < .001$). After a back vowel no significant difference was found between the participant groups ($t_1(46) = 0.87, p > .3$; $t_2(13) = 2.11, p > .05$).

For all 28 items the fricative [x] was on average 15 ms shorter after front vowels, with an average length of 178 ms, than after back vowels with an average length of 193 ms ($t(13) = 2.14$, $p < .05$). For Dutch participants there was a significant correlation between duration of the target fricative and RT after front vowels ($r(14) = .7, p < .01$); but none after back vowels ($r(14) = .34$, $p > .1$). An ANCOVA on the item RTs using target duration as a covariate, however, still found no effect of context ($F_2(1, 26) = 2.27, p > .1$).

The results of Experiment 2 indicate that there were no anomalies in the materials. RTs of Dutch participants did not differ in responses to nonwords containing a front or a back vowel. It therefore remains to be explained why detection of a target segment that violated a German phonotactic constraint was facilitated for German listeners in Experiment 1 rather than inhibited as it was in previous studies (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996). Three possibilities suggested themselves as sources for the facilitation effect. One was that the listeners were attending to non-native stimuli. The second possibility was the monosyllabic structure of the items. The third possibility was direction in which the tested rule operated. Experiments 3 and 4 examined these possibilities.

Although other studies have addressed how knowledge of the phonological structure of one's native language affects the perception of spoken non-native languages (see for example Cutler, Mehler, Norris, & Seguí, 1986; Otake et al., 1996), they have not looked at the effects of native assimilation rules on foreign language processing. In Otake et al. (1996), for example, Dutch listeners were presented with Japanese materials with place assimilation violations. Although the items violated obligatory Japanese place assimilation for nasals, they did not violate Dutch place assimilation rules (because the rule in Dutch is optional). Experiment 1 is therefore the first experiment in which a native assimilation rule was violated in non-native language materials. So it remains possible that the facilitation observed in Experiment 1 is due specifically to processes that operate when listeners are presented with non-native language. Violation of a native assimilation rule may cause facilitation when listening to a non-native language (as in Experiment 1), but not when listening to one's native language.

Experiment 3 therefore investigated the same German fricative assimilation in a phoneme detection task with native speech stimuli. The German fricative assimilation rule was violated in similar nonwords, this time pronounced in German by a native speaker of German. Only German participants were tested. The question was whether the facilitation effect found in non-native listening for violation of the German fricative assimilation would still be found when participants were listening to their native language. Nonwords were used again for compatibility with Experiments 1 and 2.

Experiments 1 and 2 are the first to test violation of assimilation in monosyllables. To assess whether the findings of Experiment 1 might have been due to the monosyllabic structure of the experimental items, fricative assimilation was tested in both monosyllabic and bisyllabic items in Experiment 3. The monosyllabic items had the same structure as the nonwords used in Experiments 1 and 2. In the bisyllabic items, the target fricative [x] occurred at the onset of the second syllable. So both the number of syllables and the position of the target sound within a syllable changed.

The German fricative assimilation rule applies across a syllable boundary if the first syllable ends in a vowel (i.e., if it is an open syllable). Whereas German *rauchen*, 'smoke', is realized with a velar fricative as [rau.xən] due to the preceding back vowel, German *kriechen*, 'crawl', is realized with a palatal fricative as [kriː.çən] due to the preceding front vowel. The only exception to this rule is the diminutive suffix −*chen*, which is always realized with the palatal fricative [ç]. The diminutive form of *Frau*, 'woman', is therefore realized with a palatal fricative as [frau.çən] even though the first syllable is open and ends with a back vowel. However, if the first syllable ends in a consonant (closed syllable), the second syllable must begin with the palatal fricative [ç] regardless of whether the vowel of the first syllable is back or front. German *horchen*, 'hear', as well as *München*, 'Munich', are realized with the palatal fricative as [hɔr.çən] and [mʏn.çən].

To sum up, Experiment 3 was designed in part to investigate whether processing differences in non-native and native listening caused the facilitation effect for assimilation violation found in Experiment 1. The experiment also addressed whether syllable membership and preceding context influence this processing.

Experiment 3: Germans monitoring for [x] in German

Method

Participants. Twenty-four students of the University of Regensburg took part in the experiment for a small payment. They were all native speakers of German. None of them had participated in Experiment 1.

Materials. The experiment was based very closely on Experiments 1 and 2. Again, a list of 28 monosyllabic items was selected. The same syllable structure and the same vowels as in Experiments 1 and 2 were used. This time the items (as German nonwords) only had to fulfill German constraints on word construction. As before, 14 of the nonwords with violation of assimilation contained the front vowels /ɛ/ or /ɪ/. Fourteen more nonwords with no violation contained the back vowels /a/, /ɔ/, or /ʊ/.

In addition to the 28 monosyllabic items, 28 bisyllabic items were selected. The bisyllabic items were also nonwords of German and also matched German constraints on word construction. The first syllable of half of the items had the closed syllable structure CVC or CCVC, where the vowel was either /ɛ/ or /ɪ/ and the syllable coda either of the sonorant consonants /n/ or /l/ (such as *[plɛnxən] or *[ʃpɪlxən]). The other half of the bisyllabic items had the open syllable structure CV or CCV in the first syllable with /uː/, /oː/, /aː/, or /au/ as its nucleus (such as [bluːxən]). The second syllable of all 28 bisyllabic items was either [xən] or [xər]. Items with a closed first syllable violated a German phonotactic constraint because the velar fricative [x] appeared as a syllable onset following a closed syllable. Items with open first syllables contained no phonological violation. The items, forming 14 matched pairs each for mono− and bisyllabic nonwords, are listed in Appendix 2−2, p. 154.

A total of 254 filler nonwords were added to the materials. The filler material included both mono− and bisyllabic nonwords. In 28 fillers, the target fricative [x] occurred at different positions across the nonwords. The palatal fricative [ç] never occurred. Fillers contained different German vowels, including all vowels used in the target items.

Four different pseudo-randomized orders were constructed from the total set of 310 items. The items were constructed in such a way that each experimental item was preceded by at least one non target-bearing filler. Fourteen representative practice items were additionally created and were presented at the beginning of the experiment. There was a pause between the practice list and the experimental items.

Procedure. All materials were read by a female native speaker of German in a sound-proof booth and recorded on DAT tape. The speaker was also fluent in Dutch, which helped her to produce the velar [x] after front vowels naturally.

All other details were as in Experiment 1 with one exception: Digitized and labelled stimuli were transferred as individual speech files to four pseudo-randomized lists and re-recorded on DAT for presentation. A portable computer with NESU experiment control software was used for the timing and data collection.

Participants were tested one at a time in a sound-proof room. Instructions were given in the same manner as in Experiment 1, except that the participants were told that they would hear German stimuli. Response times were measured from the onset of each target nonword. Each participant heard the practice list first, followed, after a short pause, by one of the four experimental lists. The experiment lasted about 16 minutes.

Results

Prior to statistical analysis, RTs, which were originally measured from the onset of the items, were adjusted so as to measure from the onset of the target fricative [x]. Missed responses and one response to a monosyllabic item, which was slower than 1500 ms, were treated as errors. Mean RTs and mean error rates for the monosyllabic and bisyllabic nonwords are given in Table 2−3.

Table 2−3. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants in Experiment 3 to the velar fricative [x] after back or front vowels in monosyllabic and bisyllabic German nonwords.*

	Monosyllabic		Bisyllabic	
Measure	Back vowel [bɔxt]	Front vowel *[bɪxt]	Back vowel [bluːxən]	Front vowel *[blɪnxən]
RTs	513	488	535	497
Errors	2.1%	1.5%	3.9%	2.4%

As in Experiment 1, German listeners detected [x] more quickly when a phonotactic constraint was violated than when no violation occurred. RTs of monosyllabic and bisyllabic items taken together were 524 ms in the legal context and 492 ms in the illegal context. A combined analysis with both monosyllabic and bisyllabic items showed that the difference in RTs was significant by participants and by items ($F_1(1, 23) = 18.44$, $p < .001$; $F_2(1, 26) = 9.69$, $p = .004$). There was no interaction between number of syllables and context (F_1 & $F_2 < 1$).

When evaluated separately, the difference in RTs for monosyllabic items was 25 ms. ANOVAs showed that this effect was significant by participants ($F_1(1, 23) = 9.60$, $p = .005$) but not by items ($F_2(1, 13) = 2.91$, $p = .1$). For 2 of the 14 monosyllabic pairs of items, average RTs were considerably slower to the illegal item (by 50 ms in one case and 38 ms in the other), contrasting with the overall pattern. When both pairs of items were excluded from the ANOVA, there was a main effect of context for participants and items ($F_1(1, 23) = 16.10$, $p = .001$; $F_2(1, 11) = 6.58$, $p < .03$). The mean RTs after exclusion of these items were 481 ms for illegal items and 519 ms for legal items.

For the remaining 12 monosyllabic pairs of items the duration of the target fricative was measured and was on average 14 ms shorter after front vowels, with an average length of 164 ms, than after back vowels, with an average length of 178 ms ($t(11) = 2.71$, $p < .02$). There was no significant correlation between duration of the fricative and RTs after either back vowels ($r(12) = .24$, $p > .5$) or front vowels ($r(12) = .48$, $p = .1$). An error analysis revealed no significant main effect.

Inspection of the means of the first and the second formant in target items with front vowels and fillers with the same vowels but no velar fricative showed that there was no lowering or backing of the vowels before the velar fricative which could have functioned as a cue for the upcoming target fricative.

In bisyllabic nonwords, listeners detected the target fricative [x] on average 38 ms faster when a phonotactic constraint was violated than when no such violation occurred. The difference was significant both by participants and by items ($F_1(1, 23) = 11.41$, $p = .003$; $F_2(1, 13) = 7.02$, $p = .02$). An error analysis revealed no significant main effect.

Duration of the target fricative in the 14 bisyllabic pairs of items did not contribute to the effect. Average lengths of the target fricative after front and after back vowels were both 131 ms.

Because there were intervening consonants between the front vowels and the velar fricatives in the bisyllabic targets (/n/ or /l/), no lowering or backing of the front vowels was expected and therefore no comparison between first and second formant was done.

Another possibility was that participants learned to detect the anomalous sequence in the course of the experiment. For Experiments 1, 2, and 3 the RTs were split up into four groups depending on the position of the corresponding item in the experiment. No significant interaction was found between position of the corresponding item in the experiment and context, with one exception: In the bisyllabic items in Experiment 3, an interaction suggested that listeners were learning to detect the fricative in the legal context, but not in the illegal context.

Thus, if violation of assimilation occurred, a facilitation effect appeared in native and non-native listening, in mono‒ and bisyllabic items, for target sounds in initial and penultimate position, and for target sounds preceded by either a vowel or a consonant. But so far facilitation for assimilation violation has been found only for German listeners, whereas inhibition was found for Japanese, Dutch, and English listeners. Another difference between the assimilation tested here and those in the previous literature lies in the direction in which the assimilation rule operates. German fricative assimilation operates progressively, whereas earlier experiments tested assimilation rules that operate regressively. Regressive assimilation narrows the number of legal second segments to a small set (few legal continuations), while progressive assimilation does not (many legal continuations) but rather explicitly excludes one certain continuation. Accordingly, listeners can have different kinds of expectations about the upcoming segment. The defeat of the expectations might be processed differently as a result. Is the facilitation effect for violation of assimilation due to the German fricative assimilation being progressive instead of regressive, or is it because the listeners are German? In Experiment 4, a German regressive assimilation rule was tested to investigate this question.

Place assimilation for nasals was chosen as the German regressive assimilation rule. This rule has been tested before for Japanese by Otake et al. (1996). Regressive place assimilation for nasals is seen as the spreading of the place feature of a stop to the preceding nasal (see Wiese, 1996), so that the nasal becomes homorganic with the following stop. Regressive place assimilation for nasals is obligatory within German syllables only for the velar stop /k/ and for the bilabial stop /p/: Thus, German *Bank*, 'bank', must be realized as [baŋk] but not as *[bank] or *[bamk]. German *Lump*, 'rogue', must be realized as [lʊmp] but not as *[lʊŋp] or *[lʊnp].

In Experiment 4, German participants had to detect either the target phoneme /k/ or the target phoneme /p/ in three different types of monosyllabic German nonwords. The experimental items were nonwords again, for comparison with Experiments 1, 2, and 3. They were monosyllabic because place assimilation for nasals in German is only obligatory within syllables. In one type of nonword, the stop was preceded by the correctly assimilated nasal, while in the two other types, the stops were preceded by two different unassimilated nasals. The German phoneme repertoire contains three nasals /n/, /ŋ/, and /m/ (which are distinctive in syllable final position).

Experiment 4: Germans monitoring for /p/ or /k/ in German

Method

Participants. Forty-eight students from the University of Regensburg, all native speakers of German, were tested. They were paid for their participation. Twenty-four of the participants had to detect the target phoneme /k/; the other 24 had to detect the target phoneme /p/.

Materials. A list of 84 monosyllabic items was selected with the help of the CELEX database (Baayen et al., 1993). All items were German nonwords, which were constructed of legal onset consonants and clusters, had a CVCC or CCVCC structure and contained short vowels only. Forty-two of the items ended in the velar stop /k/. In 14 of those, the stop was preceded by the assimilated velar nasal /ŋ/ (such as in [fɛŋk]). Fourteen items violated the nasal place assimilation, in that /k/ was preceded by the alveolar nasal /n/ (such as in *[fɛnk]). Fourteen more contained violations in that /k/ was preceded by the bilabial nasal /m/ (such as in *[fɛmk]). The other half of the 84 items ended in the bilabial stop /p/. In 14 of those nonwords, the stop was preceded by the assimilated bilabial nasal /m/ (such as in [flɔmp]). Fourteen nonwords with the velar nasal /ŋ/ showed violation of assimilation (such as *[flɔŋp]). In the last 14 nonwords, also containing violation, the alveolar nasal /n/ occurred in penultimate position (such as in *[flɔnp]). Both for the nonwords ending in /k/ and for the nonwords ending in /p/ it was possible to find 14 matched triplets (see Appendix 2–3, p. 155).

In addition, 226 fillers were added to the materials, consisting of both mono– and bisyllabic legal nonwords. In 26 items the target phonemes /k/ and /p/ occurred in a variety of positions. In the rest of the fillers the target phonemes did not occur. The three nasals /n/, /m/, and /ŋ/ occurred altogether 150 times in the fillers. The experimental items and the fillers were then used to construct four lists. Each list contained all 310 items in a different pseudo-random order, such that there was always at least one non target-bearing filler before an experimental item. A list of 14 practice items was also constructed with similar materials, which was presented at the beginning of the experiment. As before, there was a pause between the practice list and the experimental list.

Procedure. All nonwords were recorded by the same female native speaker of German as in Experiment 3. The speaker was trained beforehand to produce the illegal sequences */nk/, */mk/, */np/, and */ŋp/ correctly in monosyllabic nonwords. Although experimental items were read two or three times and the best recording was chosen, in some items a short click sound was audible between the nasal and the stop. Epenthetic stops can appear in these environments because of mis-timing of articulators (Ohala, 1995). This happened in four items containing the sequence */nk/, in four more items containing the sequence */mk/ and in five items containing the sequence */ŋp/. In order to avoid the possibility of participants responding to these instead of the actual target stops, these clicks were removed by cutting them out of the nonwords. The spliced utterances were played to two native listeners of German who reported that they sounded as natural to them as unspliced utterances.

After digitizing the speech materials using Xwaves, the nonword boundaries and the release of the stops in the experimental items were labeled. As in Experiment 3, a portable computer with NESU experiment control software was used for the timing and data collection and stimuli were played from a DAT recorder over headphones.

Participants were tested individually in a sound-proof room. The participants with /k/ as target were asked to press the button in front of them with their preferred hand as fast as possible if they detected the target stop /k/, the other participants were instructed to react to the target stop /p/. Response times were measured from the onset of each target nonword. Participants heard the practice list first, followed after a short pause by one of the four experimental list. The whole experiment lasted about 16 minutes.

Results

As in Experiments 1, 2, and 3, RTs were adjusted so as to measure in

Experiment 4 from the burst of the target phoneme. Missed responses and RTs outside the range of 100 to 1500 ms were again treated as errors. Altogether 13 responses were treated as errors because they lay outside the evaluated range of RTs. One participant who monitored for /p/ missed 67 of all stops. Because missed responses spread over all three types of nonwords, it was decided to exclude his RTs from the analysis. The results for phoneme detection of the velar stop /k/ and the bilabial stop /p/ are shown in Table 2−4.

Table 2−4. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants in Experiment 4 to the velar stop /k/ and /p/ after the three nasals /ŋ/, /n/, and /m/ in monosyllabic German nonwords.*

	/k/			/p/		
Measure	Velar nasal [fɛŋk]	Alveolar nasal *[fɛnk]	Bilabial nasal *[fɛmk]	Bilabial nasal [flɔmp]	Alveolar nasal *[flɔnp]	Velar nasal *[flɔŋp]
RTs	435	500	516	366	490	532
Errors	1.8%	4.2%	3.9%	4.0%	10.9%	7.8%

Listeners detected the target stops more slowly when they were preceded by non-homorganic nasals than when no violation occurred. A two factor mixed ANOVA was used, with the target stop as the between-participants factor and the nasal as the repeated measures factor. There was a highly significant main effect of nasal ($F_1(2, 90) = 17.68, p < .001; F_2(2, 52) = 10.46, p < .001$), but no main effect of stop was found.

Because both the bilabial and the velar nasal differ with respect to legality depending on the following stop (whereas the alveolar nasal does not), the factors nasal and target were expected to interact and indeed a highly significant interaction of nasal and target was found in the two factor mixed ANOVA ($F_1(2, 90) = 84.12, p < .001; F_2(2, 52) = 44.71, p < .001$).

For both target stops strong inhibition effects for violation of assimilation were observed when analyzed separately. With the repeated measures factor nasal as the only factor, RTs to both /k/ and /p/ still showed a highly significant main effect of preceding nasal (RTs to /k/: $F_1(2, 46) = 21.78$, $p < .001$; $F_2(2, 26) = 15.52$, $p < .001$; RTs to /p/: $F_1(2, 44) = 75.44$, $p < .001$; $F_2(2, 26) = 33.66$, $p < .001$).

An overall analysis of errors revealed a significant main effect of nasal by participants but not by items ($F_1(2, 90) = 4.25$, $p < .02$; $F_2(2, 52) = 1.88$, $p > .1$). However, it also yielded a significant effect of stop ($F_1(1, 45) = 7.53$, $p = .009$; $F_2(1, 26) = 14$, $p = .001$). Separate analyses showed a significant effect of nasal by participants for /p/ ($F_1(2, 44) = 5.79$, $p = .006$; $F_2(2, 26) = 1.81$, $p > .1$) but not for /k/ ($F_1(2, 46) = 1.31$, $p > .2$; $F_2(2, 26) = 1.77$, $p > .1$). Obviously participants had some difficulty detecting the target stop /p/. This difficulty might partly be due to the fact that in German monosyllabic words ending with p−final clusters are much less frequent than monosyllabic words ending with k−final clusters . In addition, of the various stops /p/ is most likely to be confused with other non-stops because of weaker perceptual cues (Ohala, 1996).

Planned comparisons for the reaction time data and the error rates revealed the same effects as those found in the preceding analyses.

Overall neither the length of the target stop /k/ nor of the target stop /p/ were found to have contributed to the effects discovered. Although /k/ was longer after /m/ (131 ms) than after /ŋ/ (102 ms) ($t(13) = 2.91$, $p < .02$) or /n/ (107 ms) ($t(13) = 2.23$, $p < .05$), none of the contexts yielded a significant correlation of stop length with detection time. Similarly, although /p/ was longer after /ŋ/ (99 ms) than after /m/ (81 ms) ($t(13) = 2.80$, $p < .02$), with an intermediate length after /n/ (91 ms), target length did not correlate with detection time.

To sum up, Experiment 4 replicated the inhibiting effect of regressive assimilation violation found in earlier studies. German participants detected the target stop /k/ or /p/ more slowly if the preceding nasal was not homorganic.

General Discussion

The aim of the present study was to investigate how violation of different obligatory assimilation rules affects spoken-language processing. Previous research had consistently found an inhibition effect for processing violation of obligatory assimilation. But some aspects of assimilation were not examined in those studies. First, the effects of violating native assimilation rules in a non-native language were unknown. Second, assimilation had never been tested within syllables. Third, violations of progressive, rather than regressive assimilation rules had never been tested. The present study therefore investigated the role of these factors in the processing of assimilation violations.

In Experiment 1, German listeners detected a target fricative in monosyllabic Dutch nonwords. When nonwords violated the obligatory German progressive fricative assimilation rule, a facilitation effect was found rather than an inhibition effect. For Dutch listeners, no native phonological violations were included in the target contexts, and accordingly in Experiment 2 they showed no effect of context in their reactions to the target fricative. The results of Experiments 1 and 2 are a confirmation of previous findings in the sense that the process of listening is language-specific. Earlier studies have already reported evidence that the process of listening in a non-native language is influenced by the native language of the listener (see for example Cutler et al., 1986; Otake et al., 1996).

Experiment 3 investigated whether the observed facilitation effect in Experiment 1 was due to the non-native listening task or to the monosyllabic structure of the tested items. The same German fricative assimilation rule was tested. German participants again detected a target fricative in illegal German sequences more quickly, in both monosyllabic and bisyllabic items.

Experiment 4 showed that the necessary condition for facilitation in processing assimilation violations was the nature of the assimilation rule. Violations of regressive place assimilation for German nasals showed slower, not faster, RTs to stops in illegal segment strings. This matches the effect found in the previous literature for other languages (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996).

Both regressive and progressive assimilation form phonotactically legal segment strings and violation of obligatory assimilation always results in phonotactically illegal segment strings. It is in both cases at the second involved segment that the sequence becomes illegal. So why does the violation make it easier in one case and harder in the other to recognize the second segment? Part of the answer to this question can be found in the different predictions that progressive and regressive assimilation make for the upcoming second segment, the monitoring target.

The phonotactics of German do not allow German listeners hearing the front vowels in Experiments 1 and 3 to have strong expectations for the next segment. For example, /ɛ/ in /kɛ–/ could have been followed by /f/, /k/, /l/, /m/, /n/, /ŋ/, /p/, /r/, /s/, /ʃ/, /t/, or [ç]. Hearing /kɛ–/, however, excluded [x] as the following segment. Similarly, hearing /n/ or /l/ in syllable final position in the first syllable of bisyllabic items in Experiment 3 did not provoke strong expectations about the following syllable initial segment, but excluded [x]. The forward operating assimilatory effect weakly restricted what the following segment would be, while strongly excluding one certain segment.

In Experiment 4, on the other hand, German listeners hearing the velar nasal /ŋ/ could have strong expectations for the next segment. German phonology allows only a very restricted set of segments to follow the velar nasal in coda position. In /fɛŋ–/, for instance, /ŋ/ could have been followed by only three consonants, /k/, /s/, and /t/. Thus, the regressive assimilatory effect strongly restricted what the following segment might be. This difference in set restriction effects is true for all segment strings that result from obligatory assimilation, at least in the tested languages.

Thus, although both kinds of assimilation violations form phonotactically illegal sequences, progressive and regressive assimilation impose different kinds of constraints on the second segment for the listener. Regressive assimilation results are the easier to explain. Here the listener has strong expectations as soon as the first segment is identified. Two tendencies could then be responsible for the present results. If the following legal segment can be predicted before it actually occurs, detection of the correct segment may be facilitated compared to the incorrect. Alternatively, when strong expectations are defeated by an illegal segment, inhibition of the unexpected item may result, as it does in visual attention experiments (Posner, Nissen, & Ogden, 1978).

For progressive assimilations, which defeat expectations, this explanation does not apply. Instead, the explanation seems to lie in another difference between the illegal sequences used here. Although both the illegal fricative and the illegal stop sequences have zero transitional probabilities within words, the latter types do occur across word boundaries (e.g., *dem*

Kamm, 'the comb'). The former do not: No German word begins with [x]. The fact that German listeners had never heard the sequence *[ɪx] before in their native language made this sequence a truly novel one for them. The difference between entirely novel sequences that never occur in any environment in the language and sequences which can occur in some environments in the language may be crucial.

Facilitation of processing for novel items is also reported in recent research on spontaneous visual attention; specifically, novel items are reported to cause rapid orientation (see also Christie & Klein, 1996). Johnston and colleagues have recently published several articles (Johnston & Schwarting, 1996, 1997) on this phenomenon, which they have called *novel popout*. They found that novel items were more accurately localizable than familiar items when observers had a glimpse of a scene but were not looking for anything in particular. Familiar items were visually presented words that occurred frequently and thus were expected to occur in the experiment, whereas novel items were presented once and thus were not expected to occur in the experiment (novel items were therefore not novel in an absolute sense because they were existing words in the participants' native language). According to Johnston, novel popout reflects the fact that novel items receive some sort of processing priority and that this is an important adaptive process that has evolved in early phylogenetic history (Johnston & Schwarting, 1997).

This adaptive process might explain why listeners in the present study were faster in their detection of the target fricative (not absolutely novel by itself) if it occurred in a sequence that was truly novel for the listeners. Listeners could not have had strong expectations about which segment would follow the front vowel /ɪ/ in Experiments 1 and 3, but since the following segment [x] was completely new for them in that context it stood out. A number of participants even reported after the experiment that the target fricative sometimes 'popped out'. It appears to be the combination of weak expectations and the novel popout effect that facilitated processing Experiments 1 and 3. When a particular second segment is strongly expected, novelty alone does yield facilitation, as the Japanese experiments on violation of assimilation showed (Otake et al., 1996). In Japanese, nasals never mismatch in place of articulation with a following stop, not even across word boundaries (Vance, 1987). Therefore, Japanese listeners have, for example never heard the sequence */mk/ in their native language. After hearing /m/ they very strongly expected a labial to follow and this expectation might have outweighed any novelty effects.

To sum up, facilitation for detection of a segment in a phonotactically illegal sequence has been shown when listeners' weak expectations about an upcoming segment are defeated and the sequence itself is absolutely novel for the listeners (Experiments 1, 2, and 3 of the present study). Inhibition has been shown when listeners have strong expectations about an upcoming segment which are defeated, but when the sequence itself is not absolutely novel for the listeners (Experiment 4 of the present study and Gaskell & Marslen-Wilson, 1996, 1998; Kuijpers & van Donselaar, in preparation, Otake et al., 1996). Otake et al. (1996) have also shown that detection of a segment that violates phonotactic constraints is inhibited when listeners have strong expectations about an upcoming segment and the sequence itself is a novel one for the listeners. How listeners process a segment that runs counter to weak expectations in a sequence that is not novel for the listeners has not been yet investigated. Following the explanation for the novel popout, inhibition would be expected because popout requires a truly novel sequence for the listener. It is predicted that of all possible combinations of strong and weak expectations with sequences which never occur in the language (absolutely novel) and sequences which occur in other environments (not absolutely novel), only the combination of weak expectations together with a novel sequence causes facilitation in processing phonotactically illegal sequences.

The facilitation effect for violation of progressive assimilation has been observed so far only in a phoneme monitoring task. Listeners are specifically asked to focus their attention on the recognition of a single sound. During the processing of spontaneous speech in normal conversation, single sounds are usually not the focus of attention. Admittedly the link between phoneme monitoring and real language processing is only an indirect one, and therefore it cannot be excluded that the present findings are task specific and rather represent effects on decision making than automatic optimization processes. But in both cases, phoneme monitoring and real language processing, listeners are presented with speech input which they have to process. Furthermore, there is, at least, evidence for the inhibitory effect of assimilation violation from other tasks. Koster (1987) and Gaskell and Marslen-Wilson (1996) observed inhibition for violation of assimilation in word recognition tasks. The inhibition effect is therefore not specific to phoneme monitoring, but whether the same is true for the facilitation effect is still a matter for further investigation.

So far the facilitation effect has been shown for violation of progressive assimilation. But interpretation of the effect depends on an assessment of its generality. If facilitation for phoneme recognition is due to a weak set restriction effect in combination with novel popout, then it might not be confined to assimilation violation. In fact, any other phonological restriction causing weak set restrictions and novel popout for the following segment might replicate the facilitation effect for violation of that restriction. In the same way, the inhibition effect has been observed for processes other than violation of assimilation. If anticipatory coarticulatory information fails to match the actual consonant which follows, processing has been shown to be adversely affected (Marslen-Wilson & Warren, 1994; Martin & Bunnell, 1981; McQueen, Norris, & Cutler, 1999; Whalen, 1984, 1991). In these studies, performance on cross-spliced items that contain acoustic-phonetic mismatches was worse than on matching items. In McQueen et al. (1999) for instance, Dutch items like *sloop*, 'pillowcase', were made by splicing *sloo* from *sloot*, 'ditch', and adding the /p/ burst from *sloop*. Although formant transitions in the vowel signaled an upcoming /t/, the following segment was /p/. In other words, when hearing the vowel, the set of possible segments to follow was reduced to /t/, but /p/ occurred instead. Phonetic decision to the final stop was harder in items with mismatching information than in items without mismatching information. Further investigations testing other phonological violations may show whether the facilitation effect can be replicated in the same way for other phonological restrictions. This would suggest that neither facilitation nor inhibition effects are assimilation-specific, but are rather general phoneme sequence effects. Further investigation is also necessary to determine the exact border between weak and strong set restriction effects.

The current study has shown evidence that the novel popout effect is operative not only in visual perception but also in speech perception. In most current models of speech perception, all sequences with zero transitional probability are treated equally, whether they are impossible in all environments or only in the one in question. The results reported here suggest that not all such sequences are processed in the same way. The finding of a novel popout effect for speech perception adds to our knowledge of mechanisms listeners use. It is hoped that future research will clarify the role of this effect.

Assimilation violation and spoken-language processing: A supplementary report[4]

CHAPTER 2B

Abstract

Previous studies have shown that spoken-language processing is inhibited by violation of obligatory regressive assimilation. In Chapter 2a, I replicated this inhibitory effect in a phoneme-monitoring study examining regressive place assimilation of nasals, but found facilitation for violation of progressive assimilation. German listeners detected the velar fricative [x] more quickly when fricative assimilation was violated (e.g., *[bɪxt] or *[blɪnxən]) than when no violation occurred (e.g., [baxt] or [bluːxən]). It was argued that a combination of two factors caused facilitation: (1) progressive assimilation creates different restrictions for the monitoring target than regressive assimilation does, (2) the sequences violating assimilation (e.g., *[ɪx]) are novel for German listeners and therefore facilitate fricative detection (novel popout). Chapter 2b tests progressive assimilation violation in non-novel sequences using the palatal fricative [ç]. Stimuli either violated fricative assimilation (e.g., *[baːçəl]) or did not (e.g., [biːçəl]). This manipulation does not create novel sequences: Sequences like *[aːç] can occur across word boundaries, while *[ɪx] cannot. No facilitation was found. However, violation also did not significantly inhibit processing. The results confirm that facilitation depends on the combination of progressive assimilation with novelty of the sequence.

---

[4] *Slightly adapted version of manuscript submitted for publication in Language and Speech (Weber, submitted a).*

Introduction

Phonological adjustment processes are common in spoken language. A prominent example is assimilation, in which a feature in a sound is altered under coarticulatory influence of neighboring segments. Rules of assimilation can be either optional or obligatory. If an assimilation rule is optional, both the assimilated and the unassimilated form are equally acceptable. In German, the phrase *ein Bad*, 'a bath', can be realized with an alveolar nasal as [ain#baːt] or with bilabial nasal as [aim#baːt] assimilated to the following bilabial stop /b/. If an assimilation rule is obligatory, in contrast, there is only one acceptable standard realization. In Japanese, a nasal must be homorganic with a following consonant. In *tombo*, 'dragonfly', the final nasal of the first syllable is bilabial [m] before the bilabial /b/, while in *kondo*, 'this time', the nasal is alveolar [n] before the alveolar /d/. The occurrence of such assimilatory adjustments in continuous speech can provide listeners with cues for phoneme recognition.

A number of studies have investigated optional assimilation in different languages, via phoneme detection or word recognition tasks (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Quené et al., 1998). The results of those studies have shown that spoken-language processing is neither facilitated nor interfered with by optional assimilation in a context that licenses the change, but is inhibited by assimilation in a phonologically unviable context (for a discussion of the individual studies see Chapter 2a).

Otake et al. (1996) investigated obligatory assimilation in Japanese. Place of articulation of a nasal differs as a function of the place of articulation of the following segment. Japanese listeners responded equally rapidly and accurately to moraic nasals irrespective of their place of articulation. When asked to respond to the following stop, however, the same listeners were sensitive to violation of the obligatory place assimilation. Their RTs in a phoneme-monitoring task were significantly slower in rule-violating items (heterorganic nasal-stop cluster) than in lawfully assimilated items (homorganic nasal-stop cluster). Thus, violation of regressive place assimilation for nasals inhibited spoken-language processing in Japanese.

I replicated the inhibitory effect for regressive place assimilation for nasals in German (Chapter 2a). Regressive place assimilation for nasals is obligatory within German syllables for the velar stop /k/ and the bilabial stop

/p/: Thus, German *Bank*, 'bank', must be realized as [baŋk] not as *[bank] or *[bamk]. German *Lump*, 'rogue', must be realized as [lʊmp], not as *[lʊnp] or *[lʊŋp]. German listeners were asked to listen to a list of nonwords and press a button in front of them if they detected /k/ or /p/ in any of the nonwords. They detected the target stop /k/ or /p/ more slowly in monosyllabic nonwords when the preceding nasal was not homorganic (e.g., *[fɛnk], *[fɛmk] or *[flɔnp], *[flɔŋp]) than when it was homorganic (e.g., [fɛŋk] or [flɔmp]).

However, in the same phoneme-monitoring study, I found facilitation rather than inhibition for items that violated progressive fricative assimilation instead of regressive nasal place assimilation (Chapter 2a). In progressive assimilation (1) an earlier segment affects a later one, whereas in regressive assimilation (2) a later segment affects an earlier one (examples from German).

(1)	*Licht*	[lɪçt], 'light'	front vowel shifts following fricative to palatal place of articulation
	lacht	[laxt], 'laughs'	back vowel shifts following fricative to velar place of articulation
(2)	*Bank*	[baŋk], 'bank'	velar stop shifts preceding nasal to velar place of articulation
	Lump	[lʊmp], 'rogue'	bilabial stop shifts preceding nasal to bilabial place of articulation

The palatal fricative [ç] and the velar fricative [x] stand in complementary distribution in German: The velar fricative [x] occurs after back vowels, the palatal fricative [ç] after front vowels, glides, sonorant consonants, word initially and in the diminutive suffix −*chen* (Hall, 1989). Thus, the place of articulation of a vowel specifies the place of articulation for the following fricative. It violates German fricative assimilation to realize *Licht*, 'light', with a velar fricative (*[lɪxt]).

In Chapter 2a, German participants were asked to detect the target fricative [x] in mono− and bisyllabic nonwords. They detected [x] more quickly in phonotactically illegal sequences (e.g., *[bɪxt] and *[blɪnxən]) than in legal sequences (e.g., [bɔxt] and [bluːxən]). In the bisyllabic items, *[blɪnxən] might not be considered assimilation, since the fricative follows a nasal rather than the relevant vowel. However, the fact that no interaction was found between number of syllables and phonotactic legality suggests that this condition is comparable to others in the experiment. Acoustic measurements and control experiments excluded the possibility that the facilitation effect was due to the target fricative being acoustically more prominent.

Why was processing facilitated when a progressive assimilation rule was violated and inhibited when a regressive assimilation rule was violated? It was argued that the reason lay in the combination of two factors. First, progressive and regressive assimilation restrict the set of possible later segments differently. For any two-segment string, listeners can develop expectations about what the second segment will be on the basis of the first segment. Regressive assimilation limits the set of possible second segments strongly. In German, for instance, there are only three consonants, that can follow the /ŋ/ in /fɛŋ–/ within the syllable: /k/, /s/, and /t/. Regressive assimilation therefore creates strong expectations for listeners about the identity of the upcoming segment. When regressive assimilation is violated, these strong expectations are defeated because a segment is heard that is not a member of the small set of possible continuations. Inhibition of the unexpected item can result. Progressive assimilation, on the other hand, does not allow listeners to form strong expectations about an upcoming segment, it rather acts to specifically exclude certain continuations. In German, /bɪ–/ can be followed by /f/, /k/, /l/, /m/, /n/, /ŋ/, /p/, /r/, /s/, /ʃ/, /t/, or [ç], but not by [x]. Thus, when progressive assimilation is violated only weak positive expectations are defeated. This difference in strength of expectations is true for all segment strings that result from regressive or progressive assimilation, at least in the languages tested.

The second factor responsible for facilitation lies in the degree of novelty of the illegal sequences. Although sequences like *[np], which violate nasal assimilation, can never occur within words, they do occur across word boundaries in German (e.g., *mein Platz*, 'my place', which only optionally assimilates to [maim#plats]). But sequences like *[ɪx] and *[nx], which violate fricative assimilation, not only have zero transitional probability within words, but also across word boundaries because no word has initial [x] in German. This makes the sequences truly novel for German listeners. Novel items can receive processing priority; they 'pop out' for the listeners and facilitate detection. Such a novel popout effect has been reported in research on spontaneous visual attention (Christie & Klein, 1996; Johnston & Schwarting, 1996, 1997). Johnston and Schwarting (1997) found that novel items caused rapid orientation and were more accurately localizable than familiar items. They claimed that novel popout is an important adaptive process that evolved in early phylogenetic history (Johnston & Schwarting, 1997).

If the defeat of weak expectations about the upcoming segment facilitates phoneme detection in novel sequences, then how does the defeat of weak expectations influence phoneme detection in non-novel sequences? If novel popout depends both on expectations being weak and on novel sequences, then processing should not be facilitated. But is the defeat of weak expectations sufficient to inhibit processing, as previous studies have shown for strong expectations?

The Experiments 1, 2, and 3 in Chapter 2a were the first experiments which investigated progressive assimilation. Because assimilation was only tested in novel sequences we do not know yet how processing of progressive assimilation is influenced without the additional factor novelty. The present study therefore tested the same German progressive fricative assimilation rule, but this time using the palatal instead of the velar fricative as target. Chapter 2a tested *[ɪx] as the violation condition an [ax] as the assimilated case. The current experiment tested *[aːç] (violation) and [iːç] (assimilated) instead. Here, it is still the case that listeners cannot have strong expectations about what segment will follow the vowel, but the sequences are not truly novel in German. Sequences like [aːç], although illegal within words, can occur across word boundaries, since in Standard German a small number of words, like *Chemie*, 'chemistry', *China*, 'china', *Chirurg*, 'surgeon', begin with the palatal fricative [ç] (e.g., *sah China*, 'saw China', [zaː#çina], or *zu Chemie*, 'to chemistry', [tsuː#çeːmiː]).

Thus, the two types of stimuli chosen for this experiment were items such as *[baːçəl] (back vowel−ç, assimilation violation) and [biːçəl] (front vowel−ç, no violation). Bisyllables were chosen for ease of articulation for the speaker. Because sequences like [aːç] occur across word boundaries, producing those sequences within nonwords across syllable boundaries was considered unproblematic. Listeners were not expected to detect the target fricative [ç] faster when violation of assimilation occurred. It was not clear, however, whether the defeat of weak expectations in non-novel sequences would be sufficient to slow down the detection the target fricative.

Experiment 5: Germans monitoring for [ç] in German

Method

Participants. Twenty-four students from the University of Hannover, all native speakers of Standard German, took part in the experiment. They were paid for their participation.

Materials. A list of 28 bisyllabic items, all German nonwords, was selected with the help of the CELEX database (Baayen et al., 1993; Appendix 2−4, p. 157). In 14 items the first syllable ended in one of the long front vowels /i/ or /e/, followed by the palatal fricative [ç] as the beginning of the second syllable (as in [biːçəl]). In 14 more items the first syllable ended in one of the long back vowels /a/ or /o/, followed by the palatal fricative [ç] (as in *[baːçəl]). The second syllable was [çən], [çəl], or [çər]. The syllable [çən] can be a diminutive ending in German, while [çəl] and [çər] are not morphemes. The diminutive [çən] is realized with a palatal fricative not only after front vowels but also after back vowels, leading to a very small number of words with [ç] after back vowels, such as *Frauchen*, 'mistress', [frauçən]. However, almost all German nouns with a back vowel shift to a front vowel before the diminutive (e.g., *Schuh – Schühchen*, 'shoe' – 'little shoe', [ʃuː] – [ʃyːçən]; Duden, 1989). Therefore, this difference was considered unlikely to have a significant influence on processing. All items were matched pairs, varying only in the backness of the first vowel. No phonotactic constraints except fricative assimilation were violated in these nonwords.

In addition, 252 mono− and bisyllabic filler items, also phonotactically legal German nonwords, were selected. Forty-two of the fillers contained the fricative [ç] in a variety of positions in the nonwords. From the complete set of 280 items, four pseudo-random orders were constructed, with the restriction that for at least two items before a target item, only fillers without the target fricative [ç] were used. Fourteen similar practice items were created.

Procedure. All materials were recorded onto a DAT in a sound proof booth by a female native speaker of German (not the same speaker as in Chapter 2a). The materials were transferred to a computer and measured using the Xwaves software. Items were presented over headphones using the NESU experiment control software. Participants were instructed in writing and orally to listen to the nonwords and press the button in front of them as fast as possible if they detected the target fricative [ç] in one of the nonwords. The computer timed and stored reaction times (RTs). Each participant heard the practice list first, followed by one of the four experimental lists. RTs were measured from the onset of the target fricative.

Results

Missed responses and RTs slower than 1500 ms (only four responses) were treated as errors. Mean RTs and mean error rates are given in Table 2−5. Analyses of Variance with both participants (F_1) and items (F_2) as the repeated measures were performed.

Table 2−5. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants to the penultimate alveolar fricative [ç] after front or back vowels in bisyllabic German nonwords in Experiment 5.*

Measure	Front vowel [biːçəl]	Back vowel *[baːçəl]
RTs	547	564
Errors	1.7%	0.8%

A two factor ANOVA was used, with phonotactic legality (legal front vowel and illegal back vowel) and second syllable ([çən], [çər], [çəl]) as within-participants factors. In the items analysis, the second syllable factor was a between-items factor.

No facilitation effect was found. Participants did not detect the palatal fricative more quickly when fricative assimilation was violated than when no violation occurred. However, no inhibitory effect was found either. Although

RTs were 17 ms slower to items containing illegal sequences, this difference was not significant ($F_1(1, 23) = 1.92, p > .1; F_2 < 1$). The low percentage of errors indicates that the participants had no problems performing the task. An error analysis revealed no significant effects. Neither the factor second syllable ($F_1(2, 46) = 1.48, p > .2; F_2(2, 11) = 1.77, p > .2$) nor its interaction with phonotactic legality showed any effect in RTs (F_1 & $F_2 < 1$).

In Chapter 2a, processing was facilitated when the fricative assimilation rule was violated by the velar fricative. Detection of the velar fricative was faster in bisyllabic nonwords containing violation of assimilation than in lawfully assimilated items. The RTs of German listeners were 535 ms to lawfully assimilated bisyllables (e.g., [bluːxən]) and 497 ms to violation items (e.g., *[blɪnxən]). The facilitatory effect for violation of assimilation in the velar fricative conditions was significant by participants and by items ($F_1(1, 23) = 11.41, p = .003; F_2(1, 13) = 7.02, p = .02$). A post-hoc two factor ANOVA with phonotactic legality (legal and illegal) as within-participants factor and target sound (velar versus palatal fricative) as between-participants factor was performed to compare the present results with those from the previous study. The interaction between phonotactic legality and target sound was highly significant by participants and by items ($F_1(1, 46) = 11.55, p = .001; F_2(1, 26) = 14.30, p = .001$). Thus, the pattern of responses differed significantly for violation of assimilation by the velar fricative and violation of assimilation by the palatal fricative.

General Discussion

The present study clarifies earlier work on assimilation in spoken-language processing. A number of studies have shown that spoken-language processing is inhibited by violation of assimilation (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996). The results in Chapter 2a showed that violation of assimilation can also facilitate processing under certain conditions. It was argued that facilitation occurs when two conditions are met. First, listeners must have only weak expectations about an upcoming segment, which are defeated. This is the case with violation of progressive assimilation, whereas violation of regressive assimilation rather defeats strong expectations. Second, the sequences containing violation of assimilation must be novel sequences in the language. Novel items can receive processing priority; they 'pop out' for the listeners and facilitate detection, as has been shown in visual attention research. Violation of the progressive fricative assimilation rule with the velar fricative meets both these conditions. If facilitation depends on novelty of the sequence, then no such effect should occur when progressive assimilation (weak expectations) is violated in non-novel sequences.

The present study therefore tested the same progressive assimilation rule, but using the palatal fricative to create the violation instead. Sequences in which a back vowel is followed by a palatal fricative are illegal within words, but possible across word boundaries, so they are not novel in German. Listeners detected the target fricative [ç] equally quickly whether it was preceded lawfully by a front vowel (e.g., in [biːçəl]) or illegally by a back vowel (e.g., in *[baːçəl]). Phonotactic illegality neither facilitated nor inhibited detection. If the sequence that violates assimilation defeats weak expectations but is not an entirely novel sequence, processing was not facilitated. This confirms that the previously found facilitation is indeed a novel popout effect. The fact that the illegal sequences did not significantly inhibit processing either suggests that the weak expectations created about an upcoming segment are not sufficient to create strong inhibition when defeated.

Indirectly this attests to the importance of weak expectations for any novel popout effect. Consider that when listeners have strong rather than weak expectations, a defeat of these expectations causes strong inhibition of processing. This strong inhibition might outweigh any novelty effects. Only the defeat of weak expectations, which cause no inhibition, can cause a novel popout effect. Exactly this is confirmed by Otake et al's (1996) study on regressive place assimilation of nasals in Japanese. With regressive assimilation listeners can develop strong expectations about an upcoming segment. Furthermore, the illegal sequences were novel for the Japanese listeners because nasals in Japanese never mismatch in place of articulation of a following stop, not even across word boundaries (Vance, 1987). Japanese listeners' RTs were significantly slower in rule-violating items than in lawfully assimilated items. Thus, although novel sequences were involved, the defeat of strong expectations outweighed the novelty factor, and inhibition of processing resulted.

In the other relevant studies (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation), regressive assimilation was tested with non-novel sequences. Thus, neither of the two factors responsible for facilitating processing was met, so strong inhibition of processing was shown in all four studies for assimilation violation.

To sum up, the present study sheds new light on the role of the factors strength of expectations and novelty of the sequence in processing phonotactically illegal sequences. It was confirmed that only the defeat of weak expectations in novel sequences causes significant facilitation. The defeat of weak expectations of non-novel sequences did not facilitate processing, but it was not sufficient to inhibit processing, either.

Although violation of assimilation always results in phonotactically illegal sequences, the combined results show that not all illegal sequences are processed in the same way. Rather, we know now that under various circumstances processing of these illegal sequences can either inhibit or facilitate processing or not influence it at all. Hay, Pierrehumbert, and Beckman (in press) also found that not all phonotactically illegal sequences are equivalent in a judgment task where English participants had to judge auditorily presented nonsense words as possible additions to the English vocabulary. Different sequences, all with zero transitional probability in the tested environment, were not treated equally by listeners. Rather, listeners' judgments on well-formedness of phoneme combinations were related to the overall frequency of phoneme combinations in the language. The results of the present study show that such variation among phonotactically illegal sequences applies to processing as well.

The influence of phonotactic constraints on phoneme detection

CHAPTER 2C

Abstract

The results in Chapters 2a and 2b have shown that listeners are sensitive to the violation of native assimilation rules in both native and non-native spoken language. This study expands the inquiry to the violation of phonotactic constraints in syllable onsets and its influence on phoneme detection. Native German and Dutch listeners detected the alveolar stop /t/ in German nonwords. German listeners had no knowledge of Dutch, but Dutch listeners had a fair knowledge of German. Half of the German nonwords violated a German phonotactic constraint (e.g., [stɪm]), the other half violated a Dutch phonotactic constraint (e.g., [ʃtɪm]). German listeners detected /t/ faster when a native phonotactic constraint was violated. This suggests that German phonotactics mark a syllable boundary right before the target phoneme, which makes it easier to detect /t/. Dutch listeners detected /t/ equally fast whether it was preceded by /s/ or by /ʃ/. Although Dutch listeners might be sensitive to both German and Dutch phonotactics, a follow-up experiment suggests that this may not be the explanation, since Dutch listeners had problems distinguishing German /s/ and /ʃ/.

Introduction

Are listeners just sensitive to native assimilation rules in non-native spoken-language processing (as previous research has already shown), or are they also sensitive to other phonological regularities, such as phonotactic constraints? Phonotactic constraints differ between languages just as assimilation rules do. The phonotactic constraints of a language restrict which phonemes and sequences of phonemes are permissible in different syllabic positions in that language. For example, the sequence /st/ is not permissible as a syllable onset in German (no German word begins with /st−/), but is permissible in syllable-coda position (e.g., German *Last*, 'burden', /last/). In Dutch, however, the sequence /st/ is permissible as both syllable onset and coda (e.g., Dutch *straat*, 'street', /strat/ and *lijst*, 'frame', /lɛist/).

Just as with assimilation, violation of phonotactic constraints always results in phonotactically illegal sequences (although the same sequences may be legal in a different syllable position or across syllables). Assimilation is a phonetically natural pattern in which a feature of a phoneme is taken over from a neighboring phoneme. Universally, two segments that have undergone assimilation share a feature rather than differ in a feature. Because assimilation is phonetically motivated, rules of assimilation are very similar across languages in which they apply. Phonotactic constraints on the other hand are arbitrary. Whereas one language allows a certain syllable onset but not another syllable onset, a second language may impose exactly the reverse constraints on syllable onsets. Phonotactic constraints may provide less useful information for listeners than assimilation rules because they are highly language specific.

A previous study by Altenberg and Cairns (1983) has already given some evidence that listeners are sensitive to phonotactic constraints of both the native and the non-native language when they process a non-native language (for similar evidence from the visual domain see for example Jared & Kroll, 2001; Nas, 1983). They asked English-German bilinguals to rate visually presented nonwords as potential English words. Some of the nonwords were phonotactically legal in one language and illegal in the other (e.g., *smatt* is legal in English but illegal in German, *schwuf* is legal in German but illegal in English), other nonwords were either illegal in both languages (e.g., *fnoss*) or legal in both languages (e.g., *blem*). The ratings of the bilinguals were only influenced by phonotactic legality in English, which suggests that the bilinguals had excellent knowledge of English phonotactic constraints. However, in a lexical-decision experiment the bilinguals were then asked to decide whether a visually presented item was an English word or not, and this time their decisions were affected by the phonotactic legality of the stimuli in both German and English, even though German was irrelevant for the task. This suggests that the bilinguals could not suppress the influence of the language that was irrelevant for the task, during on-line processing (even though they rated these nonwords in a meta-linguistic task based on the phonological structure of one language only). Altenberg and Cairns (1983) also tested monolingual English listeners, and found that their decisions in the lexical-decision task were affected only by the phonotactic legality of the stimuli in their native language.

Experiments 6 and 7 of the present study examine whether phoneme detection in both native and non-native spoken language is influenced by the violation of phonotactic constraints. In Experiment 6, native listeners of German were presented with German monosyllabic nonwords in which a German phonotactic constraint was violated in the syllable onset. As mentioned above, in standard German the consonant cluster /st/ does not occur as a syllable onset, but is permissible in syllable-coda position. The cluster /ʃt/ on the other hand occurs both as syllable onset and coda (e.g., German *Straße*, 'street', /ʃtrasə/ and *mischt*, 'mixes', /mɪʃt/). The experiment contained items such as [stɪm], in which Germans listeners had to process the phonotactically illegal sequence /st/. Their task was to detect the target stop /t/ in the items, which were nonwords. As in Chapters 2a and 2b, the task was generalized phoneme monitoring (Frauenfelder & Seguí, 1989; Seguí & Frauenfelder, 1986). A difference in detection times for nonwords with phonotactically illegal sequences would show that native listeners are sensitive to phonotactic constraints of their native language when they listen to that language, even if the phonotactic constraints are not assimilation.

In Experiment 7, native listeners of Dutch were presented with the same German materials that were used in Experiment 6. Dutch has different phonotactic constraints than German. In standard Dutch, the consonant cluster /st/ is permissible as both syllable onset (e.g., Dutch *straat*, 'street', /strat/) and syllable coda (e.g., Dutch *lijst*, 'frame', /lɛist/). The consonant cluster /ʃt/, however, occurs in Dutch only in some loanwords as syllable coda (e.g., Dutch *luncht*, 'lunches', /lʉnʃt/) and never occurs as syllable onset (no Dutch word begins with /ʃt–/). In items such as [ʃtɪm], a phonotactic constraint of Dutch was violated (but not of German). In items such as [stɪm], a phonotactic constraint of German, the non-native language in which the items were produced, but not of Dutch, was violated. Just as for the German listeners in Experiment 6, the task of the Dutch listeners was to detect the target stop /t/ in the nonwords. A difference in detection times for nonwords with phonotactically illegal sequences would show that native listeners of Dutch are sensitive to the phonotactic constraints of either their native language or to the phonotactic constraints of the non-native language (which is the language they were listening to). No difference in detection times would show that Dutch listeners are sensitive to both (or none of the) native and non-native phonotactic constraints in non-native phoneme detection.

Experiment 6: Germans monitoring for /t/ in German

Method

Participants. Twenty-four students of the University of Regensburg in Germany were paid to take part in the experiment. They were all native speakers of German who grew up in the southern part of Germany and had no knowledge of Dutch.

Materials. Fifty-six monosyllabic items were selected with the help of the CELEX database (Baayen et al., 1993). All items were nonwords in both German and Dutch, and contained short vowels only. No phonotactic constraints of either language, except the phonotactic constraints in question, were violated in these nonwords (for Dutch phonotactics see for example Booij, 1995; for German phonotactics see for example Wiese, 1996). Only phonemes that occur in both languages were used. Fourteen of the items began with the consonant cluster /st/ (such as in [stɪm]). Whereas /st–/ is a not a permissible syllable onset in German, it is a permissible syllable onset in Dutch. Note that in Standard German the alveolar fricative /z/ must be voiced in syllable-initial position. However, in southern Germany, where the participants came from, the voiced fricative /z/ is always devoiced (see for example Russ, 1990). Still, initial [s] cannot be followed by /t/. Fourteen other items began with the consonant cluster /ʃt/ (such as in [ʃtɪm]). The cluster /ʃt–/ is a permissible syllable onset in German, but not in Dutch.

Fourteen more nonwords ended in the consonant cluster /st/ (such as in [nɔst]). In both German and Dutch, /st/ occurs as syllable coda. The other 14 nonwords ended in the consonant cluster /ʃt/. Again, /ʃt/ occurs as syllable coda in both German and Dutch, in Dutch however, the cluster occurs only in some loanwords (for an overview see Table 2−6). Note that some phonologists like Booij (1995) for example, do not consider the cluster /ʃt/ a permissible cluster in Standard Dutch. Some phonologists also claim that the Dutch phoneme inventory contains /sj/ but no /ʃ/. However, /sjt−/ is also not a permissible onset in Dutch. For neither the items with /st/ and /ʃt/ in the onset or those with /st/ and /ʃt/ in the coda was it possible to find 14 matched pairs of nonwords that differed only in the alveolar or post-alveolar fricative. This was taken into account in the statistical analyses. The nonwords are listed in Appendix 2−5, p. 158.

Table 2−6. *Phonotactic constraints of both German and Dutch used in Chapter 2c.*

		German	Dutch
Syllable onset	/st/	illegal	legal *straat* /strat/, 'street'
	/ʃt/	legal *Straße* /ʃtrasə/, 'street'	illegal
Syllable coda	/st/	legal *Last* /last/, 'burden'	legal *lijst* /lɛist/, 'frame'
	/ʃt/	legal *mischt* /mɪʃt/, 'mixes'	legal in loanwords *luncht* /lʉnʃt/, 'lunches'

In addition, 318 mono− and bisyllabic filler nonwords, also legal nonwords in both Dutch and German, were selected. Sixty-four of the fillers each contained the alveolar fricative /s/ or the post-alveolar fricative /ʃ/ in a variety of positions in the nonwords. Seventy-five of the fillers contained the alveolar stop /t/. From the complete set of 374 items, four different pseudo-random orders were constructed, with the restriction that for at least one item before a target-bearing item, only fillers without /t/, /s/, or /ʃ/ were used.

Eleven similar practice items were created and presented at the beginning of the experiment. There were three pauses, one after the practice list and two more in the experiment itself after every 125 items.

Procedure. All materials were recorded onto DAT in a sound-proof booth by a female native speaker of German. Although the cluster /st/ never occurs as syllable onset in German, it was not a problem for the speaker to produce it in items like [stɪm] because the cluster does occur as German syllable coda. Speech stimuli were down-sampled during transfer to a computer to 16 kHz.

Using the Xwaves software, the beginnings of the fricatives /s/ and /ʃ/ and the beginning of the stop /t/ were identified. Each nonword was then transferred as an individual speech file to the hard-disk of a personal computer. Stimulus presentation, timing, and data collection were performed using the NESU experiment control software.

Participants were tested one at a time in a sound-proof booth. Written instructions in German told them to listen for the target stop /t/ in a series of German nonwords, and to press the button in front of them with their preferred hand as fast as possible if they detected /t/ in any of the nonwords. Each participant heard the practice list first, followed by all experimental stimuli in one of the four pseudo-randomized orders. The experiment lasted approximately 20 minutes.

Results

Prior to statistical analysis, RTs (response times), which were originally measured from the onset of the items, were adjusted so as to measure from the onset of the respective target phoneme. Missed responses were treated as errors. Mean RTs and mean percentage of errors are given in Table 2−7.

Table 2–7. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants to the alveolar stop /t/ in Experiment 6.*

	Syllable onset		Syllable coda	
Measure	*[stɪm]	[ʃtɪm]	[nɔst]	[nɔʃt]
RTs	489	533	324	346
Errors	0%	2.9%	0.8%	3.8%

In syllable onsets, German participants detected the target stop /t/ more quickly when a German phonotactic constraint was violated (e.g., *[stɪm]) than when no such violation occurred (e.g., [ʃtɪm]). RTs were submitted to Analyses of Variance (ANOVAs), with both participants (F_1) and items (F_2) as the repeated measure. The pattern in RTs was significant by both participants ($F_1(1, 23) = 7.18, p < .02$) and items ($F_2(1, 26) = 5.56, p < .03$). German phonotactics require a syllable boundary between /s/ and /t/ when the two consonants occur as syllable onset, which puts /t/ in syllable-initial position. This is not the case for the cluster /ʃt/ because German phonotactics do not require a syllable boundary between /ʃ/ and /t/. German listeners found it easier to detect a target that was aligned with a syllable boundary according to German phonotactics than a target that was misaligned with a syllable boundary. If the difference in RTs to /t/ in syllable onsets is indeed due to a sensitivity of these listeners to the phonotactics of their native language, then no such difference should emerge in items without violation of German phonotactic constraints.

All nonwords with /t/ in syllable-final position were legal German nonwords that violated no German phonotactic constraints. The participants showed no significant difference in their RTs whether /t/ was preceded by the alveolar fricative /s/ (e.g., [nɔst]) or by the post-alveolar fricative /ʃ/ (e.g., [nɔʃt]) ($F_1(1, 23) = 2.70, p > .1$; $F_2(1, 26) = 3.38, p > .07$).

The low percentage of errors indicates that participants had few problems detecting the target phoneme /t/ in the nonwords. The fact that RTs were slower overall to /t/ in syllable onset than in syllable-final position matches a general trend in phoneme monitoring for faster detection of targets that occur later in items (see for example Frauenfelder & Seguí, 1989).

Experiment 7: Dutch monitoring for /t/ in German

Method

Participants. Twenty-four native speakers of Dutch, students at the University of Nijmegen in the Netherlands, took part in the experiment. They were paid for their participation. The Dutch participants all had a fair knowledge of German, since Dutch students receive at least 3 years of training in German as a foreign language during their secondary education.

Materials. The materials were the same as in Experiment 6.

Procedure. The procedure was the same as in Experiment 6. Dutch participants were told that they would listen to German.

Results

Mean RTs and mean percentage of errors are given in Table 2−8.

Table 2−8. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of Dutch participants to the alveolar stop /t/ in Experiment 7.*

	Syllable onset		Syllable coda	
Measure	[stɪm]	*[ʃtɪm]	[nɔst]	[nɔʃt]
RTs	489	483	310	331
Errors	1.4%	0.8%	2.6%	4.7%

For the Dutch participants, a phonotactic constraint of their native language was violated in the German materials because /ʃt−/ is not a permissible syllable onset in Dutch. However, Dutch participants did not react to the violation of Dutch phonotactic constraints. Rather, these listeners detected /t/ equally fast whether it was preceded by /s/ (e.g., [stɪm]) or by /ʃ/ (e.g., *[ʃtɪm]) (F_1 & $F_2 < 1$). Because the Dutch listeners had a fair knowledge of German (and thereby presumably also had a fair knowledge of German phonotactics), it might have been that their phoneme detection was influenced by both Dutch and German phonotactic constraints. In that case, no difference in RTs is predicted because a nonword such as [ʃtɪm] violates Dutch phonotactics and a nonword such as [stɪm] violates German phonotactics.

In syllable-final position, Dutch listeners were slower to detect /t/ when the stop was preceded by /ʃ/ (e.g., [nɔʃt]) than when it was preceded by /s/ (e.g., [nɔst]). This difference in RTs was significant by participants only ($F_1(1, 23) = 5.50$, $p < .03$; $F_2(1, 26) = 2.88$, $p > .1$). Because the consonant cluster /ʃt/ occurs only in a few loanwords as syllable coda in Dutch, its frequency of occurrence is low, whereas the cluster /st/ occurs in many Dutch words (including native vocabulary) as syllable coda. This was not the case for the German listeners in Experiment 6, since in German both /st/ and /ʃt/ occur quite frequently in syllable-coda position. The difference in frequency and loanword status may be responsible for the slower RTs of the Dutch listeners to the target stop /t/ preceded by /ʃ/.

To further analyze whether native German listeners in Experiment 6 behaved differently than non-native Dutch listeners in Experiment 7, a two factor mixed ANOVA was carried out, with language of the listener as the between-participants factor (with the two levels 'German' and 'Dutch'), and initial fricative as the within-participants factor (with the two levels '/s/' and '/ʃ/'). The factor initial fricative was within-participants in the participant analysis, but was between-items in the item analysis. The factor language was between-participants in the participant analysis, but was within-items in the item analysis. Because phonotactic constraints were only violated in syllable onsets for both German and Dutch, only nonwords with initial /st/ and /ʃt/ were included in the analysis. The interaction between language of the listener and initial fricative was fully significant by participants ($F_1(1, 46) = 5.29$, $p < .03$) and items ($F_2(1, 26) = 5.92$, $p < .03$), reflecting the fact that Germans reacted faster to the onset /st/ condition, but Dutch listeners did not. This suggests that native Dutch listeners do indeed process spoken German differently than native German listeners do.

The low percentage of errors indicates that Dutch participants had no problems detecting the target phoneme /t/ either.

Experiment 8: Germans monitoring for /s/ or /ʃ/ in German

Differences in RT patterns in Experiments 6 and 7 might have been influenced by general processing differences between the two fricatives /s/ and /ʃ/, which preceded the target stop /t/ in the nonwords. Therefore, in Experiments 8 and 9 different German and Dutch listeners were presented with the same German stimuli that were used in Experiments 6 and 7. Some listeners were asked to detect the alveolar fricative /s/, and others were asked to detect the post-alveolar fricative /ʃ/. Lack of a difference in detection times between the two fricatives would show that predicted phonotactic effects in the Experiments 6 and 7 are indeed due to the phonological information given in the nonwords, and are not due to general processing differences between the different consonants involved.

Method

Participants. Forty-eight native speakers of German, students at the University of Regensburg in Germany, took part in the experiment for monetary compensation. Twenty-four of the participants monitored for the target phoneme /s/; the other 24 monitored for the target phoneme /ʃ/. None of them had participated in Experiment 6.

Materials. The materials were the same as in Experiment 6.

Procedure. The procedure was the same as in Experiment 6, except that the target phoneme was either /s/ or /ʃ/.

Results

Mean RTs and mean percentage of errors are given in Table 2−9.

Because in a two factor ANOVA with the additional factor syllable position a significant interaction was found, separate ANOVAs for syllable-onset position and coda positions are reported. One factor ANOVAs were performed with fricative as the between-participants factor (with the two levels '/s/' and '/ʃ/'). Although German listeners detected /ʃ/ more quickly than /s/ in both syllable-onset and coda position, this difference in RTs was only significant for the coda position (initial position: $F_1 < 1$; $F_2(1, 26) = 1.19$, $p > .2$; coda position: $F_1(1, 46) = 7.35$, $p = .009$; $F_2(1, 26) = 44.78$, $p < .001$).

Table 2–9. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of German participants to the alveolar fricative /s/ and the post-alveolar fricative /ʃ/ in Experiment 8.*

	/s/		/ʃ/	
Measure	[stɪm]	[nɔst]	[ʃtɪm]	[nɔʃt]
RTs	686	558	673	432
Errors	0.8%	1.7%	3.5%	0.2%

Because it is possible that RTs might be influenced by differences in the length of the target fricatives, an analysis of the duration of the two target fricatives /s/ and /ʃ/ in syllable-initial position was performed. However, the alveolar fricative /s/ was on average only 2 ms longer, with an average length of 277 ms, than the fricative /ʃ/, which had an average length of 275 ms ($F < 1$).

Neither participants who monitored for /s/ nor participants who monitored for /ʃ/ had problems detecting the target phonemes in the nonwords, as the low error percentages indicate. Because Experiments 6 and 8 had the same experimental lists, participants who monitored for /s/ also heard all stimuli containing /ʃ/, and vice versa. German participants in general had no problem distinguishing /s/ from /ʃ/ and only very rarely responded falsely to nonwords containing the fricative they were not monitoring for. However, participants who monitored for /s/, responded falsely to nonwords with initial /ʃ/ (e.g., [ʃtɪm]) in 17.8% of all cases (for /ʃ/ in coda position only 4.7% of false responses). This unexpectedly high percentage of false alarms can probably be explained by the orthography of German. In German orthography, the sound /s/ is usually spelled as *s* and the sound /ʃ/ as *sch*, except when /ʃ/ in word-initial position is followed by another consonant, where it is transcribed as *s* (e.g., German *Straße*, 'street', /ʃtrasə/). Thus, participants hearing the onset /ʃt/ may have responded with /s/ because *s* would be the appropriate orthography for the /ʃ/ sound. The same orthographic influence might be responsible for the lack of a difference, described above, in RTs to /s/ and /ʃ/ in syllable-initial position.

Experiment 9: Dutch monitoring for /s/ or /ʃ/ in German

Method

Participants. Forty-eight native speakers of Dutch, students at the University of Nijmegen in the Netherlands, were paid to take part in the experiment. Twenty-four of the participants monitored for the target phoneme /s/; the other 24 monitored for the target phoneme /ʃ/. None of them had participated in Experiment 7.

Materials. The materials were the same as in Experiment 6.

Procedure. The procedure was the same as in Experiment 6, except that the target phoneme was either /s/ or /ʃ/. Dutch participants were told that they would listen to German.

Results

Mean RTs and mean percentage of errors are given in Table 2−10.

Table 2−10. *Mean RTs in ms, measured from target onset, and mean percentage errors for responses of Dutch participants to the alveolar fricative /s/ and the post-alveolar fricative /ʃ/ in Experiment 9.*

	/s/		/ʃ/	
Measure	[stɪm]	[nɔst]	[ʃtɪm]	[nɔʃt]
RTs	626	494	747	529
Errors	0.2%	0.5%	23.5%	3.8%

As in Experiment 8, the interaction between the factors fricative and syllable position was significant, therefore separate ANOVAs for syllable-onset position and coda positions are reported. One factor ANOVAs were performed with fricative as the between-participants factor (with the two levels '/s/' and '/ʃ/'). In contrast with the German listeners in Experiment 8, Dutch listeners detected /s/ rather than /ʃ/ more quickly in both syllable-onset and coda position. However, the difference in RTs was only marginally significant for the syllable-initial position (initial position: $F_1(1, 46) = 3.11$; $p > .07$; $F_2(1, 26) = 18.46$, $p < .001$; coda position: $F_1 < 1$; $F_2(1, 26) = 2.59$, $p > .1$).

Just as the German participants did in Experiment 8, the Dutch participants who monitored for /s/ also heard all stimuli containing /ʃ/, and vice versa. However, Dutch participants obviously had problems distinguishing German /s/ from /ʃ/. Most noticeably, Dutch participants who monitored for /s/ not only responded to nearly all nonwords containing /s/, they also responded to 89% of the items containing /ʃ/ but no /s/ (91% false responses to /ʃ/ in syllable-initial position and 87% to /ʃ/ in coda position). The percentages of false alarms were lower (14% on average) for participants who monitored for /ʃ/. Dutch /s/ is articulated somewhat further back in the mouth than German /s/ and is acoustically intermediate between German /s/ and /ʃ/. For Dutch listeners German /ʃ/ might therefore sound rather like Dutch /s/, which might explain the extremely high percentages of false alarms for nonwords containing /ʃ/ of Dutch participants who monitored for /s/. For Dutch participants who monitored for /ʃ/ on the other hand, German /s/ was easier to distinguish from German /ʃ/ because German /s/ is even more acoustically distinct from /ʃ/ than Dutch /s/ is.

These high percentages of false alarms not only make any claims about general processing differences between /s/ and /ʃ/ impossible, but they also weaken any explanation of Dutch RTs in Experiment 7 based on German and Dutch phonotactics. In Experiment 7, Dutch participants detected /t/ equally fast in syllable onsets whether /t/ was preceded by /s/ (e.g., [stɪm]) or by /ʃ/ (e.g., [ʃtɪm]). Whereas the former violates German phonotactic constraints, the latter violates Dutch phonotactic constraints. Because Dutch listeners confused German /s/ so often with German /ʃ/ in Experiment 9, it can no longer be claimed that the lack of a difference in RTs in Experiment 7 is due to the combined influence of German and Dutch phonotactics.

In addition, two factor mixed ANOVAs were carried out for the combined results of Experiments 8 and 9, with language of the listener as the between-participants factor (with the two levels 'German' and 'Dutch'), and fricative as the within-participants factor (with the two levels '/s/' and '/ʃ/'). The factor fricative was within-participants in the participant analysis, but was between-items in the item analysis. The factor language was between-participants in the participant analysis, but was within-items in the item analysis. The ANOVAs were carried out separately for fricatives in syllable-onset position and fricatives in coda position. For fricatives in syllable-initial position the interaction between language of the listener and fricative was significant only by items but not by participants ($F_1(1, 92) = 1.83$; $p > .1$; $F_2(1, 26) = 185.33$, $p = .001$). For fricatives in coda position the interaction was fully significant both by participants and items ($F_1(1, 92) = 5.96$; $p < .02$; $F_2(1, 26) = 41.83$, $p < .001$). As in Experiment 7, the interactions show that native spoken-language processing differs significantly from non-native spoken-language processing.

General Discussion

The results of Experiment 6 show clearly that phonotactic constraints of the native language influence phoneme detection when listening to that language. German listeners were faster to detect the target stop /t/ in German nonwords when /t/ occurred in syllable onsets that are phonotactically illegal in German than when /t/ occurred in permissible syllable onsets of German. A sequence of two phonemes that is phonotactically illegal within syllables can only occur across syllables (if at all), or in other words requires a syllable boundary between its constituents. For example, the cluster /st/ requires a syllable boundary between /s/ and /t/. This aligns /t/ with a syllable boundary and makes it hence easier to spot than /t/ in the syllable-onset cluster /ʃt/, where no syllable boundary is necessary. The results support earlier findings that listeners are sensitive to phonotactic constraints in their native language (see for example Altenberg & Cairns, 1983; McQueen, 1998), even though phonotactic constraints are fairly arbitrary sequencing constraints. Obviously it does not make a difference for the listener whether a sequencing constraint is phonetically motivated, like assimilation rules are, or whether it is a highly language-specific constraint with no phonetic motivation (cf. Hyman, 1999). Listeners exploit information from both kinds of sequencing constraints equally in spoken-language processing.

In Experiment 7, native listeners of Dutch were presented with the same German nonwords. Nonwords that violated German phonotactic constraints for syllable onsets were phonotactically legal in Dutch, and nonwords that were phonotactically legal in German violated Dutch phonotactic constraints for syllable onsets. Phoneme detection of native Dutch listeners who listened to German could have been influenced either by Dutch phonotactic constraints, or by the constraints of the non-native language in which the stimuli were produced, or by both. In contrast to the non-native listeners in Chapter 2a who had no knowledge of the non-native language, the non-native listeners in this study had a fair knowledge of the non-native language. This made it more likely that their processing of the non-native language was influenced not only by native phonological structure, but that they also had learned to make use of the phonological structure of the non-native language in processing. The results of the Dutch listeners seemed indeed to indicate that Dutch listeners were sensitive to both German and Dutch phonotactic constraints, since they showed no difference in their detection times for /t/ whether it was preceded by initial /s/ or by /ʃ/.

This argument was, however, considerably weakened by the results of Dutch listeners in Experiment 9. In Experiment 9, Dutch listeners were asked to detect the fricatives /s/ and /ʃ/ in the same German nonwords. The high percentage of false alarms (particularly among participants who monitored for /s/, falsely responding to items containing /ʃ/) showed that Dutch listeners often confused German /s/ with German /ʃ/. Therefore, it is not clear whether Dutch listeners in Experiment 7 perceived the fricatives that determined phonotactic legality correctly. Effects of phonotactic constraints on phoneme detection observed in Experiment 7 became indistinguishable from effects of phoneme confusion. German listeners in Experiment 8 did not show such any confusion of the two fricatives /s/ and /ʃ/. Therefore, the effects of phonotactic constraints on native phoneme detection observed in Experiment 6 are reliable.

In Experiment 7, Dutch listeners detected /t/ more quickly when it was preceded by /s/ than when it was preceded by /ʃ/ in syllable-coda position. Both /st/ and /ʃt/ occur as syllable coda in Dutch, although the latter occurs only in a few loanwords. Phoneme sequences that are permissible in a language, may still occur only rarely in that language. In some sense such low-frequency sequences are closer to illegal sequences than high-frequency sequences are. Phonotactic constraints can therefore also be regarded as gradual constraints rather than binary constraints (for the use of transitional probabilities in phoneme detection see for example McQueen & Pitt, 1996). The low-frequency cluster /ʃt/ could cue, at least weakly, a boundary for Dutch listeners before the /t/ and hence facilitate the detection of /t/. However, Dutch listeners rather detected /t/ more slowly in /ʃt/ than in /st/. The results seem to be the opposite of the results of German listeners in

Experiment 6, who did indeed detect /t/ more quickly when it was aligned with a syllable boundary in syllable onsets. Since the results of Experiment 8 showed that Dutch listeners confused German /s/ and /ʃ/, it is again not clear whether the inhibitory effect in syllable codas found for Dutch listeners is reliable.

The present experiments certainly suggest a role for phonotactics in spoken-language processing. However, the phoneme-detection task, which demands decisions about phonetic identity, is not the most direct measure of word recognition. Phonotactic constraints may not only play a role in phoneme detection but may also be important for other aspects of spoken-language processing such as segmentation. In the next series of experiments, the investigation of the role of phonotactics was continued using the word-spotting task.

Phonotactic constraints and the segmentation of native and non-native spoken language[5]

CHAPTER 3A

Abstract

Previous research has shown that listeners make use of their knowledge of phonotactic constraints to segment speech into individual words. This study investigates the influence of phonotactics on the segmentation of a non-native language. German and English listeners detected English words in nonsense sequences. German listeners also had knowledge of English, but English listeners had no knowledge of German. Word onsets were either aligned with a syllable boundary or not, according to the phonotactics of the two languages. Words aligned with either German or English phonotactic boundaries were easier for German listeners to detect than words without such alignment. Responses of English listeners were influenced primarily by English phonotactic alignment. The results suggest that both native and non-native phonotactic constraints influence lexical segmentation of a non-native, but familiar, language.

---

[5] *Slightly adapted version of manuscript submitted for publication in Journal of Memory and Language (Weber, submitted b).*

Introduction

To understand spoken language, listeners have to segment a continuous speech stream into individual words, which convey the meaning of an utterance. This is not a trivial task because spoken language does not contain the clear demarcation of word boundaries that is provided by spaces in many written language. Rather speakers produce a continuous speech signal in which word boundaries are not reliably marked with unambiguous acoustic cues. Nonetheless listeners have little trouble recognizing discrete words in a spoken utterance. Different phonological restrictions, for example, can provide information about likely word boundaries and thereby help the listener to parse the speech stream into words. The present study focuses on one type of phonological boundary marker, namely phonotactic constraints, and its use in the segmentation process of a non-native language.

In principle, lexical segmentation can be achieved by a competition process without any acoustic or phonological boundary cues. Competition between a set of candidate words is the core assumption of several models of spoken-word recognition. In Shortlist, for instance, multiple candidate words, spanning different portions of the input, are activated by the speech signal at any moment (Norris, 1994; Norris, McQueen, & Cutler, 1995; Norris et al., 1997). Only those words which best match the input are activated and allowed to compete with each other. The competition itself is a process of lateral inhibition. The candidates that provide an optimal parse of the input win the competition. Given the input *bus ticket* /bʌstɪkət/, for example, the candidates *bus*, *bust*, *stick*, and *ticket*, among others, will be activated. Candidates like *bust* and *stick* will lose the competition because they inhibit each other (/st/ cannot be part of both words) and they cannot provide a complete parse of the input (e.g., *stick* leaves /bʌ/ and /ət/ unaccounted for). The candidates *bus* and *ticket* will win the competition because they do not inhibit each other and they can account for all of the input. The boundary between *bus* and *ticket* emerges as a consequence of word recognition. Thus, segmentation can arise without explicit marking of the word boundary in the input.

But competition alone does not account for the evidence that listeners use other sources of information, when available, to help solve the segmentation problem. A number of studies provide evidence that word boundaries can be acoustically marked in the speech signal by durational differences (see for example Beckman & Edwards, 1990; Gow & Gordon, 1995; Klatt, 1975; Lehiste, 1960, 1972; Oller, 1973). Quené (1992, 1993) found that Dutch listeners exploit durational cues to make segmentation decisions when asked to choose between two alternative readings of an ambiguous two-word utterance such as /dipɪn/ (*diep in*, 'deep in', or *die pin*, 'the pin'). Dumay, Content and Frauenfelder (1999) observed reliable durational differences in French obstruent-liquid clusters due to word boundary location (C#C or #CC). In a word-spotting task, detection of a target word was delayed when the target onset did not correspond to an intended word boundary (e.g., *roche*, 'rock', in /i#kʀɔʃ/) versus when it did (*roche* in /ik#ʀɔʃ/). However, the value of acoustic cues for segmentation in general may be weak. First, none of the acoustic cues are reliably present in the speech signal, and, when present, they are often small and variable (Klatt, 1976; Lehiste, 1972; Nakatani & Dukes, 1977; Quené, 1992). Second, segment duration, as one such cue, is not only varied due to position in the word but also due to stress, emphasis, speech rate, etc..

A phonological rather than acoustic cue to likely word boundaries has been found in the metrical structure of languages. In a stress-timed language like English for example, the majority of content words begin with strong syllables (Cutler & Carter, 1987). Cutler and Norris (1988) showed that English listeners use this metrical information for segmentation by inferring a word boundary at the onset of strong but not at the onset of weak syllables. English listeners find it easier to detect a word like *mint* in a nonsense sequence consisting of a strong syllable followed by a weak syllable, like /mɪntəv/, than in two strong syllables /mɪnteɪv/. Vroomen, van Zon, and de Gelder (1996) replicated these results in Dutch, another stress-timed language. Japanese is a mora-timed language, and it has been shown that Japanese listeners are sensitive to moraic information in their language, which could be used for lexical segmentation (Cutler & Otake, 1994; McQueen, Otake, & Cutler, in press; Otake et al., 1993). Native listeners of a syllable-timed language like French or Spanish on the other hand are sensitive to syllabic structure (Cutler et al., 1986; Cutler, Mehler, Norris, & Seguí, 1992; Mehler et al., 1981; Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992).

Another boundary cue for segmentation is provided by vowel harmony. In Finnish, vowels from a front or back harmony set cannot co-occur within a word. It therefore follows that if two successive syllables contain vowels from these two mutually exclusive sets, the syllables cannot belong to the same word. In a word-spotting experiment, Finnish listeners detected words embedded in nonsense strings more quickly when there was a harmony mismatch between the context and the embedded word (e.g., *hymy*, 'smile', in /puhymy/) than when there was no harmony mismatch (e.g., *hymy* in /pyhymy/; Suomi et al., 1997). The harmony mismatch flagged the word boundary for the listeners. Vroomen, Tuomainen, and de Gelder (1998) tested the joint operation of vowel harmony and prosodic information in Finnish (Finnish has fixed word stress on the initial syllable). Again, detection of embedded words was faster with harmony mismatch (e.g., *hymy* faster in /puhymy/ than in /pyhymy/). The profit of harmony mismatch was, however, bigger for words without a pitch accent on their first syllable than for words whose first syllables were accented.

The present study concentrates on the use of phonotactic constraints as boundary cues for segmentation. Phonotactic constraints restrict which phonemes and sequences of phonemes are permissible in different syllabic positions. For example, the sequence /ml/ is not legal within syllables in English either as onset or coda, therefore a syllable boundary must be placed between /m/ and /l/. Because syllable boundaries are highly correlated with word boundaries, listeners can use their phonotactic knowledge for lexical segmentation by inserting a potential word boundary between consonants that cannot co-occur within syllables. In a word-spotting experiment, Dutch adult listeners found it easier to detect words in nonsense sequences when the word onsets were aligned with a phonotactic boundary (e.g., *rok*, 'skirt', in /fim.rɒk/) than when they were misaligned (e.g., *rok* in /fi.drɒk/; McQueen, 1998). In /fim.rɒk/, a syllable boundary is required between /m/ and /r/, since /mr/ is not permissible within a syllable in Dutch. This leaves the onset of *rok* aligned with a syllable boundary. In /fi.drɒk/, however, a syllable boundary is required between /i/ and /d/, since /fid/ is not a possible syllable in Dutch due to syllable-final devoicing. This creates /drɒk/, in which the onset of *rok* is misaligned, and hence harder to spot. In addition, McQueen manipulated the metrical structure of the nonsense strings. StrongStrong aligned and misaligned pairs (e.g., *rok* in /fim.rɒk/ and /fi.drɒk/) were, for example, contrasted with WeakStrong pairs (e.g., *rok* in /fəm.rɒk/ and /fə.drɒk/). Phonotactic alignment effects were found regardless of the metrical structure. Once phonotactics signaled a syllable boundary, the metrical structure could not provide any further boundary information.

Phonotactic constraints are not powerful enough to mark all word boundaries. In a corpus of continuous English speech, only 37% of the word boundaries could be detected on the basis of English phonotactic constraints when the speech was transcribed phonemically (Harrington, Watson, & Cooper, 1989). Therefore, the role of phonotactic constraints for segmentation may be relatively limited. However, when present, phonotactic cues, in contrast to gradient cues like segment duration, reliably mark the onset of a syllable.

Phonotactic constraints, like metrical structure and vowel harmony, differ between languages. Whereas, for example, English phonotactics allow /sl–/ as syllable onset, German phonotactics do not allow this string as either an onset or a coda (for English phonotactics see for example Giegerich, 1992; for German phonotactics see for example Wiese, 1996). For metrical information, it has been shown that listeners segment speech according to the metrical structure of their native language, even when they are listening to a non-native language that does not share the same metrical structure (Cutler et al., 1986). French listeners are sensitive to syllabic structure not only when listening to their native language, but also when listening to English, which is a stress-timed language. English listeners, on the other hand, do not show sensitivity to syllabic structure, either when listening to their native language or when listening to French (Cutler et al., 1986). Further support for a segmentation process that is influenced by language-specific cues to word boundaries comes from Vroomen et al. (1998). Finnish, Dutch, and French listeners learned to segment an artificial language. Performance was best when the phonological properties vowel harmony and word-initial stress of the artificial language matched those of the native one.

But what role do language-specific phonotactic constraints play during the process of listening to a non-native language? There is already evidence that listeners are sensitive to phonotactic constraints of both the native and the non-native language when processing a non-native language. Altenberg and Cairns (1983) tested the influence of phonotactic constraints on the processing of visually presented words in English monolinguals and English-German bilinguals. When asked to rate nonwords, which were, for example, phonotactically legal in one language and illegal in the other (e.g., *smatt* is legal in English but illegal in German) as potential English words, the ratings of both the bilinguals and the monolinguals were only influenced by phonotactic legality in English, suggesting that their knowledge of phonotactic constraints was equivalent. But when asked to decide whether a visually presented item was an English word or not (lexical decision), the bilinguals did not show the same pattern of response times as the monolinguals did. Bilinguals were affected in their decision by the phonotactic legality of the stimuli in both German and English while monolinguals' decisions were only affected by phonotactic legality in English. Although these results suggest that listeners are sensitive to the phonotactics of both the native and the non-native language, there has to date been no experimental test of the on-line use of phonotactic constraints in the lexical segmentation process of a non-native language. Do listeners segment non-native speech with the help of native phonotactic constraints, non-native phonotactic constraints, or both?

The current study addresses this question using the word-spotting task (McQueen, 1996). In Experiment 10, German listeners who were highly proficient in English were presented with spoken English nonsense sequences. Their task was to detect English nouns embedded in these nonsense sequences. The onset of the embedded nouns was either aligned with a clear syllable boundary or not. There were four different ways in which the phonotactics of English and German could relate to the syllable boundary. In one condition, neither English nor German forces a syllable boundary at the onset of the embedded word. In /fuklɛŋθ/ for example, the onset of the word *length* is not clearly aligned, since neither language requires a syllable boundary between /k/ and /l/. In both English and German, /kl–/ is a possible syllable onset (e.g., English *class* /klɑːs/ and German *Klasse*, 'class', /klasə/). Likewise, /l–/ is a possible syllable onset in both languages (e.g., English *light* /laɪt/ and German *Licht*, 'light', /lɪçt/). In another condition, English, but not German, requires a syllable boundary at the onset of the word, as for example in *length* embedded in /zarʃlɛŋθ/. Whereas /ʃl–/ is not a possible syllable onset in English, it is one in German (e.g., German *Schlitten*, 'sleigh', /ʃlɪtən/). In a third condition, German, but not English, requires a syllable boundary at the onset of the word, as for example in /joʊslɛŋθ/. Words cannot begin with /sl–/ in German, but can in English (e.g., English *sleigh* /sleɪ/). In the fourth condition, both German and English require a syllable boundary at the word onset, as for example in /funlɛŋθ/. Words cannot begin with /nl–/ in either language.

One possibility was that German listeners would use exclusively native phonotactic cues for segmentation in English. If this is the case, it should be easier for German listeners to detect English words when German requires a syllable boundary at the onset of the word (e.g., *length* in /funlɛŋθ/ and /joʊslɛŋθ/) and harder when German does not require a syllable boundary (e.g., *length* in /fuklɛŋθ/ and /zarʃlɛŋθ/). Such a result would indicate that listeners rely on native phonotactic information about possible word boundary locations even when listening to a non-native language, where, of course, this information might not help. Such a result would be consistent with previous research on metrical information and vowel harmony, where it appears that listeners continue to use native boundary cues even when listening to a non-native language that differs with respect to these properties (Cutler et al., 1986; Vroomen et al., 1998).

The second possibility was that German listeners would use exclusively English phonotactic cues for segmentation in English. Because word-spotting in a non-native language requires high proficiency in the non-native language, not only might German listeners have acquired the knowledge of English phonotactic constraints, they might also have learned to use these cues instead of German phonotactic cues for processing. In that case, they should find it easier to detect *length* in /funlɛŋθ/ and /zarʃlɛŋθ/ (where English requires a syllable boundary at the onset of *length*) and harder to detect it in /fuklɛŋθ/ and /joʊslɛŋθ/ (where English does not require a syllable boundary). Such a result would indicate that listeners have learned to use non-native phonotactic cues to segment a non-native spoken language and have learned to ignore native phonotactic cues that might mislead them.

The third possibility was that German listeners would use both native and non-native phonotactic cues for segmentation in English. Words that are aligned with a clear syllable boundary according to both German and English phonotactics should then be the easiest for German listeners to spot (e.g., *length* in /funlɛŋθ/). Words that are aligned according to only one of the two languages, either German or English, should be harder to spot (e.g., *length* in /joʊslɛŋθ/ and /zarʃlɛŋθ/). Words not clearly aligned according to both language should be the hardest (e.g., *length* in /fuklɛŋθ/). Such a result would indicate that listeners have learned to use non-native phonotactic cues to segment a non-native spoken language but that, at the same time, the use of native phonotactic cues is not suppressed.

As mentioned above, McQueen (1998) manipulated in his study the detection of embedded words when the word onset was aligned versus when it was misaligned. The sequence /fim.rɒk/ clearly aligns the onset of the embedded word *rok*, whereas the sequence /fi.drɒk/ allows only this syllabification, which misaligns the onset of *rok*. The present study on the other hand manipulates alignment versus lack of alignment. Whereas sequences like /fun.lɛŋθ/ clearly align the onset of the embedded word *length*, sequences without alignment allow two syllabifications, for example /fuk.lɛŋθ/ and /fu.klɛŋθ/. The predictions were, however, very similar to McQueen's study: Embedded words were predicted to be detected faster when the word onset was aligned with a clear syllable boundary and more slowly when not.

English nouns beginning with /l/ and /w/ were chosen for the experiment. Whereas the lateral /l/ belongs to the phoneme inventories of both languages, the labiovelar approximant /w/ does not occur in German. This difference might affect processing of the non-native language. If German listeners use German phonotactic information when listening to English, they would first have to assimilate the non-native /w/ into a native sound before any German phonotactic constraint could be applied. No German phonotactic constraint involves /w/. It is very likely though that German listeners assimilate English /w/ to German /v/. Cognates like quality/Qualität and quantity/Quantität begin with /kw/ in English but /kv/ in German. German-accented English is also often described as having no /w/ and instead having /w/ replaced with /v/. For both /l/− and /w/−initial words, in one condition, neither English nor German forces a syllable boundary at the onset of the embedded word (e.g., *wasp* in /ðikwɔsp/; not clearly aligned in English or German). To establish for sequences with /w/−initial words whether German forces a syllable boundary at the onset of the embedded words, German phonotactic constraints regarding /v/ were considered. In another condition, English, but not German, requires a syllable boundary at the onset of the word (e.g., *wasp* in /gruʃwɔsp/; aligned in English only). In a third condition, German, but not English, requires a syllable boundary at the onset of the word (e.g., *wasp* in /gɔɪtwɔsp/; aligned in German only). In the fourth condition, both languages require a syllable boundary at the word onset (e.g., *wasp* in /jarlwɔsp/; aligned in English and German).

Experiment 10: Germans spotting words in English

Method

Participants. Forty-eight native speakers of German, students of English translation and interpretation at the University of Heidelberg, were paid to take part in the experiment. They had received an average of 15 years of training in English as a foreign language beginning at a mean age of 11. They were experienced in both British English and American English.

Materials. Thirty-six mono– and bisyllabic English nouns with initial /l/ (e.g., *length* and *leather*) and 32 with initial /w/ (e.g., *wasp* and *weapon*) were selected as target words with the help of the CELEX database (Baayen et al., 1993). With a few exceptions, target words had no other English or German words embedded. If a target word did contain other embeddings, the embeddings were either not in final position (e.g., *win* is embedded in *wind*), or were vowel-initial monosyllables (e.g., *awe* is embedded in *law*). Each target word was appended to four different English nonsense monosyllables in order to create the four alignment conditions (see Appendix 3, p. 159). The final consonant of the nonsense syllable determined whether the onset of the target word was aligned with a syllable boundary or not. As much variety as possible was included in the final consonants of the nonsense syllable. The nonsense syllable ended in either /k/, /p/, or /f/ when the onset of a following word with initial /l/ was not clearly aligned according to both English and German (e.g., *length* in /fuklɛŋθ/). For words with initial /w/ the nonsense syllable always ended in /k/ for this alignment condition (e.g., *wasp* in /ðikwɔsp/). The nonsense syllable ended in /ʃ/ when the onset of a following word was aligned only according to English (e.g., /zarʃlɛŋθ/ and /gruʃwɔsp/). The nonsense syllable ended in /s/ when the onset of a following word with initial /l/ was aligned only according to German phonotactics (e.g., *length* in /joʊslɛŋθ/). For words with initial /w/ the nonsense syllable ended in either /t/ or /s/ for this alignment condition (e.g., *wasp* in /gɔɪtwɔsp/). The nonsense syllable ended in /m/ or /n/ when the onset of a following word with initial /l/ was aligned with a syllable boundary according to both English and German phonotactics (e.g., *length* in /funlɛŋθ/). For words with initial /w/ the preceding syllable ended in /l/, /n/, or /m/ for this alignment condition (e.g., *wasp* in /jarlwɔsp/). All nonsense syllables contained either a long vowel or a diphthong. They were constructed in such a way that the syllable was phonotactically legal even if the last consonant was considered as beginning of the following syllable (e.g., both /gɔɪt/ and /gɔɪ/ are legal in /gɔɪtwɔsp/). Thus, the identity of the nonsense syllable did not provide any additional segmentation cues. Each target-bearing nonsense sequence contained its target word in final position but no other English or German words.

In addition, there were 55 filler nonsense sequences that contained embedded English words in final position with an initial consonant other than /l/ or /w/. A further 251 bi– and trisyllabic nonsense sequences contained no embedded English or German words. In both target-bearing sequences and filler sequences, stress was placed on the first nonsense syllable. Four lists were constructed. Each list contained all 306 filler sequences and 68 target-bearing sequences, in a pseudo-random order, such that before each target-bearing sequence there was at least one filler that contained no embeddings. The fillers appeared in the same sequential position in all four lists. Each target also appeared in the same sequential position, but in only one of its four possible contexts in any given list. Each list contained an equal number of all four types of target-bearing sequences. Fourteen more representative practice items were added at the beginning of each list.

Procedure. All materials were recorded onto DAT in a sound-proof booth by a female native speaker of American English who is a trained phonetician, and were sampled at 48 kHz. The speaker was instructed to avoid any clear syllable boundaries in her production. The materials were then down-sampled to 16 kHz during transfer to a computer. Durations of each target word and its context were measured using Xwaves software. Items were presented in the list orders over headphones using a portable computer and the NESU experiment control software. Participants were instructed in written English to listen to the nonsense sequences and press the button in front of them as fast as possible if they detected an embedded English word at the end of one of the nonsense sequences. They then had to say the word aloud. The computer timed and stored manual responses, beginning the clock at the auditory onset of each item and stopping it at a button press. Oral responses were recorded on tape. Each participant heard the 14 practice stimuli first, followed by one of the four experimental lists. The participants were tested one at a time in a quiet room.

Results

Prior to statistical analyses, reaction times (RTs) were adjusted so as to measure from the offset of the target words (the duration of each item was subtracted from the raw RTs for that item). Because different tokens were recorded for the different alignment conditions, small differences in the duration of the target word could be excluded from the analysis by using word-offset RTs. Oral responses were also analyzed, in order to establish whether each button-press was accompanied by the appropriate oral response. Missed manual responses and manual responses that were accompanied either by no oral response or by a word other than the intended target word, as well as RTs outside the range of –200 to 2000 milliseconds (ms), were treated as errors. Thirteen times participants responded before word offset, and seven of these RTs were to the item /hirlwɑtʃ/, which was obviously particularly easy to recognize. Seven target words with particularly high error rates were excluded from the analysis, leaving 61 words for the analysis.

Mean RTs and error rates are given in Table 3−1. It can be seen that there are strong effects of phonotactic alignment for German listeners in both RTs and error rates. Detection was fastest and most accurate when words were aligned with a phonotactic boundary according to both the native and the non-native language (e.g., /jarlwɔsp/). When words were aligned according to only one of the two languages, detection was somewhat slower and less accurate (e.g., /gɔɪtwɔsp/ or /gruʃwɔsp/). When words were not clearly aligned according to either language, RTs were slowest and error rates highest (e.g., /ðikwɔsp/).

Table 3−1. *Mean RTs in ms, measured from target offset, and mean percentage errors of German participants in Experiment 10. G = German, E = English.*

Measure	Initial sound	Not clearly aligned in E or G	Aligned in E only	Aligned in G only	Aligned in E and G
		/fuklɛŋθ/	/zarʃlɛŋθ/	/joʊslɛŋθ/	/funlɛŋθ/
		/ðikwɔsp/	/gruʃwɔsp/	/gɔɪtwɔsp/	/jarlwɔsp/
RTs	/w/	678	589	563	529
	/l/	675	656	650	602
Errors	/w/	32%	20%	21%	18%
	/l/	33%	24%	25%	17%

Analyses with both participants (F_1) and items (F_2) as repeated measure were performed. A four factor mixed analysis of variance (ANOVA) was carried out, with German and English phonotactics (each with the two levels 'aligned' and 'not clearly aligned') and initial sound (/w/ or /l/) as within-participants factors, and experimental list as the between-participants factor. The factor initial sound was within-participants in the participant analysis, but was between-items in the item analysis. The factor experimental list was a factor only in the participant analysis. The German and English phonotactics factors were within-items in the items analysis. This also applies to the other experiments in this chapter. German phonotactics significantly influenced RTs of German listeners ($F_1(1, 44) = 23.93$, $p < .001$; $F_2(1, 59) = 10.11$, $p = .002$), as did English phonotactics ($F_1(1, 44) = 13.52$, $p = .001$, $F_2(1, 59) = 4.84$, $p < .04$). There was no interaction between the constraints of the two languages (F_1 & $F_2 < 1$). The response pattern did not differ for words with initial /w/ and /l/, and no three-way interaction was found between initial sound and the constraints of the two languages ($F_1(1, 44) = 2.62$, $p > .1$; $F_2 < 1$). Analyses of errors revealed very similar results (German phonotactics: $F_1(1, 44) = 35.57$, $p < .001$, $F_2(1, 59) = 10.51$, $p = .002$; English phonotactics: $F_1(1, 44) = 26.61$, $p < .001$, $F_2(1, 59) = 14.93$, $p < .001$; interaction between the constraints of the two languages:

$F_1(1, 44) = 3.56$, $p > .06$, $F_2(1, 59) = 1,25$, $p > .2$; three-way interaction between initial sound and the constraints of the two languages: $F_1(1, 44) = 1.10$, $p > .3$, $F_2 < 1$). As can be seen in Table 3−1, neither the error rates nor the RTs are unusually high for a word-spotting experiment. Apparently, these German listeners had no problems performing the word-spotting task in English.

The lack of an interaction between the constraints of the two languages shows that alignment in the native language helped segmentation as much as alignment in the non-native language did. Next, in order to compare the size of the English and German alignment effects, the results were averaged for the two conditions in which words were not clearly aligned according to German (e.g., average of /ðikwɔsp/ and /gruʃwɔsp/), as were the results for the two conditions in which words were aligned according to German (e.g., average of /jarlwɔsp/ and /gɔɪtwɔsp/). Those two averages were then subtracted from each other to calculate the size of the German alignment effect (63 ms on average). The same was done for the two conditions in which words were not clearly aligned according to English (e.g., average of /ðikwɔsp/ and /gɔɪtwɔsp/) and those in which they were (e.g., average of /jarlwɔsp/ and /gruʃwɔsp/). Again, those two averages were subtracted from each other to calculate the size of the English alignment effect (47 ms on average). Because the same data contributed to more than one condition under this method, these differences in the size of the alignment effect could not be tested statistically. However, the difference of the size of the effect was so small (16 ms) that it appears from it that words which were aligned with a phonotactic boundary in German were not detected any faster than words which were aligned with a phonotactic boundary in English.

Experiment 11: Americans spotting words in English

In Experiment 11, English listeners who had no knowledge of German were presented with the English stimuli from Experiment 10. Because they had no experience with German, only the English phonotactics should influence the segmentation process. Embedded words were predicted to be easier to detect when the onset was aligned with a clear syllable boundary according to English phonotactics (e.g., *length* in /funlɛŋθ/ and /zarʃlɛŋθ/) and harder when not (e.g., *length* in /fuklɛŋθ/ and /joʊslɛŋθ/).

Method

Participants. Forty-eight native speakers of American English, mostly students at the University of South Florida, took part in the experiment for either monetary compensation or course credit. They had no knowledge of German.

Materials. The same materials as in Experiment 10 were used.

Procedure. The same procedure as in Experiment 10 was used.

Results

Missed responses and wrong oral responses, as well as RTs outside the range of –200 to 2000 ms, were treated as errors. Twenty-one times participants responded before word offset, and seven of these RTs were again to the item /hirlwɑtʃ/, which was obviously particularly easy to recognize. The seven target words that were excluded from the analysis in Experiment 10 due to particularly high error rates were excluded again for better comparison with Experiment 10 although the error rates for these items were lower in Experiment 11. Mean RTs (from target offset) and error rates are given in Table 3−2.

An ANOVA, equivalent to that used in Experiment 10 was carried out. The influence of English phonotactics on the RTs of English listeners was highly significant by both participants and items ($F_1(1, 44) = 32.60$, $p < .001$; $F_2(1, 59) = 16.47$, $p < .001$). No other main effects or interactions were significant by both participants and items, except in the error analysis, in which the effects of both English ($F_1(1, 44) = 25.54$, $p < .001$; $F_2(1, 59) = 14.72$, $p < .001$) and German phonotactics ($F_1(1, 44) = 9.28$, $p = .004$; $F_2(1, 59) = 5.29$, $p < .03$) were significant.

Table 3−2. *Mean RTs in ms, measured from target offset, and mean percentage errors of English participants in Experiment 11. G = German, E = English.*

Measure	Initial sound	Not clearly aligned in E or G	Aligned in E only	Aligned in G only	Aligned in E and G
		/fuklɛŋθ/	/zarʃlɛŋθ/	/joʊslɛŋθ/	/funlɛŋθ/
		/ðikwɔsp/	/gruʃwɔsp/	/ɡɔɪtwɔsp/	/jarlwɔsp/
RTs	/w/	522	468	590	452
	/l/	590	541	510	480
Errors	/w/	24%	17%	24%	14%
	/l/	26%	17%	17%	11%

The influence of German phonotactics in the error analysis was unpredicted, since the English listeners had no knowledge of German. An inspection of the mean error rates suggests that the significant effect of German phonotactics is primarily due to a much lower error rate for /l/−initial words when word onsets were aligned according to German phonotactics only (17%; e.g., /joʊslɛŋθ/) compared to when word onsets were not aligned in English or German (26%; e.g., /fuklɛŋθ/). A possible explanation for this effect might be found in the special status of /s/ in processing. Fricatives in syllable-initial position have been suggested to be perceptually invariant with their context and plosives to be perceptually variant with their context (see for example Foss & Swinney, 1973; Klaassen-Don & Pols, 1983). If /s/ is processed independently from its context, detection of the embedded word in /joʊslɛŋθ/ would be more accurate. On the other hand if /k/ is processed unitarily with its context, detection would be less accurate.

On average, RTs of English listeners were faster and error rates were lower than those of German listeners in Experiment 10. This is presumably because the English listeners performed the task in their native language.

The results so far have shown that German listeners who are highly proficient in English use both German and English phonotactic constraints when segmenting English (Experiment 10). On the other hand, English listeners who have no knowledge of German use primarily English phonotactic constraints when segmenting English (Experiment 11).

Experiment 12: Americans performing lexical decision in English

Because all of the nonsense sequences were natural utterances, different tokens of the target words were used in different contexts. The alignment effects could therefore be due to differences between the targets across contexts rather than to the contexts themselves. One way to address this concern is a lexical-decision experiment. In Experiment 12, all target words from Experiment 11, together with a number of nonwords, were presented to English listeners without their original context (e.g., /lɛŋθ/ coming from /joʊslɛŋθ/). Listeners were asked to press a button whenever they heard a real word. English listeners participated in the lexical-decision experiment in order to obtain native response patterns. If acoustic differences between target tokens are at least partially responsible for the alignment effects in the word-spotting experiment, then similar effects should occur in the lexical-decision experiment. If this is not the case then no difference in response times should emerge between target words coming from different contexts.

Method

Participants. Another 40 native speakers of American English, students at the University of South Florida, took part in the experiment for either monetary compensation or course credit. None had participated in Experiment 11.

Materials. Each target word from Experiment 11 was excised from each of its four contexts using the Xwaves software. For example, /lɛŋθ/ was excised from /fuklɛŋθ/, /zarʃlɛŋθ/, /joʊslɛŋθ/, and /funlɛŋθ/. For all filler sequences from Experiment 11, the first syllable was removed, leaving either words with an initial consonant other than /l/ or /w/ (e.g., *donkey*, /dɔŋki/ coming from /huʃdɔŋki/), or leaving nonwords (e.g., /θɪʃ/ coming from /krɛnθɪʃ/). Four lists were constructed, using the same order of items as in Experiment 11. Target words appeared only once in any given list, excised from one of the four possible contexts. Each list contained target words coming from all four types of context.

Procedure. Listeners were told they would hear a list of words and nonwords. They were asked to press the response button whenever they heard a real word. As before, they were asked to repeat each word they detected aloud.

Results

Only RTs that were accompanied by correct oral responses were included in the analyses. Twelve responses which were slower than 1500 ms were excluded from the analyses. The same 61 target words analyzed in Experiments 10 and 11 were analyzed here. Mean RTs (from target offset) and error rates are given in Table 3−3.

Table 3−3. *Mean RTs in ms, measured from target offset, and mean percentage errors of English participants in Experiment 12. G = German, E = English. Note that condition labels refer to the structures in Experiments 10 and 11 from which the targets were excised.*

Measure	Initial sound	Not clearly aligned in E or G	Aligned in E only	Aligned in G only	Aligned in E and G
RTs	/w/	348	396	398	355
	/l/	360	382	369	400
Errors	/w/	11%	16%	15%	11%
	/l/	12%	17%	11%	15%

An ANOVA equivalent to that in Experiments 10 and 11 was carried out. There was no significant influence on RTs of previous phonotactic alignment in either German or English. Target words that had not been aligned with a phonotactic boundary according to German or English in Experiment 11 were detected as fast as targets that had been aligned. Although the difference among conditions may seem rather large, the variability of the data is such that no main effects or interactions reached significance by both participants and items. It is worth noting that the data are normally distributed and therefore the patterns are not due to outliers.

ANCOVAs were then performed on the data of Experiments 10 and 11, using the RTs from Experiment 12 as covariates. With the latencies and error rates of the lexical-decision experiment as covariates, the alignment effects found in Experiments 10 and 11 remained. For German listeners there was still a significant influence on RTs of both German phonotactics ($F_2(1, 58) = 10.66$, $p = .002$) and English phonotactics ($F_2(1, 58) = 9.21$, $p = .004$). Similarly, error analyses revealed a significant effect of both German ($F_2(1, 58) = 10.24$, $p = .002$) and English phonotactics ($F_2(1, 58) = 16.12$, $p < .001$). RTs of English listeners still showed a significant influence of English phonotactics ($F_2(1, 58) = 17.53$, $p < .001$), but no significant influence of German phonotactics ($F_2(1, 58) = 3.12$, $p > .08$). Error analyses, however, again revealed a significant effect of both English ($F_2(1, 58) = 15.67$, $p > .001$) and German phonotactics ($F_2(1, 58) = 5.05$, $p < .03$).

General Discussion

The results of Experiment 10 show clearly that both native and non-native phonotactic constraints influence lexical segmentation of a non-native language. German listeners were faster and more accurate to detect English words embedded in nonsense sequences when the words were aligned with syllable boundaries cued by English and/or German phonotactic constraints than when the words lacked such alignment. Native language phonotactic boundary cues were as helpful for segmentation as were non-native phonotactic boundary cues. There were alignment effects for words with initial /l/ and for words with initial /w/. Whereas the former is a phoneme of German, the latter is not. For words with initial /w/, German listeners used boundary cues for segmentation as if the English /w/ were German /v/ (as well as the boundary cues that /w/ marks in English).

In Experiment 11, English listeners were presented with the same speech stimuli. The results show clearly that English speakers' segmentation is influenced by English phonotactic constraints. For both words with initial /w/ and /l/, detection was faster and more accurate when words were aligned with a phonotactic boundary according to the phonotactics of the listeners' native language. Because the English listeners had no knowledge of German, German phonotactics did not significantly influence the listeners' detection times.

In Experiment 12, English listeners were presented with target words from Experiment 11 without their original contexts. The results of this lexical-decision experiment suggest that the acoustic differences in the target words across contexts were not sufficient to account for the phonotactic alignment effects found in Experiments 10 and 11. It appears that the differences in Experiments 10 and 11 between spotting words which were aligned with a phonotactic boundary and spotting those which were not clearly aligned were due to the nature of the contexts of the words rather than to some acoustic property of the words themselves.

The results of the present study support earlier findings that the legality of phoneme sequences is used to help solve the segmentation problem. Previous research has shown that Dutch listeners use native phonotactic cues to segment spoken Dutch (McQueen, 1998), and studies in language acquisition found that even infants show sensitivity to phonotactic legality constraints (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001). Nine-month-olds prefer to listen to speech that meets the sequencing constraints of their language over speech which does not (Jusczyk et al., 1993). They also prefer legal over illegal word boundary clusters within their own language (Friederici & Wessels, 1993). There is also robust evidence now that adult listeners are sensitive to finer grained differences in probabilities of acceptable sequences and not simply to gross differences between legal and illegal patterns (see for example Hauser, Newport, & Aslin, in press; McQueen & Pitt, 1996; Saffran, Newport, & Aslin, 1996; van der Lugt, in press; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Yip, 2000).

The present data also provide further support for the influence of the native language on the segmentation process in a non-native language. Previous studies found that the process of segmentation in a non-native language is influenced by metrical cues and vowel harmony information from the native language. Both English and French listeners show sensitivity to specific native metrical cues in speech not only when listening to their native language but also when listening to a non-native language (Cutler et al., 1986). Similarly, listeners who learned to segment an artificial language, performed best when the phonological properties of vowel harmony and word-initial stress were the same in the artificial language and their native language (Vroomen et al., 1998).

But whereas tests of metrical segmentation and vowel harmony to date have only examined the influence of either native or non-native cues, testing phonotactic constraints allows a direct comparison of the influences of native and non-native cues. A phoneme sequence can either provide the same phonotactic cue for two languages, or it can provide different cues for two languages. Therefore, the present study allowed for comparison of the degree to which both native and non-native phonotactic constraints influence the segmentation of a non-native language. Word spotting in a non-native language requires high proficiency in the non-native language. The results show that listeners who are highly proficient in the non-native language have indeed learned to use non-native phonotactic boundary cues for segmentation. However, though irrelevant, listeners continue to use native phonotactic boundary cues as well for segmenting non-native speech. This could possibly even lead to misunderstandings when, for example, German listeners, under the influence of native phonotactics, are more likely to turn a 'gray twig' into a 'great wig'.

As outlined in the introduction, current models of spoken-word recognition assume that spoken-word recognition can be achieved via a process of competition between lexical candidates. The competition process, however, can be enriched with additional information about where in the speech signal word boundaries are most likely to occur. Therefore, the use of language-specific phonological information for segmentation does not contradict the importance of competition. Multiple cues, such as metrical cues, vowel harmony and phonotactic constraints, can act to bias the activation and competition process. A word-spotting experiment in English has in fact revealed combined effects of competition and metrical structure (Norris et al., 1995). Words were harder to detect in StrongStrong strings (e.g., *mask* in /maskʌk/) than in StrongWeak strings (e.g., *mask* in /maskək/). Stress pattern effects emerged most clearly when the second syllable activated many instead of few competing candidate words. The number of competitors for the second syllable therefore modulated the metrical effect. Similarly, Vroomen and de Gelder (1995) observed both competition and metrical segmentation in Dutch with a cross-modal priming task. Lexical decisions to targets like *melk*, 'milk', were slower when the second syllable in spoken StrongStrong strings was consistent with many words (e.g., /mɛlkaːm/) than when it was consistent with few words (e.g.,

/mɛlkøːm/). Decision times were fastest when there were no second syllable competitors in StrongWeak strings (e.g., /mɛlkəm/).

Competition models of spoken-word recognition can readily model the use of phonotactic cues, metrical information and vowel harmony for segmentation. In essence, the activation levels of word candidates that are misaligned with a syllable onset in the input are penalized. Computer simulations demonstrate that the Shortlist model can account for the joint effects of number of lexical competitors and of metrical cues to likely word boundaries (Norris, 1994; Norris et al., 1995, 1997). Different boundary cues could also provide different degrees of assistance in segmentation. McQueen found, for instance, that once phonotactics signaled a syllable boundary, the metrical structure could not provide any further boundary information (McQueen, 1998). Again, these results could be simulated with the Shortlist model (Norris et al., 1997).

Shortlist, however, can only model native spoken-word recognition. In fact, there is, to my knowledge, no functioning model for non-native spoken-word recognition yet. Numerous studies have shown that processing of a non-native language is influenced by the phonological structure of the native language (see for example Cutler et al., 1986, 1992; Otake et al., 1996; Weber, 2001). These results suggest that a competition model for non-native spoken-word recognition would have to implement processes that are sensitive to language-specific phonological structure. In addition, the results of the present study suggest, that at least for phonotactic constraints such a model would have to implement both native and non-native phonotactics.

A system that is sensitive to what sequences tend to co-occur could be the basis for initial acquisition of phonotactics of the native language. If the system could continue to learn, it would start to show effects of the non-native language, without necessarily loosing the native language. This would mean that, for example, less proficient non-native listeners might not yet show an effect of the phonotactics of the non-native language in segmentation. On the other hand, highly proficient non-native listeners who stopped using their native language, might only show weak effects of the phonotactics of the native language in segmentation (cf. Flege & Frieda, 1997; Meador, Flege, & MacKay, 2000).

The only attempt so far to model non-native spoken-word recognition is BIMOLA (Bilingual Model of Lexical Access), but this model is still in development (Grosjean, 1988, 1997). One assumption of the BIMOLA model is that listeners can move quite frequently between varying points of a language mode continuum with a monolingual and bilingual language mode as its endpoints. In the monolingual language mode, one language is strongly activated, whereas the other language is activated only very weakly. In the bilingual language mode, both languages are activated, but one more than the other. The results of the present study challenge this assumption, since no increased activation for either language was found, rather both native and non-native phonotactic boundary cues were equally used for segmentation in the non-native language. BIMOLA is currently being implemented as a computer model (Léwy & Grosjean, in preparation). It remains to be established, however, whether the results of Experiment 10 could be simulated by the model.

In conclusion, the results of the present study confirm previous findings that listeners use phonotactic constraints to identify likely word boundaries in continuous speech. This study also confirms that the process of listening is language-specific. It was shown that phonotactic constraints of a non-native language influence listeners' segmentation strategies to the same degree as native phonotactic constraints do. Thus, listeners not only make use of phonotactic information about their own language, they can also use information they have learned about another language, in which they are proficient but not native, in order to segment speech of that language.

The role of acoustic cues for word segmentation in spoken English[6]

CHAPTER 3B

Abstract

This study investigates the influence of acoustic cues on the segmentation of spoken English. In a previous study (Chapter 3a), both English and German listeners detected words in nonsense sequences more easily when they were clearly aligned with a phonotactic boundary in English (e.g., *length* in /fun.lɛŋθ/ or /zarʃ.lɛŋθ/) than when they were not clearly aligned (e.g., *length* in /fuklɛŋθ/ or /joʊslɛŋθ/). Acoustic cues to boundaries could also have signaled word boundaries, especially when word onsets lacked phonotactic alignment in English. In this study first, all nonsense sequences from Chapter 3a without clear phonotactic alignment were re-recorded with two intended syllabifications (e.g., /fuk.lɛŋθ/ and /fu.klɛŋθ/). Three durational measurements were found to vary systematically with the intended syllabification. Next, these potential boundary cues were measured in the stimuli from Chapter 3a. However, only one acoustic boundary cue showed a significant correlation with the reaction times of both English and German listeners. The results suggest that word segmentation in English is influenced primarily by phonotactic constraints and only secondarily by acoustic aspects of the speech signal.

---

[6] *Adapted version of article published in Proceedings of the 6th International Conference on Spoken Language Processing (Weber, 2000).*

Introduction

The present study investigates the extent to which acoustic cues to word boundaries may have influenced the segmentation of spoken English in Chapter 3a. In the word-spotting experiment in Chapter 3a, both English and German listeners were presented with English speech stimuli. Their task was to detect any English word embedded in a list of nonsense sequences. The onset of the embedded word was either aligned with a clear syllable boundary according to English phonotactics (e.g., *length* in /fun.lɛŋθ/ or /zarʃ.lɛŋθ/) or not (e.g., *length* in /fuklɛŋθ/ or /joʊslɛŋθ/).[7] In English, /nl/ and /ʃl/ are not legal consonant clusters within syllables and therefore the sequences /fun.lɛŋθ/ and /zarʃ.lɛŋθ/ require a syllable boundary at the onset of the embedded word *length*. On the other hand, /kl–/ and /sl–/ are possible syllable onsets in English, so both the sequences /fuklɛŋθ/ and /joʊslɛŋθ/ do not require a boundary at the onset of *length*. A sequence like /fuklɛŋθ/ can be syllabified as either /fu.klɛŋθ/ or /fuk.lɛŋθ/. Both English and German listeners found it easier to spot words that were aligned with a phonotactic boundary than words that lacked such alignment.

Because a sequence like /fuklɛŋθ/ allows two syllabifications, a speaker might use acoustic cues to signal a word boundary in the absence of clear phonotactic alignment. Numerous previous studies have shown that word boundaries can be acoustically marked (see for example Lehiste, 1960, 1972; Nakatani & Dukes, 1977). Furthermore, some studies have shown the use of such acoustic boundary cues for segmentation. Quené (1992, 1993) found, for instance, that Dutch listeners use durational cues to word boundaries when asked to choose between two alternative readings of an ambiguous two-word utterance such as [dipɪn] (*diep in*, 'deep in', or *die pin*, 'that pin').

Did listeners in the word-spotting experiments in Chapter 3a use any potential acoustic boundary cues for segmentation? In sequences where English phonotactics do not force a boundary (e.g., /fuklɛŋθ/ and /joʊslɛŋθ/), the speaker may have unconsciously marked boundaries with durational cues (other than silent intervals), even though she was instructed

---

[7] *In the present study alignment according to German phonotactics is not considered (see Chapter 3a).*

to avoid any clear syllable boundaries in her production, and listeners may have used those cues for segmentation.

The present study addresses this question in two parts: First, acoustic measurements investigated how speakers realize syllabification differences, and second, correlation analyses of the established acoustic cues with the results of the word-spotting experiments in Chapter 3a were carried out.

Acoustic cues

Method

The target-bearing nonsense sequences from the word-spotting experiment that lacked phonotactic alignment were recorded again by the same female native speaker of American English, who is a trained phonetician. Items that were excluded from the analysis in the word-spotting experiment were not re-recorded. The speaker produced all sequences twice, with different intended syllable boundaries (e.g., /fu.klɛŋθ/, /fuk.lɛŋθ/ and /joʊ.slɛŋθ/, /joʊs.lɛŋθ/). The intended boundary can only be varied where phonotactics do not force a boundary because otherwise the speaker would have to produce onset clusters that do not occur in the language (e.g., /nl/ in */fu.nlɛŋθ/). Silent intervals (i.e., pauses) within the item were avoided, since silence would be a very strong boundary cue. The closure portion preceding the burst of a stop was not regarded as a pause.

Using the Xwaves speech analysis software, several potentially relevant durations were measured. Because different final consonants were used for the nonsense syllables within an alignment condition, and embedded words started with either /l/ or /w/ (see Chapter 3a, subsection Method), most acoustic measurements did not apply to all stimuli. The first measure, duration of the first syllable vowel, was however, determined for all items (122 pairs). After voiceless obstruents, vowels were defined as beginning at the onset of voicing. After voiced segments, onset of the vowels was defined as onset of the second and third formant. When the vowel in the nonsense syllables was followed by the approximant /r/ (51 times out of 122), the approximant was included in the vowel duration because the speech signal does not show a clear boundary between the vowel and /r/. Vowels were expected to be longer when word onsets were misaligned with syllable onsets (e.g., /fu.klɛŋθ/).

Second, voice onset time (VOT) was measured for the clusters /pl/, /kl/, /tw/, and /kw/ (64 pairs) from the beginning of the burst to the onset of voicing. VOT was expected to be longer when word onsets were aligned with syllable onsets (e.g., /fuk.lɛŋθ/). Third, fricative duration was measured for the clusters /fl/, /sl/, and /sw/ (58 pairs) from the end of voicing of the preceding vowel to the end of high frequency frication noise. Fricative duration was expected to be longer when word onsets were aligned with syllable onsets (e.g., /fɔs.wɛpən/). Fourth, duration of the voiced portion of /l/ was measured for the clusters /pl/, /kl/, /fl/, and /sl/ (68 pairs) because /l/ is often partially devoiced after an aspirated stop. Duration was measured from the onset of voicing to a rise in the second formant and an increase in amplitude of the third formant. Duration of voiced portion of /l/ was expected to be longer when word onsets were aligned with syllable onsets. Finally, the wave forms and spectrograms of sequences containing the clusters /pl/, /kl/, and /tw/ (64 items) were checked for signs of glottalisation.[8] Stop consonants that were intended as syllable-final (e.g., /fuk.lɛŋθ/) were expected to be subject of glottalisation.

Results

Slight glottalisation was visible in the spectrograms for only two sequences with the cluster /tw/ and not in any items with any other clusters. That the two cases of glottalisation were in /tw/ clusters is consistent with the patterns of glottalisation Pierrehumbert (1994) found; that is, glottalisation is more likely for /t/ before /w/, whereas, for example, /p/ is not glottalized before /l/. However, the low frequency of occurrence of glottalisation in the present study rendered any further investigation of this measurement unnecessary. A spectrogram showing glottalisation in a sequence with a stop consonant that was intended as syllable-final and a spectrogram showing no glottalisation are given in Figure 3−1 and Figure 3−2. Mean durations of the remaining acoustical measurements for the productions with differing intended syllabifications are given in Figure 3−3.

---

[8] *Thanks to Ken Stevens for suggesting this acoustic boundary marker.*

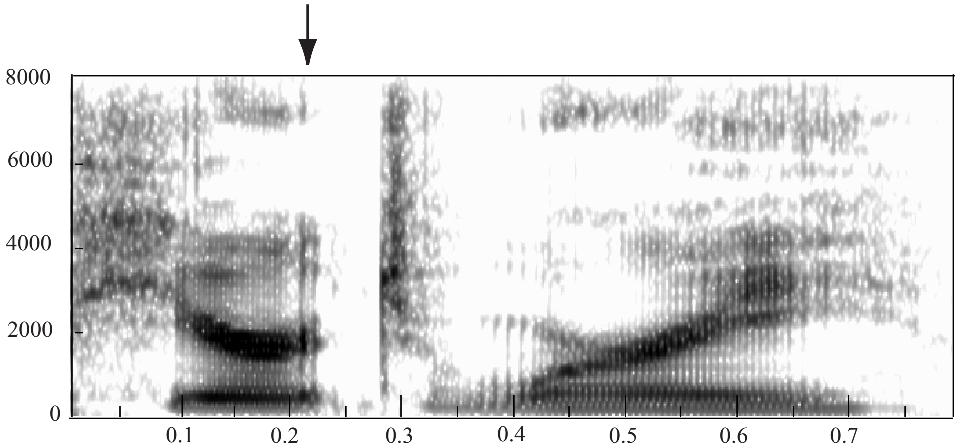

Figure 3–1. *Spectrogram of the sequence /pirtwəri/ showing glottalisation. The stop consonant /t/ was intended as syllable-final by the speaker. The arrow marks the glottalisation.*

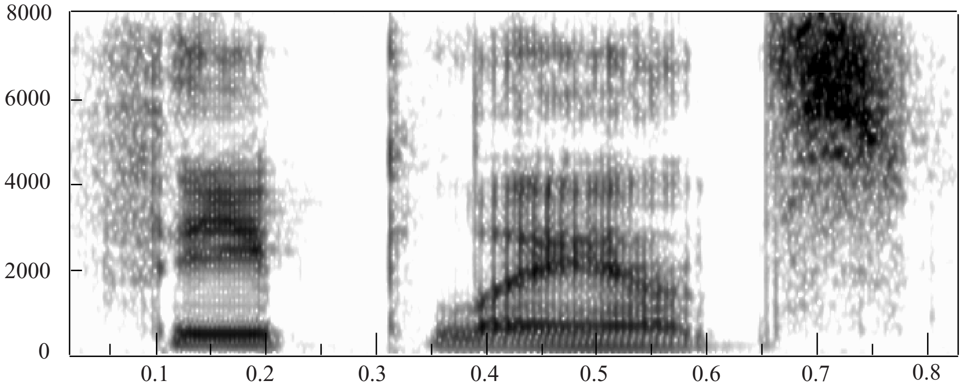

Figure 3–2. *Spectrogram of the sequence /θiplæns/ showing no glottalisation. The stop consonant /p/ was intended as syllable-final by the speaker.*

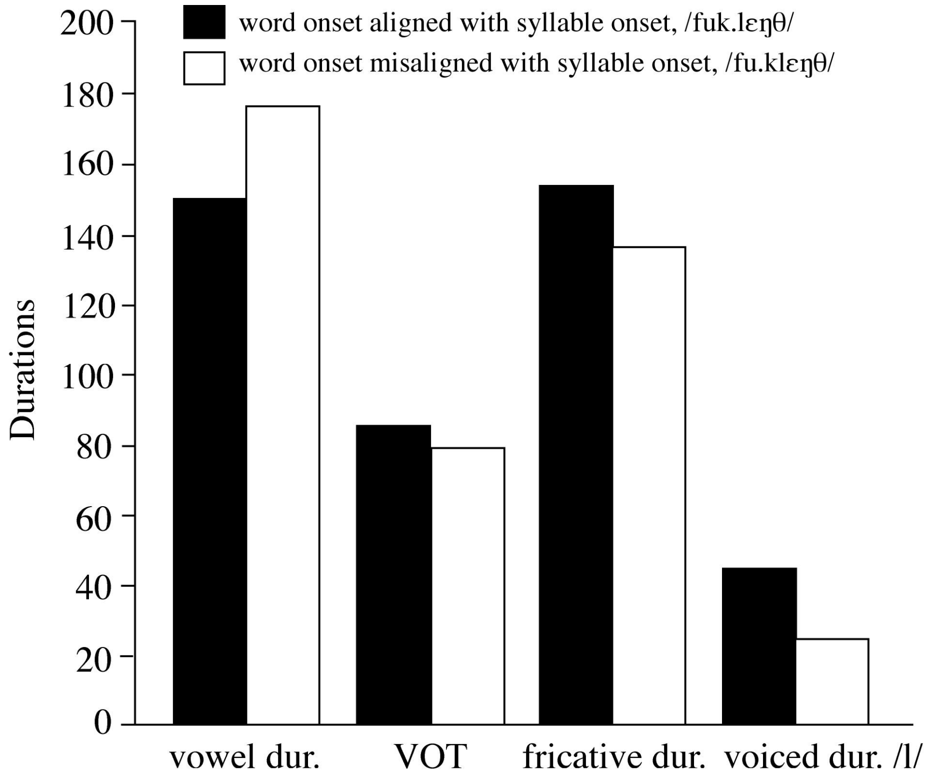

Figure 3‒3. *Mean durations in ms for the two intended syllabifications.*

Two factor mixed ANOVAs were carried out for each of the four measurements, with consonant cluster as a between-items factor, and intended syllabification with the two levels 'word onset aligned with syllable onset' and 'word onset misaligned with syllable onset' (e.g., /fuk.lɛŋθ/ and /fu.klɛŋθ/) as a within-items factor. First, the vowel duration ANOVA produced highly significant effects of intended syllabification, but no interaction with consonant cluster ($F(1, 115) = 116.89$, $p < .001$). The duration of the first syllable vowel was significantly longer when the word onset was intended to align with the syllable onset. Second, for VOT, no significant effect of intended syllabification was found ($F(1, 60) = 3.01$, $p > .08$). Third, in the ANOVA for fricative duration there was a significant interaction between intended syllabification and consonant cluster; accordingly separate ANOVAs were performed for the three consonant clusters containing fricatives. Significant effects of intended syllabification were found for the clusters /fl/ ($F(1, 10) = 24.89$, $p = .001$) and /sl/ ($F(1, 33)$

= 47.77, $p < .001$), both with longer fricatives when an alignment of word onset and syllable onset was intended, but not for /sw/ ($F(1, 12) = 1.13$, $p > .3$). Fourth, the duration of the voiced portion of /l/ was significantly longer when an alignment of word onset and syllable onset was intended ($F(1, 64) = 126.92$, $p < .001$). In summary, three durational measurements, first syllable vowel duration, fricative duration, and duration of the voiced portion of /l/ were found to vary systematically with the intended syllabification.

Correlation analyses

To investigate whether listeners used any of these potential acoustic boundary cues for segmentation in the word-spotting experiment, the three durational measurements in Figure 3−3 that showed a significant difference (first syllable vowel duration, fricative duration, and duration of voiced portion of /l/) were measured in the stimuli from the word-spotting experiment without phonotactic alignment. Correlation analyses for duration measurements (of the speech signals used for the word-spotting experiment) with RTs were then performed for the three measures. Note that what is investigated here is the correlation between durational measurements of the actual stimuli from Chapter 3a with RTs from the same chapter. Thus, the data of the two intended syllabifications above was used only to determine which acoustic features might provide boundary cues. Whenever a durational cue aligned the onset of a embedded word with a syllable onset, RTs were expected to be faster, if listeners indeed make use of the durational cues for segmentation. One would predict, for example, that RTs were faster the shorter the vowel in the first syllable, since shorter vowel duration has been established for word onsets that are aligned with syllable onsets. Similarly, one would predict shorter RTs for longer fricative durations and longer durations of the voiced portion of /l/.

The first measure, 'vowel duration of the first syllable', failed to show a significant correlation with RTs either from English or German listeners in the word-spotting experiment (English listeners: $r(122) = -.01$, $p > .8$; German listeners: $r(122) = -.09$, $p > .2$). Because the measurements above show that vowel duration does vary with intended syllabification, it is a potential cue. The reason why no correlation was found for vowel duration could lie in the parameters chosen for the measurements. When the vowel in the nonsense syllables was followed by the approximant /r/ (51 times out of

122), the approximant was included in the vowel duration. Consequently vowel durations differed noticeably depending on whether they included a following approximant or not.

The second measure, 'fricative duration', was also established as an acoustic difference the speaker uses to signal a boundary when intentionally manipulating boundary locations but it too failed to show a significant correlation with RTs for the clusters /fl/ and /sl/ for either English or German listeners (English listeners: $r(45) = .12$, $p > .4$; German listeners: $r(45) = .02$, $p > .9$). The cluster /sw/, which showed no significant durational difference for the two intended syllabifications, was excluded from this analysis.

The third measure, the duration of the voiced portion of /l/, showed a marginally significant negative correlation with RTs for English listeners ($r(68) = -.23$, $p = .058$; see Figure 3−4). For German listeners the correlation was fully significant ($r(68) = -.41$, $p = .001$; see Figure 3−5). As predicted a longer duration of the voiced portion of /l/ marked the lateral as syllable initial for both English and German listeners, so that the onset of the embedded word was aligned with this onset and thus easier to spot. When duration of voiced portion of /l/ and first syllable vowel duration were combined in a linear regression analysis with RTs, a significant correlation was found for German listeners ($F(2, 67) = 6.57$, $p = .003$) but not for English listener ($F(2, 67) = 2.05$, $p > .1$).

Using direct RTs to items for the correlation analyses means including frequency effects of the lexical item in the analysis. RTs to a high frequency word are known to be faster than RTs to a low frequency word. In order to exclude such frequency effects, all correlation analyses were in addition carried out with subtracted RTs, for which RTs to an item in the not clearly aligned context were subtracted from RTs to that item in the clearly aligned context. Only the correlation for German listeners of subtracted RTs with duration of voiced portion of /l/ was significant.

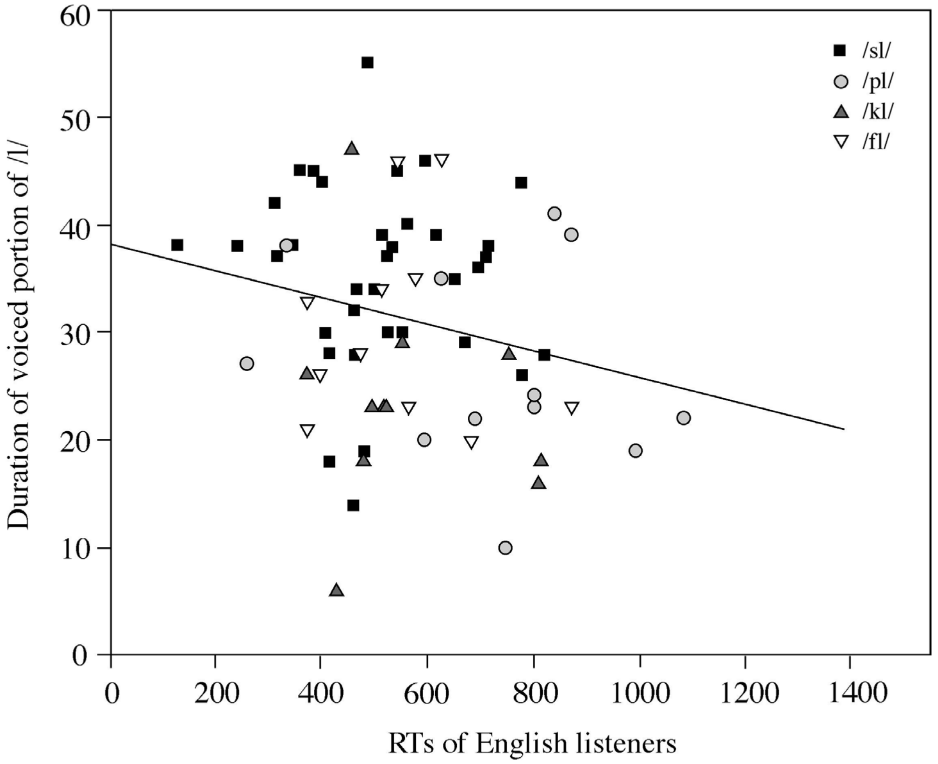

Figure 3–4. *Scatter plot of duration of voiced portion of /l/ with RTs of English listeners from the word-spotting experiment in Chapter 3a.*

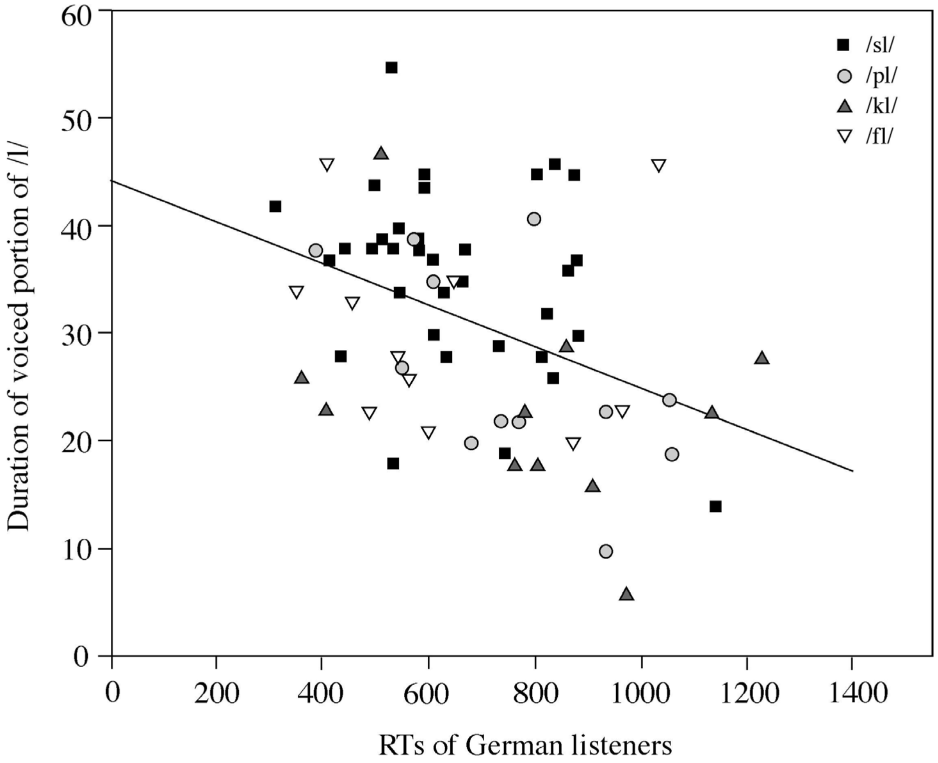

Figure 3–5. *Scatter plot of duration of voiced portion of /l/ with RTs of German listeners from the word-spotting experiment in Chapter 3a.*

General Discussion

The study in Chapter 3a showed that phonological information influences the process of spoken-word segmentation in English. Both English and German listeners find it easier to detect English words in nonsense sequences when the words are clearly aligned with a phonotactic boundary than when they are not clearly aligned. However, previous research has shown that various acoustic features, particularly segment durations, vary systematically with word boundary location, and that listeners are sensitive to such differences (see for example Dumay et al., 1999; Quené, 1992, 1993). The speaker may have marked boundaries in the stimuli for the experiments in Chapter 3a with such acoustic cues (other than silent intervals), and listeners might have used those acoustic cues for segmentation.

In the present study, acoustic analyses of nonsense sequences with two intended syllabifications were carried out in order to establish which durational parameters a speaker may use to signal word boundaries in the absence of phonotactic alignment. Vowel duration of the first syllable, fricative duration, and duration of voiced portion of /l/ were found to vary systematically with the intended syllabification in recorded nonsense sequences. These three durational parameters were then measured in the speech signals of the word-spotting experiment stimuli. However, only one durational measurement, namely duration of voiced portion of /l/, correlated with the RTs from the word-spotting experiment. Participants' perceived word segmentation may still have been affected by other acoustic boundary markers not measured in this analysis, but the parameters measured here represent the most likely options.

In general, acoustic cues may be too variable and too small to be used extensively by listeners for lexical segmentation (see for example Nakatani & Dukes, 1977). In consequence, acoustic cues other than silence may provide relatively weak assistance in segmentation. We know that speakers produce large acoustic differences when they intend different syllabifications, and listeners can probably make use of these differences (see for example Quené, 1992, 1993). But that does not mean that speakers produce these acoustic differences also in normal speech, or even in careful speech when they are not thinking about syllable boundaries. The present study showed that when speakers do not intentionally produce a particular syllabification, listeners make little use of the potential acoustic cues for segmentation. However, to what extent the stimulus productions contained potential acoustic boundary markers cannot be established. Because the speaker was not instructed to produce a particular syllabification for the stimuli that were used for the word-spotting experiments in Chapter 3a (and hence for the correlation analyses in the present chapter), no direct comparison between these stimuli and the same stimuli with a different indented syllabification was possible. Different boundary cues provide different degrees of assistance in segmentation. Phonotactic boundaries may be more powerful because they are reliable when present, whereas the observed durational cues may be less powerful segmentation cues because they are gradient and speaker and speech rate dependent.

Eye movements and the activation of native competitors in non-native spoken-word recognition

CHAPTER 4

Abstract

In this study, native Dutch participants followed spoken instructions in English to click on pictures using a computer mouse, while their eye movements were monitored. A target picture (e.g., the picture of a desk) was presented along with distractor pictures. The Dutch name of a distractor picture was either phonologically related to the English name of the target picture (e.g., English target *desk* /dɛsk/ and Dutch competitor *deksel*, 'lid', /dɛksəl/) or it was phonologically unrelated (e.g., *bloem*, 'flower' or *schommel*, 'swing'). Participants fixated distractor pictures with phonologically related Dutch names more than distractor pictures with phonologically unrelated names. The results demonstrate that listeners consider candidates of the native language for recognition of a non-native word, even though the native candidates are irrelevant for the task.

Introduction

When a Dutch listener who knows English is listening to an English sentence that contains the word *desk* /dɛsk/, is the Dutch word *deksel* /dɛksəl/, which is phonologically similar, considered as a candidate during the spoken-word recognition process? Extensive empirical evidence supports the claim that words sharing initial segments in the native language are briefly activated during the recognition of spoken words (see for example Marslen-Wilson & Zwitserlood, 1989; McQueen, Norris, & Cutler, 1994; Norris et al., 1995; Tanenhaus et al., 1995). For example, given the input *desk*, English listeners will initially activate *desk* and *dentist* among other candidates, which will then compete against each other for recognition (Allopenna, Magnuson, & Tanenhaus, 1998). Competition between candidate words is the core assumption of models of native language spoken-word recognition like the Cohort model (Marslen-Wilson, 1987), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994). The assumption of competition presumably holds as well for spoken-word recognition when listening to a non-native language that one understands well. However, when one listens to a non-native language, words from the native language might also share initial segments with the input. Candidate words of the listener's native language could compete for recognition with candidate words of the non-native language. Using the eye tracking paradigm, the present study investigates whether native candidate words are activated during the recognition of non-native spoken words and if so what the time course of that activation is with respect to the unfolding speech stream.

The eye-tracking paradigm makes use of the fact that participants make saccadic eye movements to either real objects or pictures of objects on a screen immediately after the names of the objects are mentioned in spoken instructions. It has been shown that the eye movements are closely time-locked to the referring expressions in the unfolding speech stream (for an overview of the paradigm see Tanenhaus & Spivey-Knowlton, 1996). Locations and latencies of eye movements to pictures can therefore be used to examine lexical access in spoken-word recognition. Because recording eye movements allows one to monitor the ongoing comprehension process as spoken language unfolds over time, eye movements can be used to evaluate the time course of competition effects in spoken-word recognition.

Tanenhaus et al. (1995) presented American English speaking participants with a display of objects that sometimes included two objects with initially similar names (e.g., *candy* and *candle*) and instructed them to move the objects around. They found that the mean time to initiate an eye movement to the correct object (e.g., *candy*) was longer when an object with a phonologically similar name (e.g., *candle*) was included in the display than when no such object was included. Several studies have since replicated this competition effect. For French, Dahan, Swingley, Tanenhaus, and Magnuson (2000) showed that participants fixated pictures with names sharing initial sounds with the target (e.g., target *boutons*, 'buttons', and competitor *bouteilles*, 'bottles') more than pictures with phonologically unrelated names. Allopenna et al. (1998) found clear activation not only for competitors with names that share the onset with a target (e.g., target *beaker* and onset competitor *beetle*) but also for competitors that share the rhyme with a target (e.g., target *beaker* and rhyme competitor *speaker*). Allopenna and colleagues then compared the average probabilities of fixations on the pictures with fixation probabilities derived from activations of the names of the pictures from a number of TRACE simulations. The general patterns of fixations clearly followed the general patterns predicted from the simulations using TRACE. Similarly Dahan, Magnuson, and Tanenhaus (in press) compared fixation proportions with fixation probabilities derived from TRACE activations over time for targets and competitors and found very similar shapes of functions. The results suggest that the competition effects as determined by fixation proportions to pictures can indeed be closely mapped onto activation levels of word candidates over time.

However, these studies investigated competition effects in listeners perceiving speech in their native languages. There is evidence from the visual domain, using lexical decision tasks, that word recognition in a non-native language is sensitive to phonological cross-language similarity. In an English visual lexical decision task, Dutch-English bilinguals were slower to reject nonwords that were cross-language pseudohomophones (e.g., the English nonword *SNAY* pronounced according to English spelling-to-sound conversion rules sounds like the Dutch word *snee*, 'slice') than regular nonwords (e.g., *ROLM*; Nas, 1983). Doctor and Klein (1992) found inhibitory effects of cross-language homophony for English-Afrikaans bilinguals in a lexical decision task. Similarly Dijkstra, Grainger, and van Heuven (1999) found inhibitory effects of phonological cross-language overlap in a lexical decision task with Dutch-English bilinguals. Bijeljac-Babic, Biardeau, and Grainger (1997) found inhibition of English target word recognition for French-English bilinguals not only when orthographically related English primes were presented, but also when orthographically related French primes were presented (see also Brysbaert, van Dyck, & van de Poel, 1999 for Dutch-French bilinguals). These studies have in common that they showed the influence of the native lexicon on the processing of a non-native language without actually presenting a stimulus of the native language. They demonstrated that bilinguals cannot deactivate the lexicon of the native language even when they are in a monolingual non-native situation where the native language is irrelevant.

There are currently no similar studies using auditory rather than visual lexical-decision experiments. However, a small number of studies investigated non-native spoken-word recognition using the gating paradigm, in which listeners were asked to identify words in a sentence on the basis of increasing fragments of the word (Grosjean, 1988; Li, 1996). They found activation of both the native and the non-native lexicon. These studies were conducted in code-switching situations where the participant was listening to speech input mixed from both languages. In a code-switching situation, however, one would not even expect the listener to deactivate one lexicon while using the other.

Eye tracking is a useful paradigm for investigating the topic of cross-language competition because it allows one to test the activation of the lexicon of the irrelevant language without necessarily presenting an auditory stimulus from that language. Furthermore, eye tracking allows one to investigate lexical activation over time as information from the acoustic input becomes available. Just as competition effects were shown with eye tracking for phonologically similar names of objects within a language, competition effects can be investigated for phonologically similar names of objects across languages, without presenting the competitor auditorily. Spivey and Marian (1999) monitored eye movements of Russian-English bilinguals to objects that were displayed on a table. The participants lived in the U.S., and English had been their primary language for an average of 4 years. Participants differed with respect to their self-reported language preferences at the time of the study. In separate sessions, participants were instructed in Russian and English to move objects on a table. In one condition in the Russian session, the target object was accompanied by an object whose English name shared initial sounds with the Russian target (e.g., Russian target *marku*, 'stamp' and English competitor *marker*). In another condition, the cross-language competitor was replaced by an unrelated distractor whose name bore no similarity to the target (e.g., Russian *lineika*, 'ruler'). Similarly, in the English session the English target object was either accompanied by a Russian competitor or not. Across the two sessions, participants made more eye movements to the cross-language competitors than to the unrelated distractors on average. However, when analyzed separately, significant competition from English items during the Russian session was found, but no significant competition from Russian items during the English session. In a follow-up study, Marian and Spivey (1999) found the mirror reverse pattern. They explained this asymmetry with manipulations of the language mode during the experiment and general language preferences of the participants. In the Marian and Spivey (1999) study, more effort was made to put the participants into a Russian language mode by additionally playing popular Russian songs at the beginning of the Russian session.

In addition to replicating the cross-language competition effects of Spivey and Marian (1999) and Marian and Spivey (1999) for another language pair, namely Dutch and English, the present study investigates the time course of lexical activation more closely. The use of an eye tracker allows for measurement of proportions of fixations to the different objects over time, providing information about the temporal dynamics of lexical access. With this method, one can locate the exact point in time at which a difference between fixations to the cross-language competitor and to the unrelated distractor emerges, for how long this effect lasts and what the course of it is. Furthermore, unlike in the previous studies, Dutch was always the primary language of the participants, who had learned English as a second language in school. Although the experiment was conducted completely in English, no long-term shifts in language mode were expected because the participants lived in the Netherlands at the time of the experiment. (Reduced use of native language over a longer period of time can cause a shift in the base language from the native language to the second language, which has been found to positively affect word recognition and production in a second language; cf. Flege, Frieda, & Nozawa, 1997; Meador et al., 2000). Furthermore, the participants were not aware of the relevance of the Dutch language for the experiment. This approach allows one to test exclusively for activation of native competitors in non-native spoken-word recognition, without involving listeners who may be more fluent in their second language than their first language. Robust competition effects from the native language were expected. Competition effects from the native language would suggest that listeners do not deactivate lexical candidates from the native language during the recognition of non-native spoken words.

The participants of the experiment were native Dutch listeners who were also highly proficient in English. Most educated Dutch speakers are very good at English, but are definitely Dutch-dominant. They were instructed in spoken American English to click on pictures of objects on a computer screen and then drag the pictures on top of a geometric shape on the screen. A target picture (e.g., the picture of a desk) was always presented along with three distractor pictures. The Dutch name of a distractor picture was either phonologically related to the American English name of the target picture (e.g., English target *desk* /dɛsk/ and Dutch competitor *deksel*, 'lid', /dɛksəl/) or it was phonologically unrelated (e.g., *bloem*, 'flower' or *schommel*, 'swing'). As the initial sounds of the target words were heard, the Dutch competitors were expected to be fixated more than the unrelated distractors, as a consequence of their phonological similarity in Dutch with the initial portion of the input. Assuming that the probability of fixating a picture reflects activation of the lexical representation associated with this picture, more fixations on the Dutch competitor than on the unrelated distractors would show that the Dutch competitor was activated during the presentation of the English target word.

Experiment 13: Dutch listening to instructions in English

Method

Participants. Twenty students from the University of Nijmegen took part in the experiment for monetary compensation. They were native speakers of Dutch who had lived in the Netherlands all their lives, and had normal or corrected-to-normal vision and normal hearing. They had received an average of 7.45 years of training in English as a foreign language in secondary education beginning at a mean age of 11.25.

Participants were asked to take a multiple-choice test in English after completing the eye-tracking experiment. For 20 nouns (none of which occurred in the eye-tracking experiment), they had to choose the correct definition out of three possibilities. The definitions for the nouns were taken from the Longman Dictionary of Contemporary English (1987). Most false definitions described nouns that were either phonologically or semantically related to the target noun (e.g., the definition for *brunch* was an option for the phonologically related target word *branch*, the definition for *fountain* was an option for the semantically related target word *river*). The average score was 97% correct. Thus, the particiants were highly proficient in English.

Materials. The target words consisted of 20 English nouns referring to picturable objects (e.g., *desk* /dɛsk/). Each target word was paired with a Dutch competitor. The onset of the competitor in Dutch overlapped phonemically with the onset of the target word in English (e.g., English target word *desk* /dɛsk/ and Dutch competitor *deksel*, 'lid', /dɛksəl/). Phonemic overlap stretched over two or three segments, with one exception of five segments, and was based on American English and Standard Dutch. There was no other onset overlap within or between items. For example, the name of the target item in English did not overlap with the name of that item in Dutch (e.g., the English target word *desk* /dɛsk/ is *bureau* /byro/ in Dutch). Neither did the name of the target word in English overlap with the English name of the Dutch competitor (e.g., target word *desk* /dɛsk/ and Dutch competitor *deksel* translated into English *lid* /lɪt/). Two phonologically unrelated distractors were added for each target word (e.g., *flower* and *swing*). Neither the English nor the Dutch names of the unrelated distractors (e.g., *bloem* /blum/, 'flower' /flaʊər/ and *schommel* /sxɔməl/, 'swing' /swɪŋ/) overlapped with the English target word. The pictures of a target item, its competitor and two unrelated distractors were displayed together in one trial set. The English target word was actively named in the spoken instructions, whereas the competitor and the unrelated distractors were not named. The 20 English target words, their Dutch competitors and unrelated distractors are listed in Appendix 4, p. 162.

To prevent participants from developing expectations that pictures with phonologically similar names were likely targets, 20 additional filler trials were constructed, consisting of four items each (e.g., *candle*, *ashtray*, *dress*, *pig*). In the filler trials, no phonemic overlap occurred between the names of the items in either language. For example, when the picture of a *candle* was displayed, neither the English nor the Dutch names of the other three items in that trial had initial /k/. Six representative trials were constructed as practice trials.

The pictures of the items were selected from the Snodgrass and Vanderwart (1980) and Cycowicz, Friedman, Rothstein, and Snodgrass (1997) picture sets, as well as from the *Art Explosion library* (1995). All pictures were black and white line drawings. In order to establish naming norms, 10 native speakers of Dutch were asked to name the target pictures and their competitors in Dutch and English. The agreement between participants' responses and the intended names was 91% in Dutch and 85% in English. In addition, 10 native speakers of Dutch and 10 native speakers of English were asked to rate the goodness of the pictures as pictures of the intended object, named in their language, on a scale from zero to seven. Dutch participants rated the goodness of the pictures with a mean of 5.8. English participants rated the goodness of the pictures with a mean of 6.1. Some suggestions of the participants for improvement of the pictures were taken into consideration. None of the participants from the naming or rating experiments took part in the eye-tracking experiment.

Previous research has shown that the probability of fixating a competitor that matches the acoustic information of the target word varies with its lexical frequency (Dahan et al., in press). To control potential frequency confounds in the present study, the lexical frequencies of the targets and of the competitors were counted using the CELEX database (Baayen et al., 1993). Lemma frequencies of the targets in Dutch (39.71 per million) and the competitors in Dutch (37.07 per million), as well as English word form frequencies of the targets (42.01 per million) and Dutch word form frequencies of the competitors (24.55 per million) were computed.[9] For 13 of the 20 pairs, the name of the English target had a higher frequency than the name of the Dutch competitor in the word form count. Statistical analyses revealed no significant difference between the frequency of the target words and the frequency of the competitors, either for lemma frequencies or for word form frequencies (both F−values < 1). Dahan et al. (in press) also showed that the probability of fixating unrelated distractors, which do not match the acoustic information of the target word, does not vary with lexical frequency. Therefore, in the present study lexical frequencies of the unrelated distractors were not taken into consideration.

---

[9] *Lemma frequencies represent the sum of the appropriate word form frequencies of a lexical entry (e.g., singular plus plural word form frequencies of a noun).*

The spoken instructions were recorded onto DAT in a soundproof booth by a male native speaker of American English, sampling at 48 kHz. The materials were then down-sampled to 16 kHz during transfer to a computer. Durations of the preceding contexts and the target words were measured using the Xwaves software. An instruction was, for example: "Click on the desk. Now put it on top of the diamond". The average duration of the preceding context ("click on the") was 451 ms, and that of the target word (e.g., "desk") was 575 ms long. In addition, the duration of the phonemic overlap between the English target word and its Dutch competitor (e.g., the duration of /dɛ/ in *desk* with competitor *deksel*, /dɛksəl/) was measured. Some vowels, mostly diphthongs, that differ somewhat between the two languages (e.g., the diphthong in English *bike* /baɪk/ and Dutch *bijl*, axe, /bɛil/) were considered as overlapping for the measurements. The average duration of overlap was 270 ms.

Procedure. The experiment was controlled by a Compaq 486 computer. Pictures were presented on a ViewSonic 17PS screen, and the auditory stimuli were presented over headphones using the NESU experiment control software. Participants' eye movements were monitored using a SMI EyeLink-Hispeed 2D eye-tracking system. Two cameras on a lightweight headband provided the input to the tracker. The center of the pupil and the first Purkinje image (corneal reflection) were tracked to determine the position of the eye relative to the head. Throughout the experiment, the computer recorded the onset and offset times and the spatial coordinates of the participants' fixations. The signal from the eye tracker was sampled every 4 ms. Both eyes were monitored, but only the data from the right eye were analyzed.

Participants were tested individually. At the beginning of a session they received written instructions in English, which included an example of a trial display and an explanation of the task. Participants were then seated in a quiet room approximately 60 cm in front of a monitor. After the eye tracker was calibrated, each participant was presented with the 46 trials (6 practice trials plus 20 experimental trials plus 20 filler trials). Every experimental trial was preceded by a filler trial. All pictures were presented as white line drawings on a blue background on a 5 × 5 gray grid. In each trial four line-drawing pictures and four green geometric shapes, each scaled to fit into a cell of the grid, and a cross centered in the middle, appeared on the screen (see Figure 4−1). Each cell measured 4.3 × 4.3 cm, corresponding to a visual angle of approximately 4°, which is well within the resolution of the tracker (0.1°). The positions of the target object and its competitor were randomized across trials. The positions of the geometric shapes were fixed, and participants were told this in advance. Spoken instructions started simultaneously with the appearance of the pictures on the screen. Participants were first asked to click on one of the four pictures using the mouse (e.g., "Click on the desk."), and then to move the picture on top of one of the four geometric shapes (e.g., "Now put it on top of the diamond."). Once this was accomplished, the experimenter initiated the next trial.

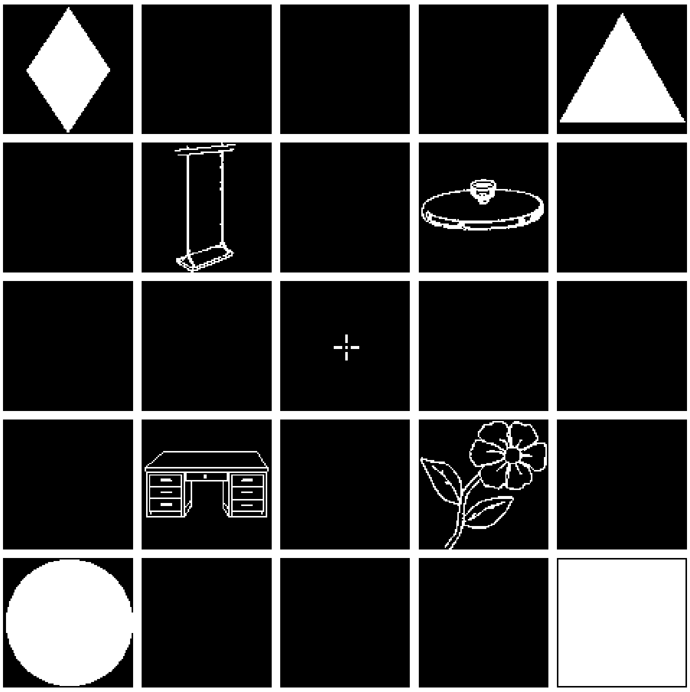

Figure 4‒1. *Example of stimulus display presented to participants in Experiment 13.*

Following Dahan et al. (2000, in press), the set of pictures was not shown to the participants before the experiment. Furthermore, there was no delay between the appearance of the pictures on the screen and the beginning of the spoken instructions. This procedure makes it less likely that participants have implicitly named the pictures beforehand. Neither were participants instructed to fixate the cross at the beginning of the trials. Therefore, participants could be fixating any of the four objects or the cross at the onset of the target word. (Only very rarely did participants fixate any other location on the screen.)

After every five trials a fixation point appeared centered on the screen, and participants were instructed to look at it. The experimenter could then correct potential drifts in the calibration of the eye tracker. The experiment lasted approximately 10 minutes.

Results

Graphical software was used to display the locations of the participants' fixations as dots superimposed on the four line drawings for each trial and each participant. Onset times and durations of the fixations were displayed in another window. Fixations on the line drawings were coded as pertaining to the target object, the competitor, or one of the two unrelated distractors. Fixations that lay clearly outside the cell of an object were coded as zero. (Usually these fixations lay on the cross in the center of the screen.) For each trial, fixations were coded from the onset of the target word until the participant had clicked with the mouse cursor on the target picture. Six trials had to be removed from the analysis, because participants clicked on an object other than the target object (1.5% of all trials). The proportions of the fixations were analyzed in 10 ms slices to provide fine-grained information about the time course of lexical activation as the speech unfolded. Figure 4−2 presents the proportions of fixations averaged over participants to the target, the competitor and the average for the two unrelated distractors in 10 ms time slices from 0 to 1000 ms after target onset.

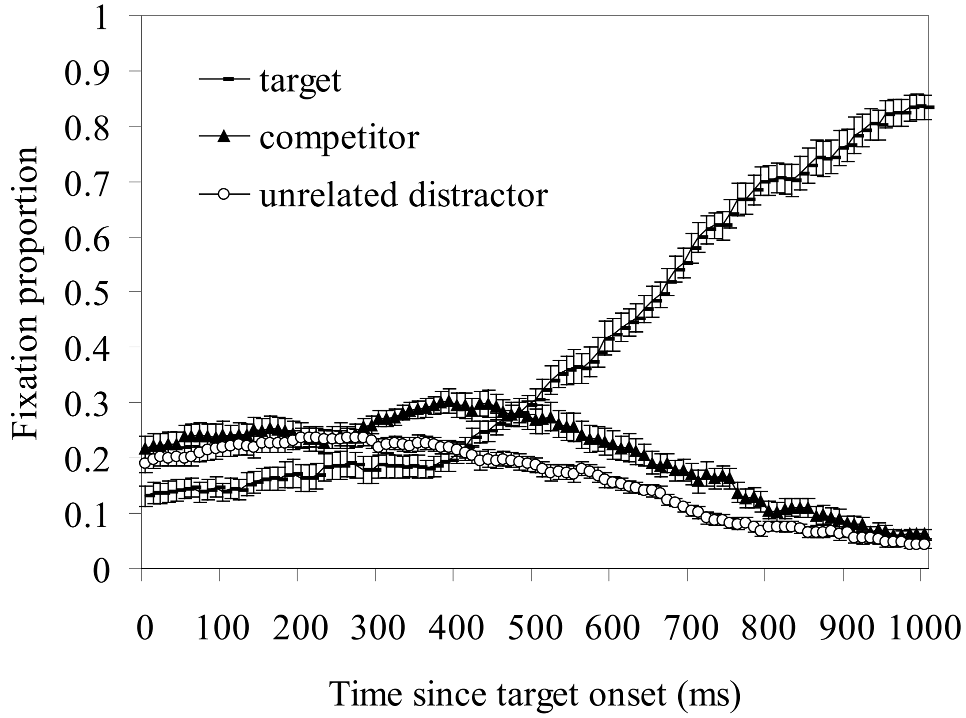

Figure 4‒2. *Fixation proportions over time for the target, the competitor and the averaged distractors in Experiment 13. Bars indicate standard errors.*

The graph shows that the probability of fixating the Dutch competitor began to diverge from the probability of fixating the unrelated distractors about 300 ms after target word onset. The probability of fixating the Dutch competitor remained greater than that of the unrelated distractors until approximately 800 ms after target word onset. It is estimated that an eye movement is typically programmed about 200 ms before it is launched (Matin, Shao, & Boff, 1993). Thus, 300 ms after target onset is approximately the point at which fixations driven by the first 100 ms of acoustic information from the target word can be seen. To compare the proportions of fixations to the competitor and to the average for the two unrelated distractors, a time window extending from 300 to 800 ms after target onset was defined. Both Allopenna et al. (1998) and Dahan et al. (2000) found competition effects that started at 300 ms. Over the 300 to 800 ms time window the proportion of fixations was 23.3% to the Dutch competitor and 16.3% to the average of the unrelated distractors. A one-factor ANOVA on the mean proportion of fixations was conducted over the 300 to 800 ms time window, with picture (with the two levels 'competitor' and 'unrelated distractors') as the within-participants factor. The competitor was fixated significantly more often than the unrelated distractors ($F_1(1, 19) = 15.76$, p = .001; $F_2(1, 19) = 5.62$, $p < .03$). This suggests that during the presentation of the English target word the Dutch competitor was activated.

In addition, a one-factor ANOVA was carried out over the 300 to 800 ms time window when the two unrelated distractors were not averaged. This was done to check for possible differences in the fixation proportions for the two unrelated distractors. The within-participants factor picture had now the three levels 'competitor', 'first unrelated distractor' and 'second unrelated distractor'. Proportion of fixations were 23.3% to the Dutch competitor, 17.9% to the first unrelated distractor, and 14.7% to the second unrelated distractor. The main effect of picture was significant by participants and items ($F_1(2, 38) = 11.64$, p < .001; $F_2(2, 38) = 4.37$, $p = .02$). Newman-Keuls tests indicated that the proportion of fixations to the competitor was higher than that to both the first unrelated distractor and the second unrelated distractor, but the proportion of fixations to the first unrelated distractor did not differ from that to the second unrelated distractor. This suggests that the competitor was fixated more than both individual unrelated distractors.

However, differences in fixations to pictures could also be influenced, for example, by different visual complexity of the pictures, making some pictures more interesting to look at than others. To examine differences in fixations to pictures before any acoustic information from the target word could influence eye movements, an ANOVA was conducted on the fixations to the target, the competitor and the average of the unrelated distractors over a time window extending from 0 to 300 ms. The fixation proportions over the first 300 ms after target onset differed significantly only by participants ($F_1(2, 38) = 8.63$, $p = .001$; $F_2 < 1$). Newman-Keuls tests indicated that the proportion of fixations to the target was lower than that to both the competitor and the unrelated distractor, but the proportion of fixations to the competitor did not differ from that to the unrelated distractor. This suggests that the difference found between fixations to the competitor and to the unrelated distractor in the 300 to 800 ms time window cannot be attributed to a general bias toward the picture of the competitor.

The graph presenting the proportions of fixations to the different pictures also shows an advantage for the competitor over the target until approximately 500 ms after target word onset (see Figure 4−2). Over a 300 to

500 ms time window this difference was significant by participants but not by items ($F_1(1, 19) = 8.14$, $p = .01$; $F_2 < 1$). The lower proportion of fixations to the target could partly be due to the general bias against the target pictures, found prior to target word onset. However, there at least two other possible explanations for this difference in fixation proportions. First, the difference between fixations to the target and the competitor might be a subjective frequency effect across languages. Although no significant difference has been found between the lexical frequency of the targets and the competitors, the Dutch participants presumably have heard the Dutch word *deksel* significantly more often in their lives than the English word *desk*. In this sense, the Dutch competitors had a much higher frequency than the English targets. A second explanation involves the phonemic overlap across the two languages. In Dutch, only the names of the competitors and not the name of the targets overlapped with the incoming acoustic information (e.g., the English target *desk* /dɛsk/ is *bureau* /byro/ in Dutch). The English target might have been activated with a delay because in the native language of the participants, the acoustic information did not match the name of the target picture. In the time window between 300 and 400 ms, the fixation proportion of the competitor rises whereas the fixation proportion of the averaged unrelated distractor falls, which shows that acoustic information of the target word is being processed at that time. However, the fixation proportion of the target does not start to rise until about 400 ms, at which point the fixation proportion of the competitor starts to fall. This suggests that the target is not considered as a candidate before 400 ms. It could be the case that just after the incoming acoustic information did not overlap with a Dutch word anymore, the English target word became activated. The average duration of phonemic overlap between the English targets and the Dutch competitors was 270 ms. An advantage of the competitor can indeed be seen until approximately 470 ms after target onset in Figure 4–2, which matches the 270 ms overlap plus the approximately 200 ms it takes before a programmed eye movement is launched. Whether there is indeed a short delay between the activation of native and non-native competitors, or whether the observed effect is simply due to a difference in subjective frequency, is still a matter of further research.

General Discussion

This study explored the recognition of non-native spoken words. In particular, it focused on whether Dutch listeners activate Dutch competitors during the recognition of English words. In order to assess the activation of Dutch words, Dutch participants were presented with a four-picture display and spoken English instructions asking participants to click on and move one of the pictures (e.g., the picture of a desk). Presented together with the target picture were a competitor picture whose Dutch name shared the initial sounds of the English target word and two unrelated distractor pictures whose names were not phonologically related to the target. Eye movements to a picture were interpreted as evidence for activation of the lexical representation associated with this picture. As the English target word unfolded over time, the Dutch competitor was fixated significantly more than the unrelated distractors. This demonstrates that during the presentation of the English target word the Dutch competitor was activated as a result of its phonological similarity to the target.

The Dutch participants were living in the Netherlands at the time of the experiment, leading their daily lives in Dutch. The robustness of the native competitor effect for non-native word recognition can probably be attributed to the strong role of the native language in the participants' daily lives. As the results by Spivey and Marian (1999) and Marian and Spivey (1999) suggest, a change in language mode and language preference can reduce the competition effect. Competition effects of the less preferred language or the language that is not the base language, tend to be weaker.

The results of the present study support the claim that during the processing of a spoken non-native language both the non-native and the native language are activated (see for example Grosjean, 1988; Li, 1996; Marian & Spivey, 1999; Soares & Grosjean, 1984; Spivey & Marian, 1999). Listeners do not appear to be able to deactivate the native mental lexicon even when they are in a monolingual non-native situation where the native language is irrelevant and possibly even misleading.

The data of the present study also seem to suggest that the activation of native and non-native candidates does not start simultaneously. Non-native candidates were activated with a delay in time compared to native candidates. However, since the present study is the first study that shows the time course of activation of cross-language competition, further research is necessary to investigate whether indeed the activation of native candidates starts prior to the activation of non-native candidates.

Models for non-native word recognition must take the processing of two languages into account. Currently the BIA (Bilingual Interactive Activation) model (Dijkstra & van Heuven, 1998), a model of visual-word recognition, is the only implemented model of non-native word recognition. The BIA model consists of four layers of nodes: letter features, letters, words, and languages. An important feature of the model is that initially word candidates from both languages are activated. The word nodes activate then their language nodes, and language nodes send top-down inhibition to word nodes from the other lexicon. The language nodes can thereby collect activation of all words from one lexicon and suppress words from the other lexicon.

Spoken language, in contrast to written language, is a temporal signal, and models of auditory word recognition have to take this difference into account. At this moment there is, however, no functioning model of non-native auditory word recognition. Both the results of the present study and previous findings suggest that a competition model for non-native spoken-word recognition would have to include lexical items from both languages, with activation of one language set not suppressed during recognition of words from the other language. (Although the present results suggest that there might be a timing difference in the activation of words from both languages.) The only attempt so far to model non-native spoken-word recognition is BIMOLA (Bilingual Model of Lexical Access), but this model is still in development (Grosjean, 1988, 1997). BIMOLA distinguishes a feature level that is common to both languages, and a phoneme and a word level, each with two independent subsets for both languages. The subsets for the two languages are then enclosed in one larger subset. One assumption of the BIMOLA model is that language nodes are not necessary. Instead of language nodes, top-down preactivation based on external information about the listener's language mode and higher linguistic information activates words of the appropriate lexicon. BIMOLA is currently being implemented as a computer model (Léwy & Grosjean, in preparation). It remains to be established whether the results of Experiment 13 could be simulated by this model. The general issue of whether a model of non-native word recognition requires the flow of top-down information, as it is assumed in BIMOLA, or whether bottom-up processes are sufficient, still needs to be investigated.

Models of native spoken-word recognition differ with respect to top-down information flow (Norris, McQueen, & Cutler, 2000).

In conclusion, the results of the present study confirm the activation of the native language during non-native spoken-word recognition using the eye tracking paradigm. How frequently participants looked at distractor objects was a function of the objects' names in the native language even though listeners knew their native language was irrelevant to the task. The results demonstrate that listeners initially consider candidates of the native language for recognition of a non-native word, even though the native candidates might be irrelevant for the task, and that bilingual listeners who know more than one language cannot choose which language to use in parsing speech, but instead activate relevant words from both languages.

Conclusions

CHAPTER 5

Understanding speech requires the recognition of individual words (or more exactly of lexically represented units) in the continuous speech stream. Listeners have to solve two problems in order to achieve word recognition: the problem of variability and the problem of segmentation. The word recognition process is already complex enough for listeners with utterances in their native language; it is of course even harder when the input is in a foreign language. Different factors, related to the speaker, the listener or the spoken items themselves have been shown to influence non-native word-recognition (see for example Bradlow & Pisoni, 1999). This thesis investigated how listeners process both native and non-native languages, and whether phonological differences between languages bring about differences in the way speech is processed. In this final chapter, first the empirical results of the thesis are summarized and discussed with respect to their implications for theories and models of non-native spoken-word recognition, and second, the results of simulations of some non-native experimental data with Shortlist are described, and resulting implications for models of non-native spoken-word recognition are discussed.

Summary of results

The experiments in Chapters 2a and 2b investigated how violation of different obligatory assimilation rules affects both native and non-native spoken-language processing. In Experiments 1 and 2, both German and Dutch listeners heard Dutch monosyllabic nonwords containing the velar fricative [x], which undergoes assimilation in German but not in Dutch. German listeners detected the velar target fricative [x] faster when nonwords violated the obligatory German progressive fricative assimilation rule (e.g., *[bɛxt]) than when no such violation occurred (e.g., [bɑxt]; Chapter 2a). Dutch listeners showed no such effect, since none of the materials violated Dutch phonology. The results demonstrated that listeners are sensitive to the violation of an assimilation rule of their native language, even while listening to some other, non-native, language. All sounds in the materials belonged to the inventories of both Dutch and German. Only the combination of sounds made some sequences language-specific. Whereas the sequence [ɛx], for example, is permissible within syllables in Dutch, it is not a possible sequence of German. Listeners could not suppress the influence of native phonology, and consequently processing of the non-native language was influenced by violations of a native assimilation rule. German listeners in Chapter 2a had no knowledge of Dutch. Therefore, it is not surprising that their performance was not influenced by the phonology of the non-native language. However, the results suggest that when listeners encounter a new language, rather than turning off all language-specific strategies for processing, they continue to make use of the phonological structure of their native language. In Experiment 3, the facilitation effect for violation of German progressive fricative assimilation was replicated for native listening with German materials (monosyllabic and bisyllabic nonwords).

However, in previous studies on native listening, participants found it harder rather than easier to process violations of assimilation (Gaskell & Marslen-Wilson, 1996, 1998; Koster, 1987; Kuijpers & van Donselaar, in preparation; Otake et al., 1996; Quené et al., 1998). Therefore, an experiment on native listening was carried out to clarify the origin of the facilitation in Chapter 2a. The cause of facilitation in the present experiments was found in two factors. First, the current studies tested a progressive assimilation rule rather than the regressive ones previous studies had used. Progressive assimilation (i.e., an earlier segment affects a later one) creates weak restrictions about the monitoring target (the later segment), whereas regressive assimilation (i.e., a later segment affects an earlier one) creates strong restrictions. It was argued that only the defeat of strong restrictions (violation of regressive assimilation) can cause inhibition. Second, the sequences violating German assimilation (e.g., *[ɛx]) were novel sequences for German listeners. The sequences do not even occur across word boundaries in German. It seems that this novelty, in combination with only weak expectations being violated, facilitated detection (novel popout). No facilitation was found when either one of the two factors proposed as causing facilitation did not apply. Violation of a German regressive assimilation rule caused inhibition rather than facilitation of processing. German participants detected stop consonants more slowly in sequences that violated regressive nasal place assimilation (e.g., *[fɛnk] and *[fɛmk]) than in sequences that contained no such violation. Violation of the German progressive fricative assimilation rule using the palatal fricative [ç] rather than the velar fricative [x] also did not result in facilitation (Chapter 2b). Although a sequence like *[aːç] also violates German progressive fricative assimilation, it is not a truly novel sequence because the sequence can occur across word boundaries in German (e.g., *sah China*, 'saw China', [zaː#çina]).

Obviously, the strength of restrictions and the novelty of the sequence influence the processing of phonotactically illegal sequences. Most current models of speech perception treat all phonotactically illegal sequences (sequences with zero transitional probability) equally, whether they are impossible in all environments or only in the one in question. The results in Chapters 2a and 2b, however, suggest that not all these sequences are processed in the same way. Whereas listeners may only rarely encounter sequences that are truly novel when they listen to their native language (e.g., speech errors or foreign accented speech), they may encounter sequences that are novel in their native language quite regularly when they listen to a non-native language. Therefore, the findings of Chapters 2a and 2b (together with previous findings on phonologically conditioned variation) are of special interest for models of non-native spoken-language processing. The results suggest that models of non-native spoken-language processing should be sensitive to the violation of phonologically conditioned variation in the listener's native language. In a more specific way, the results suggest that sequences which are phonotactically illegal in the listener's native language do not necessarily inhibit the processing of a non-native language, but may even facilitate it.

The experiments in Chapter 2c investigated whether violation of phonotactic constraints (rather than violation of assimilation rules) influences both native and non-native phoneme detection, also. Native German and Dutch listeners detected the alveolar stop /t/ in German nonwords. German listeners had no knowledge of Dutch, but Dutch listeners had a fair knowledge of German. Half of the German nonwords violated a German phonotactic constraint (e.g., [stɪm]), the other half violated a Dutch phonotactic constraint (e.g., [ʃtɪm]). German listeners detected /t/ faster when a native phonotactic constraint was violated. This suggests that German phonotactics mark a syllable boundary right before the target phoneme, which makes it easier to detect /t/. Dutch listeners detected /t/ equally fast whether it was preceded by /s/ or by /ʃ/. Although Dutch listeners might be sensitive to both German and Dutch phonotactics, a follow-up experiment suggests that this may not be the reason for the lack of a difference in RTs because Dutch listeners had problems distinguishing German /s/ and /ʃ/ in a follow-up experiment. Whereas the results of the native listeners in Chapter 2c clearly indicate sensitivity to phonotactic constraints in phoneme detection, the results of the non-native listeners are less clear.

The experiments in Chapter 3 investigated the influence of phonotactic constraints on the segmentation of a non-native language. Native listeners of German who had an excellent knowledge of English detected English words in nonsense sequences (Chapter 3a). Word onsets were either clearly aligned with a syllable boundary or not, according to the phonotactics of German and English. In one condition, neither English nor German forced a syllable boundary at the onset of the embedded word (e.g., *length* in /fuklɛŋθ/). (Both /kl–/ and /l–/ are permissible syllable onsets in English and German.) In another condition, English, but not German, required a syllable boundary at the onset of the word (e.g., *length* in /zarʃlɛŋθ/). (/ʃl–/ is a permissible syllable onset in German, but not in English.) In a third condition, German, but not English, required a syllable boundary at the onset of the word (e.g., *length* in /joʊslɛŋθ/). (/sl–/ is a permissible syllable onset in English, but not in German.) In a fourth condition, both German and English required a syllable boundary at the word onset (e.g., *length* in /funlɛŋθ/). (/nl–/ is not a permissible syllable onset in either language.) Words clearly aligned with either German or English phonotactic boundaries were easier for German listeners to detect than words without such clear alignment. English boundary cues were as helpful for segmentation as were German boundary cues. English listeners who had no knowledge of German were presented with the same English stimuli. Responses of English listeners were influenced primarily by phonotactic alignment in English.

The results of Chapter 3a show clearly that both native and non-native phonotactic constraints influence lexical segmentation of a non-native language. In contrast with those in Chapter 2a, the German listeners in Chapter 3a were highly proficient in the non-native language. Listeners in Chapter 3a still used phonotactic cues supplied by their native language for segmenting the non-native language, even though the boundary cues specified by the native language may harm word recognition in the non-native language. However, listeners did not rely exclusively on native language phonotactic boundary cues for the segmentation of a non-native language, rather they were able to make use of phonotactic boundary cues specific to the non-native language, too. Previous studies found that the process of segmentation in a non-native language is also influenced by other cues from the native language, such as metrical cues and vowel harmony (see for example Cutler et al., 1986; Cutler & Otake, 1994; Otake et al., 1993; Vroomen et al., 1998). For models of non-native spoken-word recognition this suggests that processes should be implemented which are sensitive to a number of native phonological boundary cues. However, the results of Chapter 3a further suggest that at least for phonotactic boundary cues, such a model would have to implement non-native phonotactic cues, as well.

In Chapter 3b, the extent to which acoustic cues to word boundaries might have influenced segmentation in Chapter 3a was investigated. In sequences in Chapter 3a where English phonotactics did not force a boundary (e.g., /fuklɛŋθ/ and /joʊslɛŋθ/), the speaker may have unconsciously marked boundaries with acoustic cues, and listeners may have used those cues for segmentation. Previous studies have shown that word boundaries can be acoustically marked (see for example Lehiste, 1960, 1972; Nakatani & Dukes, 1977), and that listeners can use such acoustic boundary cues for segmentation (Dumay et al., 1999; Quené, 1992, 1993). In Chapter 3b, acoustic measurements on new recordings investigated how the speaker for the word-spotting experiment realized syllabification differences when producing a particular syllabification intentionally (e.g., /fu.klɛŋθ/ versus /fuk.lɛŋθ/ and /joʊ.slɛŋθ/ versus /joʊs.lɛŋθ/). Three durational measurements (first syllable vowel duration, fricative duration, and duration of the voiced portion of /l/) were found to vary systematically with the intended syllabification. These three durational measurements were then taken on the stimuli from the word-spotting experiment without phonotactic alignment. For the word-spotting experiment the speaker did not intentionally produce a particular syllabification. Correlation analyses of the acoustic measurements of the stimuli of the word-spotting experiments in Chapter 3a with the reaction times experiments were carried out. Only duration of voiced portion of /l/ correlated with the RTs from the word-spotting experiments for both English and German listeners, and that only weakly. The results suggest that when speakers do not intentionally produce a particular syllabification, word segmentation in English is influenced primarily by phonotactic constraints and only secondarily by acoustic aspects of the speech signal.

Whereas the experiments in Chapters 2 and 3 were primarily about low-level effects in processing speech, in Chapter 4 high-level effects involving the lexicon were investigated. Experiment 13 in Chapter 4 investigated whether listeners activate native candidate words along with non-native candidate words during the process of recognizing spoken words in a non-native language. Dutch participants followed spoken instructions in English to click on pictures using a computer mouse, while their eye movements were monitored. A target picture (e.g., the picture of a desk) was presented along with distractor pictures. Some distractor pictures had names in Dutch with the onset phonologically similar to the English name of the target picture (e.g., English target *desk* /dɛsk/ and Dutch competitor *deksel*, 'lid', /dɛksəl/), and others were phonologically unrelated to the English target (e.g., *bloem*, 'flower' or *schommel*, 'swing'). Eye movements to a picture were interpreted as evidence for activation of the lexical representation associated with this picture. As the spoken English target words unfolded over time, the Dutch competitors with similar onsets to the target word were fixated significantly more than the unrelated distractors. The results support the claim that during the processing of a non-native spoken language, both lexical entries of the non-native and the native language are activated (see for example Grosjean, 1988; Li, 1996; Marian & Spivey, 1999; Soares & Grosjean, 1984; Spivey & Marian, 1999). Listeners do not appear to be able to deactivate the native mental lexicon even when they are in a monolingual non-native situation where the native language is irrelevant and possibly even misleading. The results suggest that models of non-native spoken-word recognition have to include lexical items from both languages, with activation of one language set not suppressed during recognition of words from the other language.

Simulations with Shortlist

Every model of spoken-word recognition has to do justice to both the stability and the flexibility of human performance. As outlined in the introduction, numerous studies have shown that listeners can, for example, exploit the phonological structure of their native language for the recognition of speech in that language (see for example Costa, et al., 1998; Gaskell & Marslen-Wilson, 1996, 1998; van Donselaar et al., 1999). In other words, language-specific phonological structure can positively influence the stability with which listeners understand their native language. The core assumption of models of native spoken-word recognition is a competition process between candidate words. However, in the light of the available findings, such models should also implement processes that are sensitive to the phonological structure of the listener's native language in addition to the process of competition. Shortlist is one model of native spoken-word recognition that indeed allows phonological information to modulate the competition process. In this section, a preliminary attempt to model non-native spoken-word recognition based on Shortlist is reported.

The results of the present thesis, together with previous findings (see for example Cutler et al., 1986; Koster, 1987; Otake et al., 1996), support the claim that models of non-native spoken-word recognition should also consider phonological structure, both that of the listener's native language and that of the non-native language. Further, such models should allow access to the lexicons of each language.

The advantage of models that have been computationally implemented is that they can be used to simulate experimental data. Shortlist is such a model of native spoken-word recognition, and it includes sensitivity to the phonological structure of the listener's native language. Previous simulations with Shortlist have shown that the model can capture the simultaneous operation of competition and phonotactic effects in the processing of a native language (Norris et al., 1997). McQueen (1998) showed that Dutch listeners find it easier to detect Dutch words in nonsense sequences when the word onsets are aligned with a phonotactic boundary marked in the input (e.g., *rok*, 'skirt', in /fim.rɒk/) than when they are misaligned (e.g., *rok* in /fi.drɒk/; the voiced stop /d/ must be syllable initial in Dutch). Simulations with Shortlist using the experimental stimuli as input showed that the activation levels of target words which were aligned with phonotactic boundaries (e.g., *rok*, 'skirt', in /fim.rɒk/) were considerably higher than those of target words which were misaligned with boundaries (e.g., *rok* in /fi.drɒk/). Shortlist can also successfully simulate other word boundary effects, such as effects of metrical information (see Norris et al., 1995, 1997).

In Shortlist, processing is carried out segment by segment. Every time a new segment is presented to the model, the evidence in the signal is re-evaluated, and a new shortlist of candidate words is generated. Depending on the degree of match, words may be added to the shortlist or deleted from it. The Possible Word Constraint (PWC; Norris et al., 1997) in Shortlist disfavors parses that leave an impossible residue between the end or beginning of a candidate word and a known boundary. An impossible residue is anything that is not a syllable. A boundary can be marked by silence, but also by other cues, such as phonotactic boundary cues. For instance, the word *apple* is penalized in the nonsense sequence /fæpəl/ because the parse would leave the single consonant /f/, which is an impossible residue, between *apple* and the closest boundary (which is the silence preceding the nonsense sequence). Similarly, the Dutch word *rok*, 'skirt', is penalized in the nonsense sequence /fi.drɒk/ because the closest boundary is marked by phonotactics before the voiced stop /d/, and parsing *rok* would therefore leave the single consonant /d/. On the other hand, *rok* in the nonsense sequence /fim.rɒk/ is not penalized, since the onset of *rok* is aligned with the closest phonotactic boundary which is right before the /r/: Parsing *rok* would therefore leave no impossible residue. However, Dutch words with initial /ɒ/, which are misaligned with a boundary are penalized in /fim.rɒk/, because they would leave the single consonant /r/ as impossible residue. Because candidates which are misaligned with a boundary are penalized, the PWC makes it easier for words which are aligned with a boundary to win the competition and be recognized. Since Shortlist is able to simulate phonotactic boundary effects for native word recognition, it may in principle also be able to simulate the phonotactic boundary effects for non-native word recognition found in Chapter 3a. In the absence of a computationally implemented model of non-native spoken-word recognition, Shortlist was used to simulate the recognition of the materials from Chapter 3a.

In Chapter 3a, German listeners detected English words in nonsense sequences. Word onsets were either clearly aligned with a syllable boundary or not, according to the phonotactics of English and German. Words clearly aligned with either English or German phonotactic boundaries were easier for German listeners to detect than words without such clear alignment. That is, words which were not clearly aligned with a syllable boundary according to at least one of the two languages were harder for German listeners to detect than words which were clearly aligned according to either language or both languages. For the simulations, the nonsense sequences with embedded target words from Chapter 3a (see Appendix 3, p. 159) were transcribed phonemically. The transcriptions were then used as input to the model. In the experiments in Chapter 3a, detection of the target word was delayed when the onset of a target word was not clearly aligned with a phonotactic boundary in either language. Therefore, only word boundaries that were clearly phonotactically determined according to both English and German were cued in the input with a marker (e.g., *length* in /fun.lɛŋθ/; /nl/ is not permissible within syllables in both English and German). In cases where either both languages, or only German, or only English, did not clearly determine a phonotactic boundary, the nonsense sequences were not marked in the input (e.g., *length* in /fuklɛŋθ/, /zarʃlɛŋθ/, /joʊslɛŋθ/). The English lexicon that was used for the simulations was based on British English, and therefore some small adjustment had to be made to the original stimuli because they were recorded in American English (e.g., American English /oʊ/ was now transcribed as British English /əʊ/).

The length of the target words in Chapter 3a varied between two and six segments. Shorter words will be recognized earlier by Shortlist than longer words because all the information the model needs for recognition will be available earlier. Phonotactic effects on word recognition occur during processing and not after the word is recognized. If mean activation functions were computed for all words ranging from two to six segments, the shorter words would weaken any phonotactic effect for the longer words. Therefore, the results below represent only the 46 target words with three and four segments (out of the 61 that were analyzed in Chapter 3a).

The performance of the model was compared segment by segment for the target words, followed by three slices of silence. The model was run using the PWC as described in Norris et al. (1997), on an English lexicon containing more than 25,000 entries. The mean activation functions for the target words with three and four segments (e.g., /wik/ and /lɛŋθ/) are shown in Figure 5. Simulations were also run separately for target words that were shorter than three segments or longer than four. The activation functions looked similar to those in Figure 5.

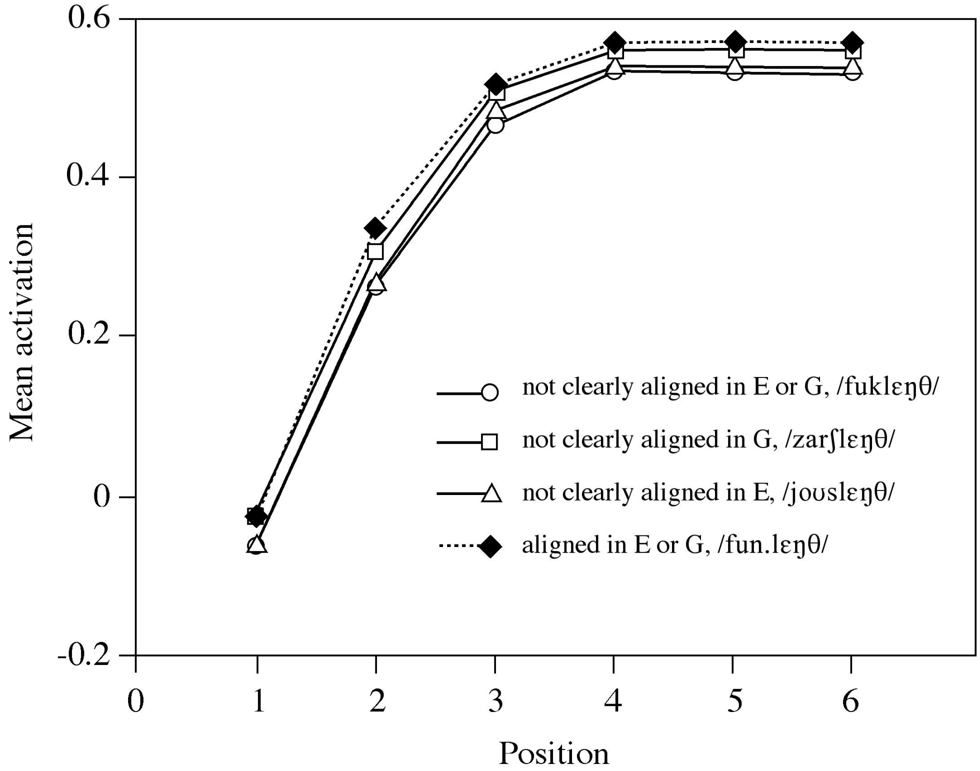

Figure 5. *Mean target activation levels for the materials from Chapter 3a simulated in Shortlist. Activation functions are shown for target words when neither English nor German forced a syllable boundary at the onset of the embedded word (e.g., length in /fuklɛŋθ/); when English, but not German, required a syllable boundary at the onset of the embedded word (e.g., length in /zarʃlɛŋθ/); when German, but not English, required a syllable boundary at the onset of the embedded word (e.g., length in /joʊslɛŋθ/); and when both German and English required a syllable boundary at the onset of the embedded word (e.g., length in /fun.lɛŋθ/). The activation functions are aligned relative to the first phoneme of the embedded word. Thus, 0 is, for example, the /l/ in either /fuklɛŋθ/ or /zarʃlɛŋθ/. G = German, E = English.*

The competition process and its time course were investigated by looking at the mean activations of the target words at different segments. In planned comparisons, the mean activation levels of target words that were clearly aligned with a phonotactic boundary (e.g., *length* in /fun.lɛŋθ/) were for all segment positions significantly higher than those of target words that were not clearly aligned with a phonotactic boundary (e.g., *length* in /fuklɛŋθ/, /zarʃlɛŋθ/, and /joʊslɛŋθ/). This simulates the experimental results of

German listeners well. German listeners in Chapter 3a found it harder to detect English words which were not clearly aligned with a syllable boundary according to at least one of the two languages than words which were clearly aligned with a syllable boundary according to either language or both languages. Thus, Shortlist seems to be able to simulate to a large extent the use of both native and non-native phonotactic boundary information in non-native word recognition. In a sequence like /fun.lɛŋθ/, the word *length* and other candidate words with initial /lɛ/ were not penalized by the PWC because the onset of *length* was aligned with a phonotactic boundary; however, all words with initial /ɛ/ were penalized because they would leave the impossible residue /l/. This pruning of competitor words will tend to make it easier for *length* to dominate the shortlist, and attain a higher level of activation sooner. In a sequence like /fuklɛŋθ/ on the other hand, both words with initial /l/ (e.g., *length*) and words with initial /kl/ (e.g., *cleric*) were not penalized in the simulations because both leave the possible residue /fuk/ and /fu/ respectively. Both words with initial /l/ and initial /kl/ competed for recognition, and there was therefore no advantage for words with initial /l/.

In addition, simulations were run in which word boundaries that were clearly phonotactically determined according to only one of the two languages (either English or German) were marked in the input (e.g., /zarʃ.lɛŋθ/ and /joʊs.lɛŋθ/). The mean activation levels of target words which were clearly aligned with a phonotactic boundary according to one but not the other language were now very similar to the activation levels of target words which were clearly aligned with a phonotactic boundary according to both languages (e.g., /fun.lɛŋθ/). Just as in a sequence like /fun.lɛŋθ/, the word *length* was not penalized by the PWC in sequences like /zarʃ.lɛŋθ/ or /joʊs.lɛŋθ/ because the onset of *length* was now aligned with a phonotactic boundary.

However, the activation functions in Figure 5 (where only word boundaries were marked in the input which were clearly phonotactically determined according to both English and German) of words which were not clearly aligned with a phonotactic boundary in German (e.g., *length* in /zarʃlɛŋθ/) were higher than the activation functions of words which were not clearly aligned in English (e.g., *length* in /joʊslɛŋθ/). (Statistically this difference was only significant for segment position 1 and 2). This does not simulate the experimental data of German listeners well. German listeners detected words which were not clearly aligned with a phonotactic boundary according to German (e.g., *length* in /zarʃlɛŋθ/) just as fast as words which were not clearly aligned with a phonotactic boundary according to English (e.g., *length* in /joʊslɛŋθ/). One reason why the activation levels of words that were not clearly aligned with a phonotactic boundary in German were higher than predicted on the basis of the experimental data can be found in the lexicon that was used for the simulations. The simulations were run on an English lexicon. Whereas there are English words with initial /sl/ in the English lexicon (e.g., *sleigh*), there are no English words with initial /ʃl/ in the lexicon; therefore no candidate words with initial /ʃl/ can compete against words with initial /l/ for recognition. The fewer competitors there are in the lexicon, the higher the activation of the available competitors tends to be. Because no words with initial /ʃl/ are in the English lexicon but words with initial /sl/ are, words with initial /l/ will be activated more highly in a sequence like /zarʃlɛŋθ/ than in a sequence like /joʊslɛŋθ/.

When the simulations are compared with the experimental data in Chapter 3a, the discrepancy described above implies that the German listeners also activated German words with, for example, initial /ʃl/ as in Schlitten, 'sleigh', during the recognition of English words. Informal further simulations were then conducted with Shortlist in which approximately 3,000 German words (mostly words with initial /ʃl/ and /ʃv/, but also words with initial /pl/, /kl/, /fl/, and /kv/) were added to the English lexicon. German /v/ was replaced with English /w/ for the simulations, so that German words could compete with English words in a sequence like /gruʃwɔsp/. These simulations indicate that by adding German competitors, the activation levels for words that were not clearly aligned with a phonotactic boundary in German drop (e.g., *length* in /zarʃlɛŋθ/). More direct evidence that listeners do indeed activate native competitors during the recognition of non-native words was found in Chapter 4 of the present thesis, where Dutch participants, when instructed to click on a target picture (e.g., the picture of a desk), fixated distractor pictures with Dutch names phonologically related to the English target name (e.g., Dutch *deksel*, 'lid', /dɛksəl/ begins similarly to English *desk* /dɛsk/) more than distractor pictures with phonologically unrelated names.

Implications for a model of non-native spoken-word recognition

The simulations with Shortlist have shown that only a model specific to non-native spoken-word recognition can adequately simulate the experimental data. One important reason for this is that phonemic inventories differ between languages. Two phonemes that are distinctive in one language may not occur at all in another language, or may occur, but only as allophones (contextual variants) of a single phoneme. As a result, non-native listeners may be less reliable in processing phonemes that only occur in the non-native language than in processing phonemes that also occur in their native language. One possibility is that non-native listeners 'assimilate' those non-native phonemes to a native phoneme category. Consider the results of German listeners in Chapter 3a for English words with initial /w/. English distinguishes between words with initial /w/ and /v/ (cf. *whale* /weɪl/ and *veil* /veɪl/). German, however, has only /v/ as in *Wal*, 'whale', /val/. For English words with initial /w/, German listeners used boundary cues for segmentation as if the English /w/ were German /v/. This is indirect evidence that German listeners assimilated English /w/ to /v/. The results of Dutch listeners in Chapter 4 when they processed spoken English are also relevant to this issue. The English target word *stamp* /stæmp/, for example, activated the Dutch word *stekker*, 'plug', /stɛkər/. This suggests that Dutch listeners assimilated English /æ/ to /ɛ/, which is not unexpected because /ɛ/ but not /æ/ occurs in Dutch.

The perception of non-native phonemes has been subject to a large amount of research (for an overview see Strange, 1995). In investigations of phoneme perception, listeners are usually either asked to label stimuli one at a time, using phonetic labels provided by the experimenter (identification task), or they are presented with two or more stimuli sequentially and asked to make comparative judgments about the physical identity or difference of the stimuli (discrimination task). These studies provided already ample evidence that non-native listeners are indeed often inaccurate in their perception of non-native phonemes. Inaccurate phoneme perception (and resulting phoneme confusion) may in turn lead to the activation of lexical competitors that would not feature in the native listener's set of candidate words. Perception of non-native phonemes has to be explored for every pair of native and non-native languages separately because phonemic differences between languages are always specific to a language pair. A model of non-native spoken-word recognition has to take these phoneme confusions into consideration. For example, to simulate the experimental data of Chapter 3a adequately, an English sequence like /ʃw/ has to allow the activation of German words with initial /ʃv/. This might not be the case for Dutch listeners perceiving English because Dutch listeners probably do not assimilate English /w/ to Dutch /v/, but rather to Dutch /ʋ/. Doubtless more of these 'assimilations', not only for consonants but also for vowels, must also be implemented in a model for English word recognition by German listeners.

However, to a large extent, a model of non-native spoken-word recognition could presumably be an extension of native spoken-word recognition models. Like models of native word recognition, it will also have to explain the processes by which listeners get from a highly variable speech signal to the recognition of discrete words. Universal aspects of language processing, like the process of competition, the way information flows between different levels of processing, and the units of perception therefore presumably hold for both a model of native and a model of non-native spoken-word recognition. The PWC (Norris et al., 1997) has been established as another universal constraint in language processing (Norris, Cutler, McQueen, Butterfield, & Kearns, 2000), and it should therefore also be implemented in a model of non-native spoken-word recognition. In addition, language-specific aspects of native spoken-word recognition can be taken over. Processes that are sensitive to English phonotactic boundaries have, for example, been implemented in Shortlist for native English word recognition. A model for English word recognition by Germans could take these processes over, and then further add processes that are sensitive to German phonotactic boundaries. Similarly, a model for English word recognition by Dutch could implement processes that are sensitive to English phonotactic boundaries, and then add processes that are sensitive to Dutch phonotactic boundaries. However, German listeners will confuse English phonemes differently than Dutch listeners; phoneme confusions must be established for every pair of native and non-native languages separately in order to model non-native spoken-word recognition adequately.

The proficiency of the listener in the two languages should also play an important role in a model of non-native spoken-word recognition. Speech perception is known to be modified by the amount of non-native language experience (see for example MacKain, Best, & Strange, 1980; Mochizuki, 1981). Thus, non-native speech perception by listeners with very little knowledge of and experience in the non-native language differs from non-native speech perception by listeners who are highly proficient in the non-native language. Presumably only the latter have learned to make extensive use of phonological information from the non-native language for processing, whereas the former will primarily be influenced by the phonological structure of their native language in processing. For listeners with a very low proficiency in the non-native language, a model of non-native spoken word-recognition should therefore presumably implement processes that are exclusively sensitive to the phonological structure of the listener's native language, whereas for listeners with a high proficiency in the non-native language such a model should allow processes that are sensitive to the phonological structure of both the native and the non-native language. The latter is at least true for phonotactic constraints, as the experiments in Chapter 3a have shown. Note that the influence of different levels of proficiency on non-native word recognition does not imply that a model of non-native spoken-word recognition must necessarily include top-down feedback. Different levels of proficiency may cause nonspecific top-down effects that are not due to the on-line influence of lexical representations on lower-level representations.

For other segmentation cues, like metrical information, the model may have to implement different processes. Cutler et al. (1992) have argued that segmentation procedures based on metrical information are mutually exclusive in the segmentation of a non-native language. Even listeners with a high degree of bilingualism did not employ a segmentation procedure based on the different rhythmical structures of two languages. Cutler et al. (1992) found that French-English bilinguals, who grew up with both languages but were dominant in French, were able to suppress syllabic segmentation, which is typical of French, when they were listening to English. However, the French-dominant listeners were not using instead a segmentation procedure that is based on stress, as it is typical of English. These results suggest that a model of non-native spoken-word recognition should allow the suppression of processes that are sensitive to native metrical information at least for listeners who are highly proficient in the non-native language.

Listening to a non-native language differs from listening to a native language. The results of this thesis support the claim that the phonological structure of the native language influences they way in which non-native spoken language is processed. It was shown that different processes, such as phoneme detection, segmentation strategies, and activation of candidate words, are affected by phonological information of the listeners' native language. The level of proficiency in the non-native language was found to restrain whether phonological information of that language was used for processing, too. Although these results provide important information for a model of non-native spoken-word recognition, more research has to be done before a fully implemented computational model of non-native spoken-word recognition will exist. Such a model must always be specific to a pair of languages.

References

Allopenna, P., Magnuson, J., & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, **38**, 419-439.

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.

Art Explosion library [Computer software]. (1995). Calabasas, CA: Nova Development Corporation.

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

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, (pp. 152-178). Cambridge: Cambridge University Press.

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

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Bradlow, A. R., & Pisoni, D. P. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, **106**, 2074-2085.

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.

Christie, J., & Klein, R. (1996). Assessing the evidence for novel popout. *Journal of Experimental Psychology: General*, **125**, 201-207.

Cole, R. A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing*, (pp. 67-116). London: Academic Press.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech*, (pp. 133-164). Hillsdale, N.J.: Erlbaum.

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, **11**, 635-645.

Cooper, R. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, **6**, 84-107.

Costa, A., Cutler, A., & Sebastián-Gallés, N. (1998). Effects of phoneme repertoire on phoneme decision. *Perception & Psychophysics*, **60**, 1022-1031.

Cutler, A. (1995). Spoken word recognition and production. In J. Miller & P. D. Eimas (Eds.), *Handbook of Perception and Cognition*, (Vol. 11, pp. 97-136). San Diego: Academic Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133-142.

Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385-400.

Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381-410.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113-121.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824-844.

Cycowicz, Y., Friedman, D., Rothstein, M., & Snodgrass, J. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, **65**, 171-237.

Dahan, D., Magnuson, J., & Tanenhaus, M. (in press). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*.

Dahan, D., Swingley, D., Tanenhaus, M., & Magnuson, J. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, **42**, 465-480.

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

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

Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating competitor effects with the bilingual interactive activation model. *Psychologica Belgica*, **38**, 177-196.

Doctor, E., & Klein, D. (1992). Phonological processing in bilingual word recognition. In R. J. Harris (Ed.), *Cognitive processing in bilinguals*, (pp. 237-252). Amsterdam: Elsevier.

DUDEN: Deutsches Universal Wörterbuch A-Z (1989). Bearbeitet vom wissenschaftlichen Rat unter der Mitarbeit der Dudenredaktion unter Leitung von Günther Drosdowski. 2., völlig neu bearbeitete und erweiterte Auflage. Mannheim, Wien, Zürich: Dudenverlag.

Dumay, N., Content, A., & Frauenfelder, U. H. (1999). Acoustic-phonetic cues to word boundary location: Evidence from word spotting. In *Proceedings of 14th International Conference of Phonetic Sciences*, (pp. 281-284). Berkeley, CA: University of California.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, **4**, 99-109.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, **27**, 143-165.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, **25**, 169-186.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*, (pp. 257-284). Amsterdam: North Holland.

Foss, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, **8**, 457-462.

Foss, D. J., & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *Journal of Verbal Learning and Verbal Behavior*, **12**, 246-257.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

Frauenfelder, U. H., & Seguí, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory and Cognition*, **17**, 134-140.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, **54**, 287-295.

Gaskell, M., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 144-158.

Gaskell, M., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 380-396.

Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge: University Press.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 344-359.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, **3**, 233-274.

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic perspectives*, (pp. 225-254). New Jersey: Lawrence Erlbaum Associates.

Gussenhoven, C. (1992). Illustrations of the IPA: Dutch. *Journal of the International Phonetic Association*, **22**, 45-47.

Hall, T. (1989). Lexical phonology and the distribution of German [ç] and [x]. *Phonology*, **6**, 1-17.

Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, **3**, 367-382.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, **78**, B53-B64.

Hay, J., Pierrehumbert, J., & Beckman M. (in press). Speech perception, well-formedness and the statistics of the lexicon. In *Proceedings of 6th Conference on Laboratory Phonology*. York, Great Britain: University of York.

Hyman, L. M. (1999). The limits of phonetic determinism in phonology. *Handout of talk presented at the annual meeting of the Linguistic Society of America*. Los Angeles, California.

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.

Johnston, W., & Schwarting, I. (1996). Reassessing the evidence for novel popout. *Journal of Experimental Psychology: General*, **125**, 208-212.

Johnston, W., & Schwarting, I. (1997). Novel popout: An enigma for conventional theories of attention. *Journal of Experimental Psychology: Human Perception and Performance*, **23**, 622-631.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, **32**, 402-420.

Klaassen-Don, L. E. O., & Pols, L. C. W. (1983). Dependency of processing of consonants and vowels in CV syllables. *Speech Communication*, **2**, 170-172.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129-140.

Klatt, D. H. (1976). Linguistic use of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208-1221.

Kohler, K. (1990). German. *Journal of the International Phonetic Association*, **20**, 48-50.

Koster, C. (1987). *Word recognition in foreign and native language: Effects of context and assimilation*. Dordrecht: Foris Publications.

Kuijpers, C., & van Donselaar, W. (in preparation). Phonological variation and phoneme identification in Dutch.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, **5**, (Suppl. 5), 1-54.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018-2024.

Léwy, N., & Grosjean, F. (in preparation). *The computerized version of BIMOLA – A bilingual model of lexical access*.

Li, P. (1996). Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language*, **35**, 757-774.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Longman Dictionary of Contemporary English (2nd ed.). (1987). London, Longman Group UK Limited.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing*, **19**, 1-36.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, **2**, 369-390.

Marian, V., & Spivey, M. (1999). Activation of Russian and English cohorts during bilingual spoken word recognition. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, (pp. 349-354). Mahwah, NJ: Erlbaum.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition, *Cognition*, **25**, 71-102.

Marslen-Wilson, W. D. (1990). Activation, competition and frequency in lexical access. In G. T. M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, (pp. 148-172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1-71.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, **15**, 576-585.

Martin, J., & Bunnell, H. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, **65**, 559-567.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perceptual Psychophysics*, **53**, 372-380.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, **38**, 465-494.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, **78**, 91-121.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, **86**, 287-330.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, (pp. 3-35). Hillsdale: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, **23**, 1-44.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

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**, 375-405.

McQueen, J. M. (1996). Word spotting. *Language and Cognitive Processes*, **11**, 695-699.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21-46.

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621-638.

McQueen, J. M., Norris, D. G., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 1363-1389.

McQueen, J. M., Otake, T., & Cutler, A. (in press). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*.

McQueen, J. M., & Pitt, M. A. (1996). Transitional probability and phoneme monitoring. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the 4th International Conference on Spoken Language Processing*, (Vol. 4, pp. 2502-2505).Wilmington, DE: University of Delaware and Alfred I. duPont Institute.

Meador, D., Flege, J. E., & MacKay, I. R. (2000). Factors affecting the recognition of words in a second language. *Bilingualism*, **3**, 55-67.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298-305.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, **9**, 283-303.

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, **76**, 165-178.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714-719.

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.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189-234.

Norris, D. G., Cutler, A., McQueen, J. M., Butterfield, S., & Kearns, R. (2000). Language-universal constraints on the segmentation of English. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes*, (pp. 43-46). Nijmegen, The Netherlands: Max-Planck-Institute for Psycholinguistics.

Norris, D. G., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1209-1228.

Norris, D. G., McQueen, J. M., & Cutler A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, **23**, 299-370.

Norris, D. G., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34**, 191-243.

Ohala, J. (1995). A probable case of clicks influencing the phonology of some European languages. *Phonetica*, **52**, 160-170.

Ohala, J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, 1718-1725.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54**, 1235-1247.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258-278.

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. H. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustic Society of America*, **100**, 3831-3842.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllable structure of spoken words. *Journal of Memory and Language*, **32**, 373-389.

Pierrehumbert, J. (1994). Knowledge of variation. *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society*, **2**, 232-256.

Posner, M., Nissen, M., & Ogden, W. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. Pick & E. Saltzman (Eds.), *Modes of Perceiving and Processing Information*, (pp. 137-157). Hillsdale: Erlbaum.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, **20**, 331-350.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, **94**, 2027-2035.

Quené, H., van Rossum, M., & van Wijck, M. (1998). Assimilation and anticipation in word recognition. In *Proceedings of 5th International Conference on Spoken Language Processing*, (paper no. 113). Sydney, Australia.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception, Part 2: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **89**, 60-94.

Russ, C. V. J. (Ed.) (1990). *The dialects of modern German*. London: Routledge.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, **35**, 606-621.

Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31**, 18-32.

Seguí, J., & Frauenfelder, U. H. (1986). The effect of lexical constraints on speech perception. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive abilities: Symposium in memoriam Hermann Ebbinghaus*, (pp. 795-808). Amsterdam: North-Holland.

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 174-215.

Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and bilingual speech mode: The effect on lexical access. *Memory & Cognition*, **12**, 380-386.

Spivey, M., & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, **10**, 281-284.

Strange, W. (Ed.) (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422-444.

Tanenhaus, M., & Spivey-Knowlton, M. (1996). Eye-tracking. *Language and Cognitive Processes*, **11**, 583-588.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, **268**, 1632-1634.

Trubetzkoy, N. (1939). *Grundzüge der Phonologie*. Göttingen: Vandenhoek Ruprecht.

Vance, T. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.

van der Lugt, A. H. (in press). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*.

van Donselaar, W., Kuijpers, C., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language*, **41**, 59-77.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, **40**, 47-62.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 98-108.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, **38**, 133-149.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24**, 744-755.

Weber, A. (2000). Phonotactic and acoustic cues for word segmentation in English. In B. Yuan, T. Huang, & X. Tang (Eds.), *Proceedings of the 6th International Conference on Spoken Language Processing*, (Vol. 3, pp. 782-785). Beijing, China: Beijing International Convention Center.

Weber, A. (2001). Help or hindrance: How violation of different assimilation rules affects spoken-language processing. *Language and Speech*, **44**, 95-118.

Weber, A. (submitted a). Assimilation violation and spoken-language processing: A supplementary report. *Language and Speech*.

Weber, A. (submitted b). Phonotactic constraints and the segmentation of native and non-native spoken language. *Journal of Memory and Language*.

Whalen, D. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

Whalen, D. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, **50**, 351-360.

Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.

Wurzel, W. (1980). Phonologie: Segmentale Struktur. In K. Heidolph, W. Flamig, & W. Motsch (Eds.), *Grundzüge einer deutschen Grammatik*, (pp. 898-1003). Berlin: Akademie-Verlag.

Appendices

Appendix 2−1

Dutch target materials used in Experiments 1 and 2 in Chapter 2a.

Targets with violation of the German progressive fricative assimilation.

[bɛxt], [blɪxt], [frɛxt], [frɪxt], [hɪxt], [kɛxt], [klɪxt], [krɛxt], [knɪxt], [nɛxt], [pɛxt], [pɪxt], [prɛxt], [ʋɛxt]

Targets without violation of the German progressive fricative assimilation.

[bɑxt], [blɑxt], [blɔxt], [brɔxt], [frɔxt], [hɑxt], [hɔxt], [kɑxt], [knɑxt], [nɔxt], [plɔxt], [prɔxt], [rɑxt], [ʋɔxt]

Appendix 2–2

German target materials used in Experiment 3 in Chapter 2a.

Monosyllabic targets with violation of the German progressive fricative assimilation.

[bɪxt], [dɛxt], [drɛxt], [kɛxt], [kɪxt], [mɛxt], [nɛxt], [pɪxt], [ʃɛxt], [ʃpɪxt], [tɛxt], [tɪxt], [vɛxt], [zɛxt]

Monosyllabic targets without violation of the German progressive fricative assimilation.

[bɔxt], [dʊxt], [draxt], [kaxt], [kʊxt], [mɔxt], [nʊxt], [pʊxt], [ʃɔxt], [ʃpʊxt], [tʊxt], [taxt], [vɔxt], [zɔxt]

Bisyllabic targets with violation of the German progressive fricative assimilation.

[blɪnxən], [bɛlxər], [dɛlxən], [fɛnxən], [krɪnxən], [lɪnxər], [mɛlxən], [nɛlxər], [plɛnxən], [rɛlxən], [ʃlɪnxər], [ʃpɛnxər], [ʃpɪlxən], [tɪnxən]

Bisyllabic targets without violation of the German progressive fricative assimilation.

[bluːxən], [boːxər], [doːxən], [fuːxən], [kruːxən], [luːxər], [mauxən], [nauxər], [plaːxən], [raːxən], [ʃloːxər], [ʃpauxər], [ʃpoːxən], [taːxən]

Appendix 2−3

German target materials used in Experiment 4 in Chapter 2a.

Targets with violation of the German regressive place assimilation for nasals.

Alveolar-velar sequences:

[bɛnk], [bɪnk], [blʊnk], [fɛnk], [flʊnk], [frʊnk], [lʊnk], [pɛnk], [plɪnk], [pʊnk], [ʃɪnk], [ʃlʊnk], [ʃrʊnk], [tɛnk]

Alveolar-bilabial sequences:

[bɛnp], [blɛnp], [dɔnp], [fɛnp], [flɔnp], [glɛnp], [lɛnp], [lɔnp], [ʃɛnp], [ʃlɔnp], [tɛnp], [kʊnp], [tʊnp], [zʊnp]

Bilabial-velar sequences:

[bɛmk], [bɪmk], [blʊmk], [fɛmk], [flʊmk], [frʊmk], [lʊmk], [pɛmk], [plɪmk], [pʊmk], [ʃɪmk], [ʃlʊmk], [ʃrʊmk], [tɛmk]

Velar-bilabial sequences:

[bɛŋp], [blɛŋp], [dɔŋp], [fɛŋp], [flɔŋp], [glɛŋp], [lɛŋp], [lɔŋp], [ʃɛŋp], [ʃlɔŋp], [tɛŋp], [kʊŋp], [tʊŋp], [zʊŋp]

Targets without violation of the German regressive place assimilation for nasals.

Velar-velar sequences:

[bɛŋk], [bɪŋk], [blʊŋk], [fɛŋk], [flʊŋk], [frʊŋk], [lʊŋk], [pɛŋk], [plɪŋk], [pʊŋk], [ʃɪŋk], [ʃlʊŋk], [ʃrʊŋk], [tɛŋk]

Bilabial-bilabial sequences:

[bɛmp], [blɛmp], [dɔmp], [fɛmp], [flɔmp], [ɡlɛmp], [lɛmp], [lɔmp], [ʃɛmp], [ʃlɔmp], [tɛmp], [kʊmp], [tʊmp], [zʊmp]

Appendix 2−4

German target materials used in Experiment 5 in Chapter 2b.

Targets without violation of the German fricative assimilation rule.

[biːçəl], [bleːçər], [deːçən], [gleːçər], [heːçəl], [keːçəl], [kiːçən], [liːçən], [meːçəl],[niːçər], [peːçər], [piːçən], [tiːçəl], [tiːçən]

Targets with violation of the German fricative assimilation rule.

[baːçəl], [bloːçər], [daːçən], [glaːçər], [hoːçəl], [koːçəl], [kaːçən], [laːçən], [moːçəl], [noːçər], [paːçər], [poːçən], [toːçəl], [taːçən]

Appendix 2–5

German target materials used in Experiments 6, 7, 8, and 9 in Chapter 2c.

Targets with initial /st/.

[stalm], [starç], [starm], [stɛlm], [stɛrm], [stɪlm], [stɪm], [stɪmp], [stɪn], [stɪrk], [stɔlm], [stɔn], [stɔrk], [stɔrn]

Targets with initial /ʃt/.

[ʃtak], [ʃtalm], [ʃtarm], [ʃtɛmp], [ʃtɪlm], [ʃtɪm], [ʃtɪmp], [ʃtɪn], [ʃtɪp], [ʃtɪrk], [ʃtɪrm], [ʃtɔl], [ʃtɔlm], [ʃtɔn]

Targets with final /st/.

[blɪst], [blɔst], [flast], [fɔst], [frast], [kɛst], [knɛst], [knɔst], [krɔst], [nast], [nɪst], [nɔst], [plɛst], [vɔst]

Targets with final /ʃt/.

[bɔʃt], [flaʃt], [flɪʃt], [fɔʃt], [fraʃt], [knɔʃt], [krɪʃt], [lɔʃt], [mɛʃt], [nɔʃt], [plɛʃt], [plɪʃt], [prɔʃt], [vɔʃt]

Appendix 3

English target materials used in Experiments 10 and 11 in Chapter 3a. E = English, G = German. Those targets excluded from the analysis are marked **.

Embedded words with initial /l/.

Not clearly aligned in E or G	Aligned in E only	Aligned In G only	Aligned in E and G	Embedded word
/θiplæns/	/ðiʃlæns/	/blɔɪslæns/	/dʒimlæns/	lance
/zarplɔft/	/prarʃlɔft/	/forslɔft/	/fumlɔft/	loft
/narplɪrɪk/	/diʃlɪrɪk/	/nɔslɪrɪk/	/dʒarnlɪrɪk/	lyric
/bɔklɛd/	/rarʃlɛd/	/birslɛd/	/rinlɛd/	lead**
/mɔklɔs/	/nuʃlɔs/	/tirslɔs/	/dɔɪnlɔs/	loss
/guklardʒ/	/giʃlardʒ/	/ðaʊslardʒ/	/jɔɪnlardʒ/	large
/garflɛʒər/	/frarʃlɛʒər/	/fərslɛʒər/	/garnlɛʒər/	leisure**
/jiflɛtər/	/ruʃlɛtər/	/puslɛtər/	/paʊnlɛtər/	letter
/narflup/	/griʃlup/	/hoʊslup/	/marnlup/	loop
/nuplɪft/	/narʃlɪft/	/rirslɪft/	/pumlɪft/	lift
/jarplord/	/kriʃlord/	/mɔɪslord/	/fumlord/	lord
/farplɔndri/	/parʃlɔndri/	/borslɔndri/	/pərnlɔndri/	laundry
/ʃorklaf/	/priʃlaf/	/krirslaf/	/gɔɪnlaf/	laugh
/fuklɛŋθ/	/zarʃlɛŋθ/	/joʊslɛŋθ/	/funlɛŋθ/	length
/jorklʊntʃ/	/riʃlʊntʃ/	/faʊslʊntʃ/	/harnlʊntʃ/	lunch
/buflɪst/	/farʃlɪst/	/gɔɪslɪst/	/waʊnlɪst/	list
/parflɛvəl/	/dʒiʃlɛvəl/	/buslɛvəl/	/hɔɪnlɛvəl/	level
/darflʌst/	/grarʃlʌst/	/pɔɪslʌst/	/parnlʌst/	lust
/marplɔ/	/θiʃlɔ/	/grirslɔ/	/dʒimlɔ/	law
/darplɑdʒɪk/	/jiʃlɑdʒɪk/	/hɔɪslɑdʒɪk/	/fumlɑdʒɪk/	logic
/jiplɔn/	/tarʃlɔn/	/θirslɔn/	/wɔɪnlɔn/	lawn
/ruklidʒən/	/darʃlidʒən/	/buslidʒən/	/punlidʒən/	legion
/fuklʌv/	/briʃlʌv/	/paʊslʌv/	/zarnlʌv/	love
/dɔklɪnən/	/dʒiʃlɪnən/	/kɔɪslɪnən/	/ðaʊnlɪnən/	linen

/rarflɛft/	/trarʃlɛft/	/jorslɛft/	/rarnlɛft/	left
/marflɛðər/	/buʃlɛðər/	/hirslɛðər/	/ʃaʊnlɛðər/	leather
/trarflaʊndʒ/	/krarʃlaʊndʒ/	/nɔɪslaʊndʒ/	/marnlaʊndʒ/	lounge
/fuplɑk/	/garʃlɑk/	/mərslɑk/	/pumlɑk/	lock
/rarplɛnz/	/frarʃlɛnz/	/torslɛnz/	/dʒimlɛnz/	lens
/garplæg/	/drarʃlæg/	/rərslæg/	/zɔɪnlæg/	lag
/wuklʌk/	/briʃlʌk/	/dorslʌk/	/farnlʌk/	luck
/nɔklig/	/karʃlig/	/jɔɪslig/	/jaʊnlig/	league
/gorklɛkʃər/	/θrarʃlɛkʃər/	/mɔɪslɛkʃər/	/mɔɪnlɛkʃər/	lecture
/krarflɑdʒ/	/fuʃlɑdʒ/	/dɔɪslɑdʒ/	/rərnlɑdʒ/	lodge
/tarflaʊs/	/driʃlaʊs/	/foʊslaʊs/	/parnlaʊs/	louse
/giflɛpərd/	/dʒarʃlɛpərd/	/gaʊslɛpərd/	/faʊnlɛpərd/	leopard

Embedded words with initial /w/.

Not clearly aligned in E or G	Aligned in E only	Aligned In G only	Aligned in E and G	Embedded word
/zorkworm/	/flarʃworm/	/nutworm/	/pirlworm/	warm
/mɔkwɛl/	/narʃwɛl/	/mɔɪtwɛl/	/ʃirlwɛl/	well
/plukwor/	/farʃwor/	/mirtwor/	/narlwor/	war
/ðikwɔsp/	/gruʃwɔsp/	/gɔɪtwɔsp/	/jarlwɔsp/	wasp
/horkwʊmən/	/fuʃwʊmən/	/θaʊswʊmən/	/tarlwʊmən/	woman
/jikwɛðər/	/darʃwɛðər/	/dɔswɛðər/	/pɔɪnwɛðər/	weather
/rukwərm/	/griʃwərm/	/grərswərm/	/zaʊnwərm/	worm
/nɔkwaɪf/	/riʃwaɪf/	/blirswaɪf/	/pumwaɪf/	wife
/θikwund/	/larʃwund/	/gaʊtwund/	/girlwund/	wound
/fukword/	/tarʃword/	/tʃutword/	/narlword/	ward
/borkwɔk/	/fruʃwɔk/	/nirtwɔk/	/darlwɔk/	walk
/plukwɪʃ/	/glarʃwɪʃ/	/lɔɪtwɪʃ/	/jarlwɪʃ/	wish
/bukwʊl/	/ðiʃwʊl/	/gorswʊl/	/mirlwʊl/	wool**
/pukworp/	/ruʃworp/	/drirsworp/	/hərnworp/	warp
/gukwərd/	/guʃwərd/	/glɔɪswərd/	/fɔɪnwərd/	word
/θɔkwɪn/	/nuʃwɪn/	/fərswɪn/	/fumwɪn/	win

/dukworf/	/priʃworf/	/hɔɪtworf/	/rarlworf/	wharf**
/nukweɪ/	/guʃweɪ/	/putweɪ/	/parlweɪ/	way
/ðikwæg/	/giʃwæg/	/lirtwæg/	/kirlwæg/	wag
/jɔkwɪŋ/	/klarʃwɪŋ/	/fɔɪtwɪŋ/	/rarlwɪŋ/	wing**
/gorkwərs/	/muʃwərs/	/glirswərs/	/birlwərs/	worse**
/horkwɛpən/	/rarʃwɛpən/	/fɔswɛpən/	/ʃɔɪnwɛpən/	weapon
/frukwʊd/	/plarʃwʊd/	/mirswʊd/	/funwʊd/	wood
/ʃorkwum/	/puʃwum/	/plurswum/	/dʒimwum/	womb**
/bukwɛt/	/kluʃwɛt/	/mɔɪtwɛt/	/nirlwɛt/	wet
/lukwɑtʃ/	/jiʃwɑtʃ/	/zɔɪtwɑtʃ/	/hirlwɑtʃ/	watch
/jɔkwɛdɪŋ/	/parʃwɛdɪŋ/	/hirtwɛdɪŋ/	/dirlwɛdɪŋ/	wedding
/dɔkwəri/	/buʃwəri/	/pirtwəri/	/marlwəri/	worry
/prukwik/	/luʃwik/	/krərswik/	/farlwik/	week
/mukwɪntər/	/prarʃwɪntər/	/nərswɪntər/	/dərnwɪntər/	winter
/zorkwɪnd/	/parʃwɪnd/	/θrirswɪnd/	/θaʊnwɪnd/	wind
/jikwɪdθ/	/pruʃwɪdθ/	/pərswɪdθ/	/pumwɪdθ/	width

Appendix 4

English target words and Dutch competitor words used in Experiment 13 in Chapter 4.

English target words	Dutch competitor words	Unrelated distractors in Dutch and English	
carrot /kɛrət/	kerk /kɛrk/	fluitje	spiegel
	'church'	'whistle'	'mirror'
bowl /boʊl/	boom /bom/	auto	snor
	'tree'	'car'	'moustache'
kitten /kɪtən/	kist /kɪst/	moer	borstel
	'chest'	'nut'	'brush'
knife /naɪf/	nijlpaard /nɛilpart/	schelp	vogelhuis
	'hippopotamus'	'shell'	'birdhouse'
pie /paɪ/	pijl /pɛil/	deur	kous
	'arrow'	'door'	'stocking'
seatbelt /sitbɛlt/	citroen /sitrun/	kleed	pot
	'lemon'	'rug'	'pot'
meat /mit/	mier /mir/	tafel	beker
	'ant'	'table'	'cup'
shark /ʃɑrk/	sjaal /ʃal/	föhn	berg
	'scarf'	'hairdryer'	'mountain'
light bulb /laɪt bʌlb/	lijst /lɛist/	strik	vogel
	'frame'	'bow'	'bird'
stamp /stæmp/	stekker /stɛkər/	ezel	bril
	'plug'	'donkey'	'glasses'
desk /dɛsk/	deksel /dɛksəl/	bloem	schommel
	'lid'	'flower'	'swing'
spine /spaɪn/	spijker /spɛikər/	hoed	raam
	'nail'	'hat'	'window'
flashlight /flæʃlaɪt/	fles /flɛs/	klerenhanger	muis
	'bottle'	'coat hanger'	'mouse'
money /mʌni/	mand /mɑnt/	den	aardbei
	'basket'	'pine'	'strawberry'

closet /klɒzɪt/	klomp /klɔmp/ 'wooden shoe'	paddestoel 'mushroom'	oog 'eye'
bike /baɪk/	bijl /bɛil/ 'axe'	meisje 'girl'	wolk 'cloud'
lake /leɪk/	lepel /lepəl/ 'spoon'	muts 'cap'	afvalemmer 'trashcan'
spring /sprɪŋ/	sprinkhaan /sprɪŋkhan/ 'grasshopper'	tomaat 'tomato'	been 'leg'
duck /dʌk/	dak /dɑk/ 'roof'	schaar 'scissors'	vliegtuig 'plane'
leaf /lif/	libel /libɛl/ 'dragonfly'	hand 'hand'	knoop 'button'

Summary

In order to understand spoken language, listeners have to recognize individual words in the speech stream. This is not a simple task, first, because the acoustic realization of a given word is highly variable (strictly speaking every utterance is unique, even if the same utterance is uttered twice by the same speaker), and second because speakers do not pause between words of an utterance but rather produce a continuous speech stream. Nevertheless, listeners have to find word boundaries in order to segment the speech stream into individual words. Adult listeners possess a repertoire of strategies that they can use to facilitate the recognition of spoken words in their native language. For adult listeners in their native language the process of spoken-word recognition is highly efficient and seems to be quite effortless. However, when one listens to a second, non-native language, it can become more apparent how complex the process of spoken-word recognition in fact is. Many of the structural factors that listeners use to process their native spoken language are specific to that particular language and do not apply to a second, non-native language.

The research reported in this thesis focuses on the processing of phonetic sequences in both native and non-native spoken language. The phonology of a language restricts, among other things, which phonemes can be combined to sequences within a syllable. For example, the sequence /sl/ can occur within syllables in English, as in the word *sleigh*; the sequence /pf/, on the other hand, never occurs within syllables in English. In German, on the other hand, the sequence /sl/ never occurs within syllables whereas /pf/ does (e.g., German *Pferd*, 'horse'). In other words, information for spoken-word recognition coming from phonetic sequences can differ between languages. Adult listeners have learned early in life to make use of such phonological regularities that are specific to their native language. That is, they have tuned their perception to their native language. But listeners can also learn a second, non-native, language later in life. To what extent can they decide to ignore information coming from the phonological structure of the native language when they listen to a non-native language? And to what extent can they make use of information coming from the phonological structure of the non-native language? In other words, can listeners choose between listening to the non-native language with their native or their non-native "ear"?

In Chapter 2, the influence of the legality of phonetic sequences on phoneme detection was investigated for both native and non-native spoken-language processing. First, obligatory assimilation rules were tested. The violation of an obligatory assimilation rule results in illegal sequences of segments. For instance, in German the vowel /ɪ/ can be followed by the palatal fricative [ç] but not by the velar fricative [x]. It would be illegal in German to pronounce the word *Licht*, 'light', with a velar fricative as in [lɪxt]. In Dutch, however, *licht*, 'light', pronounced with the velar fricative is legal. German listeners were asked to press a button as soon as they detected the velar fricative [x] in a list of Dutch nonwords. (The German listeners had no knowledge of Dutch.) In some of the nonwords the velar fricative occurred in sequences which are illegal in German. German listeners detected the velar fricative faster in Dutch nonwords when it occurred in sequences that were illegal for German than when it occurred in sequences that were legal for German. Dutch listeners showed no such effect, since none of the materials violated Dutch phonology. The results therefore suggest that when listeners encounter a new language, rather than turning off all language-specific strategies for processing, they continue to make use of the phonological structure of their native language. In follow-up experiments the sensitivity to violation of obligatory assimilation rules was replicated for native listening (see Chapters 2a and b).

In addition, violation of phonotactic constraints (rather than assimilation rules) was tested. Phonotactic constraints differ between languages just as assimilation rules do, and just as with assimilation, violation of phonotactic constraints always results in phonotactically illegal sequences. However, two segments that have undergone assimilation share a feature rather than they differ in one (e.g., /ɪ/ is a front vowel and the palatal fricative [ç] is also more front than its velar counter part [x]). Thus, assimilation is phonetically motivated, but any phonetic motivation for phonotactic constraints is more abstract. Phonotactic constraints might therefore provide less useful information for the listeners than assimilation rules do, since they are highly language specific. Whereas the results from German listeners in Chapter 2c clearly indicate sensitivity to phonotactic constraints in phoneme detection when they listen to German, the results of the Dutch, non-native, listeners were less clear.

In Chapter 3, the role of phonotactic constraints was investigated for a different level of processing, namely segmentation. Phonotactics constrain which sequences occur within syllables in a language. The sequence /sl/, for example, is legal in English (e.g., *sleigh*), but illegal in German (no German syllable and therefore no German word begins with /sl/). When a sequence is illegal within syllables, this implies that there must be a syllable boundary between the two segments of the sequence. Syllable boundaries are highly correlated with word boundaries and a word boundary between /s/ and /l/ is therefore likely in German, though not in English. Likewise, /ʃl/ cues a possible word boundary for English but not for German. A sequence like /fl/, on the other hand, does not cue a boundary for either German or English, since there are both English and German words beginning with /fl/ (e.g., English *flight* and German *Flug*, 'flight'). In the experiments in Chapter 3 native listeners of German who had excellent knowledge of English detected English words in nonsense sequences. The onsets of the embedded English words were either aligned with a boundary or not, according to the phonotactic constraints each of English and German (e.g., *length* in /fuklɛŋθ/, /zarʃlɛŋθ/, /joʊslɛŋθ/, /funlɛŋθ/). English boundary cues were as helpful for segmentation as were German boundary cues. English listeners who had no knowledge of German also participated in the experiments, and their reactions were influenced primarily by phonotactic cues of English. Although, in contrast with Chapter 2, the non-native listeners in Chapter 3 were highly proficient in the non-native language, they still used phonotactic cues supplied by their native language for segmenting the non-native language, even though the boundary cues specified by the native language may interfere with word recognition in the non-native language. Obviously, they could not simply choose to ignore information from the phonological structure of their native language. However, these listeners were also able to make use of phonotactic boundary cues specific to the non-native language.

In Chapter 4, the role of native and non-native lexical competitors in non-native spoken-word recognition was investigated. Words sharing initial segments in the native language are briefly activated during the recognition of spoken words. Given, for example, the input *desk*, English listeners will activate *desk* and *dentist* among other candidates, which will then compete against each other for recognition. Chapter 4 tested whether listeners still activate native candidates when they listen to a non-native language. Dutch participants followed spoken instructions in English to click on pictures using a computer mouse while their eye movements were monitored. A target picture (e.g., the picture of a *desk*) was presented along with distractor pictures. Some distractor pictures had names in Dutch with the initial segments similar to the English name of the target picture (e.g., English target *desk* /dɛsk/ and Dutch competitor *deksel*, 'lid', /dɛksəl/). Eye movements to a picture were interpreted as evidence for activation of the lexical representation associated with this picture. Dutch listeners briefly fixated Dutch competitors with similar onsets significantly more often than unrelated distractors. In other words, when a Dutch listener who knows English is listening to an English sentence that contains the word desk /dɛsk/, the Dutch word *deksel*, 'lid', /dɛksəl/, which is phonologically similar, is considered as a candidate during the spoken-word recognition process. Listeners do not appear to be able to deactivate the native mental lexicon even when they are in a non-native situation where the native language is irrelevant and possibly misleading. Clearly this application of native procedures to nonnative input does not promote listening efficiency.

Adult listeners command a repertoire of procedures appropriate for spoken-language processing in their native language. However, when they encounter a second, non-native language they still apply the native procedures to the new input irrespective of whether this facilitates processing or renders it less efficient. Highly proficient non-native listeners seem to succeed in using non-native procedures as well, at least for some aspects of spoken-word recognition (as has been shown for phonotactic constraints in segmentation). In conclusion, it can be said that the efficiency with which we can process spoken language in our native language reduces the efficiency in listening to a second language learned later in life, at least to the extent that the first and the second language differ in aspects of phonological structure used in spoken-language processing.

Samenvatting

Om gesproken taal te begrijpen moeten luisteraars de individuele woorden in de spraakstroom herkennen. Dit is om twee redenen geen eenvoudige taak. Ten eerste is de akoestische realisatie van een woord zeer variabel; strikt genomen is iedere uiting uniek, zelfs wanneer dezelfde uiting tweemaal door dezelfde spreker wordt geproduceerd. Ten tweede pauzeren sprekers niet tussen de woorden, maar produceren ze een continue spraakstroom. Desondanks moeten luisteraars woordgrenzen vinden om de spraakstroom in individuele woorden te segmenteren. Volwassen luisteraars beschikken over een repertoire van strategieën om de herkenning van gesproken woorden in hun moedertaal te vereenvoudigen. Het proces van gesproken woordherkenning verloopt voor volwassen luisteraars zeer efficiënt en schijnbaar moeiteloos. Bij het luisteren naar een tweede, vreemde taal kan echter merkbaar worden hoe complex het proces van gesproken woordherkenning in feite is. Veel van de structurele factoren die luisteraars gebruiken om hun gesproken moedertaal te verwerken zijn specifiek voor de taal in kwestie en niet van toepassing op een tweede, vreemde taal.

Het onderzoek beschreven in deze dissertatie richt zich op het verwerken van fonetische reeksen in spraak in de moedertaal en in een vreemde taal. De fonologie van een taal bepaalt onder andere welke fonemen tot een reeks gecombineerd kunnen worden binnen een lettergreep. De reeks /sl/ kan bijvoorbeeld in het Engels voorkomen binnen een lettergreep, zoals in het woord *sleigh*, 'slee', terwijl de reeks /pf/ in het Engels nooit binnen een lettergreep voorkomt. In het Duits daarentegen komt de reeks /sl/ nooit voor binnen een lettergreep, maar /pf/ wel (bijv. het Duitse woord *Pferd*, 'paard'). Met andere woorden, de informatie die fonetische reeksen bieden voor gesproken woordherkenning kan per taal verschillen. Volwassen luisteraars hebben jong geleerd om gebruik te maken van fonologische regelmatigheden die specifiek zijn voor hun moedertaal. Zij hebben hun waarneming afgestemd op hun moedertaal. Maar luisteraars kunnen ook op latere leeftijd een tweede, vreemde taal leren. In hoeverre kunnen ze informatie die afkomstig is van de fonologische structuur van de moedertaal verkiezen te negeren bij het luisteren naar een vreemde taal? En in hoeverre kunnen ze informatie afkomstig uit de fonologische structuur van de vreemde taal gebruiken? Met andere woorden, kunnen luisteraars bij het luisteren naar de vreemde taal kiezen tussen de “oren” van hun moedertaal en die van hun vreemde taal?

In Hoofdstuk 2 is de invloed van de legaliteit van fonetische reeksen op foneem-detectie onderzocht voor spraakverwerking in zowel de moedertaal als de vreemde taal. Eerst zijn verplichte assimilatieregels getest. De overtreding van een verplichte assimilatieregel leidt tot illegale reeksen van segmenten. In het Duits kan de klinker /ɪ/ bijvoorbeeld worden gevolgd door de palatale fricatief [ç], maar niet door de velare fricatief [x]. Het zou in het Duits illegaal zijn om het woord *Licht*, 'licht', met een velare fricatief uit te spreken, als in [lɪxt]. In het Nederlands daarentegen is de uitspraak van *licht* met een velare fricatief legaal. Duitse luisteraars die geen Nederlands kenden werd gevraagd op een knop te drukken zodra zij de velare fricatief [x] hoorden in een serie Nederlandse nonwoorden. In sommige nonwoorden kwam de velare fricatief voor in reeksen die illegaal zijn in het Duits. Duitse luisteraars vonden de velare fricatief in Nederlandse nonwoorden sneller wanneer deze voorkwam in reeksen die in het Duits illegaal zijn, dan wanneer deze voorkwam in reeksen die in het Duits legaal zijn. Bij Nederlandse luisteraars trad een dergelijk effect niet op, aangezien het materiaal geheel in overeenstemming was met de Nederlandse fonologie. De resultaten suggereren dan ook dat luisteraars die geconfronteerd worden met een nieuwe taal, in plaats van alle taal-specifieke strategieën uit te schakelen, gebruik blijven maken van de fonologische structuur van hun moedertaal. In vervolgexperimenten is de gevoeligheid voor overtreding van verplichte assimilatieregels voor luisteren naar de moedertaal gerepliceerd (zie Hoofdstukken 2a en 2b).

Naast de overtreding van assimilatieregels is ook de overtreding van fonotactische restricties getest. Evenals assimilatieregels verschillen fonotactische restricties per taal en net als bij assimilatie leidt overtreding van fonotactische restricties altijd tot (in dit geval fonotactisch) illegale reeksen. Bij assimilatie wordt een kenmerk van een segment naar een ander segment verspreid, waardoor beide het kenmerk delen (bijv. de klinker /ɪ/ is 'voor' en de palatale fricatief [ç] is ook meer 'voor' dan zijn velare pendant [x]). Assimilatie is dan ook fonetisch gemotiveerd, terwijl fonetische motivatie voor fonotactische restricties abstracter is. Fonotactische restricties zouden de luisteraar daardoor minder bruikbare informatie kunnen bieden dan assimilatieregels, aangezien deze eerste sterk taal-specifiek zijn. Terwijl de resultaten van de Duitse luisteraars in Hoofdstuk 2c duidelijk wijzen op gevoeligheid voor fonotactische restricties in foneem-detectie bij het luisteren naar Duits, zijn de resultaten van de Nederlandse luisteraars bij het luisteren naar het Duits als vreemde taal minder eenduidig.

In Hoofdstuk 3 is de rol van fonotactische restricties voor een ander niveau van spraakverwerking, namelijk segmentatie, onderzocht. Fonotactische restricties bepalen welke reeksen er voorkomen binnen lettergrepen in een bepaalde taal. De reeks /sl/ is bijvoorbeeld legaal in het Engels (zoals in *sleigh*, 'slee'), maar illegaal in het Duits (geen enkele Duitse lettergreep en daarmee geen enkel Duits woord begint met /sl/). Als een reeks illegaal is binnen een lettergreep moet er een lettergreepgrens zijn tussen de twee segmenten van de reeks. Lettergreepgrenzen correleren sterk met woordgrenzen en een woordgrens is dan ook waarschijnlijk tussen /s/ en /l/ in het Duits, maar niet in het Engels. Op dezelfde manier wijst /ʃl/ op een mogelijke woordgrens in het Engels, maar niet in het Duits. Een reeks als /fl/ daarentegen wijst noch in het Duits, noch in het Engels op een grens, aangezien er zowel Engelse als Duitse woorden zijn die beginnen met /fl/ (bijv. het Engelse woord *flight*, 'vlucht', en het Duitse woord *Flug*, 'vlucht'). In de experimenten in Hoofdstuk 3 moesten luisteraars met Duits als moedertaal en een uitstekende kennis van het Engels Engelse woorden vinden in nonsensreeksen. Het begin van de ingebedde Engelse woorden viel al dan niet samen met een grens volgens zowel de Engelse als de Duitse fonotactische restricties (bijv. *length*, 'lengte', in /fuklɛŋθ/, /zarʃlɛŋθ/, /joʊslɛŋθ/, /funlɛŋθ/). Engelse aanwijzingen voor een grens waren even bruikbaar voor segmentatie als Duitse. Ook Engelse luisteraars zonder kennis van het Duits namen deel aan de experimenten. Hun reacties werden voornamelijk beïnvloed door fonotactische aanwijzingen uit het Engels. Terwijl de vreemde taal-luisteraars in Hoofdstuk 3, in tegenstelling tot die in Hoofdstuk 2, zeer vaardig waren in de vreemde taal, maakten ze toch gebruik van fonotactische aanwijzingen uit de moedertaal voor het segmenteren van de vreemde taal, hoewel de aanwijzingen voor een grens in de moedertaal onverenigbaar kunnen zijn met woordherkenning in de vreemde taal. Kennelijk konden ze niet simpelweg beslissen om informatie uit de fonologische structuur van hun moedertaal te negeren. Toch waren deze luisteraars ook in staat gebruik te maken van de fonotactische aanwijzingen voor een grens specifiek voor de vreemde taal.

In Hoofdstuk 4 is de rol van woordkandidaten uit de moedertaal en de vreemde taal bij het herkennen van spraak in de vreemde taal onderzocht. Woorden in de moedertaal die een beginsegment delen worden kort geactiveerd tijdens het herkennen van gesproken woorden. Bij de input *desk*, 'bureau', zullen Engelse luisteraars bijvoorbeeld *desk*, 'bureau', *dentist*, 'tandarts, en andere woordkandidaten activeren. Deze kandidaten zullen vervolgens onderling wedijveren om herkend te worden. In Hoofdstuk 4 is onderzocht of luisteraars woordkandidaten uit de moedertaal activeren bij het luisteren naar een vreemde taal. Nederlandse deelnemers volgden Engelse gesproken instructies om op afbeeldingen te klikken met een computermuis terwijl hun oogbewegingen werden geregistreerd. Een doel-afbeelding (bijv. van een *desk*, 'bureau') werd gepresenteerd tegelijk met afbeeldingen die als afleider fungeerden. Sommige afleider-afbeeldingen hadden Nederlandse namen die in het beginsegment overeenkwamen met de Engelse naam van de afbeelding (bijv. het Engelse doel *desk* /dɛsk/ en de Nederlandse afleider *deksel* /dɛksəl/). Oogbewegingen naar een afbeelding werden geïnterpreteerd als evidentie voor activatie van de lexicale representatie behorende bij de afbeelding. Nederlandse luisteraars keken vaker even naar Nederlandse woordkandidaten met een overeenkomstig begin dan naar ongerelateerde afleiders. Met andere woorden, wanneer een Nederlandse luisteraar die Engels kent naar een Engelse zin luistert die het woord *desk* /dɛsk/ bevat, wordt het fonologisch overeenkomstige Nederlandse woord *deksel* /dɛksəl/ als kandidaat beschouwd bij het woordherkenningsproces. Luisteraars lijken niet in staat het mentale lexicon van de moedertaal te deactiveren, zelfs in situaties waarin de moedertaal irrelevant en mogelijk misleidend is. Deze toepassing van processen uit de moedertaal op input in een vreemde taal is duidelijk niet bevorderlijk voor efficiënt luisteren.

Volwassen luisteraars beschikken over een repertoire van procedures die geschikt zijn voor de verwerking van spraak in de moedertaal. Geconfronteerd met een tweede, vreemde taal blijven zij deze procedures van de moedertaal echter toepassen, ongeacht of dit een efficiënte verwerking van de tweede taal bevordert of tegengaat. Luisteraars met een grote vaardigheid in een vreemde taal lijken erin te slagen ook procedures uit de vreemde taal te gebruiken, tenminste voor bepaalde aspecten van gesproken woordherkenning (zoals getoond voor fonotactische restricties in segmentatie). In conclusie kan worden gesteld dat de efficiëntie waarmee luisteraars spraak in de moedertaal kunnen verwerken tegelijkertijd een beperking vormt voor de efficiëntie waarmee geluisterd wordt naar een vreemde, later geleerde taal, tenminste wanneer de eerste en tweede taal verschillen in aspecten van de fonologische structuur die gebruikt worden bij het verwerken van gesproken taal.

Curriculum Vitae

Andrea Weber was born in München, Germany, on August 24, 1972. After she had graduated from the Carl-von-Linde Gymnasium in 1991 in Kempten, she studied in the program of linguistics and educational sciences, and in the program of speech sciences and communication at the University of Regensburg, Germany. In 1995 she received her certification in speech sciences and communication at the University of Regensburg. From 1995 to 1996 an Erasmus scholarship enabled her to study linguistics at the University of York, Great Britain. In 1996 she received her MA in linguistics and educational sciences at the University of Regensburg, Germany. In 1997 she was awarded a scholarship from the German Max-Planck-Gesellschaft to prepare her Ph.D. thesis at the Max-Planck-Institute for Psycholinguistics in Nijmegen, the Netherlands. From 1999 to 2000 she was a research associate at the Graduate Center of the City University of New York, USA.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsaggital articulography. *Niels Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organizing artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba, an Oceanic Language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language. *Eva Schultze-Berndt*

15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*

16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*