MPI SERIES
49

MPI SERIES
IN PSYCHOLINGUISTICS

PHONEME INVENTORIES AND PATTERNS OF SPEECH SOUND PERCEPTION

Anita Wagner

ISBN 978-90-76203-31-7

Anita Wagner
PHONEME INVENTORIES AND PATTERNS OF SPEECH SOUND PERCEPTION

Phoneme inventories
and
patterns of speech sound perception

ISBN: 978-90-76203-31-7

Cover design: Ponsen & Looijen bv, Wageningen
Cover illustration: "To each their own Babel" by Ambra Neri,
Gregory Nazairo Kibbelaar and Anita Wagner;
Humans inspired by La Linea da Osvaldo Cavandoli
Printed and bound by Ponsen & Looijen bv, Wageningen

Phoneme inventories and patterns of speech sound perception

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 23 juni 2008
om 13:30 uur precies

door
Anita Eva Wagner
geboren op 28 februari 1975
te Katowice (Polen)

Promotor:	Prof. dr. Anne Cutler
Co-promotor:	Dr. Mirjam Ernestus

Manuscriptcommissie:	Prof. dr. Rob Schreuder
	Prof. dr. Ann Bradlow (Northwestern University)
	Dr. Silke Hamann (Universität Duesseldorf)

Promotiecommissie:	Prof. dr. Ulrich Frauenfelder (Université de Genève)
	Prof. dr. Vincent van Heuven (Universiteit Leiden)
	Dr. Kevin Russell (University of Manitoba)
	Dr. Natasha Warner (University of Arizona)

The research reported in this thesis was supported by the NWO SPINOZA grant "Native and Non-Native listening" of the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) to Anne Cutler, and by the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

Für meine Eltern

Acknowledgments

At times one may wonder what keeps a PhD student working on a dissertation for years. A motivation to indulge in such a project lies in the addictive power of an environment that constantly encourages, supports and stimulates learning. I was lucky to be in just such an environment, and want to take the opportunity to thank the people who created it.

Essential for such an environment are the people who stay curious, who visibly enjoy their work, and appreciate hearing about other people's work. The Comprehension Group assembles such people. It is an addictive experience (and one gets used to it very quickly) to be surrounded by people who share their knowledge so readily, whose doors are always open, and who give criticism as well as instant help, if needs must also in the "very last minute". Thank you, Comprehension Group. Thank you, MPI.

Anne Cutler sustains this atmosphere of learning and sharing with her untamable scientific curiosity. I want to thank you, Anne, for creating an inspiring surrounding, for promoting me with trust, for your very clear comments, for accepting though not overlooking weaker points, and for not missing out the dialogue. Thank you, Anne.

I learned so much through working together with Mirjam Ernestus. As my day-to-day supervisor she showed me, among many other things, a contagious joy of analysing data, and how to see the patterns in them. I want to thank you, Mirjam, for introducing me to R, and for patiently making time for discussion.

Natasha Warner re-joined the group in a period that turned out to be the very final phase of my PhD. It might even be that this period turned out to be my final because of her persistence in allaying doubts and in pointing to the big picture. Natasha, I also want to thank you for explaining tricky differences between the meanings of words.

I am very grateful to have met colleagues, and friends, who were a constant source of motivation, diversion, and support. In particular, thank you Keren Shatzman, Martijn Goudbeek, Elizabeth Johnson, Petra van Alphen and Attila Andics. You welcomed every single question that popped up, and you would always come up with very practical hints and advises, straight to the point, or to another important one. Thank you.

I owe thanks to the participants of my experiments for putting the patterns in my data; and I want to thank the ones who helped me conducting my experiments in different parts of Europe. For hosting and supporting me, I thank, Prof. A. Garnham (University of Sussex), Mirjam Broersma (at that time in Brighton), Dr. J. Tambor (Uniwersytet Śląski w Katowicach), Dr. L. Tagliapietra (Università di Trieste), and Prof. Nuria Sebastián-Gallés (Universitat de Barcelona). Nuria Sebastián-Gallés welcomed me in her grup in Barcelona, and hosted me there for six months. I want to thank the GRNC for their hospitality, for their questions from a different perspective, and for the rafting experience. I owe thanks to Xavier, the technical wizard in this group. His fan construction saved my data by cooling down my laptop that suffered from the climate.

Speaking of the technical support, I want to thank the MPI Technical Group. Thank you Ad and Herbert for keeping things (and progressive correspondence) running; and thank you Tobias and Johan for your help, in particular, when I was collecting my data abroad.

Furthermore, I want to thank Ambra Neri, for drawing the Babel tower, con una mano volante (se, per caso, questa espressione non esista in italiano, pian piano la facciamo esistere). Thank you, Gregory, for your enthusiastic and pertinent help with the cover. Petra, thank you for turning the summary into a samenvatting.

Very special thanks go to the ones who provided the basics for inner sanity: Nina, Fem, Keren, E. (thank you for the music…), Frank, Paula, Petra (…and dance, dance, dance…), Ambra, Pamela (…the homey feeling…), Federico, Zab (…and the songs we were singing), Dennis, Martijn, (ridiculously good ideas), my paranimphen (for the two most special ways of grounding in times of mayhem), Claudia, Fermin, Broer-jam (authentic perpetual astonishment), Bego, Fattima, Carles (for gracing Barcelona), Seb (for your trust and real hard laughter).

Bedanken möchte ich mich auch noch bei dem „Anita Hilfs Fond“ und der Ladicorp. Das sind zwei wunderbare Einrichtungen, denen ich einiges schuldig bin und vieles verdanke. Insbesondere möchte ich mich bedanken für die unnachgiebigen Einführungen und Ausführungen zu dem Konzepte der Balance.

Zum Schluß möchte ich meinen Eltern danken. Ich weiß nicht wie man sich für das Vertrauen, welches ihr mir gebt bedanken könnte. Eure Unterstützung, Zuspruch und Wärme sind bedingungslos, und sträuben sich dagegen in Worte gefaßt zu werden. Danke.

Contents

Introduction

CHAPTER 1

Infants are able to become native in whatever language surrounds them. This implies that they can tell apart all possible speech sounds. Their presumably "universal" acoustic sensitivity starts attuning to the ambient language within the first year of exposure, and, as Ladefoged (1990, p. 343) put it, 'once a language has been learned one is living in a room with a limited view'. Such a 'limited view' can cause difficulties for understanding and learning foreign languages, but there is also a bright side to it: It is the manifestation of how the human perceptual system optimizes its processing to work quickly, accurately and efficiently to fit the requirements of one's native language.

The capacity to process native language in an effortless, automatic way results from detailed language-dependent specialization at a low level of perception. Listeners can hear certain differences between their native language and a foreign language. They are, however, not aware that they and listeners from different native backgrounds never perceive one and the same reality in speech. When English or Dutch listeners hear the Polish word *pstrąg*, meaning trout, they will hear that at least one sound is foreign. Most of them will also be able to assign this combination of sounds rather to Polish than to French. Yet, they will not be aware that they may pick up different aspects of the same sound combination. It seems trivial to state that native Polish listeners will perceive *pstrąg* in a different way than non-native listeners, but also English and Dutch listeners might differ in what acoustic information they pick up. Their perception is optimized to serve different languages. This dissertation is about cross-language differences that arise at the low level of automatic uptake of information about speech sounds.

People structure the world in a way which is optimally adapted to their surrounding environment. Experience with the immediate surroundings shapes human perception such that, within these surroundings, perception combines cognitive accuracy with economy. The senses of a person are continuously stimulated by an abundant amount of information. To reduce and structure such an information overflow, people learn to recognize objects and events by relying on the most telling features. It is, for instance, difficult to recognize faces of people from a different ethnic origin. The lack of contact with people with different facial features means that people do not need to attend to details which individuate these faces. As a consequence, people do not perceive dissimilarities among unfamiliar faces, while they are very good in recognizing familiar faces (Levin, 2000). Furthermore, just seeing more unfamiliar types of faces is not enough to learn how to recognize them (Ng & Lindsay, 1994). Rather, people need to discover new facial features which may not be informative in their own environment but might provide just the relevant cues to individuate unfamiliar faces.

The native language can shape human perception at several levels. Some argue that how languages label colors affects the way people perceive shades of one and the same color. Russian speakers, for instance, have distinct labels for light blue and dark blue. English listeners can differentiate two shades of blue, but compared to Russian speakers, they seem to make a less categorical distinction between shades of blue (Winawer, Witthoft, Frank, Wu, Wade & Boroditsky, 2007). When parsing sentences, English speakers orient themselves mostly to the order of words, whereas Italian speakers depend more on the agreement between parts of the sentence (Bates, Devescovi & D'Amico, 1999). In speech, listeners differ in how they find beginnings and ends of words (Cutler & Norris, 1988; Cutler, Mehler, Norris & Segui, 1986; Otake, Hatano, Cutler & Mehler, 1993), or in their knowledge about which speech sounds can co-occur to form native words (Weber, 2002). It becomes most apparent just how persistent language-specific selection of features is, when listeners learn foreign speech sounds.

Spanish and American English listeners, for instance, differ in how they distinguish the new non-native vowel contrasts /y/ and /oe/ (Goudbeek, Cutler & Smits, 2008). These two new front rounded vowels differ in their duration and in their spectral characteristics. Spanish and American English listeners base their distinction on features which are informative in their own language. American English listeners make use of both dimensions, because both provide reliable information in their native language. For Spanish listeners, however, duration does not provide information in their native language while spectral characteristics do. These listeners distinguish /y/ and /oe/ on the basis of the spectral characteristics only. Consider also Japanese listeners' notorious difficulty to differentiate the two sounds /r/ and /l/. Japanese adults can learn to distinguish these two sounds (Bradlow, Pisoni, Akhane-Yamada & Tohkura, 1997; Bardlow, Akhane-Yamada, Pisoni & Tohkura, 1999). Yet, they draw this distinction by selectively relying on different acoustic information than native English or German listeners (Iverson et al., 2003). In this way they miss the cues which most efficiently individuate /r/ and /l/.

Nowadays, the major part of the European population learns to communicate in English. When teaching a foreign language like English, teachers might find themselves in the position of instructing speakers from various native backgrounds. In the classroom, neither teachers nor students are likely to be aware of how differently they apprehend what they hear. Listeners may thus differ in how they perceive the difference between the two English words *sick* and *thick*. This distinction may be easy for Spanish listeners, because the contrasts /s/ and /θ/ translate into very similar native contrasts. German and Polish listeners will realize that the initial sound in *thick* is different from any native sound, but German listeners would deem it as very similar to /s/ while Polish listeners might find it more similar to /t/ or /f/. These listeners thus appear to apprehend different features of these sounds.

Theories of second language perception, like Best's Perceptual Assimilation Model (Best, McRoberts & Sithole, 1988) or Flege's Speech Learning Model (1995), describe how perception of similarities between foreign and native sounds relates to listeners' phoneme inventories. The question addressed in this dissertation is not how listeners differ in their perception of foreign speech sounds, but whether and how they differ in the way they extract information, even about native sound categories. The underlying assumption is that native language shapes listeners' perception, such that all native speech sounds can be identified efficiently. Listeners of seven different backgrounds are compared in how they apprehend the same speech signals.

Listening to native speech

When infants listen to the speech surrounding them, they hear acoustic signals with constantly changing and co-occurring acoustic patterns. It is generally assumed that these statistical occurrences of acoustic events help infants to deduce which sounds have a function in their language (Anderson, Morgan & White, 2003). Speech sounds that can turn the meaning of a word into a different word are assigned to distinct categories, and infants acquire a set of native phonemes. What develops when infants acquire a native phoneme inventory is a language-specific perceptual space.

Listeners' perceptual space is defined by all contrastive sounds in their language. Once a perceptual space has developed, differences between speech sounds are no longer perceived solely on the basis of their acoustic properties. Listeners then differentiate sounds which are functionally equivalent, despite acoustic variations, from sounds that constitute distinct phonemes. Since languages have different phoneme inventories, listeners also have different perceptual spaces. Boundaries between distinct speech sounds may divide listeners' perceptual spaces into sections, which depend on the number of all native speech sounds.

Furthermore, listeners' perceptual space plays a decisive role in how similarities and dissimilarities between speech sounds are perceived. Perceptual

distances between acoustic events within a category are shrunk, and listeners' sensitivity to acoustic variability within a category is reduced; perceptual distances between categories are stretched and listeners are more sensitive to acoustic variability between their phoneme categories (Kuhl, 1991). What is perceived as similar thus depends on the entire set of native contrasts.

It is unclear which levels of speech processing are altered by one's native language. It is generally agreed that the native language does not alter auditory processing, but there is evidence for different neuronal organization between listeners of different languages (e.g., Näätänen et al., 1997). Language-specific perceptual strategies are attributed to the level of attention by some researchers (Pisoni, Lively & Logan, 1994), while others, for instance Kuhl (2000), see adult listeners as "neuronally committed". It appears that language exposure alters perception on levels which lie somewhere between pre-attention and general sensitivity. Since this is yet an open question, both terms, sensitivity and attention, will be used interchangeably throughout this dissertation.

Listening to non-native speech

When listening to a foreign language, listeners apply their native perceptual strategies. They then fail to perceive acoustic differences between foreign and native speech sounds, and are unaware that they assimilate new speech sounds to references in their native perceptual spaces (Best, 1994). In this way, foreign speech sounds are perceived as equivalent to one or many native sound categories. Cross-language research has documented three factors which play a role in erroneous mapping of foreign speech sounds: phonemic, phonetic, and psychoacoustic (e.g., Polka, 1991; Werker & Logan, 1985).

At the phonemic level, listeners differ in which speech sounds are contrastive in their native language. Spanish listeners, for instance, do not distinguish the vowels /e/ versus /ɛ/, which for Catalan listeners clearly distinguish the male name *Pere* from the

word *pere* (pear). Catalan-Spanish bilinguals whose first language was Catalan draw this distinction automatically. Bilinguals whose parents spoke Spanish with them, however, perceive these two sounds as equivalent (Sebastián-Gallés, Echeverría & Bosch, 2005).

At the phonetic level, listeners differ in their knowledge about how speech sounds may vary acoustically while still belonging to the same category. Such variability can arise from dialectal differences, phonotactic rules, or from the modifications that speech sounds undergo when they co-occur with other sounds. For instance, Spanish listeners implicitly know how the sound /p/ varies when it occurs in *par, pera, por*, or *puro*. They do not know, however, what their /p/ would be like if it occurred with /æ/, as in the English word *patsy*, because this vowel is not part of their phoneme inventory. The processing of speech sounds is sensitive to listeners' implicit knowledge about the acoustic variance within a category. Spanish listeners, who have four times as many consonants as vowels, are aware that more consonants can alter a vowel, than vowels can alter the acoustic realization of a consonant. Knowing this, they are more cautious when they identify vowels in the context of various consonants than when they identify consonants in the context of various vowels. Dutch listeners, on the contrary, have a balanced vowel-consonant ratio, and do not show such a difference (Costa, Cutler & Sebastián-Gallés, 1998).

At the lowest – the psychoacoustic – level, listeners differ in how they attend to and weigh acoustic information, below the level of the phoneme. This is illustrated by the previously cited example of Japanese listeners, who do not rely on the cues that distinguish /r/ and /l/ for native English listeners. Another example is a study by Rochet (1991). This study reports that Brazilian Portuguese speakers perceive the French front-rounded vowel /y/ as similar to the front-unrounded vowel /i/, while Canadian English listeners hear it as more similar to the back rounded vowel /u/.

Most studies investigating the effects of native phoneme inventory compare two listener groups. A speech sound target establishes a native distinction for one group of listeners but is not phonemic for the other group (e.g., Best et al. 1988; Broersma,

2005; Flege, 1984). The aim of this dissertation is to find differences in phoneme perception at the low level of attention and integration of acoustic cues for native categories. Therefore, all listeners are compared on the identification of speech sounds that are phonemic in their language, but differ in the number of similar sounds that can compete with the target for identification. For example, most languages have a /s/-like sound, but they differ in the number of additional similar sounds. Do all listeners distinguish /s/ in the same way? Or does the presence of more similar contrasts make listeners select other cues to individuate a /s/?

Among listeners whose speech perception is shaped by different languages, there may be differences in processing, but there may also be regularities based on universal perceptual strategies. The following section will more generally describe how listeners may identify speech sounds. Three aspects will be discussed which could account for similarities in phoneme perception among all listeners. Common patterns among listeners may be attributed to the properties of the signal they hear, to general mechanisms of speech sound perception, or to general tendencies across phoneme inventories.

PERCEPTION OF SPEECH SOUNDS

Patterns in the speech signal

In general, the recognition of speech sounds starts with sensory processing. At this stage, all listeners rely on the analysis of their peripheral auditory system. The auditory system reacts to changes of energy in the air molecules which are the physical constitution of speech. These changing air compressions are the consequence of the modifications of a speaker's vocal tract, and they result in a complex acoustic signal. Such an acoustic signal bears an abundant amount of information scattered across its dimensions: frequency, intensity and time. The psychoacoustic view on speech perception assumes that in order to extract the meaning of words, listeners recognize patterns in the signal. When listeners extract the meaning of words, they may

automatically identify individual speech segments. For the identification of phonemes listeners would then extract acoustic cues from all the dimensions of the signal, and map these into their mental representations.

There are acoustic patterns which clearly distinguish speech sound classes, like vowels from fricatives. Within these patterns, there are cues which specify individual speech sounds, for instance /s/ versus /ʃ/. Some of these patterns can be linked to the steady part of articulation of these sounds, and are termed static cues. Sounds in speech, however, are not produced in isolation. Speech sounds mingle into syllables, syllables further concatenate with other syllables to form words and phrases. This concatenation of segments affects their exact acoustic manifestation. For instance, an ambiguous noise between /s/ and /ʃ/ can be recognized as /s/ if it precedes the vowel /u/, and as /ʃ/ if it precedes the vowel /a/ (e.g., Mann & Repp, 1980; Smits, 2001; Whalen, 1981). Acoustic cues resulting from the coarticulation of sounds are shorter than static cues. Coarticulatory cues contain mutual information about adjacent segments, and can also be termed transitional cues.

The question whether static or transitional cues provide more information for listeners has been subject to a long lasting debate (e.g., Kewley-Port, Pisoni, Studdert-Kennedy, 1983; Ohde & Ochs, 1996; Stevens & Blumstein, 1978,1981). Traditionally, static cues have been viewed as more robust because they are longer, and contain information specific only to one speech segment. Transitional cues have been seen as increasing the variance in the acoustic form of speech sounds. Coarticulatory information cannot be assigned to only one segment, and varies depending on factors like the speed of uttered sequences (Picheny, Durlach & Braida, 1989), the style of speech or the clarity of a speaker (Bradlow, 2002). The way speech sounds mutually affect one another, however, is lawful and perceptually informative (e.g., Beddor & Krakow, 1999; Manuel, 1990).

Relevant for the studies in this dissertation is the fact that coarticulation shows language-specific patterns, and depends partly on the distribution of contrasts in

phoneme inventories. More contrasts may constrain the production of individual speech sounds. To maintain the distinctiveness among these contrasts speakers may have to articulate more precisely, and their language may tolerate less coarticulation (Manuel, 1990). As a consequence, listeners may differ in the coarticulatory patterns they have been exposed to. They might thus also differ in the way they can make use of coarticulatory information in speech perception. The informativeness of transitional versus static cues may thus depend on the distributions of contrasts in phoneme inventories. The following section describes the main acoustic characteristics, static and transitional, that can contribute to the perception of the speech sounds which are the identification targets in the present dissertation. These are vowels, voiceless fricatives and voiceless stop consonants.

Vowels

Vocalic acoustic patterns result from an articulation with a relatively open vocal tract. The acoustic signal of vowels shows a relatively harmonic distribution of energy across frequency bands. The frequencies of these concentrations of energy, termed formants, reflect the resonances of the vocal tract and represents the static cues for vowel identification. Vowels can be distinguished from each other on the basis of these static cues (Strange, 1989).

Transient movements of formants from and into their steady-state values serve as dynamic cues to vowels and their perceptual relevance has been shown when the steady-state portion of vowels is deleted (Strange, 1999). The duration of formant transitions is largely dependent on the speaking rate, but usually less than 50 milliseconds (Furui, 1986; van Wieringen & Pols, 1995). Of interest for the present study is that the exact onsets and offsets of transitions are dependent on adjacent consonants (Delattre, Lieberman & Cooper, 1954). They thus provide mutual information about the vowel and the neighboring consonant.

Fricatives

Fricatives are characterized by high-frequency noises of a relatively long duration. This acoustic pattern results from a narrow constriction in the vocal tract. The distribution of energy across the frequencies reflects the location of the articulatory constriction. The frequencies of energy peaks in the noise spectrum are the static cues for fricatives (Stevens, 1998). These static cues and the intensity of the noise have been shown to provide sufficient information to distinguish all English fricatives. (e.g., Heinz & Stevens, 1961; Hedrick & Ohde, 1993; Jongman, 1989; Jongman, Wayland & Wong., 2000).

Dynamic cues to fricatives are contained in the vowel portion adjacent to fricatives, and in the slight modifications of the fricative spectrum as a function of the neighboring vowels. The salience of static cues differs between fricatives, and formant transitions can provide additional information for less distinct fricatives, like /f/ and /θ/ (Harris, 1958). As argued by Whalen (1989) and Smits (2001), dynamic cues to fricative identification can be perceptually integrated with the cues in the static noise spectrum.

Stop consonants

Stop consonants are abrupt and short acoustic events, resulting form a complete constriction within the vocal tract. The acoustic features of voiceless stop consonants are: a silent interval of about 60-120 milliseconds corresponding to the closure, followed by a 5-10 millisecond high intensity noise resulting from the release of the constriction. The distribution of energy in the release bursts have been shown to provide the most relevant acoustic cues for stop consonants (e.g., Blumstein, 1981; Stevens, 1998).

Transitional cues are found where the closure and the release burst merge with the surrounding vowel. Formant transitions following the burst have consistently been

shown to provide reliable cues to place of articulation of stop consonants (e.g., Liberman, Delattre, Cooper & Gerstman, 1954; Sussman, Fruchter & Sirosh, 1998).

Patterns of perception

Related to the question whether static or transitional cues provide more information for listeners is the issue of whether some acoustic patterns could invariantly specify speech sounds. There have been attempts to find invariant properties in the signal, most notably by Stevens, as formulated in his Quantal Theory of Speech Perception (Stevens, 1972, 1989). This theory acknowledges that some acoustic events have a bigger perceptual impact than others. This is attributed to general auditory mechanisms. Some acoustic events thus appear to create perceptually salient and robust contrasts. Speech sounds which are characterized by such robust acoustic features are assumed to be more frequent in the phoneme inventories (e.g., Schwartz , Boë, Vallée & Abry, 1997).

There may, however, be no acoustic patterns which are invariant cues for all listeners. Alternatively, listeners may make use of all acoustic cues in the signal (Diehl & Kluender, 1987). Nonetheless, there are perceptual patterns, which are shared among listeners of a language and vary between languages. Speech perception theories like Nearey's empiricist approach to sound perception (Nearey, 1997) account for language-specific weighting of cues, while still acknowledging that some acoustic patterns might be auditorily preferred. This view implies that listeners might have a 'choice' in their selection of acoustic cues.

Listeners can make 'choices' from a multiplicity of cues. In the absence of static cues like the release burst of a plosive, listeners can identify stop consonants by relying on the information in the formant transitions (Dorman, Studdert-Kennedy & Raphael, 1977). Furthermore, listeners can also 'choose' from cues that mutually contribute to the identity of more than one segment. The four words *bat, bet, bad* and *bed*, are distinct because of the quality of the vowels /a/ and /e/, and because of the

plosives /t/ versus /d/. The vowel formants and their dynamic movements into the consonant cue at the same time the identity of the vowels and the place of articulation of the stop consonant. The duration of the vowels also contributes to their identity, while at the same time it is a cue for the voicing distinction between the /t/ and /d/ (Mermelstein, 1978). Finally, listeners weigh acoustic cues in language-specific ways. The duration of the vowel contributes to the distinction between *bet* and *bed* for English listeners (Crowther & Mann, 1992), Dutch listeners partly but inconsistently rely on the duration of the vowel (Broersma, 2005), and Arabic listeners do not make use of duration at all (Crowther & Mann, 1994).

To sum up, although there may be no cues which invariantly signal speech segments for all listeners, there may be acoustic distinctions which are generally easier to perceive. The perceptual robustness of these distinctions may give them a favored status among phoneme inventories. Acoustic features of such speech sounds might thus, in line with Stevens' view, form natural boundaries in the distinctions between speech sounds. The question addressed in this dissertation is whether listeners differ in their 'choices' of acoustic cues, when identifying the same speech sounds. To assure that all listeners are able to identify the same sounds, even though they would produce them differently, the identification targets used in this dissertation are the most frequent segments in the world's phoneme inventories. These are the point vowels /a i u/, the fricatives /f/ and /s/, and the stop consonants /p t k/.

Patterns among phoneme inventories

Phoneme inventories contain subsets of a 'universal' set of speech sounds. The International Phonetic Alphabet lists 114 articulatorily possible sounds and 31 modes in which some sounds can be secondarily modified. Twenty-eight of these speech sounds are vowels and 86 are consonants, the two main building blocks of words. The size of phoneme inventories can thus differ a great deal. On the one extreme there is the language !Xu with the largest phoneme inventory of about 110 distinctions, and on

the other extreme there are the languages Rotokas or Mura, with only 12-15 different speech sounds (Maddieson, 1984). These examples quickly illustrate what different occurrences of acoustic patterns listeners of Rotokas have been exposed to compared to !Xu listeners. The perceptual space of !Xu speakers contains nearly ten times as many phoneme boundaries as the perceptual space of Rotokas listeners. Acoustic events which belong to one category in Rotokas might be members of many different categories for !Xu speakers, and !Xu listeners might show greater sensitivity to acoustic differences within Rotokas' phoneme categories.

Despite the diversity in a 'universal' phoneme inventory, there are some co-occurring patterns in the way languages set up their phoneme inventories. Selection of speech sounds may be guided by competing demands of articulatory economy and perceptual distinctiveness (e.g., Liljencrants & Lindblom, 1972). Regarding the size of phoneme inventories, languages need to have a sufficient number of contrasts to create perceptually distinct words. Fewer contrasts lead to longer words, and more homophones in the lexicon (Cutler, Mister, E., Norris & Sebastián-Gallés, 2004; Nettle, 1995). As a consequence, the distinction between words may be more demanding. More phonemic contrasts allow for shorter words, less homophones, and an easier disambiguation in the lexicon. The processing of individual speech segments may, however, cause greater articulatory effort or perceptual complexity. As most exhaustively documented in the UCLA Phonological Segment Inventory Database (Maddieson, 1984), most languages distinguish between 20-37 phonemes, with typically approximately 2/5 of these vowels, and 3/5 consonants.

The distribution of speech sounds also appears to be motivated by the demand of perceptual distinctiveness at a minimum articulatory effort (Schwarz, Boë, Vallée & Abry, 1997). Systematicities among vowels inventories are documented by numerous studies (e.g., Disner, 1983; Jongman, Fourakis & Sereno, 1989; Liljencrants & Lindblom, 1972). Respecting the demand of perceptual distinctiveness, all languages will contain the cardinal point vowels /a i u/ before distinguishing other vowel qualities. These three vowels are produced at the extreme ends of the articulatory

system, are thus located at the edges of a global vowel space, which grants their perceptual robustness.

Similarities have also been observed among the consonantal systems (Lindblom & Maddieson, 1988). The most frequent stop consonants in phoneme inventories are /p t k/, for fricatives it is the pair /f/ and /s/ (Maddieson, 1984). The group of consonants, however, is more heterogenic compared to vowel systems. Consonants differ in manner, in place of articulation, and in secondary articulations. Some languages show an accumulation of consonant contrasts at certain places of articulation. Some areas in these listeners' perceptual spaces are thus more crowded. Consonant targets which are articulated close to one another may share acoustic features. To compensate for a greater perceptual similarity between contrasts listeners may rely on cues from coarticulation to back up the information conveyed by the static cues.

What are the effects of different sizes and distributions of contrasts in phoneme inventories? Such effects can be found in the way speakers produce speech sounds and how they perceive them. Production and perception can affect each other mutually. For production, Nettle (1994) reports that the number of vowels versus consonants in a language has an effect on the average volume of speech. More vowels in a language, thus a greater number of sonorous sounds, permit a softer production. Languages with fewer vowels may compensate for fewer sonorous sounds by increasing the intensity of speech. The acoustic vowel spaces of languages with more vowel contrasts can be expanded compared to languages with fewer vowels. This has been shown for American English listeners, who have more vowel contrasts than Spanish listeners (Bradlow, 1995). Interestingly, the absolute position of the boundaries between the point vowels /a i u/ is similar for the American English vowel space and for the Spanish vowel space (Bradlow, 1996). The number of speech sounds also affects coarticulation, as has been shown for vowels by Manuel (1990). Furthermore, listeners use language-specific patterns of coarticulation when identifying speech sounds (Beddor & Krakow, 1999).

THE CURRENT STUDY

The languages tested in this dissertation are British English, Catalan, Castilian Spanish, Dutch, German, Italian, and Polish. All these languages have the vowels /a i u/, the voiceless fricatives /f s/ and the stop consonants /p t k/ as phonemes, but they differ in the number of similar phonemes which compete with these targets. For vowels, English listeners distinguish approximately 20 vowel qualities, Dutch and German listeners make distinctions between 15-16 vowels, while Spanish listener distinguish only five different vowels. As a consequence, when identifying the point vowels /a i u/ Dutch and English listeners will have to exclude more acoustically similar competitors. Furthermore, Spanish listeners may tolerate a greater acoustic variability within vowel categories.

Regarding the fricatives, Polish is the language with the highest number of distinct categories. It contains eleven fricative phonemes, eight of which are articulated at palatal places of articulation. English contains the acoustically similar fricative pair /f/ versus /θ/. These fricatives are also part of the Spanish inventory, though in total it distinguishes only half as many fricatives as English. The perceptual spaces for fricatives might be expanded for listeners with more fricative contrasts. Alternatively, listeners may 'choose' to rely on more cues, to accurately distinguish the greater number of contrasts. In addition, coarticulation of vowels and fricatives might be more informative for listeners who have more fricative contrasts. These listeners may be used to more careful realisation of fricatives because speakers of their native language have to maintain distinctiveness among a greater number of contrasts.
The smallest difference in the distribution of contrasts among the languages tested occurs for the stop consonants. All of these languages contrast six phonemes. However, the languages differ in the number of vowels which can co-occur with plosives. All listeners have been exposed to different co-occurrences of stop

consonants and vowels. They may thus differ in their knowledge about the potential variability within these plosive categories.

The question addressed in this thesis is how such differences in the make-up of phoneme inventories affect listeners' perception of native speech sounds. The presence of more sound categories may reduce the perceptual saliency of similar sounds, and result in: (1) No differences between listeners, because each phoneme may be identified independently of other contrasts; (2) Longer processing times and lower accuracy in identification of contrasts in more crowded perceptual areas; (3) Different strategies in selecting and weighting acoustic cues to compensate for a reduced perceptual saliency in more crowded perceptual areas; (4) Different windows of integration of cues to perceptually less distinct contrasts; (5) Differences in the temporal uptake of cues specifying individual contrasts or phonological features. In the latter case, as the speech signal evolves, listeners may have different perceptual images of one and the same acoustic reality.

Structure of this thesis

Chapter 2 presents experiments designed to test how the number of contrasts in a native language affects the speed and accuracy of listeners' phoneme identification. Visual perception is affected by the number of alternative choices (set-size effect) and by the similarity among the alternatives (e.g., Palmer, Verghese & Pavel, 2000; Theeuwes, 1992). The perceptual strength of an object in a display is reduced if more objects, or similar objects, compete with the target for identification. If this is a general pattern of perceptual processing, the set-size effect may also translate to phoneme identification. In that case, listeners who have more categories with similar acoustic properties might identify the target more slowly and less accurately. In contrast to visual perception, however, the number and similarity of competitor contrasts cannot be manipulated within participants. The native language determines the number of competitors. They are thus not presented in a display, but are an internal representation of phonemes.

Chapter 3 investigates how the presence of similar fricative sounds in a native phoneme inventory affects listeners' reliance on transitional cues. Do listeners who have more fricatives in their phoneme inventories rely more on transitional information? The languages compared are Dutch and German which have spectrally distinctive fricatives, and English, Polish and Spanish which contain perceptually similar fricative contrasts. Additional contrasts within the phoneme inventories of the latter languages may crowd the perceptual space of the fricatives /f/ and /s/. This may thus create the need to rely on more acoustic cues, such as transitions, in order to accurately distinguish all native fricative contrasts.

Chapter 4 further investigates whether listeners who rely on transitional cues for fricative identification find cues to fricatives earlier in the signal than listeners who rely on static cues. Listeners with more similar fricative contrasts may optimize their perceptual strategies to gain the necessary information as soon as possible. The experiment in Chapter 4 further queries whether cross-language differences in the reliance on coarticulatory cues are specific to fricatives, or whether they generalize to other phoneme types. Chapter 5 summarizes the results and discusses their implications for native and non-native listening.

Identification of phonemes: Differences between phoneme classes and the effect of class size

CHAPTER 2

Wagner, A. & Ernestus, M. (2008), *Phonetica, 65,* 106-127.

Abstract

This study reports general and language-specific patterns in phoneme identification. In a series of phoneme monitoring experiments, Castilian Spanish, Catalan, Dutch, English, and Polish listeners identified vowel, fricative, and stop consonant targets that are phonemic in all these languages, embedded in nonsense words. Fricatives were generally identified more slowly than vowels, while the speed of identification for stop consonants was highly dependent on the onset of the measurements. Moreover, listeners' response latencies and accuracy in detecting a phoneme correlated with the number of categories within that phoneme's class in the listener's native phoneme repertoire: More native categories slowed listeners down and decreased their accuracy. We excluded the possibility that this effect stems from differences in the frequencies of occurrence of the phonemes in the different languages. Rather, the effect of the number of categories can be explained by general properties of the perception system, which cause language-specific patterns in speech processing.

INTRODUCTION

Listeners are able to focus on individual speech sounds and identify them in an effortless and largely accurate manner. Here we investigate whether identification of speech sounds varies among sound classes and among listener groups with different sets of contrastive speech sounds. We compare the identification of speech sounds between vowels, fricatives, and stop consonants and across listeners with a vowel- or fricative-rich repertoire versus listeners with fewer categories in these speech sound classes.

Models of speech perception vary in the role they ascribe to individual speech sounds, and whether they incorporate a level of prelexical phonemic processing (e.g., Norris, McQueen &Cutler, 2000; McClelland & Elman, 1986, Johnson, 2004). While the instant activation of phonemes in speech processing is controversial, daily observations, such as the occurrence of spoonerisms and puns in languages, and phonemically based orthographic systems, show listeners' effortless ability to focus on individual speech sounds. Furthermore, several studies have demonstrated that listeners' perception adjusts rapidly to speaker-specific phoneme realisations (Norris, McQueen & Cutler, 2003, Eisner & McQueen, 2005), and such adjustments spread to other instances of these phonemes in new words (McQueen, Cutler & Norris, 2006). Moreover, brain imaging studies have shown the existence of neuronal traces of phoneme representations (Näätänen, Lehtowski, Lennes, Cheour, Houtilainen, Livonen, Vainio, Alku, Ilmoniemi, Luuk, Allik, Sinkkonnen & Alho, 1997). Also, reports on brain-damaged patients show that listeners may have a normal ability to recognize individual speech sounds even though their lexical representations are disrupted (e.g., Martin, Breedin & Damian, 1999).

Commonly, speech sounds are divided into two main classes: vowels and consonants. These two groups are the alternating building blocks of words. They differ in their phonological function, with vowels forming the centres and consonants

forming the margins of syllables. The different phonological functions of vowels and consonants are reflected in different contributions of these speech sound classes to word recognition: Vowels appear to restrict lexical selection less than consonants (Cutler, Sebastián-Gallés, Solar-Vilageliu & van Ooijen, 2000; Bonatti, Pena, Nespor & Mehler, 2005). Cutler and colleagues, for instance, showed that speakers tend to change vowels rather than consonants when they are asked to turn pseudo-words into existing words. Further indications for differences in processing between vowels and consonants come from aphasic patients: Patients may be hampered in the production of only one of these classes, suggesting that vowels and consonants are processed by distinct neural mechanisms (Caramazza, Chialant, Capasso & Micelli 2000, but see Sharp, Scott, Cutler and Wise, 2005 for a different view in perception).

In acoustic terms, stop consonants are very different from vowels, and this acoustic difference forms the basis of the explanation for categorical perception of stop consonants versus continuous perception of vowels. In a series of identification and discrimination experiments, Liberman and colleagues (e.g. Liberman, Harris, Hoffman & Griffith, 1957) observed that listeners perceive stop consonants categorically (i.e., do not distinguish between different realisations of the same phoneme), whereas differences in the precise quality within a vowel category are perceived easily (more continuous discrimination). The perception of intraphonemic acoustic variation in stop consonants is less in correspondence with the actual fine-grained variation in the acoustic signal than the perception of subtle acoustic differences within a vowel category.

Pisoni and Tash (1974) suggested that vowels and consonants differ in the way they are encoded in auditory and phonetic memory. As argued by Pisoni (1973), two modes of memory play a role in phoneme discrimination and identification: auditory memory, where detailed perceptual traces are stored but decay fast, and phonetic memory, where the acoustic signal is assigned to phonemic categories. Stop consonants, because of their shorter and more abrupt acoustic properties, leave traces

in auditory short-term memory that decay faster compared to the traces of longer and continuous acoustic events like vowels. As a consequence, the traces of vowels are longer available for retrieval, and they allow detailed and more continuous discrimination. When discriminating stop consonants, listeners rely more on the information in phonetic memory, where the signal has been labeled and assigned to a phonemic category.

If the difference between categorical and continuous perception is due to the acoustic properties of the segment, the large group of consonants should also show within-group differences, as this group contains a heterogeneity of phonemes with many different acoustic properties. This is indeed the case. Healy and Repp (1982) conducted identification, discrimination, and labeling experiments with vowels and fricatives, and found that, in contrast to stop consonants, both vowels and fricatives are not categorically perceived. The discrimination precision was even higher for fricatives than for vowels.

Two decades later, Mirman, Holt and McClelland (2004) investigated the processing of non-speech sounds. Listeners categorised non-speech materials, which contained either steady-state sounds resembling simplified vowels or fricatives, or sounds with transient properties similar to consonants like stop consonants, or both. It appeared that listeners cannot discriminate rapidly changing sounds belonging to the same category, while they can easily perceive subtle acoustic variation within the boundaries of a category for steady state sounds. The authors conclude that this supports the hypothesis that vowels and fricatives are identified differently from stop consonants because of their acoustic properties. Rapidly changing sounds, such as stop consonants, tend to be discriminated according to their phonemic labels, while steady-state sounds, such as vowels and fricatives, tend to be discriminated in an acoustically more detailed manner.

Differences between vowels, fricatives, and stop consonants are also reflected in response latencies in phoneme monitoring experiments. Foss and Swinney (1973) reported slightly longer response times to fricatives than to stop consonants. Similarly,

Savin and Bever (1970) found that listeners identify an initial phoneme in nonsense syllables faster if it is a stop consonant than if it is a fricative, while vowels are detected even more slowly. Rubin, Turvey, and van Gelder (1976) observed similar differences between word-initial /b/ and /s/ and Morton and Long (1976) between word-initial plosives and non-plosives, which included fricatives, glides, and a nasal. Finally, Van Ooijen (1994) showed that the position of the phoneme in the word may play a role, at least if the stimulus is an existing word. She found that vowels were detected more slowly than stop consonants and fricatives, especially in word-final position.

Note that the studies summarised above all investigated phoneme recognition with native speakers of English. A study by Cutler and Otake (1994) is exceptional in this respect. It compared the identification of nasal consonants and vowels by Japanese and English listeners. English listeners detected vowels significantly more slowly and less accurately than nasals, independently of whether these sounds were presented in English or in Japanese words. Japanese listeners, on the other hand, did not recognise vowels more slowly than nasals. Cutler and Otake argued that Japanese listeners are not slower in identifying vowels than consonants because, in contrast to English listeners, they have only few vowels in their phoneme inventory with which a target vowel can be confused. Language-specific properties may thus obscure or induce seemingly general differences between phoneme classes, since listeners' perception is shaped by their experience with their native speech sound categories.

Also Costa, Cutler and Sebastián-Gallés (1998) have reported that the number of phonemes in the native inventory plays a role in phoneme identification. The authors described a phoneme monitoring experiment with Dutch and Spanish participants. Listeners detected vowel or consonant targets in CVCVCVCVCV strings, in which the vowel or the consonant preceding the target was either constant over the stimulus or varied between syllables (e.g., for the target /p/ *ku su tu su pu* versus *ko se to si pu*). Dutch listeners, whose language has an approximately balanced vowel to consonant ratio, were delayed to the same extent by variation in the consonantal

context for vowels as by variation in the vocalic context for consonant targets. In contrast, Spanish listeners, whose phoneme repertoire has four times as many consonants as vowels, showed a greater effect of variation in the consonantal than in the vocalic context. Costa and colleagues explained this difference between Dutch and Spanish by arguing that listeners are aware of the influence that co-occurring phonemes have on the exact realisation of a phoneme. For consonants, this variation is smaller in Spanish than in Dutch, as Spanish has only five, instead of 16 vowels.

Combining the findings in these studies on the processing of speech sounds, we formulated two hypotheses. Both hypotheses may affect listeners' identification of speech sounds simultaneously. The first hypothesis states that speech sound classes require different recognition times. This hypothesis is based not only on the differences in acoustic properties between the sound classes but also on differences in phonological and lexical function. As mentioned above, vowels play a smaller role in lexical processing than consonants, and reaction times may therefore be longer for vowels.

The second hypothesis is that differences between the speech sound classes will be modulated by the number of categories within these classes in the listener's native phoneme repertoire. Listeners with a higher number of categories within a certain speech sound class will identify a target of that class more slowly than listeners whose native repertoire does not contain as many categories in that class. If this hypothesis is correct, the number of categories should be taken into account in order to ascertain general differences between vowels, stop consonants, and fricatives.

Importantly, the second hypothesis is based on general processes of categorisation, which are not restricted to auditory perception. When participants make decisions, like for instance about the identity of a colour or shape, their processing time is longer when they have more alternative choices (e.g., Hick 1952, Nosofsky 1997, Schweickert 1993, Theeuwes, 1992). In order to make clear that our second hypothesis is not specific for speech processing, we will use the term "category" to refer to phonemes. Categories instantiate listeners' knowledge, which may be

formulated in terms of phonemes, and which is established during speech development. Note that even though the effect of the number of categories would result in language-specific performance, it would affect listeners of all languages in the same way.

The question arises whether the relevant categories are indeed the phonemes. Many phonological and psycholinguistic models (e.g., Norris et al., 2000, McClelland & Elman, 1986) assign an important role to the phoneme, which is a theoretical construct. Listeners, however, can also distinguish between allophones of the same phoneme (e.g., between the palatal and the uvular fricative in German, see Lipski 2006), and these allophones may therefore play an important role in speech processing as well. Hence, the number of relevant categories may be the number of phonemes or the number of distinguishable speech sounds. We decided to focus on phonemic categories in the current study. The most important reason is that there is not sufficient data to determine which sounds can be distinguished by which listeners.

Different from the studies mentioned above, our study examined five listener groups of different native backgrounds (in previous studies maximally two groups had been tested). If indeed phoneme classes differ in the speed and accuracy of identification due to their function and acoustic manifestation, the same differences should be found for all languages. However, as the listener groups differ in their number of categories for these phoneme classes, we nevertheless expect differences between the language groups, as a function of these numbers of categories.

Naturally, the languages of the listeners also differ in many other respects, in addition to their phoneme inventories, and these differences contribute to differences in speech processing. Examples are the languages' stress patterns, syllable structures, and phonotactic restrictions. These language-specific characteristics might make it difficult to find clear general differences between phoneme classes and a role for the number of categories.

In order to investigate how listeners' perception is shaped by both general and language-specific factors, we have to make sure that all listeners can use their native

listening strategies. One possibility is to present each listener group with natural materials produced by a native speaker of their own language. This, however, would introduce an additional source of variability, as all language groups would then be presented with different stimuli. Another possibility is to present all listeners with synthetic stimuli, which has frequently been done in cross-linguistic research (e.g., Iverson, Kuhl, Akhane-Yamada, Diesch, Tohkura, Kettermann & Siebert, 2003, Bradlow, 1996). With synthetic stimuli, however, we run the risk of presenting listeners with impoverished stimuli. Previous findings show that listeners differ in their selection of, or attention to, acoustic cues, depending on their native language (Iverson et al., 2003; Wagner, Ernestus & Cutler, 2006), and synthetically generated materials may fail to represent especially those cues relevant only for some groups of listeners.

We decided to take advantage of the assumption that listeners, when presented with a foreign language, assign the foreign sounds to their most similar native categories. We presented all listeners with the same naturally produced materials, consisting of segments which are phonemic in all languages to be tested. Variability between language groups was further reduced by choosing nonsense words as materials. In such a way, we restricted potential lexical effects, and created conditions under which listeners focus more on the acoustic surface form of the materials.

An experimental paradigm that can reveal processes at the level of speech sounds by means of nonsense words is phoneme monitoring. In this paradigm, listeners are presented with lists of words, sentences, or nonsense words, and are asked to detect target phonemes. The measured reaction times and accuracy can give us insight into speech processing, including general and language-specific patterns in speech perception (for an overview see Connine & Titone, 1996). Thus, with this paradigm, language-specific strategies can be revealed, and have previously been reported, for speech perception (Cutler & Otake, 1994, Costa et al., 1998).

Phoneme monitoring is a much-used paradigm that has contributed to the investigation of a wide range of questions, regarding both the prelexical and the lexical level of speech processing. Results obtained with phoneme monitoring have been

replicated by means of other experimental paradigms, especially auditory lexical decision, such as the role of a word's frequency of occurrence (Phoneme monitoring: Dupoux and Mehler, 1990; Lexical Decision: Luce, 1986) and phonological similarity effects (Phoneme monitoring: Foss & Dowell, 1971; Lexical Decision: Luce, 1986).

When participants listen for a target phoneme in nonsense words, they compare the incoming signal with their mental representation of the target. Naturally, languages differ in their exact acoustic manifestation of the phonemes, and, as a consequence, if participants listen to words produced by a speaker of a foreign language, they will probably hear not the best examples of their speech sound categories. Nonetheless, they will extract acoustic cues which are relevant for the identity of the speech sound, and will rely on general acoustic cues to this segment (according to Stevens' 2002 acoustic landmarks), in addition to selecting cues in a language specific way (e.g., Iverson et al., 2003; Wagner et al., 2006). Importantly, in contrast to discrimination experiments, in phoneme monitoring experiments listeners are asked to assign the auditory stimulus to a mental representation as fast as possible. In such speeded categorisation tasks, listeners' reaction times have been shown to be hardly affected by goodness of stimulus' category (Miller 2001, Flege, Munro & Fox 1994).

EXPERIMENT

Languages compared

We compared listeners of five different languages: one Slavic language (Polish), two Germanic languages (Dutch and British English), and two Romance languages (Catalan and Castilian Spanish). Among the many differences between these languages, the focus in this study is on the numbers of categories for the three speech sound classes - vowels, fricatives, and stop consonants. Table 1 displays the phonemes in these classes in the five languages.

Dialectal variations within a language add or eliminate some phonemes for certain listener groups. Also, due to language-specific phonotactic rules, phonemes may differ in their frequency of occurrence, and their occurrence may be restricted to certain contexts. For instance, Spanish listeners acquire four different fricatives in their native language, but one of them, the /x/, seldom occurs in word-final position (e.g., see LEXESP, Sebastián-Galles, Cuetos, Carreiras & Martí, 2000). Furthermore, phonological descriptions of even the same language variety may list different numbers of phonemes. The numbers in Table 1 can be considered as averages of the proposed numbers and as the numbers that most authors agree on. We followed Carbonell and Llisterri (1992) for Catalan, Martinez-Celdran, Fernandez-Planas, Carrera-Sabate (2003) for Castillian Spanish, Booij (1995) for Dutch, Ladefoged (2001) for British English, and Rothstein (1993), and Zygis and Hamann (2003) for Polish.

	Vowels	Stop consonants	Fricatives
Catalan	i e ɛ ɑ ə ɔ o u (8)	p b t d k g (6)	f s z ʃ ʒ (5)
Dutch	i y ɪ e ʏ ø ɛ ə u o ɔ ɑ a ɛi œy ɔu (16)	p b t d k (5)	f v s z x h (6)
English	i ɪ ɛ æ ɑ u ʊ ɔ ʌ ɒ eɪ ɜ aɪ aʊ aə ɔɪ ɪə ɛə əʊ ʊə (20)	p b t d k g (6)	f v θ ð s z ʃ ʒ h (9)
Polish	i ɛ a ɨ ɔ u ɛ̃ ɔ̃ (8)	p b t d k g (6)	f v s z ʃʲ ʒʲ ʂ ʐ̩ ɕ ʑ x (11)
Spanish (Castilian)	i e a o u (5)	p b t d k g (6)	f θ s x (4)

Table 1: Phonemic categories for vowels, stop consonants, and fricatives in the five languages tested. The numbers of categories are given between brackets.

For some languages, the vowels include diphthongs. The definition of diphthong has been subject to a debate among phoneticians for decades (cf. Gottfried, Miller & Meyer, 1993). In the present study, only diphthongs which are consistently described as consisting of two vowel qualities were taken into account. Hence, we counted diphthongs as different vowel categories only for Dutch and British English (Ladefoged, 2001, Fry, 1979, Booij, 1995, Rietveld & Van Heuven 2001:71). Some descriptions of the phoneme inventories of Spanish and Catalan also contain the notion of diphthongs, but these diphthongs are formed by one of the glides /j/ or /w/ and a vowel (e.g., Martinez-Celdran et al., 2003, Green, 1990). For the same reasons the British English diphthong /ju/ as in *hue*, was not counted as a vowel category for English.

The variation in the numbers of categories among the languages is evident. For instance, if we consider the fricatives, we see that Polish listeners discriminate nearly twice as many categories as Catalan, Dutch or Spanish listeners. With respect to the vowel categories, British English listeners distinguish approximately four times as many vowels as Spanish listeners. The smallest variation among the languages appears in the distribution of stop consonants. The number of categories is treated as an independent variable in the analyses, and is given in brackets in Table 1.

Naturally, these languages also differ in the exact realization of the phonemes. For instance, Spanish and Catalan speakers produce stop consonants without aspiration, Dutch and Polish speakers with little aspiration, and English speakers with long aspiration following the burst. Similarly, the vowels in these languages differ in their average formant values (see, e.g., the chapters on the relevant languages in IPA 1999, Bradlow 1995). The fricative targets in the present study (/s/ and /f/) show the least variation among the standard variants of the languages tested. For a more detailed description of the acoustic properties of fricatives in these languages see Jongman, Sereno, Wayland and Wong (1998) for English, Rietveld and Van Heuven (2001) for Dutch, Jassem (1965) for Polish, and Borzone and Massone (1981) for Spanish. Note

that, however, as described above, phoneme monitoring will hardly be affected by variation at this low phonetic level.

Materials

We created 60 words consisting of three, and 60 words consisting of four consonant (C) – vowel (V) syllables. The consonants were of the set /p t k f s/, and the vowels of the set /a i u o e/. Each phoneme occurred only once per word. These CV strings were nonsense words in all the languages tested.

In these 120 critical items, the target phonemes, /p t k f s a i u/, were always in the final syllable (e.g., /p/ or /u/ could be the target in *fasipu*). Each consonant appeared as target in 15 nonsense words, forming a syllable with each of the three vowels /a i u/ in five nonsense words. Similarly, each vowel appeared as a target in combination with one of the consonants /p t k f s/ in three nonsense words. Appendix A lists all the critical items and the corresponding target phonemes.

In addition to these critical items, 15 nonsense words were created for each target phoneme in which the target appeared in the penultimate syllable, and 15 nonsense words in which the target was missing. Ten practice items were created as well, which familiarised listeners with the experimental situation before the actual test period started.

A male Spanish speaker read the list of stimuli with primary stress on the first syllable. He was instructed to produce the words as if they were existing Spanish words. Thus, the plosives in the materials were unaspirated, the vowels were produced according to Spanish qualities and quantities, and the fricatives were labiodental /f/, and apical alveolar /s/. Recordings were made in a sound attenuated room directly to a computer, and then down-sampled to 22.05 kHz (16 bit resolution).

Procedure

Participants sat in a sound attenuated room in front of a computer screen. They were presented with the stimuli over headphones. The trials were blocked by target phoneme, with the order of blocks counterbalanced among participants. Every block of stimuli was followed by a break, the duration of which was controlled by the participants themselves.

The experimentator informed all participants orally about all targets before the experiment started. In addition, during the experiment a letter appeared on the computer screen designating the current target sound. The English listeners also heard this target over their headphones at the beginning of the block, because of the large grapheme-phoneme discrepancy in English. These auditorily presented phoneme realisations were recorded by a phonetically trained speaker, who produced the targets following the phrase “Press the button as soon as possible when you hear an ..”. The target phonemes were realized as a labiodental [f], an alveolar [s], unaspirated stop consonants, and the vowels [a], [i], [u]. The speaker produced the vowels as close as possible to their cardinal positions. A small group of native listeners of the languages tested judged that these vowels sounded like good examples of vowels in their language.

Participants were instructed to press a key as soon as they recognized the target phoneme in the aurally presented materials. From the onset of each item, listeners had 2000 ms to respond. Failures to respond, and response latencies over 2000 ms, were defined as timeout errors. The experiment was self-paced: The next stimulus was presented 1000 ms after the participant’s response or, in case of a timeout, 3000 ms after the onset of the previous trial, and it was preceded by a beep tone.

For the analyses, we measured the reaction times from the onsets of the target sounds. These onsets were determined visually on the basis of the waveform and spectrogram of the signal. For the vowels, the onset was defined as the onset of voicing. For fricatives, the onset was the offset of voicing in the preceding vowel. The

onset of stop consonants is more difficult to define. In previous studies the onset was defined as the onset of the burst (but cf. Cutler & Otake, 1994). There are, however, reasons to measure reaction times from closure onset, as the closure itself is a cue to manner and as the preceding vowel provides information about place of articulation. By measuring the reaction times from closure onset, a fairer comparison is possible between stop consonants and fricatives, which are also measured from a point directly following the formant transitions in the preceding vowel. In the present study reaction times were therefore measured first from the onset of the closure. In supplementary analyses, we included reaction times measured from the release burst in order to compare our data with previous results.

Participants

Twelve native Dutch speakers were recruited from the subject pool of the Max Planck Institute in Nijmegen. In addition, 12 Spanish native speakers who were spending an exchange period in Nijmegen participated in this experiment. Furthermore, nine Catalan listeners were tested at the Universidad de Barcelona, 12 native speakers of Polish at the Universitet Śląski in Katowice, and 12 native speakers of British English at the University of Sussex in Brighton UK. Care was taken that the listener groups were as homogenous as possible with respect to dialectal background. In particular, only those Spanish exchange students were recruited whose native dialect did not belong to the group of dialects spoken in Catalonia. None of the participants reported any speech or hearing disorders. Their participation was rewarded with a small amount of money or with credits needed for their studies.

Results

REACTION TIMES

Reaction times (RTs) shorter than 100 milliseconds and longer than 1500 milliseconds were excluded from the analysis (0.8% of the data). Table 2 shows the mean RTs for the three phoneme classes, and the five listener groups.

One way to analyse the RTs would be to just compute the averages for the different languages and phoneme classes and analyse these averages for effects of

	Vowels	Stop Consonants	Fricatives
Catalan	435	480 (396)	500
Dutch	475	522 (436)	442
English	524	554 (467)	538
Polish	589	626 (540)	648
Spanish	557	655 (570)	635

Table 2: The average response times (in milliseconds) for the three phoneme classes and the five languages. For stop consonants reaction times were measured from both the onset of the closure (first number) and from the onset of the release burst (second number, in brackets).

phoneme class and number of categories. Such an analysis, however, would not be very reliable. The averages would not only reflect the effects of phoneme class, number of categories, structural differences between the languages (e.g., syllable structure and stress patterns), but also reflect differences between the average speeds

of the different groups of participants resulting from their familiarity with the experimental task.

Instead of comparing the average RTs of the different language groups for the three phoneme types, we analysed the data by means of multilevel regression models (e.g., Venables & Ripley, 2002; Baayen in press). We inserted Language, but also Participant and Item, as crossed random effects. This implies that the model computes different intercepts for each combination of language, participant and item. In other words, it partials out the effects of these factors while computing the effects of the fixed predictors of interest. This enormously reduces the variance in the data. As a consequence, this model is able to detect patterns in the data that are not easily visible in simple scatter plots. Moreover, the inclusion of Language, Participant, and Item as random effects allows us to generalize the observed effects of the fixed predictors over languages, listeners, and words.

The two main variables of interest are the Phoneme Class of the target and the Number of Categories in its class in the participant's language. We considered the log of the Number of Categories, instead of the bare number of categories, since preliminary analyses showed a non-linear relation between the RTs and the number of categories. However, Phoneme Class of a target predicts its number of Categories to some extent, in particular for the plosives (see Table 1). The two predictors are thus collinear and just entering both of them into the model may lead to misleading results (cf. Chatterjee, Hadi & Price, 2000). We therefore orthogonized the two variables as follows. We ran a simple linear model predicting the log number of Categories as a function of Phoneme Class. The residuals of this model are highly correlated with the log Number of Categories ($r = 0.829$, $p < .0001$), but display no relationship with Phoneme Class. We entered these residuals (henceforth: Residuals of the Number of Categories: RNC) together with Phoneme Class as fixed effects in the multilevel regression model for the reaction times. A potential interaction between Phoneme Class and RNC was excluded from the initial model, as it could not provide

meaningful results: The languages hardly differ in their numbers of categories for stop consonants, while they differ strongly in their number of vowels.

Table 3 lists the statistics for this initial model. Both Phoneme Class ($F(2, 5636) = 23.75$, $p<.001$) and RNC, representing the Number of Categories ($F(1, 5636) = 27.80$, $p<.001$), were significant. Additional analyses showed that participants' reactions were significantly faster to vowels (mean reaction time: 526 ms) than to fricatives (565 ms, $F(1, 3529) = 19.19$, $p<.001$) and stop consonants (577 ms, $F(1, 4075) = 40.46$, $p<.001$), which did not differ from each other ($p>.05$). The effect of

Fixed effects:	
- Intercept (Fricative, number of categories = 0):	564.81
- Stop consonant	14.57 (-71.01)
- Vowel	-38.86
- RNC	45.86 * RNC
Random effect of Language:	
Catalan	-49.67
Dutch	-56.67
English	-7.84
Polish	48.89
Spanish	65.30
Degrees of freedom:	5639

Table 3: Estimated values for the fixed effects and the random effect of language in the model for the reaction times. For stop consonants, the first number refers to the reaction times measured from the closure onset, while the second number in brackets refers to the measurements from onset of the release burst.

RNC showed that a higher number of categories slowed down listeners' responses. The left panel of Figure 1 shows the modeled relation between the plain number of categories and the RTs for a listener monitoring a fricative (only the intercept changes for vowels or stop consonants). The relation is non-linear and shows attenuation of the effect of the number of categories at higher numbers: An additional phonemic category has a bigger impact on the response latencies if only few categories are present in the phoneme class than if there are already many categories. Note that this significant positive relationship between response latencies and number of categories is not obvious from the average response times listed in Table 2. The reason for this is that the participants of the five languages differed in their average reaction times. The model partials out this variance by means of the random effects of Language and Participant.

To examine a possible difference in the effect of the number of categories on the identification of fricatives and vowels (the languages tested do not differ in the number of categories for stop consonants), we conducted a second analysis. Again, we modeled the RTs as a function of Phoneme Class and RNC but now also included a potential interaction between these two factors. As in the first analysis Language, Participant, and Item were included as crossed random factors. Stop consonants were excluded from this analysis. This second analysis showed significant main effects of Phoneme Class ($F(1, 3528) = 19.03$, $p<.001$) and RNC ($F(1, 3528) = 27.14$, $p<.001$). The interaction between Phoneme Class and RNC also emerged as significant ($F(1, 3528) = 5.57$, $p<.05$). Further analyses showed that RNC affects both classes, but the effect is bigger for the vowels than for the fricatives.

In the analyses reported above, the reaction times to stop consonants were measured from closure onset. This implies that the reaction times include a period of silence that precedes the crucial cues carried by the release burst. Moreover, our data now cannot be directly compared with the results of previous studies (e.g., Foss & Swinney, 1973; Savin & Beaver, 1970). We therefore also ran the analyses with the reaction times for the stop consonants measured from burst onset. Again we excluded

reaction times shorter than 100 ms and longer than 1500 ms. The analyses show again an effect of RNC ($F(1, 5635) = 28.53$, $p<.001$), with a higher number of categories leading to slower responses. Phoneme Class ($F(2, 5635) = 32.61$, $p <.001$) was also significant. Reactions to stop consonants were now the fastest (mean RT for stops: 492 ms, difference with vowels: $F(1, 4075) = 14.88$, $p<.001$; difference with fricatives: $F(1, 3667) = 77.78$, $p<.001$). This is as expected, since the reaction times to stop consonants were on average 80 ms shorter than in the previous analysis.

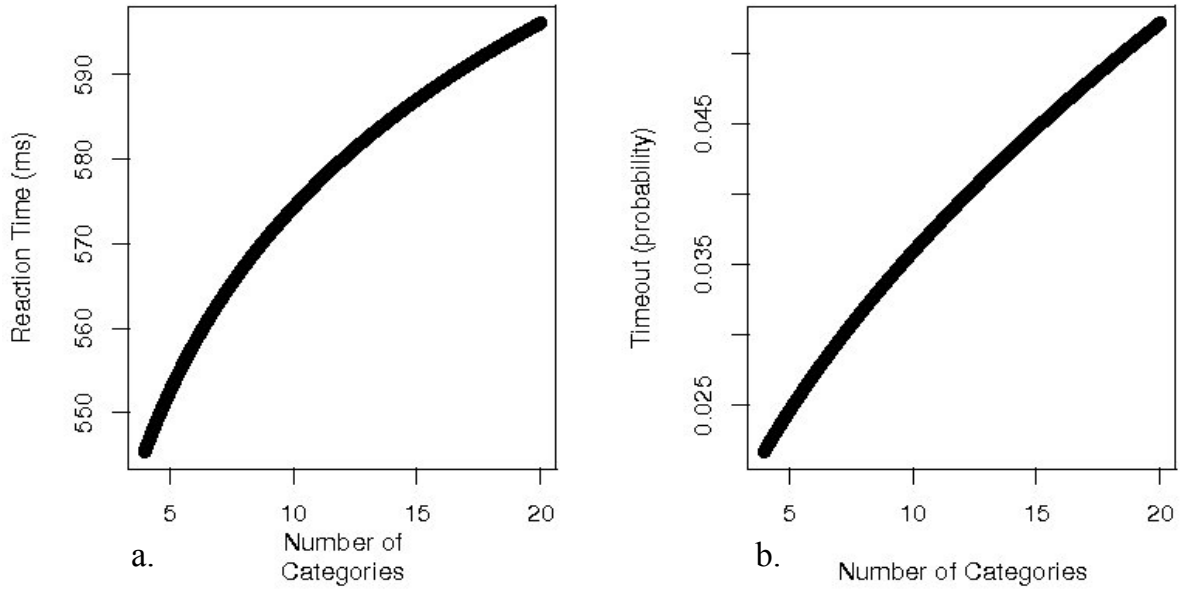


Figure 1: The modeled relation between the plain Number of Categories and (a) Reaction Times and (b) the probability of a timeout error for a listener detecting a fricative.

ERRORS

Table 4 displays the absolute numbers of timeouts and non-timeouts for the phoneme classes and for the five language groups. The percentages of timeouts are given within brackets. We modeled the probability of a timeout with a generalized

	Vowels	Stop Consonants	Fricatives
Catalan	6/238 (2.46 %)	1/238 (0.42 %)	3/177 (1.67 %)
Dutch	46/422 (9.83 %)	24/463 (4.93 %)	13/347 (3.62 %)
English	104/428 (19.55 %)	60/473 (11.26 %)	45/332 (11.94 %)
Polish	26/454 (5.42 %)	7/473 (1.46 %)	19/341 (5.28 %)
Spanish	48/459 (9.47 %)	66/565 (12.43 %)	20/370 (5.13 %)

Table 4: The absolute numbers of timeouts and non-timeouts for the three phoneme classes and the five languages. The percentages of timeouts are given in brackets.

multi-level model, with Language, Item, and Participant as random factors. The predictors considered as fixed effects were Phoneme Class, and RNC, representing the number of categories.

Both RNC ($F(1, 6164) = 33.38$, $p<.001$) and Phoneme Class ($F(2, 6164) = 10.14$, $p<.001$) appeared significant (see Table 5 for the effect sizes). Further analysis showed that listeners missed more vowels than fricatives ($F(1, 3895) = 16.17$, $p<.001$) or stop consonants ($F(1, 4498) = 12.09$, $p<.001$), but showed no difference between these two latter classes ($p>.1$). To illustrate the effect of the number of categories, the right panel of Figure 1 shows the predicted probability of a timeout error for participants monitoring fricatives as a function of the plain number of categories. For the other phoneme classes only the intercept changes.

Table 5 also shows the random effects of the languages. The languages clearly differed in their mean percentages of timeouts. For instance, the English listeners missed more targets than the Catalan listeners. One reason for this may be that the listener groups differed in their familiarity with the experimental task. The English participants, for instance, had less experience with psycholinguistic experiments than the Catalan participants. Another reason may be that the pronunciation of the Spanish

speaker is more native like to the Spanish and Catalan listeners than to the other listener groups. We investigated potential effects of speaker in a control experiment.

Fixed effects:	
- Intercept (Fricative, number of categories = 0):	-3.46
- Stop consonant	0.14
- Vowel	0.60
- RNC	0.82 * RNC
Random effect of Language:	
Catalan	-1.59
Dutch	0.06
English	0.54
Polish	-0.34
Spanish	0.82

Table 5: Estimated values for the fixed effects and the random effect of language in the model for the timeout errors

CONTROL EXPERIMENT

In the main experiment all listeners were presented with the same materials realised by a Spanish speaker. The Spanish participants were thus listening to native realisations of the nonsense words, whereas the other participants heard foreign pronunciations. It has been shown that when listening to speech on a low phonetic level, as in phoneme monitoring, listeners apply their native listening strategies, which are defined by their phonology and their phoneme inventories (Costa et al., 1998) and are hardly affected by the exact acoustic realisation of the materials (e.g., Cutler & Otake, 1994; Wagner

et al., 2006). Nevertheless, it is possible that the materials were processed in different ways by native and non-native listeners.

In order to test whether the effects of the class of the phoneme and the number of categories are present independently of the precise acoustic realisations of the stimuli, we ran a control experiment. In this experiment, Spanish and Dutch listeners were presented with materials realised by a native speaker of Dutch.

Materials and Procedure

A native speaker of Dutch recorded the experimental stimuli, in addition to some new fillers. The materials were very similar to those in the main experiment, but lacked stimuli with /k/ as a target. The Dutch speaker was asked to produce good examples of the Dutch phonemes. Thus, the Dutch stop consonants were realized with a short period of aspiration, the target vowels sounded like Dutch phonetically short /i u/ or long /a/, the fricatives were again the labiodental /f/, and alveolar /s/. Recordings were made in a sound attenuated room directly to a computer, and then down-sampled to 22.05 kHz (16 bit resolution). The procedure of the experiment was as in the main experiment.

Participants

Ten new Dutch speakers from the subject pool of the Max Planck Institute, and ten new Spanish exchange students in Nijmegen were recruited to take part in the control experiment. None had participated in the main experiment, and none had any known speech or hearing disorders.

Results

Table 6 presents the average reaction times and the numbers and percentages of timeouts for these two new groups of participants, broken by Phoneme Class.

REACTION TIMES

Reaction times to stop consonants were analyzed as in the main analysis, thus measured from closure onset. The data from the main experiment were pooled with the data from the control experiment. We then analyzed the data for all Spanish listeners for an effect of Speaker. We entered Speaker together with Phoneme Class as fixed effects in a multilevel regression model for the reaction times, with Item and Participant as crossed random factors. Note that we could not investigate the effect of the number of categories in this analysis, since this number is completely predictable given the class of the phoneme (as there is only one language). The effect of Phoneme Class emerged as significant ($F(2, 2069) = 33.78$, $p<.001$): Spanish listeners identified vowels significantly faster than stop consonants ($F(1, 1476) = 76.26$, $p<.001$) and fricatives ($F(1, 1374) = 34.41$, $p<.001$), while there was no difference between stop consonants and fricatives ($p>.1$). More importantly, the effect of Speaker was not statistically significant, neither was its interaction with Phoneme Class ($p>.1$). We then performed the same analysis for the two groups of Dutch participants and attested an effect of Phoneme Class ($F(2, 1987) = 13.28$, $p < .001$). Both Dutch groups identified vowels faster than stop consonants ($F(1, 1420) = 14.25$, $p<.001$). Fricatives were also identified significantly faster than stop consonants ($F(1, 1265) = 30.46$, $p<.001$), but there was no difference between vowels and fricatives ($p>.1$). Importantly, also this analysis showed no significant effect of Speaker ($p>.1$) and no significant interaction ($p>.1$). These data suggest that the Spanish and Dutch listeners were not affected by whether they were familiar with the exact acoustic realizations.

We also ran another analysis, which addresses more directly whether the effect of the number of categories is robust against different acoustic realizations. In this analysis the data for the Spanish and Dutch participants from the main experiment were removed from the data set and we only kept the data from the Spanish and Dutch participants in the control experiment. Reaction times were modeled as depending on RNC and Phoneme Class, with Language, Participant and Item as cross-random factors, as we did for the data of the main experiment. Both RNC ($F(1, 4669) = 4.24$

p=.03) and Phoneme Class (F(2, 4669) = 17.20, p <.001) were again significant, showing that both effects are robust and that the exact realizations of the materials are not decisive.

		Vowels	Stop consonants	Fricatives
Dutch	RT	505	560	492
	errors	39/311 (11.14%)	14/236 (5.6%)	7/223 (3.04%)
Spanish	RT	552	613	625
	errors	23/327 (6.57%)	17/233 (6.8%)	12/228 (5%)

Table 6: The mean reaction times (RT) and the absolute numbers of timeouts and non-timeouts for the Spanish and Dutch listeners in the control experiment with a Dutch speaker. Reaction times to stop consonants were measured from the onset of the closure. The percentages of timeouts are given in brackets.

ERRORS

We analyzed the errors of the control experiment in the same steps as we analyzed the reaction times. The analysis of all Spanish listeners as well as the analysis of all Dutch listeners showed neither a main effect of Speaker nor any interaction with Speaker. The analysis of the data set of the main experiment with the Spanish and Dutch listeners replaced by the Spanish and Dutch listeners from the control experiment revealed main effects of both Phoneme class (F(1, 5089) = 11.18, p<.001) and RNC (F(1, 5089) = 19.84, p<.001). Participants made more errors for vowels and more errors if the number of categories in the phoneme's class was higher. In conclusion, the control experiment shows that the effects of Phoneme Class and Number of

Categories are inherent to phoneme monitoring and independent of whether the listeners hear a native or a non-native pronunciation of the nonsense words.

PHONEME FREQUENCIES

The correlation of the reaction times and the timeout errors with RNC suggest that the speed and ease of phoneme identification depend on the sizes of listeners' phoneme repertoires. However, there may be an alternative explanation for our results.

The targets in our experiments are the most common phonemes in the world's languages, but the relative frequencies of occurrence of these phonemes vary across languages. Importantly, a language with more categories in its phoneme inventory may make less use of the phonemes tested in our experiments. In other words, there may be a confound between the number of categories and the frequencies of occurrence of the phonemes in the languages. Hence, the attested effect of the number of categories might actually be an effect of frequency of occurrence, and listeners with a smaller number of categories may be faster and more efficient in identifying phonemes just because of the more frequent occurrences of these phonemes in their language.

One might assume that listeners are so proficient in recognising their native phonemes that their performance is at ceiling, and that frequency cannot influence their performance in phoneme monitoring. Nevertheless, there are results pointing in the direction that phoneme frequency does play a role in phoneme identification. Warner, Smits, McQueen, and Cutler (2005) examined the effects of phoneme frequency on listeners' guesses about the identity of a segment in a gating study where listeners heard increasing portions of phonemes in a random order. A correlation was observed between phoneme frequency and listeners' decisions, when little acoustic information about the segment was available (that is, at short portions of the signal). For longer portions this correlation decreased gradually. Hence, faster and more

accurate identifications may be due to higher phoneme frequencies, instead of lower numbers of categories, in the present task as well.

To investigate this issue, we carried out two types of analyses. First, we examined whether the number of categories within a phoneme class is correlated with the frequencies of occurrence of its phonemes. Second, we reanalysed the response latencies and timeout errors including frequency as an additional predictor.

From the set of the languages tested, phoneme frequencies could be determined for Dutch, English and Spanish, as phonemically transcribed databases of words are available for these languages, which also include information about the token frequencies of the words (CELEX for Dutch and English, see Baayen, Piepenbrock & van Rijn, 1993; LEXESP for Spanish, see Sebastián-Galles, et al. 2000). We calculated the frequencies of occurrence of the phonemes per million phonemes, taking into account the token frequencies of the words (token frequency) or just counting every word once (type frequency).

We computed the correlation of the log number of categories for the phonemes with their log frequencies in the three languages. We found no correlation with the log token frequency of the phonemes. However, the log number of categories appeared highly correlated with the log type frequency ($r = -0.54$, $p<.01$). Unsurprisingly, listeners with fewer categories in their native phoneme repertoire make more frequent use of these phonemes in the words in their vocabulary.

In the second analysis, we included the log token and type frequencies of the phonemes as additional predictors in our model for the response latencies, for the three languages for which frequency values were available. In this analysis, the log phoneme frequencies were no significant predictors. The variance in the reaction latencies is explained by Phoneme Class and RNC but not by phoneme frequencies.

For the timeout errors, however, we observed a main effect for the log token frequency of the phonemes. A higher phoneme frequency implied fewer errors ($F(1, 4180) = 4.33$, $p<.05$). Importantly, the main effect of RNC was still significant ($F(1,$

4180) = 22.26, p<.001). This is as expected, as RNC, reflecting the log number of categories, was not correlated with the log token frequency of the phonemes.

In conclusion, the attested effect of the number of categories is not a frequency effect in disguise. In addition, our results support the view that phoneme monitoring may be affected by the frequencies of occurrence of the phonemes. However, the effect appears to be limited to participants' accuracy and not to extend to response latencies.

GENERAL DISCUSSION

This study investigated how listeners' speed and accuracy in phoneme identification is affected by the class of the speech sound (vowel, stop consonant, fricative) and by the number of categories within this class in the listeners' native phoneme repertoire. In a phoneme monitoring experiment with nonsense words, native listeners of five different languages (Castilian Spanish, Catalan, Dutch, British English, and Polish) identified vowels, fricatives, and stop consonants that represent phonemes in all the five languages. The results show that listeners identified vowels more quickly than fricatives. There was no difference between fricatives and stop consonants if reaction times to stop consonants included the interval of the closure. If the reaction times were measured from burst onset, however, as in previous studies, stop consonants were identified more quickly than fricatives and vowels. Phoneme class also affected participants' accuracy: consonants were identified more accurately than vowels. Furthermore, we found an effect of the number of categories: a phoneme is recognised faster and more accurately if it has fewer competitors belonging to the same class in the listener's phoneme inventory.

The present study extends previous research on differences between phoneme classes to more languages. Whereas nearly all previous findings are based on English (e.g., Foss & Swinney, 1973, Healy & Repp, 1982, van Ooijen 1994, Morton & Long

1976), we studied listeners of Romance languages, Germanic languages, and a Slavic language. Our study replicates the finding that stop consonants, with reaction times measured from burst onset, are recognised faster than fricatives (Foss & Swinney, 1980; Rubin, Turvey & van Gelder, 1976; Morton & Long, 1976).

Several studies, summarised in the Introduction, have attributed this difference between fricatives and stop consonants to mechanisms of auditory processing. Stop consonants would be processed more categorically: Due to their acoustic properties, their perceptual traces would decay faster, such that their recognition would be mainly based on traces in phonetic memory. Fricatives, on the other hand, would be perceived more continuously and processed on the basis of the more detailed traces in the auditory memory. As a consequence, stop consonants may be labeled faster than fricatives.

Our results show that there is an alternative explanation, which lies in the decision about the onset of the measurements of the reaction times. Stop consonants consist of the silent interval of the closure and of the abrupt release burst. Most previous studies have measured the response latencies for stop consonants from the release burst. However, the silent interval of the closure provides cues to the manner of articulation of the consonant and its duration may provide information about place of articulation. Moreover, the onset for reaction times for fricatives is set immediately after the formant transitions in the preceding vowel. By measuring the reaction times for stop consonants from the release burst, that is, much later than the end of the formant transitions, there is no fair comparison possible between fricatives and stop consonants. Obviously, a conclusion about which phonemes are identified more slowly depends very much on the onset of measurement for the reaction times. If we measure from closure onset, we see that labeling phonemes based on phonetic memory (stop consonants) or auditory memory (fricatives) does not necessarily lead to differences in identification times.

We also found that fricatives were in general identified more slowly than vowels. This seems to be in contrast to the findings of Savin and Bever (1970) and van

Ooijen (1994), who reported that vowels are detected more slowly than fricatives for Dutch and English. This contrast is only apparent. Table 2, listing the average reaction times for the different languages and phonemes classes, shows that also in our experiment, Dutch listeners detected vowels more slowly (mean reaction time: 475 ms) than fricatives (442 ms) and that there is hardly any difference between the two phoneme classes for the English listeners tested. Catalan, Polish and Spanish listeners, on the contrary, recognised vowels more quickly than fricatives. These differences between listener groups demonstrate the effect of the numbers of categories within the three phoneme classes that substantially vary among the languages tested (see below). After the effect of the number of categories is partialled out, vowels were in general recognised faster than fricatives.

In the Introduction, we formulated a hypothesis about ease of identification of vowels versus consonants on the basis of their function in lexical processing. Since vowels have been shown to constrain lexical selection to a lesser extent (e.g., Cutler et al. 2000), they might also be identified more slowly and less accurately. Regarding the accuracy of identification we found that listeners indeed made more errors on vowels than on consonants. Regarding the response latencies, however, we found exactly the opposite of what we predicted: Vowels were identified more quickly than consonants. One possible explanation may be that participants were less cautious in their reactions to vowels, exactly because vowels restrict lexical selection to a lesser extent than consonants. This would lead to faster responses but also to more errors.

Another explanation for the fast responses to the vowels may lie in their acoustic manifestation. More acoustic cues are present in the preceding context for vowels than for fricatives. For instance, whereas the formant transitions following consonants are generally assumed to be perceptually more relevant than the preceding formant transitions (e.g., Stevens & Blumstein, 1978), important cues for the identity of the vowel are present in the preceding consonant (e.g., Whalen, 1981) and even the preceding vowel (e.g., Manuel, 1990). This acoustic difference between vowels and fricatives may be determinative and interfere with any other effects.

We now turn to the role of the number of categories that we have documented for phoneme monitoring. A higher number of categories slowed participants down and made them less accurate. Since the number of categories within a phoneme class is language-specific, its effect yields language-specific patterns in phoneme recognition. Table 2 shows that there were roughly three rankings of the phoneme classes among the five languages. First, Catalan, Polish, and Spanish showed the basic pattern which emerged from our statistical analyses with number of categories as a predictor: Vowels were recognised faster than fricatives and stop consonants (as measured from closure onset). Second, the English participants recognized vowels as slowly as fricatives and stop consonants. This pattern is in line with the high number of vowels in this language, which makes listeners recognize them more slowly. Finally, in Dutch, vowels were recognized slightly more slowly than fricatives. This is in line with the high number of vowels also in this language, in combination with a relatively low number of fricatives.

These results illustrate that cross-language research is necessary to gain insight into speech processing. Studies investigating only one language tend to attribute differences between phoneme classes to the acoustic properties of the speech segments. This however is not the only factor contributing to listeners' phoneme identification, since, as we have shown, a major factor is the number of categories in the listener's native language. This factor can only be documented by comparing several languages.

Interestingly, the effect of the number of categories appeared to be greater for vowels than for fricatives (and possibly stop consonants). One explanation for this difference is in line with the fast responses and lower accuracy that we attested for vowels (see above). Vowels restrict lexical selection to a lesser extent, therefore listeners may generally pay less attention to vowels, and as a consequence be more sensitive to factors inhibiting identification.

The effect of the number of categories may stem from general properties of the perception system. A higher number of categories within a class implies a higher

number of choices, which generally impedes the process of decision making (e.g., Nosofsky 1997; Medin, Goldstone & Markman, 1995). This holds especially if the choice options are highly similar (Foss & Dowell, 1971). For instance, in visual perception, the search for an object on a display is slowed down both by a higher number of alternatives (set size effect) and a greater similarity among these alternatives (Palmer, Verghese & Pavel 2000; Theeuwes 1992).

Note that in the experiments in the visual domain, the alternatives are in general all present on the display, and therefore their number and similarity can be manipulated within participants. In phoneme monitoring, the alternatives are the categories in the participants' native phoneme repertoires. All these categories affect participants' phoneme monitoring, even though they are not all incorporated in the materials of the experiment. In consequence, number and similarity of categories cannot be manipulated within participants, and need cross-linguistic investigation.

The effect of number of categories can be explained by several aspects of identification processing. First, a greater set of phonemes involves the exclusion of more potential candidates in the search for a mental representation to match the presented signal. This implies a greater combined probability of incorrect candidates. Second, a higher number of categories implies that the perceptual space will contain more boundaries, and more sounds will be positioned at boundaries. In consequence, more sounds may be ambiguous and might therefore be harder to classify. Third, a greater set of phonemes implies that more phonemes share acoustic features, and fewer features distinguish a phoneme from its competitors. According to several categorization models (e.g., Ashby 2000, Nosofsky 2005), the degree of similarity between the alternatives affects reaction times and categorization accuracy. Hence, both the number of categories in a class and the similarity between speech sound categories may have contributed to our results.

The perceptually similar categories are not necessarily phonemes sharing the manner of articulation (phoneme class), but may also share other acoustic features. For instance, the voiced bilabial stop consonant /b/ may be perceptually similar to the

voiced bilabial fricative /v/. Further research is necessary to determine the precise effects of the number of categories in a phoneme class and the numbers and types of similar phonemes in the language belonging to different classes. Note that for such research, the degree of similarity between every pair of phonemes needs to be established separately for each language, as this degree might not only be determined by the phonemes' acoustics, but also by the phonotactic constraints in the language.

Furthermore, in the present study, we have made the simplified assumption that the categories that affect speech identification represent different phonemes. In line with this assumption, the numbers of categories that we used in our statistical analyses are the numbers of native phonemes in the different classes. However, listeners are also capable of discriminating allophonic variants of the same phoneme (e.g., Lipski, 2006). Future research has to show whether the number of distinctive speech sounds in a class is a better predictor than the number of phonemes. Note that this research will only be possible once we know which speech sounds listeners of the different languages can distinguish.

The study by Costa et al. (1998) shows that in phoneme monitoring listeners are aware of the acoustic variation in the realisation of a phoneme that can be induced by co-occurring native phonemes. Our study shows that in addition listeners are aware of the acoustically similar phonemes in their native language. Apparently, listeners' identification of phonemes is affected by the acoustic variability within the category of a phoneme, and also by the number of categories within the phoneme's class.

Interestingly, no difference was found between participants listening to a native or to a non-native speaker. This result shows that when listening to phonemes in nonsense words, a situation which resembles listeners' first contact with a foreign language, listeners assimilate speech sounds to their own native categories. The exact acoustic realization of a speech sound, which is phonemic in the native language, hardly affects listeners' identification.

The effect of the number of categories is also present in subsets of the data. It is also significant if we do not take into account the stop consonants or exclude one of

the five languages (e.g., Spanish or English). Probably, the effect would have been even stronger if the languages had been more similar in their syllable structure, stress patterns, phonotactic constraints, etc. Of course it is impossible to control for such differences between languages. The robustness of the effect of Number of Categories suggests that this effect is inherent to phoneme monitoring.

Importantly, the effect of number of categories cannot be explained by the frequencies of occurrence of the phonemes in the respective languages. As may be expected, in the languages for which frequency counts are available (Dutch, English, and Spanish), the frequencies of the phonemes in the languages' vocabularies are negatively correlated with the numbers of categories in their classes. Thus, languages with fewer phonemes in a given class use these phonemes more frequently in their words. We investigated whether the frequencies of the phonemes were predictors for the response latencies for the Dutch, English, and Spanish participants, in addition to the number of categories, but this was not the case. For the accuracy, we also found that incorporating the frequencies of the phonemes did not reduce the effect of the number of categories. Hence, the effect of the number of categories is not a frequency effect in disguise.

Nevertheless, we observed that phoneme frequency played a role in participants' accuracy: Participants made fewer errors for phonemes that occur more often in their speech (word tokens). However, the role of frequency appears minor as it only surfaces in the number of errors.

Given that higher numbers of categories in phoneme classes slow listeners down, one might expect that small numbers of categories form a preferable pattern among the languages of the world. Indeed, despite the great variation in the number of phonemes in the world's phoneme inventories, ranging from 11 (Rotokas) to 141 (!Xu), more than 70% of languages have between 20 and 37 segments (Maddieson, 1984). In natural speech interactions listeners' purpose is not to identify phonemes, but to recognize words to apprehend their meanings. Listeners would be hindered by lower numbers of phonemic categories, as this would lead to longer words, a higher number

of words embedded in other words (Cutler, Mister, Norris, Sebastián-Gallés , 2004), and higher neighborhood densities, which inhibit lexical access (e.g., Vitevitch & Luce, 1999).

In conclusion, this study documented two sources of variance in phoneme identification which affect listeners of different native backgrounds in the same way: (1) the acoustic and functional properties of the phoneme, and (2) the number of native categories within the phoneme's class. We found these general patterns across five languages, despite the many differences between the languages, for instance, in syllable structure, phonotactic constraints, and stress patterns, which might hide general patterns. The effect of the number of categories in the listener's native phoneme inventory proves to be another consequence of listeners becoming experts in their native phonology. Native speech sounds establish mental references early in speech development, and permanently divide listeners' perceptual space into distinct sound categories. While listeners do not focus on speech sounds in natural speech interactions, individual sounds enter listeners' focus of attention, for instance, when listening to speech in noise, when listening to a speaker with an unusual pronunciation, or when acquiring a foreign language. It may be especially under these conditions that phoneme class and the number of native categories within a class affect speech processing.

Appendix: List of materials used in the experiment

Consonant target	The following context		
	/a/	/i/	/u/
/f/	tekufa	kotafi	tipefu
	tasifa	tusafi	tokafu
	sokifa	pinesafi	posefu
	tilekofa	temupafi	pilotafu
	sinotufa	tenosafi	simokafu
/k/	posika	petuki	pitaku
	tufika	tusaki	sepiku
	pomiteka	palufoki	tenifaku
	finesoka	femoseki	timafeku
	fiselika	temisuki	petisaku
/p/	kesupa	tefupi	kitepu
	sefupa	fusopi	sikapu
	tekipa	talokepi	tafipu
	felukipa	tenasupi	kenosapu
	selukipa	senokapi	kosefipu
/s/	tekusa	pakesi	tepisu
	tikusa	petasi	fekatisu
	pifunesa	tukesi	kopesu
	telikusa	pomekasi	pokefisu
	pilufesa	fukeposi	tilokasu
/t/	tekuta	pakuti	fisetu
	fekuta	kopati	sakitu
	pilefuta	kosati	pemakitu
	fipokuta	pakofuti	felosatu
	simofeta	sekafuti	senoketu

Vowel target	The preceding context				
	/f/	/k/	/p/	/s/	/t/
/a/	petufa	petoka	tofepa	pitosa	pifota
	telisufa	tosuka	fekosipa	fukopesa	pomisuta
	tesupifa	fenusoka	sotipa	tokesa	siputa
/i/	talemofi	telufaki	senufopi	pakotesi	fokesuti
	pasufi	tolepuki	kefopi	fatusi	sefuti
	pomekufi	setuki	tesopi	tukesi	sokati
/u/	tepifu	tomiseku	fekipu	pafisu	fopitu
	sakomifu	paseku	sikapu	fakipesu	finesatu
	somatefu	sutileku	semalipu	tenifasu	pisatu

Formant transitions in fricative identification: The role of native fricative inventory

CHAPTER 3

Wagner, A., Ernestus, M., & Cutler, A. (2006). *Journal of the Acoustical Society of America, 120,* 2267-2277.

Abstract

The distribution of noise across the spectrum provides the primary cues for the identification of a fricative. Formant transitions have been reported to play a role in identification of some fricatives, but the combined results so far are conflicting. We report five experiments testing the hypothesis that listeners differ in their use of formant transitions as a function of the presence of spectrally similar fricatives in their native language. Dutch, English, German, Polish, and Spanish native listeners performed phoneme monitoring experiments with pseudo-words containing either coherent or misleading formant transitions for the fricatives /s/ and /f/. Listeners of German and Dutch, both languages without spectrally similar fricatives, were not affected by the misleading formant transitions. Listeners of the remaining languages were misled by incorrect formant transitions. In an untimed labeling experiment both Dutch and Spanish listeners provided goodness ratings that revealed sensitivity to the acoustic manipulation. We conclude that all listeners may be sensitive to mismatching information at a low auditory level, but that they do not necessarily take full advantage of all available systematic acoustic variation when identifying phonemes. Formant transitions may be most useful for listeners of languages with spectrally similar fricatives.

INTRODUCTION

Do formant transitions contribute to listeners' identification of fricatives? These dynamic cues are crucial for the identification of stops, but despite decades of research (Harris,1958; Heinz & Stevens, 1961; LaRiviere, Winitz & Herriman, 1975; Jongman, 1989; Jongman, Wayland & Wong, 2000), no clear answer has emerged for fricatives. Salient static cues are present in the fricative spectrum, and may suffice for phoneme identification. We report a study which contributes to this discussion by testing the hypothesis that the contribution of formant transitions is language-specific, and depends on the presence of spectrally similar fricatives in the listener's native phoneme inventory.

Fricatives are produced with a narrow constriction in the oral cavity. The turbulence of the airflow passing this constriction generates the characteristic sound of frication. The exact location of the narrow passage and the size and form of the cavity in front of the constriction define the acoustic characteristics of the fricative (Stevens, 1998). These energy peaks and minima in a fricative's spectrum serve listeners as primary cues for fricative identification (Stevens, 1998). The salience of those spectral poles, however, differs among fricatives, and previous research (e.g., Harris, 1958) suggests that listeners need additional cues to identify some but not all fricatives. Whereas sibilants have very pronounced spectral peaks and are identified primarily on the basis of these poles, dental and labio-dental fricatives have a more diffuse energy spectrum and may require additional cues for accurate identification. Two contextual sources of such cues have been found (Whalen 1981): formant transitions, which may be perceptually integrated with cues from the fricative spectrum; and the quality of the surrounding vowels, including the resulting slight modifications of the fricative spectrum itself.

It is unclear, however, whether formant transitions indeed contribute to the identification of fricatives, since the results from previous research are conflicting. Harris (1958) studied the identification of English fricatives in different vocalic

contexts. In a fricative categorization experiment, she presented American students with natural tokens of CV-syllables containing the fricatives /f v θ ð s z ʃ ʒ/ combined with the vowels /a ɪ u e/. These syllables were spliced such that every fricative was combined with every vowel as produced in the context of each of the fricatives. Thus, the formant transitions in some tokens contained misleading information with respect to the identity of the fricative. Participants accurately categorized /s/ and /ʃ/ in the combination of just the frication part from the sibilant with each of the vowels, independently of the fricative context from which these vowels were extracted. In contrast, stimuli with frication from /f/ or /θ/ were often confused with each other. In fact, the /f/ tended to be categorized as /f/ only when combined with a vowel originally produced after /f/, but as /θ/ when followed by any other vowel. Apparently, the English listeners recognized the sibilants /s/ and /ʃ/ by their frication part alone, while the dental fricatives /f/ and /θ/ were accurately categorized only when followed by correct formant transitions.

Similar results were obtained by Heinz and Stevens (1961) with synthesized English voiceless fricatives. American listeners identified /s ʃ f θ/ in isolation, and achieved satisfactory identification rates for /s ʃ/, but they could not distinguish between /f/ and /θ/. The identification scores improved when the fricatives were combined with the synthetic vowel /a/, including approximated transition movements; especially the distinction between /f/ and /θ/ was more reliably perceived.

More recent studies, however, failed to replicate these results. Jongman (1989) asked English listeners to identify fricatives by listening either to portions of the frication alone, or to the whole frication, or to complete syllables (all eight English fricatives except /h/, produced by an American speaker with the vowels /a i u/). A portion of the frication longer than 40 ms appeared to be sufficient for listeners to identify all fricatives accurately, including the oft-confused fricatives /f/ and /θ/. No

improvement of fricative identification resulted from inclusion of the vowel. Jongman, Sereno, Wayland & Wong (1998) further supported this conclusion in a production study. They analyzed the variances of locus equations (Fruchter & Sussman, 1997) of English fricatives followed by the vowels /i e ae a o u/ as produced by twenty speakers. On this parameter /f v/ differed significantly from /s z ʃ ʒ θ ð/, but the three places of articulation represented in the latter set did not differ. Jongman et al. concluded that locus equations cannot sufficiently cue fricative place of articulation.

LaRiviere, Winitz & Herriman (1975), too, queried the role of formant transitions in fricative identification. They compared identification of syllables made up of /f θ s ʃ/ and /a i u/, with the identification of the same syllables with deleted formant transitions. Listeners could reliably identify all fricatives in transitionless syllables, and the authors thus concluded that formant transitions do not necessarily contribute to fricative identification. LaRiviere et al. also found that /θ/ was the most difficult fricative to identify. They explain possible, but not necessary, perceptual benefit from the following vowel as arising from the information that it carries about the speaker's vocal tract, which contributes to the process of speaker normalization.

Klaassen-Don (1983) also found no evidence that formant transitions contribute to fricative identification. In a gating experiment with Dutch fricatives, she presented naturally produced CV and VC strings including the fricatives /f v s z ʃ x/ and the vowels /a i u/. The syllables were produced in isolation or were excerpted from running speech. Formant transitions proved to be valuable cues for liquids and stops, but their contribution in fricative identification was negligible. Klaassen-Don reached the conclusion that "vowel transitions do not contain perceptually relevant information about adjacent fricatives in Dutch" (Klaassen-Don, 1983, p.79).

Finally, in a series of production and perception experiments, Borzone de Manrique and Massone (1981) investigated the identification of Argentinian Spanish fricatives by native listeners. The perceptual power of the most prominent noise frequency bands was tested by band-pass filtering the fricatives /s f ʃ x/. The

identifications showed that /s/ is the most robust fricative, whereas /f/ requires a wide noise band to be accurately identified. In further experiments, the authors concentrated on the role of the vocalic environment for fricative identification by Argentinian listeners. Their stimuli consisted of frication and vocalic parts spliced out of naturally produced CV syllables and of transitionless CV syllables, which they constructed by combining natural fricatives and vowels produced in isolation. For Argentinian listeners the frication part alone was sufficient to identify all fricatives, with the exception of the velars /x ɣ/. The absence of transitions in the vowel biased the listeners to the fricative that is realized with the least transition movements into the following vowel. For instance, the formant transitions following /f/ are shorter before /u/ than before /i/, and the authors observed a higher number of /f/ categorizations for syllables consisting of frication and /u/ rather than frication and /i/.

In short, the literature shows that formant transitions proved to be useful cues in some experiments but of little use in others. Importantly, the experiments involved listeners of different native languages. We hypothesize that the solution to the conflicting results is that listeners' attention to formant transitions for fricative identification is language-specific, and modulated by the presence of perceptually similar fricatives in the native phoneme inventory. Languages differ widely in how many fricatives they include, and how similar these fricatives are. More fricatives in a given perceptual space may reduce the distinctiveness of individual fricatives. To maximize the distinctiveness of fricatives in denser perceptual spaces, listeners may learn to integrate additional cues to attain accurate percepts of these fricatives.

If listeners of different native languages indeed differ in the use they make of transitional cues, we can further ask whether listeners who do exploit transitional information do so for all native fricatives, or only for contrasts which are perceptually similar. Listeners' language experience may tune the perceptual system to select relevant cues efficiently for each fricative: If more salient cues suffice to distinguish a given phoneme contrast, native listeners may make no use of the information in

formant transitions. Thus our second hypothesis is that attention to formant transitions can be restricted to those fricatives that are difficult to distinguish spectrally. The fricative pair /f θ/ seems, on the evidence cited above, to be difficult to distinguish for English listeners. For Argentinian listeners, without /f θ/ in their native phoneme inventory, a different pair of fricatives appears to be potentially confusable: /x ɣ/. We assume that listeners will learn the most efficient way to identify all native fricatives, and that it might not be beneficial for them to use the cues in formant transitions for fricatives that can be identified accurately on the basis of the fricative spectrum alone.

In the present study, listeners of different languages heard pseudo-words containing either coherent or misleading information in the formant transitions surrounding fricatives. In four experiments participants performed phoneme monitoring, a task that has been used to investigate a wide range of psycholinguistic issues (see Connine & Titone, 1996, for a review). In phoneme monitoring, listeners hear spoken input, e.g., lists of words, nonwords, or syllables, and respond as soon as they detect a pre-specified target phoneme. Phoneme monitoring is especially promising as a paradigm for testing our hypothesis because it has been shown to be sensitive to formant transitions: Detection of a phoneme is more difficult when its context is cross-spliced and thus bears mismatching coarticulatory information (Martin & Bunnell, 1981; McQueen, Norris & Cutler, 1999). Moreover, the task is sensitive to cross-language differences in speech processing. Otake, et al. (1996) and Weber (2001) showed effects of language-specific phonotactic constraints in phoneme monitoring for nasals and fricatives respectively. Similarly, with the same task Costa, Cutler and Sebastián-Gallés (1998) showed that processing of acoustic variation is affected by native phoneme inventory constitution.

If listeners depend on formant transitions in fricative identification, then mismatching formant transitions should increase errors and slow reaction times in phoneme monitoring. In contrast, listeners whose fricative identification is governed

mostly by the primary static cues in the noise spectrum should be less affected by misleading formant transitions, either in reaction speed or error rate.

We tested five languages: German and Dutch, which both have only spectrally distinct fricatives, and Spanish, English, and Polish, which all have pairs of fricatives in which the distribution of noise peaks across the spectrum is very similar, so that the members of the pair are perceptually less distinctive. Spanish and English contrast with Polish with respect to which spectrally similar fricatives appear in the phoneme inventory. Table I sketches the fricative inventories of the five languages.

	Labio-dental		Dental		Alveolar		Post-Alveolar		Retroflex		Alveolo-palatal		Velar	Glottal
Dutch	f	v			s	z	(ʃ)						x	h
German	f	v			s	z	ʃ	ʒ					x	h
Spanish	f		θ		s								x	
English	f	v	θ	ð	s	z	ʃ	ʒ						h
Polish[1]	f	v			s	z	ʃʲ	ʒʲ	ʂ	ʐ	ɕ	ʑ	x	

TABLE I. The fricative inventories of the languages studied according to the place of articulation.

Experiment I contrasted Spanish with Dutch and German. Spanish, as we saw, has the confusable pair /f θ/. The spectra of the labio-dental and dental fricatives are relatively flat; the energy is distributed in each case across frequencies from circa 2 kHz to 10 kHz with no defined spectral peaks (Jongman 2000). We therefore expected

[1] Polish post-alveolar fricatives /ʃ ʒ/ are traditionally described as laminal alveolar (Jassem, 2003), and the alveolo-palatal /ɕ ʑ/ are considered as their palatalized counterparts. Hamann (2003) argues that Polish post-alveolar fricatives should be considered as retroflex; in addition Zygis and Hamann (2003) claim that the alveolo-palatal and the palatalized post-alveolar fricatives in Polish should be considered two separate sounds, as they are distinguished by native and non-native listeners. This view is adopted in our description of the Polish fricative repertoire.

Spanish listeners to pay more attention to formant transitions than Dutch or German listeners, whose languages contain no spectrally similar fricatives. The fricatives in the experiment were the labio-dental /f/ and the alveolar /s/. Since of these only /f/ is spectrally confusable with another fricative in Spanish, we further expected Spanish listeners to be particularly affected by mismatching formant transitions for /f/.

EXPERIMENT I

Method

Materials

Three- and four-syllable pseudo-words made up of the phonemes /p b t d k f s a i u e/ (e.g. *tikusa* and *dokupafi*) were recorded by a native speaker of Dutch. Note that no fricatives other than /f/ or /s/ appeared in the stimuli. The fricative identification was part of a larger phoneme monitoring experiment with various phonemes as targets. Only the results for the fricative targets will be reported here.

We created 12 pseudowords with the target /f/ and 12 pseudowords with the target /s/. The fricatives were preceded and followed by /a i u/. The target appeared always in the last syllable; stress was always on the first syllable. In addition, for every target fricative 12 filler items were created with the fricative in the penultimate syllable, and 12 filler items without the fricative.

The stimuli were recorded in a sound-attenuated room directly to computer and down-sampled to 22.05kHz (16 bit resolution). With Praat software cross-spliced and identity-spliced versions of the pseudo-words were created. Identity-spliced fricatives were replaced by the same fricative taken from another token of the same pseudo-word (e.g., /s/ in *tikusa* by /s/ of another *tikusa*). Cross-spliced fricatives were replaced by the other fricative produced in the same context (e.g. /s/ in *tikusa* by /f/ from *tikufa*). Segmentation points for the fricatives were defined visually, on the basis of oscillograms and sonagrams. The end of harmonic structure of the preceding vowel

and the beginning of harmonic structure in the fading noise of the fricative were defined as the splicing points. At zero-crossing points the coherent stochastic noise parts of the fricative were excised. The spliced stimuli were examined auditorily to ensure that no audible discontinuities had resulted from the manipulation.

Procedure

Participants sat in a sound-attenuated room in front of a computer screen, and heard both cross-spliced and identity-spliced stimuli over headphones. Each pseudo-word appeared only once in a session. Trials were blocked by target phoneme, with the order of blocks counterbalanced across participants. Participants were informed orally about the possible targets in advance; during the experiment a letter on the computer screen designated the current target. Participants were instructed to press a key immediately upon detecting in the nonword the sound represented by the displayed letter. Every target block of stimuli was followed by a break, the duration of which was controlled by the participants. From item onset, listeners had 2000 ms to respond. Failures to respond, and responses over 2000 ms, were defined as timeout errors. The experiment was self-paced: The next stimulus was presented 1000 ms after the participant's response or timeout, and it was preceded by a beep tone.

Participants

Eighteen Dutch regular students, and 21 German and 23 Spanish exchange students from the Radboud University Nijmegen took part in this experiment. They were paid for their participation. None reported any speech or hearing disorders.

Results

Two items, one for each fricative target, were missed by more than 40% of the participants and therefore excluded from the analysis. The average timeouts (mean percentages of targets not correctly detected within 2000 ms) and reaction times (RTs) for the remaining items for the three languages, the two fricatives and the two splicing conditions are shown in Table II.

	Fricative	Dutch	German	Spanish
Mean percentage of Timeouts	/s/ identity-spliced	4.3% (4/93)	1.8% (2 /115)	2.7% (4/170)
	/s/ cross-spliced	4.3% (4/93)	3.5% (4 /115)	2.7% (3/167)
	/f/ identity-spliced	2.0% (2/93)	1.8% (2 /115)	4.6% (6/169)
	/f/ cross-sliced	2.1% (2/93)	1.0% (1/115)	45.2% (55/145)
Mean RT	/s/ identity-spliced	488.22	440.82	544.04
	/s/ cross-spliced	512.23	428.8	562.52
	/f/ identity-spliced	531.27	442.22	618.6
	/f/ cross-spliced	540.50	475.67	666.8

TABLE II. Average percentages of Timeouts and mean RTs in ms for the three languages and the two fricatives in both splicing conditions in Experiment I. The absolute numbers of Timeouts and the total numbers of trials are given in brackets.

Timeouts: We analyzed the Timeouts by means of a loglinear analysis with the number of timeouts and nontimeouts for each stimulus as the dependent variable and Language (Dutch, German, and Spanish), Splicing (identity-splicing and cross-splicing), and Fricative (/s/ and /f/) as independent variables. All main effects were significant (Language: $F(2,129) = 30.22$, $p<0.001$; Splicing: $F(1,127) = 33.47$,

$p<0.001$; and Fricative: $F(1,128) = 29.16$, $p<0.001$). These main effects were modulated by an interaction between Language and Fricative ($F(2,125) = 15.48$, $p<0.001$). Importantly, we also observed the hypothesized interactions between Language and Splicing ($F(2,123) = 6.63$, $p<0.001$), and between Language, Fricative and Splicing ($F(2,120) = 4.29$, $p<0.015$). Splicing did not affect the number of timeout errors for the Dutch and German listeners, but the Spanish listeners were severely disturbed by misleading formant transitions ($F(1,41) = 48.42$, $p<0.001$). The effect of Splicing for Spanish was restricted to /f/ (interaction between Splicing and Fricative for Spanish $F(1,40) = 11.32$, $p<0.001$).

RTs: Latencies were measured from onset of the target fricative, defined as onset of the disharmonic structure in the stimulus waveform. Latencies below 150 ms were excluded from analysis (0.3% of the data). Analyses of variance were conducted for Participants (F1) and Items (F2), with Language, Splicing, and Fricative as independent variables.

The main effects of Language and Fricative were significant in both analyses (Language: $F1(2,58) = 7.14$, $p<0.01$, $F2(2,105) = 55.42$, $p<0.001$; Fricative: $F1(1,174) = 31.49$, $p<0.001$, $F2(1,21) = 8.53$ $p<0.01$), while Splicing was significant only in the analysis by Participants ($F1(1,174) = 5.29$, $p<0.05$). The interaction of Language with Fricative was significant in the analysis by Participants ($F1(2,174) = 31.60$, $p<0.001$). More importantly, in the analysis by Participants we also observed the interaction between Language and Splicing ($F1(2, 174) = 5.12$, $p<0.01$) This interaction failed to reach significance in the analysis by Items.

Summary and discussion

We found language-specific patterns in the use of formant transitions in fricative identification. Only Spanish listeners were affected by misleading formant transitions. Apparently, they were attending to cues that were neglected by the Dutch and German

listeners. Recall that the German and Dutch phoneme repertoires do not contain spectrally similar fricatives, while Spanish includes the two spectrally similar fricatives /f/ and /θ/. Even though /θ/ was not in the stimulus set, Spanish listeners paid attention to the formant transitions for /f/. They did not do so for /s/, which is spectrally distinct from the other fricatives in Spanish. These data support the hypothesis that listeners make use of formant transitions especially for fricatives that are spectrally similar to other fricatives in their native phoneme repertoire. Further, the results indicate that listeners do not necessarily take advantage of all acoustic information transmitted in the signal. The German and Dutch listeners showed no effects of the mismatching information that led Spanish listeners into errors.

However, Dutch participants had the advantage of listening to native phoneme realizations, while the Spanish listened to a foreign realization. The fact that German listeners showed the same pattern of results as the Dutch listeners may reflect a closer resemblance of German phonemes to Dutch than to Spanish phonemes. An alternative explanation for the cross-language differences might therefore be that listeners pay attention to more or to different cues when listening to a foreign pronunciation.

Experiment II was designed to test this second explanation. Experiment II and Experiment I differed principally in the native language of the speaker who recorded the stimuli: Dutch in Experiment I, Spanish in Experiment II. In Experiment II, the Spanish listeners were thus presented with a familiar pronunciation, while the Dutch and German listeners were confronted with an unfamiliar realization of phonemes.

EXPERIMENT II

Method

Materials and Procedure

The stimulus set from Experiment I was now recorded by a native speaker of Spanish. In addition, 30 new fillers were created for each target with the target in the

penultimate syllable or with the target missing. These fillers did not contain the phonemes /b/ and /d/, since Spanish phonotactics allows voiced bilabial and alveolar stops only in certain positions, and these consonants would therefore lead to a marked pronunciation by the Spanish speaker. The procedure was as in Experiment I.

Participants

Twenty-four Dutch regular, and 24 German and 24 Spanish exchange students from the Radboud University Nijmegen were paid to take part in this experiment. None had participated in Experiment I, and none had any known speech or hearing disorders.

Results

We defined and analyzed timeout errors and reaction latencies in the same way as in Experiment I. No data point was below 150ms, the common phoneme monitoring cutoff value (see, e.g., McQueen et al., 1999), and therefore no reaction time data were excluded from the analysis. Table III shows the results of this experiment.

***Timeouts*:** All main effects were significant (Language: $F(2,177) = 28.32$, $p<0.001$; Splicing: $F(1,176) = 28.49$, $p<0.001$; Fricative: $F(1,175) = 42.50$, $p<0.001$). These main effects were modulated by interactions of Language and Splicing ($F(2,173) = 5.39$, $p<0.001$), Language and Fricative ($F(2,171) = 13.68$, $p<0.001$), and Splicing and Fricative ($F(1,170) = 6.3$, $p<0.05$). The interaction between Language, Splicing, and Fricative narrowly missed significance ($F(2,168) = 2.4$, $p<0.1$). Splicing affected the number of timeout errors for the Spanish listeners ($F(1,58) = 38.4$, $p<0.001$) only, and especially for the detection of /f/ (interaction of Splicing and Fricative for Spanish $F(1,56) = 10.41$, $p<0.001$). These results replicate those of Experiment I

.

	Fricative	Dutch	German	Spanish
Mean percentage of Timeouts	/s/ identity-spliced	0% (0/180)	2.2% (5/180)	1.1% (2/172)
	/s/ cross-spliced	0% (0/180)	2.7% (4/180)	0% (0/173)
	/f/ identity-spliced	1.1% (2/180)	1.1% (0/178)	2.3% (4/172)
	/f/ cross-sliced	1.6% (3/180)	2.2% (4/180)	27.4% (47/173)
Mean RT	/s/ identity-spliced	461.54	474.34	474.93
	/s/ cross-spliced	463.05	490.815	473.89
	/f/ identity-spliced	550.67	569.06	601.93
	/f/ cross-spliced	552.43	568.67	661.33

TABLE III. Average percentages of Timeouts and mean RTs in ms for the three languages and the two fricatives in both splicing conditions in Experiment II. The absolute numbers of Timeouts and the total numbers of trials are given in brackets.

RTs: The main effects of Language, Splicing and Fricative were significant in both the Participant and the Item analyses (Language: $F1(2,58) = 7.2$, $p<0.01$, $F2(2,112) = 11.56$, $p<0.001$; Splicing: $F1(1,207) = 5.79$, $p<0.05$, $F2(1,140) = 4.94$, $p<0.05$; Fricative: $F1(1,207) = 42.45$, $p<0.001$, $F2(1,28) = 25.45$, $p<0.001$). The interaction of Language and Fricative was significant only in the analysis by Participants ($F1(2,207) = 9.27$, $p<0.001$, $F2(2,140) = 8.63$, $p<0.001$).

Summary and discussion

Experiment II further supports the hypothesis that Spanish listeners are affected by misleading formant transitions for fricative identification, while German and Dutch listeners are not. We ascribe these language differences to the different structures in

the phoneme inventories of these languages, more precisely to the presence or absence of spectrally similar fricatives. Moreover, the finding that the Spanish only appeared to attend to formant transitions surrounding the labio-dental fricative /f/ supports the hypothesis that the use of these cues is restricted to spectrally similar fricatives.

We obtained the same results for stimuli produced by a Dutch speaker (Experiment I) and by a Spanish speaker (Experiment II). Thus, Experiments I and II together suggest that the native language of the speaker, or, in other words, the listeners' familiarity with the presented realization of the phonemes, does not alter the role of formant transitions in listeners' identification. We conclude that listeners also apply the native strategy when listening to a foreign pronunciation.

To explore further whether the presence of acoustically similar fricatives in a language's phoneme repertoire results in attention to formant transitions, we performed a third experiment with English native listeners. Since English is a Germanic language, it is in many respects more like Dutch and German than like Spanish. However, English has, like Spanish, both labio-dental /f/ and the spectrally similar dental fricative /θ/ in its phoneme inventory. If our hypothesis is correct, English listeners should also attend to transitional cues, in particular for /f/.

EXPERIMENT III

Method

Materials and Procedure

The materials were as in Experiment II, i.e., the stimuli recorded by a native speaker of Spanish. The procedure and data analysis were as in the preceding experiments, with the exception that the target phoneme was not presented on screen. Grapheme-phoneme correspondences are often ambiguous in English; thus /f/ can be spelled as in "foal" or as in "phone", /s/ can also be represented by the letter "c", as in "cedar", and the letter "s" can stand for /s/, as in "basic", for /z/, as in "cousin", or for nothing, as in

"debris". Therefore we specified the target in recorded instructions at the beginning of every block of pseudo-words, instead of in visual target representations.

Participants

Twenty-seven students from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex took part in this experiment. They were native speakers of English and none reported any speech or hearing disorders.

Results

Mean timeouts and RTs are shown in Table IV.

Fricative	/s/ identity-spliced	/s/ cross-spliced	/f/ identity-spliced	/f/ cross-sliced
Mean percentage of Timeouts	6.2 (11/177)	9.3 (16/176)	9.3 (16/175)	17.4 (30/173)
Mean RT	562.43	560.37	611.14	627.3

TABLE IV. Average percentages of Timeouts and mean RTs in ms for the English listeners and the two fricatives in both splicing conditions in Experiment III. The absolute numbers of Timeouts and the total numbers of trials are given in brackets.

Timeouts: Both Splicing (cross-spliced versus identity-spliced items) and Fricative (/s/ versus /f/) were significant (Splicing: $F(1,58) = 5.76$, $p<0.05$; Fricative: $F(1,57) = 5.95$, $p<0.05$). The interaction did not reach significance. The English listeners missed more items in the cross-spliced condition, and more /f/ than /s/.

RTs: 0.4% of the data was below 150 ms, and was excluded from the analysis. Only Fricative was significant in both analyses ($F1(1,78) = 12.66$, $p=0.001$, $F2(1,56) = 2.89$, $p<0.05$). Listeners responded less rapidly to /f/ than to /s/.

Summary and discussion

English listeners also appear to pay attention to formant transitions. The crucial interaction between Fricative and Splicing was not significant, and therefore at this point we cannot decide with certainty whether English listeners make use of transition cues only for identification of /f/. However, the data suggest that English listeners, like Spanish listeners, are particularly affected in the case of /f/ (note that the effect of cross-splicing, though statistically robust for both fricatives for these listeners, was twice as strong in the timeout errors for /f/ as for /s/ – 87% increase as opposed to 47%). Both English and Spanish listeners have learnt to distinguish between /f/ and /θ/, two highly confusable fricatives. This apparently made them more attentive to the additional acoustic cues in the formant transitions.

Previous research has shown that the labio-dental fricative is hard to identify on the basis of spectral characteristics alone (Harris 1958, Jongman et al. 1998). So far we have shown that some listeners attend to transitional cues for this fricative. Our hypothesis, however, is that listener's use of transitional information in fricative identification reflects not just inherent distinctiveness of fricatives, but the presence of spectrally confusable pairs in the native fricative inventory. On this hypothesis, even fricatives which are generally easy to identify should encourage use of transitional information in a language which contains more fricatives with similar spectra.

The /s/ has been shown to be perceptually very salient because of the acoustic make-up of its noise spectrum (Wang & Bilger, 1973). During the articulation of /s/ air jets are created as the airflow passes the edges of the teeth; this results in relatively high intensity peaks in the high-frequency range of the spectrum, which serve as reliable cues and makes this fricative acoustically robust. Listeners should nevertheless

also exploit formant transitions to identify /s/, we predict, if other fricatives are close to /s/ in their native perceptual space.

We tested this in Polish, which has 11 fricatives [f v s z ʃʲ ʒʲ ʂ ʐ ɕ ʑ x]. The dental fricative is not present, so that /f/ is acoustically distinct from all other fricatives. The presence of the post-alveolar, alveolo-palatal, and palatal retroflex fricatives may, however, reduce the perceptual saliency of /s/. In acoustic terms, the /s/ typically has energy peaks in the frequency range between 3 and 7 kHz. The post-alveolar /ʃ/ exhibits energy peaks in the frequencies between 1.5 and 5 kHz, while the Polish alveolopalatal /ɕ/ has its energy maxima in the range between 2 and 6 kHz. Finally, the retroflex Polish fricative shows its high energy peaks around 1 and 4 kHz (Jassem, 1968). This concentration of several fricatives with energy distributions in the same spectral range might hinder the identification of these fricatives in Polish. We therefore expect Polish listeners to pay attention to formant transitions for /s/.

EXPERIMENT IV

Method

Materials and Procedure

Materials were as in Experiment II and III, procedure was as in Experiment II, and data analysis was as in all the preceding experiments.

Participants

Twenty-four students at the Uniwersytet Śląski in Katowice, all native Polish speakers, were paid to take part in this experiment. None reported any speech or hearing disorders.

Results

Table V shows the average Timeouts and RTs.

Timeouts: Both main effects were again significant: Splicing ($F(1,58) = 10.19$, $p<0.01$) and Fricative ($F(1,57) = 21.92$, $p<0.001$). The interaction between Fricative and Splicing narrowly failed to reach significance ($F(1,56) = 3.73$, $p<0.06$). More timeouts occurred for the cross-spliced items, and for /s/ (9.16 % versus 1.6% for /f/). Furthermore, the effect of splicing appeared smaller for /f/ than for /s/.

RTs: The main effect of Fricative was significant in the analysis by Participants only ($F1(1,69) = 5.65$, $p<0.05$). As Table V shows, the Polish RTs were relatively long.

Fricative	/s/ identity-spliced	/s/ cross-spliced	/f/ identity-spliced	/f/ cross-sliced
Mean percentage of Timeouts	5.5 (10/180)	12.7 (23/180)	0 (0/180)	3.3 (6/180)
Mean RT	652.09	654.54	688.1	676.6

TABLE V. Average percentages of Timeouts and mean RTs in ms for the Polish listeners and the two fricatives in both splicing conditions in Experiment IV. The absolute numbers of Timeouts and the total numbers of trials are given in brackets.

Summary and discussion

Like Spanish and English listeners, Polish listeners are affected by misleading formant transitions. The phoneme repertoires of all three languages contain spectrally similar fricatives, and the results are thus in line with our hypothesis that listeners learn to

direct their attention to subtle acoustic cues for fricative identification if required by their native phoneme repertoire. Furthermore, we can reject the possibility that listeners only take advantage of formant transitions in order to identify the spectrally diffuse and therefore perceptually less salient labio-dental fricative. Even though we found no significant interaction between Splicing and Fricative for Polish listeners, the error data indicates that in contrast to all the other listener groups Polish listeners missed four times as many cross-spliced /s/-items than /f/-items. Especially the spectrally salient /s/ requires attention to formant transitions if this fricative can easily be confused with other fricatives in the listeners' phoneme repertoire.

On which level may such language-specific differences occur? We used the term attention to refer to listeners' learned selection of acoustic cues for phoneme identification, without assuming that listeners differ in sensitivity at the auditory level. Differences in sensitivity would imply that Dutch and German have "lost" such sensitivity. However, listeners are known to display sensitivity to foreign-language contrasts which fall entirely outside the range of the native phoneme repertoire (Best, McRoberts & Sithole, 1988). Thus the effects that we have observed may reflect strategic listening choices which have no implications for the underlying sensitivity. If so, Dutch listeners, too, may perceive the acoustic mismatches if their attention is drawn to them. We tested this possibility in Experiment V.

Furthermore, the phoneme inventories we have tested differ in whether or not they offer an alternative category in the case of an ambiguous fricative of a particular kind. In Experiment V we also tested the effects of this response availability. We used an untimed open-choice identification task, with Dutch and Spanish listeners. If no response alternatives are given, participants are expected to choose a phoneme category from their native inventory. Spanish listeners may identify at least some of the cross-spliced /f/-tokens as /θ/. Dutch listeners, in contrast, should identify all tokens of cross-spliced /f/ as /f/. By asking subjects to judge the goodness-of-fit of the

stimuli, we examined the extent to which both Dutch and Spanish listeners perceive mismatch effects of cross-splicing.

EXPERIMENT V

Method

Materials

Materials were the target-bearing VCV-strings of all 60 items used in Experiment II, including the identity-spliced and cross-spliced targets (e.g., from the experimental item *tikufa* we presented the fragment *ufa*).

Procedure

Participants, seated in a sound-attenuated room, were presented with the VCVs over headphones. They were instructed to write down the intervocalic consonant, and to judge on a scale from 1 to 8 whether it was a poor or a good example of this consonant. After the test, participants identified the letters they used to describe the consonants by writing down a native example word containing each letter used.

Participants

Thirty-one students from the Radboud University Nijmegen took part in this experiment. 14 were native Dutch regular students, and 17 were native Spanish exchange students. They were paid for their participation. None reported any speech or hearing disorders.

Results

Dutch listeners always identified each of the stimuli as either /f/ or /s/. Spanish listeners, on the other hand, showed greater response variance. Five of the 17 Spanish listeners reported hearing exclusively /f/ and /s/, while the remaining 12 participants included other consonants in their responses. All cross-spliced /s/ were identified as /s/, but the responses for /f/ varied, including /b/, /d/, /m/ and, most frequently, the dental fricative /θ/. One item was identified by none of these 12 Spanish participant as /f/, but as a poor example of /θ/. All in all nine cross-spliced /f/ were identified by at least five Spanish participants as a consonant belonging to a category other than /f/.

The average ratings for the items which were correctly identified as either an /s/ or an /f/ were: for identity-spliced /s/, Dutch 3.95, Spanish 4.81; for cross-spliced /s/, Dutch 3.94, Spanish 4.67; for identity-spliced /f/, Dutch 3.78, Spanish. 4.53; for cross-spliced /f/, Dutch 3.01, Spanish 3.73. We analyzed the averaged ratings in an Analysis of Variance. We found main effects of Language ($F(1,56) = 120.77$, $p<0.001$), Splicing ($F(1,56) = 21.96$, $p<0.001$), and Fricative ($F(1,56) = 37.01$, $p<0.001$) and an interaction between Splicing and Fricative ($F(1,56) = 15.25$, $p< 0.001$). In general Spanish listeners rated the stimuli as better examples than Dutch listeners, probably because they were presented with their native phoneme realizations. The cross-spliced /f/ items were rated as poorer examples than the identity-spliced /f/ by both listener groups.

Discussion

Experiment V showed that the acoustic mismatch in the cross-spliced /f/ tokens turned them into poorer instances of /f/. While Dutch listeners just perceived these /f/ tokens as poorer members of the /f/ category, Spanish listeners identified some of these tokens as belonging to another category, most frequently as a /θ/. Thus the availability of an alternative category may be a crucial factor in determining whether the mismatch

between fricative noise and formant transitions results in the perception of a different category. Although Dutch listeners seem to accept the cross-splicing as allophonic variation of /f/, the goodness ratings showed that they too were sensitive to the acoustic mismatch.

We reanalyzed the Timeout errors from Experiment II, including for /f/ only the six items which the Spanish participants had always identified as /f/ when cross-spliced. In this new analysis, the significant three-way interaction between Language, Splicing, and Fricative no longer reached significance. This may be because that three-way interaction had been principally carried by the nine items which produced variable responses in Experiment V; alternatively, of course, it could simply result from reduction of statistical power.

In an additional analysis we included the average Dutch ratings as a predictor for the Spanish Timeout Errors in Experiment II. Splicing remained statistically significant ($F(1,57) = 42.12$, $p < 0.001$). This result suggests that even though Dutch listeners perceive the acoustic manipulation in the stimuli, the cross-splicing of the /f/ is definitely more harmful for the Spanish than for the Dutch listeners.

GENERAL DISCUSSION

Many studies have investigated the contribution of formant transitions to fricative identification. Some studies reported robust effects whereas others failed to find any perceptual relevance of formant transitions for fricatives. In four phoneme detection experiments, we tested the hypothesis that attention to formant transitions as cues for fricative identification differs as a function of the presence of perceptually confusable fricatives in the listeners' native language. The targets in the detection experiments were /s/ and /f/ surrounded by either misleading (cross-splicing condition) or by coherent (identity-splicing condition) formant transitions. The stimuli were presented to Dutch, German, Spanish, English, and Polish listeners.

Our results support the hypothesis. First, target fricatives surrounded by misleading formant transitions were missed more often than fricatives with coherent formant transitions. This finding confirms previous work (Harris, 1958; Heinz & Stevens, 1961) showing that English listeners attend to formant transitions for some fricatives. More importantly, however, we observed a language-specific pattern of taking these acoustic cues into account for phoneme identification. Native listeners of Dutch and German, both languages without spectrally confusable fricatives, were not affected by misleading formant transitions. In contrast, listeners of Spanish and English, languages with the spectrally similar labio-dental /f/ and dental /θ/ fricatives, and Polish, a language with spectrally similar sibilants, were affected by misleading formant transitions.

On the basis of the languages in which we found formant transitions to be used, we further queried whether attention to formant transitions is restricted to the spectrally similar contrasts only or whether it generalizes to non-confusable fricatives. We found that transition cues were restricted to /f/ for the Spanish listeners. For Polish listeners, the crucial interaction between Splicing and Fricative narrowly failed to reach significance ($p=0.053$). But, as shown in Table V, the effect of splicing was greater for /s/ than for /f/. For English, the interaction between Splicing and Fricative did not reach significance, even though the effect is numerically greater for /f/ than for /s/. This may indicate that English listeners were also affected by misleading formant transitions for /s/. This is not incompatible with our hypothesis, if we take into consideration that English, in contrast to Spanish, has a post-alveolar fricative category, which is spectrally more similar to /s/ than to /f/. Thus, with respect to our second hypothesis, we can tentatively conclude that attention to formant transition is restricted to spectrally similar fricative categories. Which fricatives are spectrally similar, of course, is a function of all fricative contrasts in a language, and their distribution in the perceptual space.

The pattern in our data, and in English in particular, might of course also have been affected by the particular splicing manipulation we applied to our stimuli. The frication noises of /f/ and /s/ differ in several ways; most importantly, /f/ has a flat diffuse spectrum, while /s/ shows prominent energy peaks. The spectra of /f/ and /θ/, and of /s/ and /ʃ/, however, show more similarities; cross-splicing within these pairs might well show effects with English listeners. Whalen (1981) found that English listeners' categorization of an ambiguous synthetic fricative noise as either /s/ or /ʃ/ was influenced by formant transitions. In his experiment, a synthetic 10-step noise continuum was combined with coherent or inappropriate natural vocalic portions, including formant transitions. Interestingly, the formant transitions contributed to listeners' decision only at those steps of the noise continuum which modeled noise spectra with energy peaks appropriate for natural /ʃ/ or /s/-spectra. This suggests that for English listeners fricative noise with spectral peaks in combination with mismatching formant transitions may have a similar effect to the mismatching transitions to /f/. In our study, however, the difference between the cross-spliced pairs apparently overrode a potential confusion for the English listeners. Further research could investigate whether mismatching information in formant transitions to /s/ might also mislead English listeners – for example, into classifying an input as post-alveolar.

Importantly, the Polish data suggest that the acoustic make-up of a fricative by itself does not determine the use of formant transitions. Even though /s/ has salient acoustic characteristics (Harris, 1958; Strevens, 1960; Jassem, 1965) which make it perceptually very robust, Polish listeners were affected in particular for this fricative. Thus, the crucial factor in the use of formant transitions appears to be the acoustic make-up of a fricative in relation to all other fricatives in the phoneme inventory.

The present results indicate that listeners integrate cues in a language-specific way. The information conveyed in formant transitions appears to play a crucial role in determining fricative categorization for Spanish, English and Polish listeners. This language-specific way of selecting cues for attention does not seem to be a strategy

that a listener can easily adapt to the requirements of the situation, or to the experimental situation. The stimulus set in our experiments did not contain the dental fricative /θ/. That is, a direct distinction between the two confusable fricatives /f/ and /θ/ was not necessary for efficient performance within the experimental situation. Nonetheless, the Spanish and English listeners were substantially misled by incorrect formant transitions for /f/. Similarly, the Polish listeners were misled by incorrect formant transitions for /s/, even though the palatal fricatives, which in Polish might be confused with /s/, were not present in the experiment. This suggests that for listeners of these languages, formant transitions are part and parcel of the fricative categories.

We have distinguished "attention" from "sensitivity" to formant transitions. Experiment V showed that Dutch listeners perceive an acoustic difference between the identity- and cross-spliced items. They rated cross-spliced /f/-tokens as poorer examples of /f/, though in phoneme monitoring these poorer examples were not responded to significantly differently from the better examples. We assume that the attunement to a native language does not have any consequences on a low auditory level: sensitivity is unaffected. All listeners may perceive acoustic mismatches between formant transitions and noise spectrum, but language experience determines whether this information is attended to in fricative identification. Experiment V shows that the mismatching information in the transitions led Spanish listeners into the percept of a different fricative; the availability of more fricative categories encourages attention to subtle cues such as formant transitions. Where there is no alternative category – as in the case of Dutch – mismatching information in formant transitions may be treated as just allophonic variation. Thus what Spanish listeners in Experiment V could identify as a dental fricative or even as a stop, Dutch listeners simply judged to be /f/. The number of possible choices for identifying an ambiguous stimulus has an effect on the distinctiveness of categories, and thus on listener s' response options. Recall that the goodness ratings of the Dutch listeners in Experiment V did not suffice

to explain the errors made by Spanish listeners, however. Thus the Dutch and Spanish listeners differed in how mismatching information affected fricative identification.

Primary cues are defined by some researchers (e.g., Stevens & Blumstein, 1981) as invariant acoustic properties which are independent of the phonetic context and sufficient to evoke the percept of a given phoneme. Secondary cues, in contrast, are context-dependent cues, exploited by listeners to support primary cues when needed, for instance in difficult listening conditions. We have shown that a context-dependent cue can also make an important and systematic contribution to fricative identification. Spanish listeners missed over 25% of the /f/-tokens which were surrounded by misleading formant transitions. The selection of primary and secondary cues appears to be language- and phoneme-specific, and depends on the degree to which cues enable listeners to distinguish native phoneme categories accurately and efficiently. Even though other acoustic characteristics, such as the generally higher intensity of the fricative noise, are used by listeners to distinguish sibilants from other fricatives, Polish listeners appear to use cues in the formant transitions, simply because of the number of confusable sibilants in their native phoneme repertoire.

In our experiments, listeners did not categorize or discriminate pairs of fricatives. In phoneme monitoring, participants react as soon as they recognize the target, and they do so only if the acoustic stimulus matches their abstract memory of the target. Reduced or mismatching information – here, the cross-spliced formant transitions – led Spanish, English and Polish listeners into errors. Most previous studies of fricative perception have used untimed identification tasks. Results showed that Argentinian listeners could use transition information for some fricative contrasts Borzone de Manrique & Massone, 1981), Dutch listeners apparently did not use it (Klaassen-Don, 1983), while English listeners appeared to use transition information in some studies (Harris, 1958) but not in others (Jongman, 1989). We cannot exclude the possibility that with unlimited response time listeners may be able to extract more information from static cues than they do in a running-speech situation, and that characteristics of particular experiments may have been more versus less encouraging

to such strategies. A task such as categorization (Whalen, 1981)[2] for example, could induce a different listening strategy. In categorization listeners assign an acoustic signal to one or another category, and it is reasonable to assume that the mental representations of these categories, including the acoustic cues which distinguish between them, are in listeners' focus of attention, and might not need to be retrieved with every stimulus. This could affect both response accuracy and reaction times.

Adult listeners are specialized in identifying their native phonemes. An efficient way of selecting acoustic cues is thus another feature of language-specific processing which children must acquire in the course of their language development. In the same way that children learn to distinguish only native language contrasts (e.g. Werker & Tees, 1999; Sebastián-Gallés & Soto-Faraco, 1999), children must learn to be parsimonious with their attention to the subtle details of the acoustic signal and with the selection of relevant cues. Research by Nittrouer and colleagues (Nittrouer & Miller, 1997a,b; Nittrouer, 2002) shows that there is indeed a developmental shift in the relevance of the cues conveyed by the frication and by the dynamics in the formant transitions for fricative identification. American English speaking children between four and seven years of age show a developmental decrease in their weighting of formant transitions and a developmental increase in their weighting of the noise characteristics for /s/ and /ʃ/. On the other hand, another study by Nittrouer (2001) showed that American English speaking children and adults are more similar in assigning weight to formant transitions for the distinction between the labio-dental and the dental fricatives. Thus, the developmental shift is restricted to the contrasts which are sufficiently characterized by the static cues alone. Nittrouer argues that the attention/sensitivity to dynamic cues diminishes when children learn which cues carry "phonetic informativeness" in their native language.

[2] Note that Whalen's (1981) research also showed effects of context vowels on the identification of fricatives. We in fact included the context vowels as factors into our analyses. As these results did not prove to be language-specific, however, we do not report them in detail.

Children's speech perception differs even up to 10 years of age from adults' speech perception (Elliot & Katz, 1980). Nittrouer´s Developmental Weighting Shift Theory contrasts with, for instance, explanation in terms of auditory cortex maturation (Sussman, 2001). Most of the data relevant to this debate come so far from English, and we suggest that the debate would profit from additional data from other languages, for instance, the five languages of the present study. Our results show that children will reorganize their sensitivity to formant transitions in a language-specific: way to spectrally similar fricatives. English, Spanish and Polish children should keep their attention to formant transitions, whereas Dutch and German children will not.

The shift in attention during language socialization entails that a listener would have to re-acquire, or reorganize attention to these cues in order to attain a native-like perception in a second language. Previous research (Repp, 1981; Hazan, Iverson & Bannister, 2005) suggests that listeners can indeed direct attention to otherwise unused phonetic cues, at least after being exposed to sufficient training. Future research will have to determine how rapidly speakers of a language without perceptually similar fricatives can learn to take advantage of formant transitions to efficiently distinguish between perceptually similar fricatives in a second language.

Are fricatives perceived only on the basis of the static characteristics of their fricative spectrum, or do formant transitions also play a role? A large number of studies have addressed these questions, but the pattern of results, as we demonstrated in the Introduction, has been contradictory. Previous studies have examined the question in different languages; and language-specific phonology may be the key to whether listeners rely solely on spectral cues to fricative identity, or also attend to transition information. Even though all listeners will always make use of information in the fricative spectrum, for listeners of some languages formant transitions also play a crucial role for some of their native fricatives. Mismatching acoustic information in formant transitions may be perceived by all listeners at a low phonetic level, but the use of this information for the identification of a given fricative seems to depend on

whether the spectral characteristics of its frication suffice to distinguish this fricative from all other fricatives in the listener's language.

Cross-language differences in the uptake of cues for place of articulation

CHAPTER 4

A slightly adapted version of this paper has been submitted to
Journal of the Acoustical Society of America (Wagner, A., in revision)

Abstract

Cross-language differences in use of coarticulatory cues for the identification of fricatives have been demonstrated in a phoneme detection task: Listeners with perceptually similar fricative pairs in their native phoneme inventories (English, Polish, Spanish) relied more on cues from vowels than listeners with perceptually distinct fricative contrasts (Dutch and German). The present gating study examined the time-course of cue uptake to further investigate whether cross-language differences in the reliance on coarticulatory cues result in: (1) Temporal differences in the uptake of cues to place of articulation for fricative identification; (2) Cross-language differences in the uptake of cues also for plosive identification; (3) Earlier or later uptake of information from coarticulatory cues preceding or following the consonant. Dutch, Italian, Polish and Spanish listeners identified fricatives and plosives in gated CV and VC syllables. The results showed cross-language differences in the temporal uptake of information for fricative identification: Spanish and Polish listeners extract information about the place of articulation from shorter portions of VC syllables. No language-specific differences were found in the use of coarticulatory cues for plosive identification, suggesting that higher reliance on coarticulatory cues does not generalise to other phoneme types. Furthermore, the language-specific differences for fricatives were based on preceding coarticulatory cues.

INTRODUCTION

To identify individual phonemes listeners integrate acoustic information which is spread across the utterance. During the time course of a spoken utterance several acoustic cues become available to specify features of a speech sound, and listeners select information in language-specific ways (e.g. Crowther and Mann, 1992). The native phonology and the make-up of the phoneme inventory set up a language-specific distribution of informative cues (Holt and Lotto, 2006), and alter listeners' perception at a very low phonetic level (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Ketterman, and Siebert, 2003). Such a subconscious selection and integration of cues appears to be guided by the demand to optimally distinguishing all native phonemes. Phoneme inventories differ in their subsets of distinctions for places of articulation, and listeners may show language-specific optimisations in the uptake of information specifying this feature. Reliance on different cues may result in differences in the temporal uptake of information. Do listeners of different native backgrounds gain detailed information at different time points? This paper presents a study which examined the temporal uptake of information for place of articulation in a cross-linguistic gating experiment.

The speech signal contains an overabundance of acoustic information. Some acoustic events may contribute to the perception of combinations of phonemes, or to individual phonemic categories, or may carry information specifying phonological features, like place of articulation. Listeners extract information from acoustic events which are internal to the articulatory constriction, and from coarticulatory cues. Internal cues, such as the spectrum of the burst, closure duration or duration of the frication provide at the same time information about manner and place of articulation (e.g., Wright, Frisch and Pisoni, 1995). Coarticulatory cues, often described in terms of formant transitions, reflect the changing configurations of a speaker's vocal tract, and can provide a reliable source of information to place of articulation.

The manifestation of coarticulatory cues is not independent of manner of articulation. In the case of plosives, the relevance of formant transitions has been acknowledged across decades of research (Liberman, Delattre, Cooper and Gerstman, 1954; Delattre, Liberman and Cooper, 1955; Sussman, Fruchter, Hilbert and Sirosh, 1998). The information in the formant transitions, and the release burst both provide information about a plosive's place of articulation (e.g., Dorman, Studdert-Kennedy and Raphael, 1977). The contribution of formant transitions for plosive identification has been formulated in the concept of locus (Stevens and House 1956), Lindblom's (1963) locus equations, and in Sussman et al.'s (1998) view of locus equations as universal and invariant cues to place of articulation.

Locus equations, capturing the onset of F2 at stop release in relation to the F2 in the vowel, are supposed to provide a measure of coarticulation between consonants and vowels (Krull, 1989). Studies comparing locus equations with articulographic (Löfquist, 1999) or electropalatographic (Tabain, 2000, 2002) measurements of coarticulation, however, do not always show a correlation between coarticulation and the acoustic measure locus equations. For instance, studies by Tabain (2000, 2002) report high correlation between electropalatographic measurements of coarticulation and locus equations for voiced stops, a less good correlation for voiceless stops, and a poor correlation between coarticulation and locus equations for fricatives. This study also shows that fricatives and plosives do not differ in the amount of coarticulation, but in the degree in which coarticulation is captures by locus equations.

In the case of fricatives, studies based on acoustic measurements show indeed a small contribution of transitional cues to distinctions of places of articulation. Jongman (1998) and colleagues (Jongman, Wayland and Wong, 2000) show that multiple acoustic cues, such as spectral peak location, noise duration, or amplitude contribute to an invariant specification of all English places of articulation for fricatives: The inclusion of the vowel portion appears not to improve fricative identification.

Perceptual studies, however, show that the vocalic portion improves listeners' identifications for some fricatives, but is less relevant for others. In a study by Harris (1958) American English listeners categorised natural tokens of fricative vowel syllables consisting of /f v θ ð s z ʃ ʒ/ and /a i u e/. The fricatives were combined to syllables with every vowel as produced in the context of each of the fricatives. Some of the syllables thus contained vocalic information which was incoherent with the information in the frication noise. Participants accurately categorised the sibilants disregarding the information in the vowel. The fricatives /f/ and /θ/ were often confused, and their identification improved when the frication was presented with the coherent vocalic portion. Heinz and Stevens (1961) obtained similar results with synthesised fricatives.

Furthermore, shifts in the perception of a fricative's noise as a function of coarticulation appear to be caused not only by the formant transitions to adjacent vowels but also by the vowel itself. Mann and Repp (1980), and Whalen (1981) showed that listeners' identification of a syllable consisting of an ambiguous synthetic noise between /s/ and /ʃ/ combined with natural vowels is affected by both the transitions and the vowel. An ambiguous noise is more often identified as an /s/ when it is followed by the rounded vowel /u/ than by /a/. Listeners adjust their perceptual evaluation of the frication depending on the roundness of the adjacent vowel.

A study by Hedrick and Ohde (1993) investigated listeners` perception of places of articulation for fricatives as a function of the duration of the frication, of the quality of the vowel, of the formant transitions, and of the amplitude of the frication relative to the amplitude of the vowel. This study showed best identifications resulting from amplitude comparisons between the frication and the onset of the vowel. The information gain resulting from the amplitude comparisons overrode the coarticulatory effects of the vowel and of the formant transitions.

Whereas coarticulatory effects for plosives can be described in terms of formant transitions, more sources of information are evaluated by listeners in fricative

identification. The spectrum of the noise, the formant transitions and the vowel seem to jointly affect the perception of fricatives (Repp, 1982). Information resulting from the coarticulation can contribute to the perception of place of articulation both for plosives and for fricatives, but how coarticulation represents itself appears to vary between the two phoneme types. Hence, different phoneme types may demand different acoustic cues for the same phonological feature place of articulation.

The search for cues to phonological features becomes even more complicated when language-specific differences in the use of cues are considered (e.g., Crowther and Mann, 1992). For instance, the evaluation of transitional cues for the distinction of /r/-/l/ has been shown to differ between Japanese, German and English listeners (Iverson et al., 2003). Japanese listeners appear to be more sensitive to F2-onset and less sensitive to the contribution of F3-onset, which is the most valuable cue for English listeners. When learning English, Japanese listeners will compensate for this lower sensitivity by attending to other cues. When listeners extract information for phonological features from different acoustic cues, they may accordingly also extract this information at different time points over the course of the acoustic signal.

For fricatives, cross-language differences in the reliance on coarticulatory cues are reported in a study by Wagner, Ernestus and Cutler (2006). In a phoneme monitoring study Wagner et al. compared the identification of fricatives among Dutch, English, German, Polish, and Spanish listeners. All listeners were presented with natural, cross-spliced, or identity-spliced materials, such that the fricative targets were surrounded by vowels containing either coherent or mismatching information. For example, /s/ as target in the nonsense word *tikusa* was either replaced by the /s/ of another token of *tikusa* in the identity-spliced condition, or it replaced /f/ in the nonsense word *tikufa* in the cross-spliced condition. While Dutch and German listeners did not show any impediment in fricative identification due to conflicting cues, English, Polish, and Spanish listeners showed a significant drop-off in their identification due to mismatching vocalic context. Moreover, English and Spanish listeners were hindered in the identification of particularly the labio-dental fricative,

while Polish listeners made more errors when identifying the acoustically salient alveolar fricative /s/.

An explanation for these language-specific differences lies in the fricative repertoires of the languages tested. The Dutch and German inventories have fricative contrasts that are spectrally very distinct, while English and Spanish have the spectrally similar fricatives /f/ and /θ/, and Polish distinguishes four palatal sibilants. Because Spanish, English and Polish listeners have learned to draw boundaries between contrasts which are perceptually more similar, the vocalic portion adjacent to the frication plays a role in their fricative categorisation. The distinctiveness of spectrally similar fricative pairs may be perceptually enhanced by integrating more cues from coarticulation.

Wagner et al.'s study suggests that listeners differ in how they extract information specifying place of articulation for fricatives. This study also suggests that listeners of some languages disregard systematic acoustic variation in the signal. In this case, listeners may differ in the amount of information they have about a speech sound at different time points as the utterance unfolds. The phoneme monitoring study leaves open the question of whether English, Polish, and Spanish listeners relied on the cues preceding or following the frication. In this study listeners were presented with conflicting vocalic portions surrounding the frication. If listeners relied on the vowel portion preceding the frication they may extract information about place of articulation earlier; If listeners relied on the cues in the vowel following the frication then the information specifying the place of articulation may be extracted later in time.

For plosives, it is generally assumed that formant transitions following the burst are more relevant cues to place of articulation than transitions preceding the closure (e.g., Stevens and Blumstein, 1978, Fujimura, Macchi and Streeter, 1978). Greater perceptual effect of post-consonantal formant transitions can be explained as a recency effect. In natural speech, when listeners accumulate pre-consonantal cues, consonant-inherent cues, and post-consonantal cues all coherently specifying the same place of

articulation, listeners may just rely on the information which comes in last. Such recency effects can explain the greater impact of pre-consonantal formant transitions in studies with conflicting pre- and post-consonantal formant transitions in stop consonant identification (e.g., Fujimura, Macchi and Streeter, 1978).

Studies investigating the effect of order of presentation for fricatives also suggest a greater perceptual relevance of post-consonantal vowels. Mann and Soli (1991) compared the contribution of formant transitions across fricative-vowel (FV) and vowel-fricative (VF) syllables. Formant transtions in FV- syllables showed a greater effect on listeners' identification than VF transitions. In a second experiment, listeners were presented with materials containing FV and VF formant transitions played in reversed order. The order of presentation and not an intrinsic difference between FV and VF formant transitions proved to determine which information affected fricative identification. Also a study by Nittrouer, Miller, Crowther and Manhart (2000), using the same paradigm, showed that adult listeners make more use of the information from the formant transitions following the frication than of the transitions preceding the frication. In this study a modest effect of order of presentation was found. VF transitions in the used materials, however, appeared to contain less information about the fricative than formant transitions following the frication.

It is thus unclear whether post-consonantal formant transitions provide per se more information than pre-consonantal, or whether listeners benefit more from the most recent information. An argument for a greater influence of the most recent information can be found in a study by Whalen (1981). This study reports that the evaluation of coarticulatory cues on vowel-fricative-sequences decays when a silent interval is inserted between the vowel and frication. Whalen argues that the silence segregates the two sources of information, and reduces the effect of coarticulation.

The present study was designed to investigate the hypothesis that listeners whose native phoneme inventory requires the differentiation between more places of articulation for fricatives will optimise their listening strategies to gain information

from more coarticulatory cues. Three issues are addressed by the present study: (1) Whether differences in attention to coarticulatory cues for fricative identification result in differences in the temporal uptake of information specifying place of articulation; (2) Whether cross-language difference in the reliance on transitional cues are specific to fricatives or whether they hold also for the perception of place of articulation for plosives; (3) Whether the uptake of additional coarticulatory information is based on earlier or later uptake of cues. If all listeners benefit more from the most recent information, it appears plausible that listeners who are in need of more sources of information give attention cues which are less relevant for listeners who can distinguish all native fricatives on the basis of the frication noise. Listeners from four different native backgrounds are compared in a gating experiment.

Gating experiments have been frequently used to study the temporal uptake of information (Grosjean 1980; Smits, 2000; Smits, Warner, McQueen and Cutler, 2003, Warner et al., 2005). In gating experiments listeners are presented with truncated portions of the acoustic signal, that is with portions from which certain temporally distributed cues have been cut off, usually without otherwise manipulating or synthesising the signal (for an overview see Grosjean 1996). This procedure allows assessing which acoustic information becomes available with different segments of the signal. Results from gating studies show that in spite of the existence of perceptually critical points, when listeners are asked to identify a segment, they base their decision on temporally distributed cues, even though they have not yet identified the entire segment. The gating technique can be used in two different ways: forward gating and backward gating. In forward gating listeners are presented with parts of the signal preceding the truncation point, while in backward gating listeners hear portions of the signal following the truncation point. Studies using backward gating have shown the relevance of cues in the vowel portions following, and studies using forward gating the relevance of acoustic events preceding the constriction of a consonant (e.g., Smits, 2000, Smits, et al, 2003 Warren and Marslen-Wilson, 1987). The gating paradigm thus

allows addressing the question of cross-language differences in the temporal uptake of coarticulatory cues preceding or following a consonant.

EXPERIMENT

Language compared

The languages compared in this study are Dutch, Italian, Polish and Spanish. Dutch, Polish and Spanish were also among the languages tested in the phoneme monitoring experiment by Wagner, et al. (2006). In that study, mismatch between vowels and frication impeded fricative identification for Polish and Spanish listeners, who have perceptually similar fricative pairs in their native fricative repertoires, but not for Dutch or German listeners, who have only perceptually distinct fricative contrasts. Italian allows to test the generalizability of the Dutch and German pattern to another language with only perceptually distinct fricative contrasts. All languages have three places of articulation for plosives: labial, alveolar, and velar, but they differ regarding the distribution of fricatives in their phoneme inventories.

Dutch distinguishes fricatives at four places of articulation, the labiodental /f v/, the alveolar /s z/ a velar /x/ and a glottal /h/. The Italian fricative inventory contains five spectrally distinct fricatives at three place of articulation: labiodental /f v/, alveolar /s z/, and palatal /ʃ/. The Polish fricative inventory contains 11 categories at six places of articulation, among them are four coronal places of articulation for sibilants: alveolar /s z/, postalveolar /ʃʲ ʒʲ/, the retroflex /ʂ ʐ/, and alveolo-palatal /ɕ ʑ/. In acoustic terms the fricative /s/ typically has energy peaks in the frequency range between 3 and 7 kHz, the post-alveolar fricative /ʃ/ exhibits energy peaks in the frequencies between 1.5 and 5 kHz. The alveolopalatal /ɕ/ has energy peaks between 2 and 6 kHz, and the retroflex Polish fricative /ʂ/ has energy maxima around 1 and 4 kHz (Jassem, 1968; Lipski, 2006). The coronal Polish fricatives thus show an overlap of energy peaks across their noise spectra. Spanish contrasts fricatives at four places of

articulation, but among them are the spectrally similar labio-dental /f/ and dental /θ/. The labio-dental and dental fricative are very similar with respect to their average spectral means, and their spectral variance (Jongman et al., 2000). The spectra of both fricatives are relatively flat with a distribution of energy across a wide range of frequencies from circa 2 to 10 kHz.

Predictions on the basis of Wagner et al.'s (2006) phoneme monitoring experiment can be formulated as follows. First, in fricative identification, listeners with several similar places of articulation (Polish and Spanish) should show an earlier or later uptake of acoustic information concerning place of articulation than listeners with distinct fricative categories (Dutch and Italian). Second, because these languages do not differ in their places of articulation for plosives, the experiment allows to test whether reliance on coarticulatory cues is specific to fricatives or whether it shows language-specific differences in reliance on transitional cues in general. Only if sensitivity to coarticulatory cues generalises across phoneme types will differences be observed with stops. In such a case, the differences would be expected to resemble those of fricatives.

Methods

Materials

The stop consonants [k p t] and the fricatives [f s] were combined with the point vowels [a i u] to create fifteen CV syllables and fifteen VC syllables, e.g. *af, ip, ut* for forward-gated materials, and *pa, su, ki* for backward-gated materials. The materials were recorded by a Dutch speaker in a sound-attenuated room directly into computer and down-sampled to 22.05 kHz (16 bit resolution).

Gating procedure. The gated materials were constructed with Praat software. The points of truncation were defined visually on the basis of the waveform and a wideband spectrogram. In order to explore the relevance of the vocalic portion

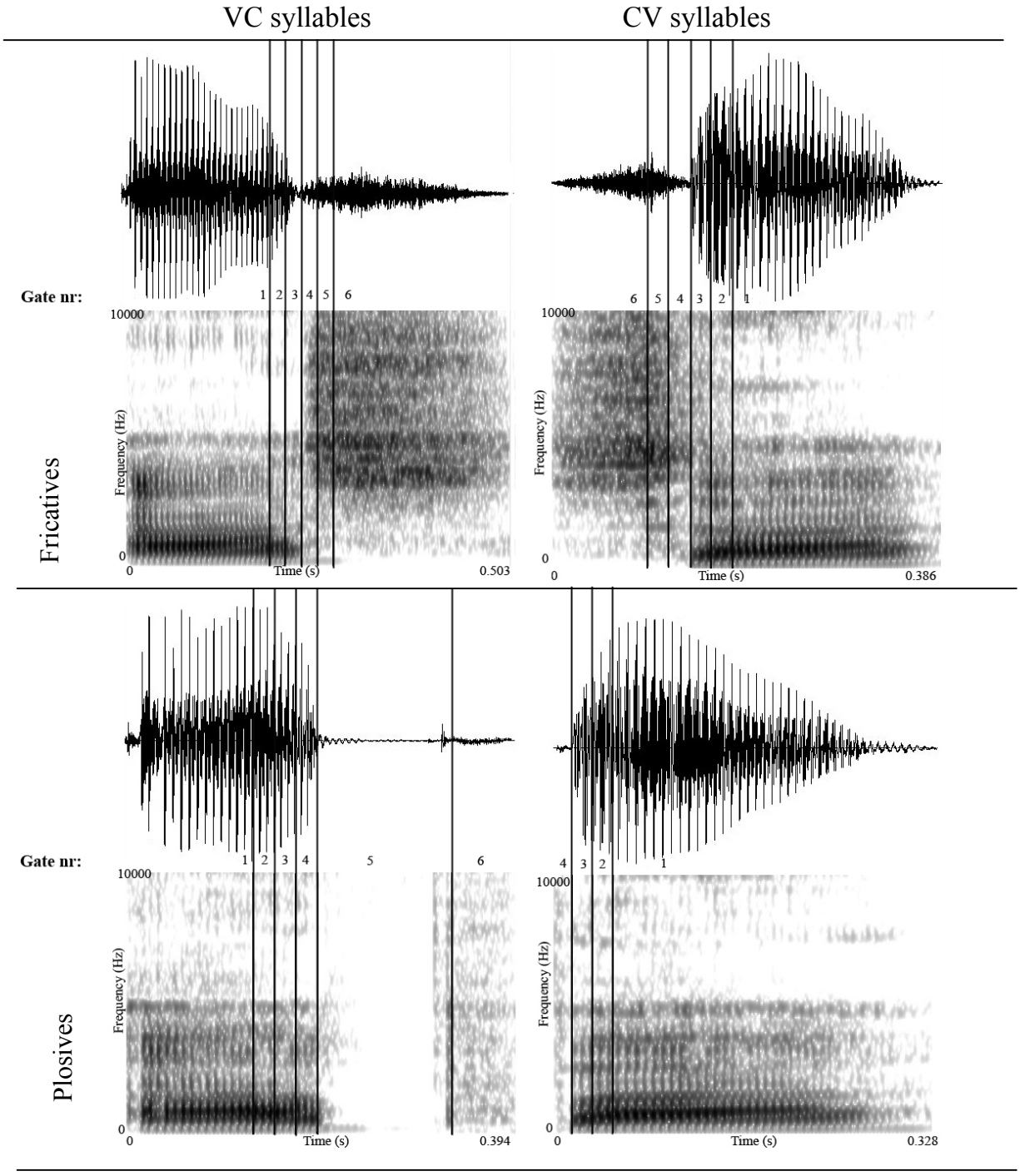


Figure 1: The placement and the number of gates for fricative and plosive targets in forward-gated (VC) and backward-gated (CV) syllables. Displayed are waveforms and spectrograms of the syllables /as/, /sa/, /ap/, and /pa/.

preceding and following the consonant, backward gating was used for CV syllables and forward gating was used for VC syllables. Figure 1 illustrates the placement of truncation points for the materials, and displays the number of gates for both fricative and stop targets.

First, the offset (for VC syllables) and onset (for CV syllables) of voicing in the vowel were defined as onset or offset of the consonantal constriction in the syllables. Second, truncation points in the vowel portion adjacent to these consonantal constriction were added in 20 milliseconds increments. The shortest gates thus contained a portion of the vowel starting (for CV syllables) or ending (for VC syllables) 40 milliseconds after (before) onset of the consonantal constriction. In a third step the consonants were gated. At this, the longer and continuous acoustic properties of fricatives versus the shorter and abrupt acoustic properties of plosives caused differences in the exact duration of gates. For fricatives, also the frication noise was gated in two 20 milliseconds increments, and the longest gate contained the entire frication noise. For plosive targets in VC syllables the first truncation point within the consonantal portion contained the closure and the release burst, and the longest gate included also a period of aspiration. For plosives in CV syllables only one gate was created within the consonantal portion. This gate contained the release burst preceding the vowel since the silent interval of the closure was not included, and syllable initial plosives were produced without aspiration. As a consequence, for fricatives the longest gates contained a frication of on average 120 milliseconds, while the longest gates for plosives in CV syllables included only the short release burst, and in VC syllables the silent interval of the closure, the release burst and a short period of aspiration. Note that only four gates were created for plosives in CV syllables. This truncation procedure resulted in six gates for fricatives and six gates for stop consonants in VC syllables, and in six gates for fricatives, but only four gates for plosives in CV syllables. In total, this resulted in 90 truncations (5 targets * 3 vowels* 6 gates) in VC sequences and 72 truncations (3 plosive targets * 3 vowels * 4 gates, and 2 fricative targets * 3 vowels * 6 gates) in CV sequences.

Previous gating experiments have shown that truncating the signal can add clicks or external noises to the gated segments (Pols, and Schouten, 1978). To avoid this, and to minimise previously observed response bias for labials (Smits, ten Bosch, and Collier, 1996) a 500 Hz square wave replaced the deleted parts of the signal, and a linear ramp within a window of 5 milliseconds was applied at the truncation points and the square wave. Ten practice syllables with [d n m] as targets and with the vowels [e u] were constructed to familiarise the participants with the gating task.

Procedure

The experiment was carried out in four different locations. At each, listeners sat in sound attenuated booths, and were presented with the materials over headphones. In advance, participants were instructed in their native language. The task was to label the consonant preceding (CV syllables) or following (VC syllables) the vowel, as one of /f k p t s/. From the onset of each item listeners had ten seconds to respond, and were instructed to guess in cases where they were very uncertain. The responses were given on a key board by pressing keys labeled with one of the five response options. The stimuli were blocked into forward-gated and backward-gated materials. Within these two blocks the materials were presented in random order. For each subject a different pseudo-random order of presentation was chosen. Between the two blocks was a pause, the duration of which was controlled by the participants themselves.

Participants

Fifteen native Dutch speakers from the subject pool of the Max-Planck-Institute for Psycholinguistics in Nijmegen, nine native Italian students at the University of Trieste, 15 Polish native speakers, students at the Universitet Slaeski in Katowice, and 18 Spanish native speakers, students at the Universitat de Barcelona, participated in this

experiment. None of them reported any speech or hearing disorders. Participants received a small payment or were rewarded with course credits.

Analysis of results

The data will first be presented in terms of percentages of correct responses. Correct responses however are limited in their explanatory power, because they only show the cases in which listeners recognised both place and manner correctly. More insight might be derived from listeners' confusion patterns; with materials of very short duration listeners' guesses reflect the acoustic information which is extracted even though the phoneme is not unambiguously identified. As the interest of this study is the temporal uptake of information for the phonological feature place of articulation, an analysis of Transmitted Information (TI) as a function of gate was conducted. Transmitted information (Shannon 1948, Miller and Nicely 1952, Jongman 1989, Smits 2000) is a measure of covariance of responses and stimuli. This measure is used for categorical judgements, per phonological feature and calculated from the entire set of responses. By investigating the covariance between stimulus and response, insight is gained about the information reducing listeners' uncertainty about a response as a function of the stimulus. One advantage of this measure is that it is bias free with values next to 0 if no information from the stimulus reduced participants' uncertainty about a response, and values approaching 1 if the stimulus transmitted all the information needed for a correct response. Transmitted information is calculated on the basis of the following formula:

$$TI_{(S,R)} = \Sigma_s \Sigma_r p(s,r) * \log p(s,r)/p(s)p(r)$$

Where p (s,r) is the probability of the joint occurrence of the response r with the stimulus s, p(r) is the probability of the response and p(s) is the probability of the stimulus. The maximum of transmitted information within a set of responses is expressed as the entropy of the stimulus, and described as:

$$H(s)=- \Sigma_s p(s) * \log p(s).$$

Whereas TI expresses the transmission from stimulus to response in bits, the relative information transmitted is given by

$$TI\ rel_{(S,R)}=T_{(S,R)}/H(s).$$

Results and discussion

Correct responses

From the analyses excluded were responses which did not occur within the ten milliseconds window from the onset of the item (1,7% in CV syllables, and 4,2% in VC syllables), and the data of one Dutch participant who missed more than 40 % of all items.

Table I. Percentage of correct responses for fricatives and plosives in CV syllables for the four listeners groups.

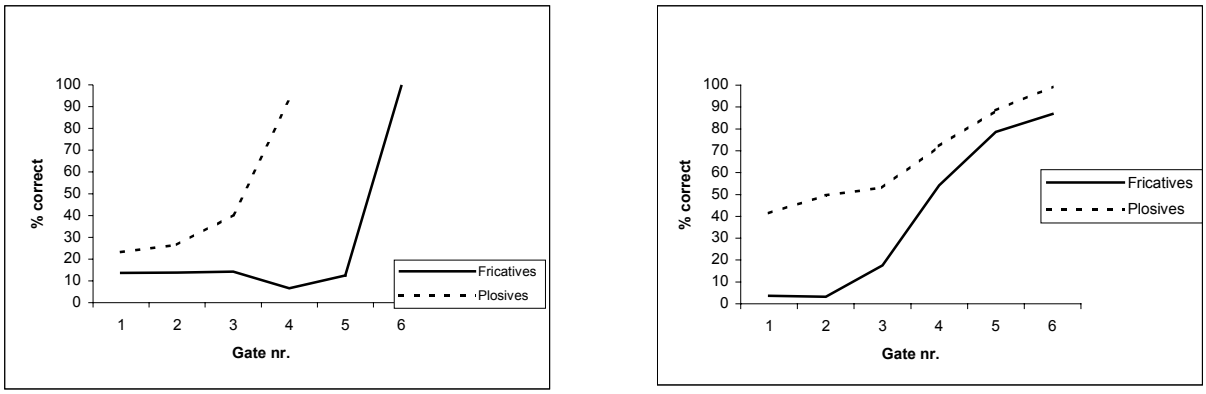
Fricatives					Plosives				
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	19.05	20.75	10.71	8.33	1	15.87	24.36	24.22	27.78
2	11.90	17.65	15.48	12.04	2	19.84	32.47	23.81	31.48
3	10.71	16.07	21.18	12.96	3	46.83	40.00	27.13	45.68
4	2.38	7.55	10.11	6.48	4	98.41	90.12	90.37	91.98
5	10.71	16.67	12.36	12.04					
6	100.00	98.15	100.00	99.07					

Table I lists the percentages of correct responses for fricatives and plosives for the four languages in CV syllables. Table II presents the same for VC syllables. The probability of a correct response was analysed in two linear multi-level models, separately for CV- and VC-syllables. Correct responses were modelled as a function of Language (Dutch, Italian, Polish, Spanish), Gate (gate 1 to 6, or gate 1 to 4 for

Table II. Percentage of correct responses for fricatives and plosives in VC syllables, for the four listeners groups.

	Fricatives					Plosives			
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	4.76	5.77	3.33	1.85	1	45.24	43.21	37.04	41.36
2	5.95	3.77	3.33	0.93	2	48.41	45.57	44.44	57.64
3	25.00	28.30	14.44	9.26	3	55.56	48.10	52.59	54.32
4	52.56	55.56	61.11	46.30	4	73.02	72.50	68.89	74.07
5	85.71	87.04	85.56	62.96	5	93.81	90.48	87.62	84.13
6	100	100	96.67	95.37	6	100	100	99.17	98.61

plosives in CV syllables) and Phoneme Type (fricative or plosive). Participant was included as a crossed random effect factor. The main effect of Language did not emerge as significant, and accordingly Figure 2 displays the correct identifications as a function of gate averaged across all listeners. Percentages of correct identifications as a function of Gate, separately for fricatives and plosives in CV syllables are displayed in the left panel of Figure 2, and the right panel displays the same for VC syllables.



a. b.

Figure 2: Percentage of correct responses averaged across all listener groups, for fricatives and plosives respectively as a function of gate in CV-syllables (2a) and VC-syllables (2b).

Fricatives. The probability of a correct response was analysed in two linear multi-level models, separately for CV- and VC-syllables. Correct responses were modelled as a function of Language (Dutch, Italian, Polish, Spanish), Gate (gate 1 to 6). Participant was again incorporated as a crossed random effect factor. In CV syllables, percentage of correct identifications rises abruptly as an effect of the presentation of the entire frication: A significant improvement in correct identifications emerged only at gate 6 ($F(1,669) = 92.1$, $p<.001$). In VC syllables, the percentage of correct identifications rises gradually: at gate 3, i.e. before the presentation of the frication, correct identifications are still at chance level. After the onset of frication, correct responses rise through gate 4 ($F(1,664) = 87.63$, $p<.001$), and gate 5 ($F(1,664) = 43.6$, $p<.001$). At gate 4, when 20 milliseconds of the frication are presented, correct responses exceed chance level, indicating that the onset of frication following the vowel is a substantial cue to the fricative identity.

Plosives. Results for plosive targets were analyzed in the same way as for fricatives. In CV syllables, the main effect of gate reached significance at gate 3 ($F(1, 980) = 19.36$, $p<.001$), during the presentation of the vowel portion including the last 20 milliseconds following the release burst. The substantial increase in correct identifications resulted from the presentation of the burst, in gate 4 ($F(1,933)=210.61$, $p<.001$). In VC syllables a significant effect of gate emerged in gate 4 ($F(1,997)=37.9$, $p<.001$), and in gate 5 ($F(1,986)=31.24$, $p<.001$). Gate 4 in VC syllables contains the vowel portion directly before the onset of the closure, and gate 5 contains the release burst.

To sum up, for fricatives, the results in CV syllables replicate Jongman's (1989) findings that listeners need at least 50 milliseconds of frication in order to identify fricatives correctly. Table I shows a drop off in correct identifications for fricatives in CV syllables in gates 4 and 5, i.e. when no more than 40 milliseconds of the frication were presented to the listeners. In VC syllables, however, listeners' identification scores improve gradually, suggesting that the first 20 milliseconds of frication do convey some information for the identity of a fricative. For plosives, a substantial

improvement in identifications in CV- and VC-syllables occurred after the presentation of the burst.

The analysis of correct responses shows a bias towards plosives. Since correct responses reflect correct identification of both manner and of place, fricatives are at disadvantage, because a short portion of the frication can be perceived as a release burst. To examine such bias an analysis of information transmitted (TI) for manner of articulation will be presented before the analysis of TI for place of articulation.

Transmitted information (TI)

Table III displays the percentages of correct responses for manner (independent of place) in CV and VC syllables. Table IV displays correct responses for place (independent of manner) for CV and VC syllables. Note that the values for correct responses are not free of the bias towards plosives. For the analysis of TI listeners responses were pooled into confusion matrices (for the confusion matrices see Appendix A, B, C, and D), separately for manner (irrespective of place of information), and for place of articulation, separately for fricatives and plosives.

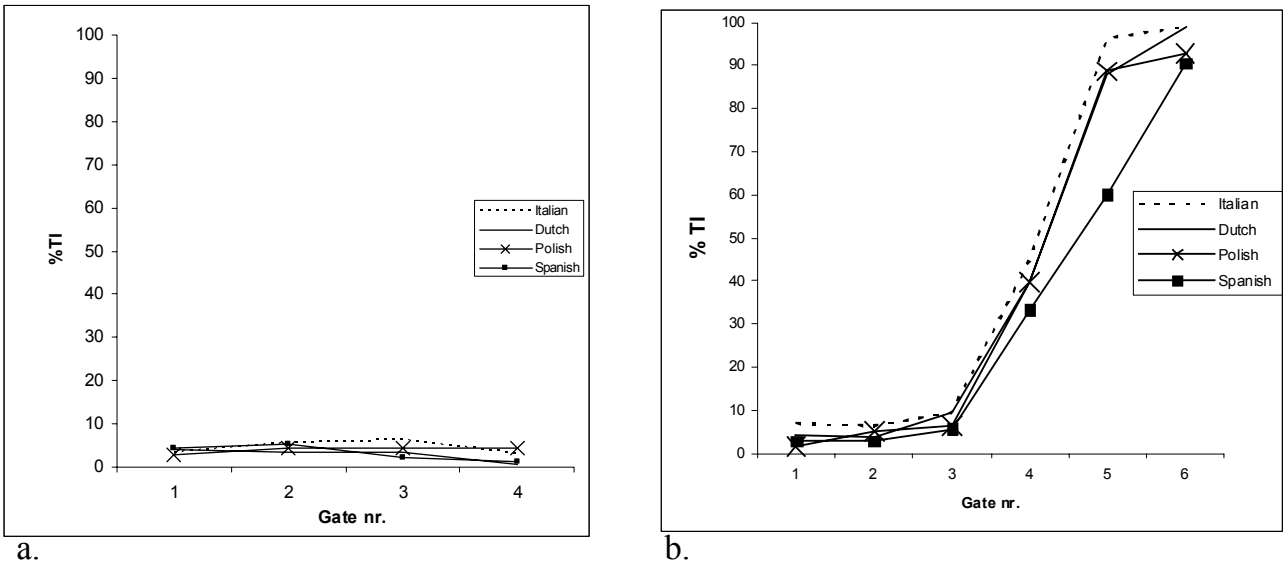
Table III. Percentage of correct identification in terms of manner of articulation (independent of place of articulation) for the four listener groups in CV and VC syllables.

CV syllables					VC syllables				
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	55.24	58.78	54.25	53.34	1	61.43	65.41	59.56	59.63
2	49.53	60.94	57.14	58.15	2	60.95	56.82	55.56	57.94
3	57.62	59.39	57.01	56.67	3	65.24	65.15	63.56	62.59
4	60.48	62.69	62.05	60	4	80.88	84.33	82.22	79.63
5	10.71	16.67	12.36	12.04	5	97.24	99.15	96.92	86.32
6	100	98.15	100	99.07	6	100	100	98.57	98.02

Table IV. Percentage of correct identification in terms of place of articulation (independent of manner of articulation) for the four listener groups in CV and VC syllables.

CV syllables					VC syllables				
Gate	Dutch	Italian	Polish	Spanish	Gate	Dutch	Italian	Polish	Spanish
1	31.90	39.69	31.60	38.52	1	49.05	45.11	42.66	46.29
2	39.05	39.84	37.62	43.70	2	50.95	50.76	51.11	57.94
3	54.29	48.87	46.73	60.37	3	59.05	55.30	53.77	56.67
4	92.38	91.04	92.86	94.44	4	74.51	73.13	77.77	77.41
5	89.29	83.33	91.01	89.81	5	92.26	89.74	89.23	83.33
6	100	98.15	100	100	6	100	100	99.05	97.22

Manner of articulation analysis



a. b.

Figure 3: Percentage of transmitted information (TI) for manner of articulation as a function of gate per language group in CV (3a) and in VC (3b) syllables

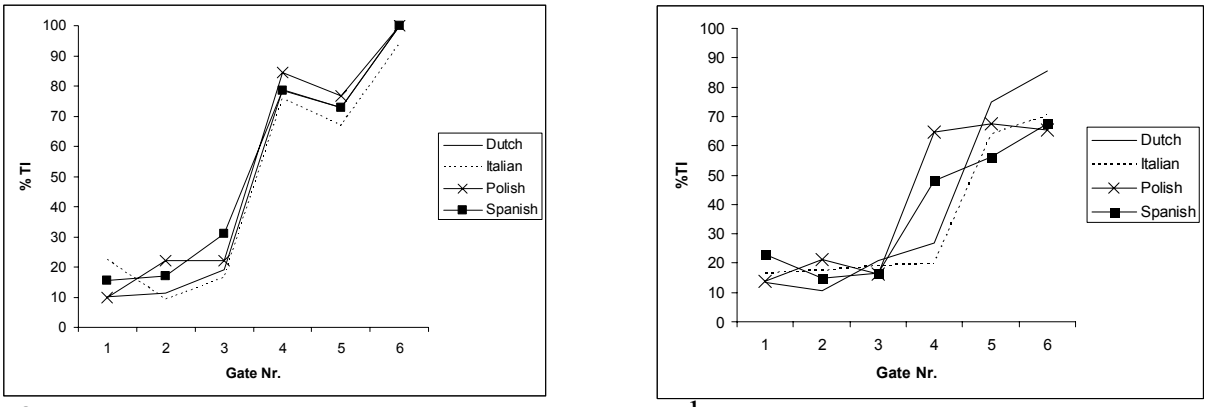
The left panel of Figure 3 shows the percentage of transmitted information (TI) for manner of articulation as a function of gate for the four languages in CV syllables,

and the right panel displays the same for VC syllables. For CV syllables, only the first four gates were included in the analysis of manner of articulation since plosive targets were presented only in four gates in CV syllables. To investigate which gates caused a significant increase in TI pair-wise Kruskal-Wallis test were conducted across two consecutive gates, for each language separately. Each comparison was evaluated at a corrected α level of 0.016 (0.05 divided by 3) in CV syllables, and of 0.01 (0.05 divided by 5) in VC syllables.

As displayed in Figure 3a, in CV syllables listeners extracted little information about manner of articulation. For all four languages no significant effect of Gate was found. Even though Table III displays correct identifications of manner of articulation of above 50 % in the first four gates, the analysis of TI reveals that these identifications were carried by the plosives. This confirms the negative bias towards fricatives. In VC syllables significant increases in TI emerged between gates 3 and 4 for all the languages, and between gates 4 and 5 for Dutch, Italian, and Polish but not for Spanish. The presence of closure versus frication proves to be the main cue for manner of articulation.

Place of articulation analysis

Fricatives. Figures 4a and 4b show the percentage of transmitted information for place of articulation as a function of Gate and Language for CV and VC syllables respectively. Again, pair-wise Kruskal-Wallis test were conducted across two consecutive gates, for each language separately, for CV and VC syllables. Each comparison was evaluated at a corrected α level of 0.01 (0.05 divided by 5 comparisons).



a. b.

Figure 4: Percentage of transmitted information (TI) for place of articulation as a function of gate per language group for fricative targets in CV (4a) and in VC (4b) syllables.

For CV syllables a significant effect of Gate for all four languages emerged between the gates 3 and 4, after the presentation of the first 20 milliseconds of frication, and between the gates 5 and 6, after the presentation of the entire frication noise. In VC syllables, pair-wise comparisons of TI between gates, separately for the four languages showed a difference between Spanish and Polish listeners on the one hand, and Dutch and Italian listeners on the other. An effect of Gate for Spanish and Polish listeners emerged only between gates 3 and 4, after the first 20 milliseconds of frication, whereas a significant increase in TI for Dutch and Italian listeners emerged only between gates 4 and 5, thus after the presentation of 40 milliseconds of frication.

The analysis of TI for place of articulation captures the place of articulation for both /s/ and /f/. The number of observations does not allow splitting the analysis into the two targets. Table 5, however, displays the percentage of correct identifications of place of articulation for the four language groups and the six gates divided into the two fricative targets, and suggests that the increase in TI for Polish and Spanish listeners is carried by more correct identification of both /f/ and /s/ in gate 4. The identification of place of articulation for Dutch and Italian listeners improves in gate 5, and in particular for /s/.

Table V. Percentage of correct identifications of place of articulation for fricative targets in VC syllables split by the two targets.

Gate	Dutch		Italian		Polish		Spanish	
	f	s	f	s	f	s	f	s
1	83	26	58	38	58	38	61	43
2	79	31	65	41	78	38	67	48
3	74	50	78	46	62	44	61	56
4	85	55	78	59	91	82	85	76
5	100	81	100	78	98	82	87	78
6	100	100	100	100	96	100	91	100

Plosives. Figure 5a presents the TI per language group as a function of gate for stop consonants in CV syllables, and Figure 5b displays the same for VC syllables. TI for plosives was analysed in the same way as for fricatives: across consecutive gates, separately for the four languages. The comparisons were evaluated at a corrected α level of 0.01 (0.05 divided by 5) in VC syllables, and of 0.016 in CV syllables (0.05 divided by 3).

In CV syllables all languages show a significant increase in TI only in gate 4, after the presentation of the burst noise. In VC syllables, Dutch and Polish listeners show a significant increase in TI between gate 3 and 4, 4 and 5, and 5 and 6. Thus the amount of information rises gradually for these listeners as a function of the presentation of the vocalic portion preceding the closure, the burst, and the aspiration. Italian and Spanish listeners, however, show a significant improvement only in gate 4 and in gate 6.

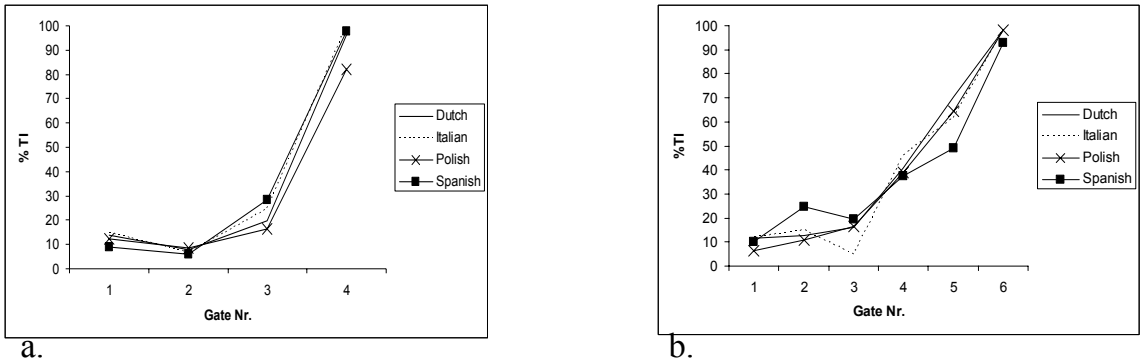


Figure 5: Percentage of transmitted information (TI) for place of articulation as a function of gate per language group for plosive targets in CV (5a) and in VC (5b) syllables.

In summary, the present data show that for fricatives, the first portion of the frication noise contains crucial information about place information. This result is in line with the data from Smits et al.'s (2003) and Warner et al.'s (2005) large-scale diphone gating experiment with Dutch listeners, which showed that for fricatives the first third of the frication contains substantial information about the place of articulation. All listeners extracted information about place of articulation on the basis of the last 20 milliseconds of frication in CV syllables. In VC syllables however, only Spanish and Polish listeners gained this information from 20 milliseconds of frication. Thus, only listeners who learned to differentiate more places of articulation appear to extract information from cues which unfold earlier in the temporal course of VC syllables.

For the voiceless plosives tested in the present study, in CV syllables all listeners gained information about place of articulation from the release burst. In VC syllables all listeners accumulate information about place of articulation when presented with the vocalic portion preceding the consonantal constriction. A difference was observed in the evaluation of the burst between Dutch and Polish listeners on the one hand, and Italian and Spanish listeners on the other. This difference can be

explained by the nature of the materials used. In this study listeners were presented with syllables recorded by a native speaker of Dutch. A Dutch speaker produces plosives with a short aspiration. Polish plosives are also produced with a short aspiration, whereas Spanish and Italian plosives are produced without aspiration. An aspiration following the burst will result in longer VOT (for VOT for the languages tested see Keating, Mikos and Ganong, 1981; Ögüt, Kilic, Engin and Midili, 2006; Rosner, Lopez-Bascuas, Garcia-Alba and Fahey, 2000), in a slower release, and a shorter closure duration (Cho and Ladefoged, 1999). The unfamiliarity with the exact manifestation of the closure and the burst in the plosives produced by a Dutch speaker may thus account for the smaller amount of information transmitted by the closure and the burst for Italian and Spanish listeners.

GENERAL DISCUSSION

The present study was designed to test the following questions: (1) whether there are differences in the temporal uptake of cues between listeners of different native backgrounds; (2) whether language-specific differences in the uptake of acoustic information differ for plosives versus fricatives; (3) and whether such differences are based on the coarticulatory cues from the vocalic portion preceding or following the consonant. Listeners with Dutch, Italian, Polish and Spanish as native languages were presented with truncated portions of CV and VC syllables with fricative and plosive targets. For plosives, a difference in the evaluation of the burst in the materials was found between Spanish and Italian listeners on the one hand, and between Dutch and Polish listeners on the other. For fricative targets an earlier uptake of information about place of articulation was found for Polish and Spanish listeners and a later uptake of information for Dutch and Italian listeners.

First, the language-specific difference in the uptake of information for fricatives confirms the findings by Wagner et al. (2006) with a different experimental paradigm:

Spanish and Polish listeners gain information about place of articulation for fricatives on the basis of information from the neighbouring vowel and a shorter portion of the frication. The present study thus supports the conclusion that listeners whose native fricative inventory contains acoustically similar and therefore perceptually less distinct contrasts pay attention to more cues in order to identify all native contrasts accurately. Moreover, this study specifies that listeners who pay more attention to cues from the adjacent vowels gather information about place of articulation earlier, from the first portion of the frication noise.

The crucial language-specific difference for fricatives was carried not by the formant transitions but by the first portion of the frication in VC syllables. One might argue that listeners of languages with perceptually similar fricatives have refined their attention to cues in the fricative spectrum rather than to coarticulatory cues in the vowel. The phoneme monitoring study by Wagner et al. (2006) however showed that mismatching information from the vocalic portion in cross-spliced VCV syllables impedes fricative identification for listeners with perceptually similar fricative contrasts. Combined, these results imply that Spanish and Polish listeners rely more on the coherence of the vowel portion with the fricative noise.

As summarised in the Introduction the effects of coarticulation for fricatives cannot be described only in terms of formant transitions but reflect rather a perceptual integration of information from the vowel with the frication noise. It is a question for further research which cues exactly were used by Polish and Spanish listeners. It may be, as suggested by Hedrick and Ohde (1993), that a comparison between the amplitude of the vowel relative to the amplitude of the frication played a more important role for Polish and Spanish listeners than for Dutch and Italian listeners.

The amount of information in formant transitions may by itself not provide sufficient information for all place of articulation distinctions for fricatives. A number of studies documented how formant transitions cue place of articulation for plosives (e.g., Liberman et al., 1954; Delattre et al., 1955, Fruchter and Sussman 1997). These studies show that slope and F2 onset of the formant transitions, as captured in the

formulation of locus equations (Sussman et al., 1998), can discriminate labial, alveolar and velar places of articulation for voiced plosives, despite an overlap of these functions for alveolar and velar distinctions. The distinction of places of articulation among fricatives however includes more contrasts. Listeners in the present study have learned to distinguish labio-dental and dental fricatives (Spanish), and four palatal places of articulation for sibilants (Polish). Fricative pairs articulated so close to one another have similar spectral characteristics, and formant transitions may not suffice to disambiguate these pairs. Even for plosives, locus equations cannot separate four coronal places of articulation in the languages Yanyuawa, and Yindjibarndi (Tabain and Butcher, 1999).

The phoneme monitoring study by Wagner et al. (2006) showed that Spanish listeners were affected by mismatching formant transitions for the identification of /f/, while the Polish listeners were affected by the misleading cues for /s/. The present data did not allow for a more fine-grained split into the two fricative categories /f/ and /s/. Together with the results from the phoneme-monitoring study the present results suggest that Spanish and Polish listeners are attentive to the coarticulatory information in the vowel when identifying a fricative, but are misled by these cues only when the spectral information following the vocalic portion cannot override the mismatching cues. The spectrum of the noise does not sufficiently disambiguate /f/ from /θ/ for Spanish listeners, and /s/ from the other palatal sibilants for Polish listeners. Thus, in the temporal course of uptake of information, Polish and Spanish listeners are attentive to the vowel portion preceding the fricative, but if the following information in the noise spectrum can disambiguate the fricative contrast (/f/ for Polish and /s/ for Spanish) mismatching information is overridden, and does not affect phoneme identification.

In line with this, Polish and Spanish listeners were not generally better in identifying fricatives than Dutch and Italian listeners. All listeners used the information in the frication but the noise spectrum may be less informative for Polish

and Spanish listeners, because it does not provide all the information needed to distinguish all native fricatives. Reliance on more coarticulatory cues thus appears to be a compensation for the lack of information individuating all native fricative contrasts in the noise spectrum.

An optimisation of the temporal uptake of cues may be in particular relevant for listeners whose native phoneme inventory demands distinction between several similar places of articulation. Cues to place of articulation are more vulnerable than cues to other phonological features: these cues have been found to be less robust in noisy conditions than the features manner and voicing (Miller and Nicely, 1955), and to be perceived less accurately than the feature manner in normal listening conditions (Warner, Smits, McQueen and Cutler, 2005).

The observed differences in the temporal uptake of cues to place of articulation to fricatives do not necessarily mean that listeners with perceptually distinct fricative contrasts never pay attention to the cues in the vowels preceding fricatives, but it suggests that they are at least selectively inattentive to this systematic acoustic variation in the signal. Just as cue weighting is a function of listeners' experience with informative acoustic variation, which is defined by the distribution of cues for all native contrasts (Holt and Lotto 2006), this inattention to coarticulatory cues for fricative identification, is a learned pattern defined by the native phoneme inventory.

Second, in the present study language-specific differences in the identification of place of articulation for plosives were found in the evaluation of the burst for Spanish and Italian listeners in VC syllables. Italian and Spanish listeners gained less information from the presentation of the vowel portion and the burst without aspiration than Dutch and Polish listeners. The participants in this study were presented with plosives produced with aspiration. Italian and Spanish plosives, however, are produced without aspiration. The articulation of plosives with aspiration may cause a shorter duration of the closure and a faster release. The release bursts in the experimental materials was thus less native-like for Italian and Spanish listeners, than for Dutch and Polish listener.

The language-specific difference for plosives, however, is not based on cross-language differences in the evaluation of coarticulatory cues. All listeners gained the same amount of information from formant transitions. Greater reliance on coarticulatory cues for fricatives appears not to generalize to other phoneme classes, and thus does not reflect language-specific greater sensitivity to coarticulatory cues in general. It is not surprising that listeners differ less in their attention to coarticulatory information for plosives, since the acoustic evidence for plosives is shorter and more abrupt, and offers therefore less space for cross-linguistic differences.

It should be noted, however, that the present results cannot exclude language-specific sensitivity to transitional cues in plosive identification for languages with more distinction in places of articulation: the participants in the present study did not differ with respect to places of articulation within plosives. If sensitivity to transitional cues is a function of listeners' native contrasts within a phoneme class - as appears to be the case for fricatives - the attention to additional cues may also vary for listeners with several perceptually similar plosive contrasts. The results for the plosives and the fricatives together give support to the hypothesis that listeners whose native phoneme inventory contains more distinctions in places of articulation rely on more cues compared to listeners with fewer places of articulation.

A third outcome of the present study is that language-specific differences were found in VC syllables and not in CV syllables. This might reflect a perceptual advantage to 'look ahead' in the acoustic signal in order to get information for less robust phonological features. In CV syllables, all listeners gained the same amount of information. If listeners benefit more from the information which comes in last, as suggested by Mann and Soli (1991), than coarticulatory information in CV syllables may be generally more accessible. Listeners who are in need of more cues may benefit from cues announcing a speech segment, thus from a source of information which can be disregarded by other listeners. Note that among the languages of the world, regressive place assimilation is more common than progressive place assimilation (Jun, 1995).

Most studies examining the effect of vowel-consonant coarticulation focus on the effect of the vowel portion following a plosive. For plosives, the formant transitions following the burst might be perceptually more relevant if listeners integrate the most recent information from formant transitions with the information in the release burst. In VC syllables, the silent interval of the closure preceding the burst might terminate the processing of the cues contained in pre-consonantal formant transitions. This however is not the case for fricatives, and thus cues from the coarticulation of vowels with fricatives can be perceptually integrated with the cues in the noise spectrum. The relevance of pre-consonantal vocalic portions thus may differ across phoneme types.

Formant transitions preceding a fricative abut a frication noise, while formant transitions preceding a stop consonant abut the silent interval of the closure. In the present study with voiceless consonant targets, the presence of either the closure or frication noise was the decisive cue for manner of articulation. Listeners did not extract information about manner of articulation on the basis of the formant transitions.

The language-specific effect for fricatives in VC syllables, combined with no languages differences in evaluation of coarticulatory cues for plosive identification, suggest the following account of language-specific uptake of acoustic information. All listeners gain some information about place of articulation from the vocalic portion in CV syllables. Listeners of languages with spectrally similar fricative pairs optimise their perceptual strategies for fricative identification with earlier attention to the cues announcing a fricative in the vowel preceding a fricative. When these listeners hear a vowel, they start accumulating information about place of articulation of the following consonant. If the vocalic portion is followed by frication noise, listeners with perceptually similar fricative pairs use this earlier information to integrate it with the cues in the frication. Listeners with spectrally distinct fricatives optimise their perceptual strategies by extracting cues to place of articulation of a fricative mostly from its noise spectrum. If the vocalic portion however fades into a silent interval of

the closure, all listeners identify the following segment as a stop, and optimise their strategy by relying on the most telling cues in the release burst.

How does this affect listeners in natural speaking conditions? Generally, higher attention to formant transitions might be the result of a compensatory mechanism, acquired by listeners who need to distinguish perceptually less distinctive contrasts. Hence, in natural native-speech interactions, greater attention to formant transitions will not result in different listening patterns between different native listeners. In adverse situations however, when listening to speech in noise, in particular to a foreign language, (Cutler, Cooke, Garcia Lecumberri and Pasveer, 2007), differences in the attention to cues spread across speech segments might cause different difficulties for listeners of different native backgrounds. Further investigation into the nature and effects of different perceptual strategies for coping with adverse listening situations could shed more light on listeners' subconscious attention to acoustic cues.

This study has documented differences in the temporal uptake of acoustic information for fricative identification among four languages. Despite many differences between the linguistic systems of the languages tested, for instance in syllable structure frequencies, similarities in the processing of vowel plosive and plosive vowel syllables were found. These similarities might reflect some generally preferred perceptual patterns, such as that information about manner of articulation is more accessible in VC syllables, and is not conveyed by coarticulatory cues. Nonetheless, language-specific differences appeared in the temporal uptake of cues to place of articulation for fricatives, showing listeners' perceptual optimisation in the uptake of a vulnerable phonological feature as a function of their phoneme inventories.

Appendix A. Confusion matrix for targets in CV syllables, for the first three gates. Percentages of responses are pooled over listeners for each of the four language groups.

Stimulus		gate 1						gate 2						gate 3					
		Response																	
		k	p	t	f	s		k	p	t	f	s		k	p	t	f	s	
Dutch	k	**21**	36	12	26	5		**21**	29	14	26	10		**36**	21	19	12	12	
	p	36	**24**	17	17	7		17	**24**	19	31	10		10	**64**	10	12	5	
	t	29	31	**2**	33	5		19	31	**14**	24	12		2	48	**40**	5	5	
	f	17	38	5	**33**	7		7	40	14	**19**	19		10	64	7	**14**	5	
	s	24	33	14	24	**5**		17	33	29	17	**5**		12	40	29	12	**7**	
Italian	k	**22**	30	22	26	0		**19**	46	15	19	0		**27**	35	19	12	8	
	p	20	**40**	16	16	8		8	**68**	16	4	4		11	**63**	4	19	4	
	t	15	50	**12**	15	8		23	42	**12**	15	8		15	37	**30**	19	0	
	f	11	41	19	**26**	4		4	56	12	**28**	0		4	70	11	**15**	0	
	s	12	31	19	23	**15**		19	54	0	19	**8**		8	38	12	35	**8**	
Polish	k	**35**	23	9	28	5		**36**	14	19	29	2		**28**	23	23	16	9	
	p	35	**30**	14	14	7		31	**24**	12	31	2		23	**23**	7	44	2	
	t	31	36	**7**	26	0		31	40	**12**	14	2		21	26	**30**	21	2	
	f	33	31	14	**19**	2		26	31	10	**29**	5		14	36	5	**40**	5	
	s	29	24	19	26	**2**		31	14	21	31	**2**		16	21	28	33	**2**	
Spanish	k	**11**	41	28	15	6		**15**	46	15	17	7		**31**	24	30	11	4	
	p	11	**57**	11	15	6		17	**50**	17	15	2		4	**65**	7	22	2	
	t	11	50	**15**	17	7		11	44	**30**	13	2		4	35	**41**	15	6	
	f	17	46	15	**13**	9		9	59	6	**17**	9		9	65	4	**22**	0	
	s	17	50	24	6	**4**		6	50	24	13	**7**		11	22	46	17	**4**	

Appendix B. Confusion matrix for targets in CV syllables, for gates 4 to 6 for fricatives, and gate 4 for stop consonants. Percentages of responses are pooled over listeners for each of the four language groups.

Stimulus		Response: gate 4						gate 5						gate 6					
		k	p	t	f	s		k	p	t	f	s		k	p	t	f	s	
Dutch	k	**100**	0	0	0	0		-	-	-	-	-		-	-	-	-	-	
	p	2	**95**	0	2	0		-	-	-	-	-		-	-	-	-	-	
	t	0	0	**100**	0	0		-	-	-	-	-		-	-	-	-	-	
	f	5	76	14	**5**	0		2	62	14	**21**	0		0	0	0	**100**	0	
	s	12	5	83	0	**0**		2	2	95	0	**0**		0	0	0	0	**100**	
Italian	k	**81**	0	19	0	0		-	-	-	-	-		-	-	-	-	-	
	p	0	**96**	0	4	0		-	-	-	-	-		-	-	-	-	-	
	t	7	0	**93**	0	0		-	-	-	-	-		-	-	-	-	-	
	f	0	77	8	**15**	0		0	44	26	**30**	0		0	0	0	**100**	0	
	s	4	7	89	0	**0**		7	0	89	0	**4**		0	4	0	0	**96**	
Polish	k	**98**	0	2	0	0		-	-	-	-	-		-	-	-	-	-	
	p	0	**76**	13	11	0		-	-	-	-	-		-	-	-	-	-	
	t	2	0	**98**	0	0		-	-	-	-	-		-	-	-	-	-	
	f	0	70	9	**20**	0		2	61	11	**25**	0		0	0	0	**100**	0	
	s	9	0	91	0	**0**		4	0	96	0	**0**		0	0	0	0	**100**	
Spanish	k	**91**	0	9	0	0		-	-	-	-	-		-	-	-	-	-	
	p	0	**87**	0	13	0		-	-	-	-	-		-	-	-	-	-	
	t	0	0	**98**	2	0		-	-	-	-	-		-	-	-	-	-	
	f	2	74	11	**13**	0		4	59	13	**24**	0		0	2	0	**98**	0	
	s	2	0	96	2	**0**		2	2	96	0	**0**		0	0	0	0	**100**	

Appendix C. Confusion matrix for targets in VC syllables, for the first three gates. Percentages of responses are pooled over listeners for each of the four language groups.

Stimulus		Response: gate 1						gate 2						gate 3					
		k	p	t	f	s		k	p	t	f	s		k	p	t	f	s	
Dutch	k	**19**	50	26	2	2		**24**	67	5	2	2		**38**	38	12	10	2	
	p	7	**83**	7	0	2		12	**79**	7	0	2		5	**83**	10	2	0	
	t	10	57	**33**	0	0		2	50	**43**	5	0		2	31	**45**	19	2	
	f	10	74	5	**10**	2		7	67	12	**12**	2		19	31	5	**43**	2	
	s	10	62	26	2	**0**		10	57	31	2	**0**		5	36	43	10	**7**	
Italian	k	**15**	48	37	0	0		**19**	54	15	8	4		**27**	54	8	12	0	
	p	7	**85**	7	0	0		4	**73**	12	12	0		4	**67**	22	0	7	
	t	4	63	**30**	4	0		7	37	**44**	11	0		4	31	**50**	8	8	
	f	19	54	19	**4**	4		19	58	12	**8**	4		15	22	7	**56**	0	
	s	4	46	31	12	**8**		11	41	41	7	**0**		8	42	46	4	**0**	
Polish	k	**18**	53	22	4	2		**20**	51	13	13	2		**29**	58	9	4	0	
	p	16	**69**	11	4	0		11	**76**	7	4	2		9	**78**	7	4	2	
	t	11	60	**24**	2	2		4	49	**38**	7	2		7	27	**51**	16	0	
	f	22	56	20	**2**	0		18	71	2	**7**	2		27	36	9	**27**	2	
	s	13	42	33	7	**4**		18	44	38	0	**0**		7	36	42	13	**2**	
Spanish	k	**28**	52	20	0	0		**31**	44	25	0	0		**31**	33	33	0	2	
	p	22	**59**	15	2	2		4	**85**	9	2	0		7	**80**	11	2	0	
	t	6	54	**37**	2	2		6	44	**48**	2	0		2	37	**52**	7	2	
	f	26	59	13	**2**	0		17	65	15	**2**	2		13	44	22	**17**	4	
	s	19	37	41	2	**2**		2	46	48	4	**0**		4	37	54	4	**2**	

Appendix D. Confusion matrix for targets in VC syllables, for gates 4-6. Percentages of responses are pooled over listeners for each of the four language groups.

Stimulus		Response: gate 4						gate 5						gate 6					
		k	p	t	f	s		k	p	t	f	s		k	p	t	f	s	
Dutch	k	**69**	21	5	5	0		**100**	0	0	0	0		**100**	0	0	0	0	
	p	0	**98**	0	2	0		0	**100**	0	0	0		0	**100**	0	0	0	
	t	2	31	**52**	10	5		0	14	**86**	0	0		0	0	**100**	0	0	
	f	7	5	10	**79**	0		0	0	0	**100**	0		0	0	0	**100**	0	
	s	2	12	23	30	**33**		2	0	10	17	**71**		4	0	0	10	**86**	
Italian	k	**67**	19	7	7	0		**100**	0	0	0	0		**100**	0	0	0	0	
	p	0	**89**	4	7	0		0	**100**	0	0	0		0	**100**	0	0	0	
	t	8	19	**62**	8	4		0	22	**78**	0	0		0	0	**100**	0	0	
	f	4	0	15	**78**	4		0	0	0	**100**	0		0	0	0	**100**	0	
	s	4	4	26	33	**33**		0	0	4	22	**74**		0	0	0	0	**100**	
Polish	k	**62**	27	4	7	0		**100**	3	0	0	3		**100**	0	0	0	0	
	p	2	**89**	2	7	0		0	**97**	0	3	0		0	**100**	0	0	0	
	t	13	18	**56**	11	2		4	18	**78**	0	0		0	0	**98**	0	2	
	f	4	4	2	**87**	2		0	0	0	**98**	2		2	2	0	**93**	2	
	s	2	2	47	13	**36**		0	0	9	18	**73**		0	0	0	0	**100**	
Spanish	k	**72**	19	9	0	0		**100**	3	6	0	0		**100**	0	0	0	0	
	p	2	**91**	6	2	0		6	**92**	3	0	0		2	**96**	2	0	0	
	t	2	30	**59**	7	2		0	26	**74**	0	0		0	0	**100**	0	0	
	f	4	7	9	**78**	2		6	4	7	**83**	0		7	0	2	**91**	0	
	s	4	6	61	15	**15**		4	4	35	15	**43**		0	0	0	0	**100**	

Summary and conclusions

CHAPTER 5

When adults listen to speech they automatically apply listening strategies which are adapted to their native language. As a consequence, adults from different native backgrounds never perceive speech in the same way. This means that learning foreign languages becomes cumbersome, but it also means that the perception of native speech is highly optimised. Listeners hearing spoken utterances are confronted with an abundant amount of acoustic information. To structure such an information overflow, listeners rely on acoustic patterns that are meaningful in their native language. To identify speech sounds listeners need to extract information which distinguishes a sound from other sounds. However, for a given speech sound, which acoustic information distinguishes it from other sounds? This depends on what other sounds are in the language. Therefore, listeners with different phoneme inventories differ in the acoustic information which makes distinctions for them.

This thesis queried whether listeners of different native languages vary in their 'choices' of acoustic patterns used in identifying speech sounds when they listen to the same spoken utterances. For instance, all listeners hearing the exclamation /psssttt/ will recognise /s/ in the occurrence. Implicitly however, listeners from different native backgrounds will have to discriminate the /s/ from different sets of similar sounds. Throughout this dissertation all listeners were presented with the same natural non-word utterances. The targets for identification were speech sounds which are contrastive in all the languages. Effects of the phoneme inventory were investigated by comparing listeners with different sets of native sounds. This chapter starts with a summary of the results and proceeds with a discussion of some implications of the results.

SUMMARY

The experiments in Chapter 2 investigated whether listeners with different numbers of categories within a phoneme class (fricative, stop consonant, and vowel) differ in the way they identify members of the class. For example, Spanish listeners have only five vowels, while English listeners distinguish about 20; Dutch distinguishes six fricatives, while Polish distinguishes eleven. Does the number of similar sound categories affect the speed and accuracy of identification? Previous studies comparing the identification speed and accuracy between phoneme classes mostly tested only one listener group, usually English speakers (e.g., Foss & Swinney, 1973; Healy & Repp, 1982). The attested differences were then attributed to different phonological functions of vowels versus consonants, and to the acoustic properties which distinguish phoneme classes. These studies consistently report that stop consonants are identified fastest, while vowels are the most difficult targets. In Chapter 2 of the current work, listeners with different numbers of categories for phoneme classes were compared to investigate whether differences in the processing of vowels versus fricatives or stop consonants are the same for all listeners; or whether the number of similar sounds in the phoneme inventory modulates identification speed and accuracy.

The results in Chapter 2 support, to a certain degree, differences between phoneme classes. In general, more errors occurred for vowels, and fricatives were identified more slowly. However, this study also showed that the effect of phoneme class depends greatly on how reaction times are measured. Stop consonants are identified fastest if the reaction times are measured from their release burst, as was customary in previous studies. When the reaction is measured from the closure onset, however, fricatives are identified faster than stops. It is not a clear-cut task to decide where a phoneme starts in spoken utterances. In addition, phoneme classes may differ strongly in the coarticulatory cues which announce them. It is thus unclear where listeners start accumulating information about an upcoming segment.

Furthermore, the experiments in Chapter 2 showed that the number of categories in the phoneme class creates language-specific rankings in speed of phoneme identification. Having more vowels in the native language slows down listeners' identification and decreases their accuracy. Just three additional vowels, as for Catalan listeners compared to Castilian Spanish listeners, can change how easily vowels are identified relative to plosives and fricatives. Additional speech sound categories thus appear to compete with the target for identification.

In other words, the comparison among five languages showed that there is no clear ranking in the identification speed and accuracy among phoneme classes that could be explained just by the acoustic properties of these phonemes. Chapter 2 established that the number of native sound categories creates differences in their processing. Chapter 3 further investigated whether the presence of perceptually similar categories leads to differences in the selection of acoustic cues, and narrowed the investigation to the perception of fricatives.

In Chapter 3, Dutch, English, German, Polish, and Spanish listeners identified the two fricative targets /f/ and /s/. The fricative repertoires of these languages differ in the number of categories which are perceptually similar to the targets. Dutch and German listeners may be able to identify these fricatives just on the basis of primary cues, which lie in the energy distribution in the frication noise, since they only have acoustically distinct fricatives. English and Spanish listeners distinguish between the labio-dental /f/ and the dental /θ/, which are actually very similar. Polish listeners have four sibilants at palatal places of articulation; hence, for Polish listeners the perceptual saliency of /s/ could be reduced. The hypothesis tested was whether listeners with similar fricatives rely more on coarticulatory information in adjacent vowels. The effect of coarticulatory information was tested by presenting listeners with materials containing either coherent or mismatching cues in the vowels surrounding the fricatives.

The results showed language-specific differences in listeners' reliance on coarticulatory information. Mismatching cues in the vowels hindered fricative

identification for English, Polish and Spanish listeners, but not for Dutch and German listeners. In addition, mismatching cues hampered the identification of /f/ for English and Spanish listeners, while Polish listeners were mostly affected in their identification of /s/. An additional experiment tested whether listeners perceived the acoustic mismatch in the materials. Both Spanish and Dutch listeners perceived acoustic mismatches in the cross-spliced items. When identifying fricatives, however, only the Spanish listeners were misled into different fricative categories; the Dutch were not.

Taken together, these results show that perceptually similar fricatives in listeners' phoneme inventories encourage attention to more subtle cues like formant transitions. Furthermore, these results suggest that attention to additional coarticulatory cues is restricted to those fricatives that have similar competitors. Also, the attention to formant transitions is not restricted to fricatives with acoustically weak features like /f/ and /θ/. Polish listeners rely on transitional information even when they identify the acoustically salient /s/. It thus appears that not the acoustic features of a fricative, but listeners' knowledge about similar sounds, guides listeners' reliance on coarticulatory cues. These results also suggest that Dutch and German listeners disregard systematic acoustic information which is used by Spanish, English and Polish listeners.

Chapter 4 examined whether such language-specific differences affect the timing of uptake of information. Specifically, three questions were addressed (1) Are there cross-language differences in the temporal uptake of cues to place of articulation? (2) Does language-specific reliance on coarticulatory cues generalise also to stop consonants? (3) Do listeners who rely on coarticulatory cues extract information from these cues earlier or later as the signal unfolds? In a gating study, Dutch and Italian listeners, whose fricatives are spectrally distinct, were compared with Polish and Spanish listeners, whose phoneme inventory contains perceptually confusable fricatives. The targets /k p t f s/ were identified from truncated vowel-consonant and consonant-vowel syllables.

The results revealed that, compared to Dutch and Italian listeners, Polish and Spanish listeners extract information specifying a fricative's place of articulation earlier in the utterance, from shorter portions of vowel-fricative syllables. No language-specific differences were found for the identification of stops. Hence, higher sensitivity to coarticulatory information does not generalise to other phoneme classes. In the stop consonant case, however, the listener groups did not differ in the number of categories. Together these results support the hypothesis that listeners optimise their extraction of cues only when their phoneme inventory demands finer distinctions between similar contrasts. In addition, listeners with perceptually similar fricatives attend to additional sources of information as soon as these are available in the signal. Listeners without perceptually similar fricatives extract information about a fricative's articulation place later in time, because their categories can be accurately distinguished on the basis of cues which occur later in the signal. Optimised reliance on cues thus means that listeners select the most telling cues, and those cues which are necessary given the native phoneme repertoire.

CONCLUSIONS

The results show that listeners from different native backgrounds apprehend speech in different ways. Throughout this dissertation listeners identified the targets /a i u f s p t k/. These speech sounds have a favoured status across phoneme inventories, which suggest that their acoustic properties make them perceptually robust. In fact, all listeners were able to identify these targets, even though most of them heard non-native realisations of them, produced by a native speaker of some other language. The number of mental representations which can compete with the target was manipulated by comparing listeners with different sets of speech sounds. The experiments simulated the situation in which listeners are confronted with a foreign language. They

hear new words, which are non-words to their ears, and some of the sounds are quite similar to sounds in their own language.

The approach followed throughout this thesis was thus comparable to studies investigating colour perception across languages. Studies in colour perception (Berlin & Kay, 1969) showed that despite differences in how languages name shades of colours, the basic colours black, white, and red, green and blue are identified most quickly. Language-specific differences in the number of category names for shades of colours, however, create perceptual differences between speakers of different languages. Russian speakers, for instance, who have a distinction between light blue and dark blue in their language, perceive these two shades of blue more categorically than English speakers (Winawer, Witthoft, Frank, Wu, Wade & Boroditsky, 2007).

The observed differences in speech sound perception can thus be attributed to different divisions into categories in listeners' perceptual spaces. Language-specific differences were found at a low phonetic level that is not consciously accessible for listeners. The comparison among listeners of various native backgrounds reveals interesting insights into how humans process the speech signal. These issues will be discussed in the following sections.

How detailed is language specific listening?

Language-specific listening has been documented at various levels of speech perception. For instance, listeners apply language-specific strategies to find beginnings and ends of words. Such segmentation of speech is partly based on the metrical structure of one's native language, but the rhythm of speech is defined by language-specific units. Dutch and English listeners rely mostly on word-initial stress (Cutler & Norris, 1988; Vroomen, van Zon & de Gelder, 1996), French and Spanish listeners rely on the boundaries of syllables (Cutler, Mehler, Norris & Sequi, 1986), and Japanese listeners use the smaller unit of a mora instead of syllables (Otake, Hatano, Cutler & Mehler, 1993). In recognizing words, listeners apply their knowledge about

which native speech sounds can co-occur, and how they assimilate to one another. English listeners, for instance, know that the sound combination /sl/, as in *slight,* can occur within a syllable while /pf/ cannot. This is reversed for German listeners, for whom /pf/ like in *Pferd* can occur within a syllable whereas /sl/ cannot. Syllable boundaries or individual speech sounds are easier to detect when they violate language-specific rules (Weber, 2002). The above mentioned examples show that what violates these rules is different for English and German listeners.

To a certain degree, listeners are aware of such regularities at the metrical, phonological or phonotactic levels of their language. Even though such language-specific strategies operate completely automatically, listeners have conscious access to such knowledge. English listeners, for instance, notice that French differs in rhythm from English. People can also assign combinations of sounds to foreign languages (e.g., Stockmal & Bond, 2002); most people would assign the combination of sounds in *schlimazel* to Yiddish, even though they do not speak this language. Listeners can also be aware of differences between native and non-native speech sounds; German listeners, for instance, know that there is a difference between the English words *think* and *sink*. As is known, German listeners have difficulties applying this knowledge. This, however, does not mean that words differing solely in these two sounds do not have different lexical representations (Cutler, Weber & Otake, 2006; Weber & Cutler, 2004). Sometimes, people can describe differences in the sound of native and foreign speech segments, and some people are very good in imitating foreign accents, because they notice differences between the sound systems of languages. The results presented in this dissertation, however, document patterns of language-specific listening which occur at a much lower level of processing. These are language-specific patterns of integration of static and transitional cues, and of extraction of information from temporally separated speech segments.

The data presented in Chapter 3 show that listeners differ in which acoustic cues they integrate to obtain the percept of a speech sound. When Spanish and Dutch listeners hear the nonsense utterance *depufa*, they differ in their "choices" of sources

of information. Dutch listeners recognise /f/ on the basis of the information in the frication noise, while Spanish listeners unconsciously grasp also the information conveyed through the adjacent vowels. Spanish listeners have learned that information in the vowel is a relevant cue to disambiguate /f/ from other native speech sounds. They are attentive to this information, even though relying just on the frication noise might be more beneficial when the cues in the vowel and frication conflict, as in the materials in Chapter 3. Spanish, English, and Polish listeners appear not to be able to ignore the mismatch, because cues in the vowels are part and parcel of their native fricative categories. Dutch and German listeners, on the other hand, disregard the mismatch, and rely on the cues in the frication because this information can reliably disambiguate all of their native fricatives.

When asked to recognise a speech sound, listeners integrate cues scattered across the utterance. How listeners find, extract and integrate acoustic information has always been a major question in speech research. The current results show that information integration entails language-specific patterns. This suggests that there might be no universal way in which listeners master this task. Interestingly, no language–specific selection of cues was observed for stop consonant identification. Hence, we can assume that listeners apply the same strategy when their phoneme inventories do not demand reliance on different cues. It thus appears that language-specific strategies are motivated by the demands, to maintain perceptual distinctiveness and processing economy.

Language-specific attention to acoustic cues has as a consequence that listeners grasp different aspects of the acoustic signal at different points in the time course of the signal. The results in Chapter 4 show that Polish and Spanish listeners, who were shown in Chapter 3 to rely on coarticulatory information, extract information about the place of articulation of /f/ in *depufa* earlier than Dutch or Italian listeners. Dutch and German listeners extract information about place of articulation of a fricative later in time. Hence, when the signal unfolds over time listeners have different perceptual

images of sounds. This shows differences in speech sound perception at a very low level of processing.

Interestingly, the results in Chapter 4 showed that Spanish and Polish listeners did extract information earlier, but not from the whole vowel. Rather, they achieved this on the basis of a short portion of the vowel-fricative syllable. One might argue that Spanish and Polish listeners have a more refined sensitivity to the information contained in the noise. The results in Chapter 3, however, showed that conflicting cues in the vowel hampered fricative identification for exactly these listeners. Together these results suggest that the differences among listeners stem from language-specific attention to the coherence between fricatives and vowels. The coherence between the vowel and fricative appears to be informative for Spanish and Polish listeners, but German and Dutch listeners are apt to disregard this source of information.

Does this imply that Dutch and German listeners are 'deaf' to the mismatch in the materials in Chapter 3? Experiment V in Chapter 3 shows that Dutch listeners, like Spanish listeners, can hear the cross-splicing manipulation in the signal, when they are asked to judge the goodness of the materials. Thus, listeners, who are able to ignore the mismatch are also able to perceive it. In other words, whereas both listeners groups are able to hear the manipulation, only listeners with perceptually confusable fricatives (English, Polish, and Spanish) were misled by this mismatch in identification. Language experience appears to determine which information is evaluated in fricative identification, but appears not to affect listeners' auditory perception of the signal.

A possible explanation for the discrepancy between auditory perception and speech sound identification is that listeners can apply different listening strategies when judging the goodness of a syllable compared to detecting a speech sound in spoken utterances. Repp (1981) showed that listeners can switch between a phonetic and an auditory mode of perception. The auditory mode entails a more detailed perception of the acoustic shape of a sound. Both Dutch and Spanish listeners can hear the acoustic mismatch when their attention is directed to the goodness of the signal. Such a detailed mode of listening, however, might not be desirable in normal listening

conditions, when the integration of acoustic cues occurs quickly, and is an unnoticed side effect of extracting meaning from words.

Whether listeners apply an auditory or a phonetic mode of listening may depend on the task. The attunement to a native language does not appear to have consequences at a low auditory level. The "choice" to attend to or ignore the mismatching information appears to play a role during the automatic detection of speech sounds in spoken utterances. Listeners whose fricative categories are sufficiently characterized by the static cues filter out such mismatch when listening to words. Attention to transitional cues is encouraged by the availability of more fricative categories. Where there is no alternative category – as in the case of Dutch – mismatching information in formant transitions may be treated as just allophonic variation.

There is, however, evidence for language-specific processing of fricatives in the auditory cortex. A study by Lipski (2006) examined the neural responses of German and Polish participants who were listening to Polish and German syllables with palatal fricatives. A behavioral discrimination test showed that German listeners were able to distinguish the non-native speech sounds /ʃ/ and /ɕ/. The neural responses, however, suggested that the behavioral discrimination relied on different neural processes. Sound categories which are native for both listener groups yielded very similar neural responses for both groups. The Polish contrast /ʃ/ versus /ɕ/, however, showed differences in the lateralization of the neural response. Lipski attributes a left hemispheric dominance for the Polish listeners to their stronger reliance on transitional information.

This suggests that language-specific integration of transitional and static cues may shape listeners' perception at very early stages of auditory processing. Different neuronal processing for native and non-native speech sounds has been reported in several studies (e.g., Näätänen, et al., 1997). Such differences have been attributed to the presence or absence of long-term memory representations of speech sounds. The study by Lipski suggests not only language-specific neural representations of speech

sounds but also different processing mechanisms taking place during the integration of static and transitional information. This implies that the effects reported in this dissertation might stem from different processing mechanisms at a very early level of auditory processing.

Interestingly, such differences do not permit predictions about listeners' behavioral discrimination capacity. Processing in predominantly the left hemisphere is ascribed to differences in the temporal processing (Belin, 1998). The left hemisphere appears to serve finer temporal resolution. Acoustic sensitivity appears not to be altered by the native language; rather listeners apply a more economic strategy in evaluating information during the highly automatic categorization of speech sounds. Further research should address the question of whether listeners' attention can be directed to coarticulatory cues, to further investigate which levels of perception are shaped by the exposure to a first language.

Throughout this dissertation, all listeners were presented with the same materials. All listeners were instructed, in their native language, to identify native sound categories. Of course, speech sound categories, even if similar, are never the same for listeners of different backgrounds. The exact acoustic realisation of individual sounds differs between languages. Language-specific realisation of speech sounds may shape the way listeners perceive speech. Throughout this thesis, two of the listener groups were always presented with native and non-native realisations of the materials. This was done to exclude the possibility that listeners can only apply their native listening strategies when presented with native realisation of phonemes. Interestingly, no differences were found in the way listeners perceived the materials produced by a native or by a non-native speaker. Hence, familiarity with the exact realisation of sounds does not affect the way listeners process the new words, similar to the materials in this thesis.

It is an open question whether such differences in the spectral and temporal integration of cues affect listeners during word recognition. The materials used in the presented experiments were nonsense words. Future research should address the

question of whether such language-specific differences also surface when listeners access the meaning of words. Differences in the selective attention to cues can have two effects. On the one hand, listeners who rely on coarticulatory cues can extract features of sounds earlier. Therefore, they might also be able to disambiguate words earlier. On the other hand, the reliance on coherence between vowels and fricatives suggests that the information in the frication noise might not be enough for these listeners. Incoherence between these two cues, or distortion of one of the sources of information might be more harmful for these listeners. Future research should thus also address the question of how reliance on different cues affects listeners' perception in adverse listening conditions.

The role of the phoneme inventory

The phoneme inventory, and the concept of a phoneme itself too, is a theoretical construct, useful as a tool in speech research. Listeners do not necessarily have conscious awareness of phonemes or phoneme inventories. The concept of a phoneme stands for a mental representation of speech sound categories, and thus for the way in which listeners group acoustic events into same, similar and different categories. These categories are established early in the speech development, and they form the native perceptual space. Such categories persistently affect speech perception, as is most obvious in the difficulty of distinguishing non-native speech sounds. The phoneme inventory stands for the entire set of categories. This concept thus includes the boundaries in listeners' perceptual space, and establishes which speech sounds are perceived as similar. Listeners' perceptual space is formed by phonetic distinctions which allow them to recognize and distinguish native words.

When listeners hear spoken utterances their goal is to understand the intended message of an utterance. When extracting the meaning of words listeners may identify individual speech sounds. The debate about the function and reality of the phoneme has a long history (e.g., Studdert-Kennedy, 1976; Lotto & Holt, 2000). Models of

speech perception vary in the role they ascribe to phonemes in speech processing (e.g., Johnson, 2004; McClelland & Elman, 1986; Norris, McQueen & Cutler, 2000). There is, however, no doubt that listeners are able to assign acoustic signals to speech sounds, and the present dissertation documents effects which can be explained by the presence of such mental categories.

Chapter 2 documents that the number of similar categories in a listener's phoneme inventory has an affect on how quickly and accurately listeners identify sounds. This suggests that listeners consider similar native speech sounds when asked to identify a target. The presence of similar sounds increases the identification time and decreases identification accuracy. At first glance, it appears to be an inefficient mechanism. Why should listeners be aware of similar perceptual entities when detecting only one? This effect, however, may stem from general properties of the perception system. A higher number of choices slows down human perception. Listeners appear to identify objects and events on the basis of similarities. Similarities among speech sounds are formed by the native phoneme inventory. More similar speech sounds cause that more features are shared among sounds, and fewer features discriminate speech sounds. This increases the joint probability of making an incorrect identification. Similar effects have been found in the visual perception. The set-size effect in visual search is an example of how human perception is affected by the presence of more and similar alternative choices.

In addition, listeners are clearly aware of which sounds are similar. People make puns and spoonerisms. Such word games would not be funny if listeners could not use their knowledge about speech sounds. People enjoy phrases like "she sells seashells by the sea shore" which build on the difficulty to keep similar sounds apart. The essence of tongue-twisters is not based on their meaning, but on the sound of their segments. People can enjoy the challenge of maintaining distinctions between similar sounds even in foreign languages. The Catalan proverb "Plou poc, però per lo poc que

plou, plou prou[3]" is appreciated also by listeners of other mother tongues. Furthermore, when linguistically-naive persons, like children or very introductory linguistics students are asked to group consonant sounds according to their similarity, they apply strategies similar to the division into phoneme classes established by linguists (Warner, personal communication). Linguistically naive listeners quickly start grouping fricatives together. Even though the instant activation of phonemes in speech perception is an object of debate, it appears unquestionable that listeners can make use of their abstract knowledge about speech sound categories.

Chapter 3 and 4 document patterns of reliance on acoustic cues. Listeners' perception appears to be shaped to maintain perceptual contrasts between speech sound categories. Similar speech sound categories create a higher density in the native perceptual space. Listeners learn to selectively rely on additional cues to enhance the reduced distinctiveness in a more crowded perceptual area. Perceptually similar categories are not necessarily phonemes sharing the manner of articulation (phoneme class), but may also share other acoustic features. For instance, the voiced bilabial stop consonant /b/ may be perceptually similar to the voiced bilabial fricative /v/. In addition, listeners can also distinguish between allophones of the same phoneme (e.g., between the palatal and the uvular fricative in German, see Lipski, 2006). This suggests that it might be possible to predict the relevance of cues to a certain degree. But the phoneme inventory does not necessarily tell us which speech sounds can be distinguished by listeners.

Selection of appropriate cues appears to be part of the perceptual learning which takes place when infants acquire a language. Once sound categories are established, they have consequences on the perception of similarity between speech sounds. The Native Language Magnet theory (Kuhl, 1991) states that a prototypical representation of a native sound category acts as a magnet, and prevents the perception of acoustic variation within a category. The perceptual space around a prototypical

[3] It does not rain a lot, but considering how little it rains, it rains enough

representation of a sound category is assumed to be warped in such a way that listeners only perceive those acoustic differences which lead to a distinct category. In this way listeners are able to cope with the immense variability in the exact acoustic realisation of speech sounds induced through varying speaker characteristics or speech sound contexts.

Perceptual learning can take place also later in life. Perceptual learning studies show that listeners can rapidly adjust to speaker-specific phoneme realisations (Eisner & McQueen, 2005; Norris, McQueen & Cutler, 2003), and that such adjustments can spread to other instances of these phonemes in new words (McQueen, Cutler & Norris, 2006). Such plasticity, however, is restricted to sound categories and to speakers (Eisner & McQueen, 2005). Moreover, it appears crucial that such perceptual learning is lexically mediated. When listeners know which sound category is to be adjusted for a given speaker they are able to accept quite some acoustic variability as an instance of a native category. The internal structure of phoneme categories, and the selection of cues to these categories, however, appear to be defined early through exposure to a language. It is an open question, which should be addressed by future research, whether the reliance on cues can also be altered through appropriate perceptual learning techniques.

Optimal processing of speech

The present results show that listeners process speech in a way that is optimally adapted to their native language. Optimal processing combines accuracy with economy. In the current experiments, only listeners with perceptually confusable fricatives relied on coarticulatory information (Chapter 3). Only listeners with subtle distinctions between places of articulation of fricatives extracted cues to place of articulation earlier (Chapter 4). The reliance on coarticulatory cues was restricted to those fricative contrasts which are perceptually confusable (Chapter 3). Reliance on coarticulatory cues did not generalize to the perception of other phoneme classes, like

stop consonants (Chapter 4). The selection of cues thus appears to be guided by the demand for accurate identification and processing economy at the same time.

Listeners, however, do not appear to be able to adapt their language-specific ways of selecting cues to the requirements of the situation, or to the experimental situation. The stimulus set in the experiments throughout this dissertation did not contain the perceptually confusable fricative /θ/. That is, a direct distinction between the confusable fricatives /f/ and /θ/ was not necessary for efficient performance within the experimental situation. Nonetheless, the Spanish and English listeners were substantially misled by incorrect formant transitions for /f/. Similarly, the Polish listeners were misled by incorrect formant transitions for /s/, even though the palatal fricatives, which in Polish might be confused with /s/, were not present in the experiment. This suggests that reliance on cues is economically targeted to the categories which are confusable, but creates an automatic pattern of cue selection, which can not easily be adapted to specific listening situations.

In general, listeners may be able to make use of all acoustic cues which are in the signal, as argued by Diehl and Kluender (1987). The automatic reliance on acoustic cues, however, shows patterns which are shared among listeners of a language, and differ between listeners with different phoneme inventories. Also, listeners can selectively disregard acoustic information in the signal. Formant transitions and coarticulatory information are inherent to the signal. The mere presence of this systematic acoustic variation, however, does not mean that all listeners exploit this information perceptually. The Orderly Output Constraint (Sussman, Fruchter, Hilbert & Sirosh, 1998) states that ordered systematic variation and a linear relationship between acoustic elements could enhance perceptual processing. In this approach, linearity and systematic variability in the acoustic signal are seen as a consequence of the evolution of the auditory and neural processing mechanisms. This view thus advocates that systematic acoustic variation might have a “universal” perceptual impact. The present results suggest that listeners can disregard systematic acoustic variation when other cues suffice to distinguish all native contrasts. It appears

that for optimal perception listeners extract exactly those cues from the signal that allow just a good-enough percept.

Language-specific cue weighting appears a logical consequence of the general perceptual mechanisms to reduce an overflow of information. There are multiple acoustic cues in spoken utterances, and listeners reduce the information by relying on the most telling ones. Several studies documented language-specific weighting of cues (e.g., Bradlow, 1995; Fox, Flege & Munro, 1995, Gottfried & Beddor, 1988). The present results contribute to this line of research by showing that the make-up of the native phoneme inventory plays a role in which cues are selected by listeners.

A shift in the weighting of acoustic dimensions appears to occur also during speech development. Infants appear to initially give more weight to transitional information (Nittrouer & Miller,1997a). The Developmental Weighting Shift Theory (Nittrouer, 2000) proposes that children rely more on transitional cues because they process speech in bigger units than adult listeners. Transitional cues establish coherence within syllables, and contribute to bigger units of perception. Later in speech development children deduce which cues offer the highest informativeness in their native language, and optimise their listening strategies. Further research should address the question of the age at which infants start adapting their listening strategies. Most likely, the attuning to an optimal cue extraction will not occur in parallel to different speech sounds. A related question is thus whether such a shift is triggered by a certain step in development.

Universals in the processing of speech sounds

The goal of psycholinguistic research is to understand how listeners process speech. Infants can become native in whatever language surrounds them. This strongly suggests that the processing of speech is based on the same mechanisms for all listeners. Comparative research across languages, however, has revealed a great many cross-linguistic differences. This suggests that there might be no universal ways in

which listeners process speech. However, all listeners may use similar types of processing mechanisms. An example for this is the use of prosodic units in speech segmentation summarised above. Even though French listeners rely on syllable boundaries, whereas Japanese listeners rely on mora boundaries, both rely on the rhythm of speech.

Which of the results presented in this dissertation can be seen as universal strategies? First, all listeners are affected by the number of phonemic categories when identifying a speech sound. Language-specific patterns in speech sound identification are based on having different numbers of categories. Nonetheless, the effect of one additional sound category appears to be the same for all listeners. A higher number of categories implies a higher number of choices, which generally impede the process of decision making (e.g., Medin, Goldstone & Markman, 1995; Nosofsky 1997). The effect of the number of categories thus appears to be similar across perceptual modalities, and appears to affect people in the same way for vision (set-size effect) and speech sound perception.

A second universal observation is that listeners optimize their listening strategies to the demands of their native phoneme inventory, which creates similar patterns among languages. In this dissertation similar patterns of cue selection were revealed for Spanish, English and Polish listeners on the one hand, and Dutch and German listeners on the other. Even though these languages differ widely, the strategies used by the listeners are similar. Another aspect of such perceptual refinement was found in Chapter 4. Polish and Spanish listeners optimized their perception of place of articulation by extracting relevant cues earlier, by attending to the coherence between the vowel and the fricative. Dutch, Italian, and German listeners also optimized their listening strategies by "choosing" cues in a more economical way. An efficient use of cues thus appears to be a general motivation in speech perception.

BROADER IMPLICATIONS

The results presented in this thesis also relate to broader topics in communication, such as computer processing of speech and second language learning. The goal of human communication is to extract meaning of spoken utterances. This is also the goal of automatic speech recognition (ASR). Both ASR and the psychoacoustic approach of speech perception assume that listeners or computers extract patterns from the acoustic signal. ASR is based on statistical learning by general algorithms to extract patterns in speech. This approach could thus mimic the way infants presumably acquire their native language. ASR algorithms, however, are based on pre-specified units, whereas infants deduce and adjust their perception units through unsupervised learning. Despite similarities to human speech recognition, ASR systems are not able to compete with human performance in speech recognition.

Current ASR systems do not model language-specific selection and weighting of cues. What distinguishes human speech recognition and automatic speech recognition is that listeners not only extract meaningful patterns in speech, they also weigh acoustic information in an economical way. Such an approach to speech perception, with optimally shaped language-specific strategies, is unlikely to be emulated by automatic speech recognisers. The improving of ASR systems is based on exposing the algorithms to more training. Improving automatic speech recognition to approach an error rate of 0% would require approximately 10.000.000 hours of speech training data (Moore, 2001). Infants have a much more efficient way to find meaningful acoustic patterns. If it were established how this is possible it would open doors to cross-fertilisation between the disciplines of automatic and human speech recognition.

In human communication the recognition of words plays the major role, but humans are also able to identify individual speech sounds, for example, in foreign or nonsense words. This gives people the option to learn new combinations of speech

sounds, new words, and new languages. It should be mentioned though, that even human listeners rarely obtain an error rate of 0% in speech recognition. It seems paradoxical, but listeners' implicit knowledge can get in their way of speech sound recognition when compared to ASR systems. Automatic speech recognition systems have been shown to outperform human recognition scores on plosives and non-sibilant fricatives (Cooke, 2006). Plosives are very context-dependent speech sounds. Several acoustic cues contribute to their identity. It is likely that an ASR system can rely on all these sources at the same time, while listeners will show language-specific reliance on fewer cues. For fricatives, an ASR system can outperform a human recogniser since the system does not have the implicit knowledge about how /f/ can easily be confused with /θ/.

In addition, research in human speech recognition acknowledges the presence of language-specific patterns of perception. Studies on human speech recognition avoid the use of materials with non-native realisation of words, because listeners might apply different listening strategies to native and non-native materials. The current results suggest that at the level of speech sound recognition, listeners appear not to be affected by familiarity with the realization of sounds. Whereas ASR systems feed global algorithms with more data, studies on human speech recognition restrict their conclusions to listeners of languages they tested. More systematic cross-language comparisons may allow discerning patterns which generally guide human processing of speech, and this may be beneficial for ASR systems.

Turning to second language learning, ASR-based systems are also used to train the pronunciation of foreign languages. Speech production is affected by speech perception. Adult listeners who are not able to detect differences between native and foreign sounds will have difficulties producing these sounds. Therefore, many people want to improve their perception of foreign speech sounds. ASR-based systems record the utterances of a learner, compare these to a native-like pronunciation by means of pre-specified similarity parameters, and give feedback to the listener. Pure feedback,

however, might be inefficient because it disregards that listeners' from different backgrounds vary in what they apprehend in speech.

Various training methods have been developed to improve listener attention to disregarded cues. Such techniques include High Variability Phonetic Training (Logan, Lively & Pisoni, 1991), in which listeners are presented with natural recording of multiple native talkers in multiple contexts. Other techniques aim at the selective enhancement of characteristic of the speech sound, with the goal of drawing listeners' attention to perceptually difficult features (e.g., Iverson et al. 2003). A comparison among such auditory training methods (Iverson, Hazan & Bannister, 2005) suggests that selective enhancement of acoustic cues does not obtain better results than training with natural stimuli. One wonders how many hours of auditory training with natural stimuli would improve listeners perception, in analogy to ASR systems.

The results in this dissertation suggest that there may be no training method that is adequate for listeners of all different native backgrounds. Listeners have their optimised strategies for listening to speech. How these strategies differ is still to be discovered. Listeners, however, do not appear to have diminished capacities to perceive differences between speech sounds when they listen to them in an auditory mode. Listeners might be able to perceive differences between speech sounds when they do not assign a function to them. Targeted auditory training, which promotes direct discrimination between native and non-native speech sounds, appears to be a promising though time-intensive method. Optimally, training techniques would be able to take advantage of listeners' sensitivity to acoustic differences when sounds are not treated as functional speech units. Possibly, such a method would, on a long term, direct listeners' attention to those acoustic features which most efficiently individuate non-native speech sounds.

At first glance, the results presented in this dissertation offer a picture for adult second language learners that is not excessively encouraging. Native listening strategies obviously hinder the learning of foreign languages, and even worse, they create differences in the perception of one and the same speech signal. There is,

however, a bright side to language-specific listening, which should not be overlooked. If the human perceptual system was not able to ‘choose’ the meaningful and necessary cues, listeners would be flooded with an unthinkable amount of information. Communication would not be possible because listeners would be distracted by the subtle differences between all the sounds which surround them. The optimization of speech sound perception allows us to focus on other aspects of speech besides its detailed acoustic manifestation.

REFERENCES

Anderson, J., Morgan, J., & White, K. (2003). A Statistical Basis for Speech Sound Discrimination. *Language & Speech, 46*, 155-182.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology 44*, 310-329.

Baayen, H. (in press). Analyzing Linguistic Data - A practical introduction to statistics. Cambridge: Cambridge University Press.

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium. Univ. of Pennsylvania.

Bates, E., Devescovi, A., & Wulfeck, B. (2001). Psycholinguistics: A cross-language perspective. *Annual Review of Psychology, 52,* 369-398.

Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: evidence for partial compensation. *Journal of the Acoustical Society of America 106*, 2868-2887.

Berlin, B., & Kay, P. (1969). Basic color terms: Their universality and evolution. Berkeley: University of California Press.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Goodman, J. C., & Nusbaum, H.C. [Eds.] *The development of speech perception: The transition from speech sounds to words* (289-304). Cambridge MA: MIT Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & Performance, 14,* 345-360.

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constrains on statistical computations. *Psychological Science, 16,* 451-459.

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

Borzone de Manrique, A. M. & Massone, M. I. (1981). Acoustic analysis and perception of Spanish fricative consonants. *Journal of the Acoustical Society of America 69*, 1145-1153.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America, 97*, 1916-1924.

Bradlow, A. R. (1996). A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish: Universal and language-specific aspects. *Phonetica, 53,* 55-85.

Bradlow, A. R. (2002). Confluent talker- and listener-oriented forces in clear speech production. In Gussenhoven, C., & Warner, N. [Eds.] *Laboratory Phonology 7* (237-274). Berlin/New York: Mouton de Gruyter.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America,101*, 2299-2310.

Bradlow, A. R., Yamada, R. A., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977-985.

Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America, 117*, 3890-3901.

Broersma, M., & Cutler, A. (in press). Phantom word activation in L2. *System*.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403,* 428-430.

Carbonell, J. F., & Llisterri, J. (1992). Illustrations of the IPA: Catalan. *Journal of the International Phonetic Association 22,* 53-56.

Chatterjee, S., Hadi, A. S., & Price, B. (2000). Regression analysis by example. New York: John Wiley & Sons.

Cho. T., and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27,* 207-229.

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

Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America, 119*, 1562-1573.

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

Crowther, C. S., & Mann, V. A. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America, 92,* 711-722.

Crowther, C. S., & Mann, V. A. (1994). Use of vocalic cues and native language background: The influence of experimental design. *Perception & Psychophysics, 55*, 513-525.

Cutler, A., Cooke, M. P., Garcia Lecumberri, M. L., & Pasveer, D. (2007). L2 consonant identification in noise: cross-language comparisons. *In Proceedings of Interspeech 07 in Antwerp, Belgium*, (1585-1588y).

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1987). The syllable's differing role in the segmentation of French and English. *Journal of Memory & Language, 25*, 385-400.

Cutler, A., Mister, E., Norris, D., Sebastián-Gallés, N. (2004). La perception de la parole en espagnol: Un cas particulier? In Ferrand, L., & Grainger, J. [Eds.] *Psycholinguistique Cognitive: Essais en l'honneur de Juan Segui* (57-74). Brussels: De Boec.

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

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

Cutler, A., Sebastián-Gallés, N., Solar-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28,* 746-755.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34,* 269-284.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 27*, 769-773.

Diehl, R. L., & Kluender, K. R. (1987). On the categorization of speech sounds. In Harnad, S. [Ed.] *Categorical Perception* (226-253). Cambridge: Cambridge University Press.

Disner, S. (1983). Vowel quality: the relation between universal and language-specific factors. *UCLA Working Papers in Phonetics, 58*,1-158.

Dorman, M. F. Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Perception & Psychophysics, 22,* 109-122.

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory & Language, 29*, 316-335.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224-238.

Elliot, L. L., & Katz, D. (1980). Children´s pure-tone detection. *Journal of the Acoustical Society of America, 67*, 343-344.

Flege, E. J. (1995) Second language speech learning: Theory, findings and problems. In Strange,W. [Ed.] *Speech perception and linguistic experience: Theoretical and methodological issues* (233-277). Timonium, MD: York Press.

Flege, E. J., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*, 3623-3641.

Foss, D. J., & Dowell, B. E. (1971). High-speed memory retrieval with auditorily presented stimuli. *Perception & Psychophysics, 9*, 465-468.

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

Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics, 55*, 597-610.

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *Journal of the Acoustical Society of America, 97*, 2540-2551.

Fruchter, D., & Sussman, H. (1997). The perceptual relevance of locus equations. *Journal of the Acoustical Society of America, 102,* 2997-3008.

Fry, D. B. (1979). The Physics of Speech. Cambridge: Cambridge University Press.

Fujimura, O., Macchi, M. J., & Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language & Speech, 21,* 337-346.

Furui, S. (1986). On the role of spectral transitions for speech perception. *Journal of the Acoustical Society of America, 80*, 1016-1025.

Gottfried, M., Miller, J. D., & Meyer, D. J. (1993). Three approaches to the classification of American English diphthongs. *Journal of Phonetics 21*, 205-229.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language & Speech 31*, 57-75.

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying nonnative speech categories. *Speech Communication, 50,* 109-125.

Green, J. N. (1990). Spanish. In Comrie, B. [Ed.] *The major languages of Western Europe* (226-249). London: Routledge.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*, 267-283.

Grosjean, F. (1996). Gating. *Language & Cognitive Processes, 11*, 597-604.

Hamann, S. (2003). The phonetics and phonology of retroflexes. Utrecht: LOT.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language & Speech 1*, 1-7.

Hazan, V., Iverson, P., & Bannister, K. (2005). The effect of acoustic enhancement and variability on phonetic category learning by L2 learners. In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception, 2005, London, UK.*

Healy, A. F. & Repp, B. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception & Performance, 8,* 68-80.

Hedrick, M. S., & Ohde, R. N. (1993). Effect of the relative amplitude of frication on the perception of place of articulation. *Journal of the Acoustical Society of America, 94*, 2005-2026.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of fricative consonants. *Journal of the Acoustical Society of America, 33*, 589-593.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*, 3059-3071.

Hume, E., Johnson, K., Seo, M., & Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. In *Proceedings of the XIVth ICPhS in San Francisco,* 1999.

International Phonetic Association (1999). Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. Cambridge: Cambridge University Press.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America 118*, 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87:* B47-B57.

Jassem, W. (1965). The formants of fricative consonants. *Language & Speech 8*, 1-16.

Jassem, W. (1968). Acoustic description of voiceless fricatives in terms of spectral parameters. In Jassem, W. [Ed.] *Speech Analysis and Synthesis* (189-206). Warsaw: Panstwowe Wydawnictwo Naukowe.

Jassem, W. (2003). Illustrations of the IPA: Polish, *Journal of the IPA 33,* 1, 103-107.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium in Tokyo, Japan (29-54)*.

Jongman, A. (1989). Duration of fricative noise required for identification of English fricatives. *Journal of the Acoustical Society of America, 85*, 1718-1725.

Jongman, A. (1998). Are locus equations sufficient or necessary for obstruent perception? *Behavioral & Brain Sciences, 21*, 271-272.

Jongman, A., Fourakis, M., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language & Speech, 32*, 221-248.

Jongman, A., Sereno, J., Wayland, R., & Wong, S. (1998). Acoustic properties of English fricatives. *Journal of the Acoustical Society of America, 103*, 3086.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108*, 1252-1263.

Jun, J. (2004). Place Assimilation. In Hayes, B., Kirchner, R., & Steriade, D. [Eds.] *Phonetically Based Phonology* (58-86). Cambridge: Cambridge University Press.

Keating, P.A., Mikos, M.J., Ganong, W.F. (1981). A cross-language study of range of voise onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America, 70*, 1261-1271.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America, 73*, 1779-1793.

Klaassen-Don, L. E. O. (1983). The influence of vowels on the perception of consonants. Doctoral Dissertation, Leiden University.

Krull, D. (1989). Second formant locus patterns and consonant-vowel coarticulation. Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS), 10, 87-108.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

Kuhl, P. K. (2000). A new view of language acquisition. *PNAS, 97,* 11850-11857.

Ladefoged, P. (1990). Some reflections on IPA. *Journal of Phonetics 18*, 335-346.

Ladefoged, P. (2001). Vowels and Consonants. Los Angeles: Blackwell Publishing.

LaRiviere, C., Winitz, H., & Herriman, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech & Hearing Research, 18,* 613-622.

Levin, D. T. (2000). Race as a visual feature: using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129*, 559-574.

Liberman, A. M., Delattre, P., Cooper, F. S., & Gerstmann, L. J. (1954). The role of consonant-vowel transition in the perception of the stop and nasal consonants. *Psychological Monographs, 279*, 68,1.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358-368.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel systems: the role of perceptual contrast. *Language, 48*, 839-862.

Lindblom, B. (1963). On vowel reduction. Royal Institute of Technology Report, 29.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In Hyman, L. M., &. Li, C. N. [Eds.] *Language, Speech, and Mind* (62-78). New York: Routledge.

Lipski, S. C. (2006). Neural correlates of fricatives across language boundaries. Doctoral Dissertation, University of Stuttgart.

Löfqvist, A. (1999). Interlocutor phasing, locus equations, and degree of coarticulation. *Journal of the Acoustical Society of America, 106*, 2022-2030

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874-886.

Lotto, A. J., & Holt, L. L. (1999). The illusion of the phoneme. *Special Phonetics Panel (ChiPhon) of 35-th Conference of the Chicago Linguistic Society*, University of Chicago.

Luce, P. A. (1986). Neighborhoods of Words in the Mental Lexicon. *Research on Speech Perception. Technical Report No. 6.*

Maddieson, I. (1984). Patterns of sounds. Cambridge: Cambridge University Press.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of [ʃ]-[s] distinction: I. Temporal factors. *Perception & Psychophysics, 28*, 213-228.

Mann, V., and Soli, S.D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics*, *49*, 399-411.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88*, 1286-1298.

Martin, R. C., Breedin, S. D., & Damian, M. F. (1999). The relation of phoneme discrimination, lexical access, and short-term memory: a case study and interactive activation account. *Brain & Language 70*, 437-482.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America 69*, 559–567.

Martinez-Celdran, E., Fernandez-Planas, A. M., & Carrera-Sabate, J. (2003). Illustrations of the IPA: Castilian Spanish. *Journal of the International Phonetic Association 33 (2),* 255-259.

Massaro, D. W. (1975). Perceptual images, processing time and perceptual units in speech perception. In Massaro, D. W. [Ed.] *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics* (Vol. 4, 125-149). New York etc.: Academic Press.

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

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science 30*, 1113-1126.

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

Medin, D. L., Goldstone, R., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review, 2,* 1-19.

Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics, 23*, 331-336.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27*, 338-352.

Miller, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorization. *Language & Cognitive Processes 16*, 683-690.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech-sounds: Differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America, 116*, 1198-1207.

Moore R. K., & Cutler A. (2001). Constraints on theories of human vs. machine recognition of speech. *In Proceedings of the SPRAAC Workshop on Human Speech Recognition as Pattern Classification, Max-Planck-Institute for Psycholinguistics, Nijmegen*.

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning & Verbal Behaviour, 15*, 43-51.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature 385*, 432-434.

Nearey, T. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America, 101*, 3241-3254.

Nettle, D. (1994). A behavioural correlate of phonological structure. *Language & Speech, 37*, 425-429.

Nettle, D. (1995). Segmental inventory size, word length, and communicative efficiency. *Linguistics 33*, 359-67.

Ng, W., & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25,* 217-232.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America, 112*, 711-719.

Nittrouer, S., & Miller M. E. (1997a). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 101*, 572-580.

Nittrouer, S., & Miller M. E. (1997b). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America, 101*, 2253-2266.

Nittrouer, S., Miller, M.E., Crowther, C.S., and Manhart, M.J. (2000). The effect of segmental order on fricative labelling by children and adults. *Perception & Psychophysics, 62*, 266-284.

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

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgement. In Marley, A.A.J. [Ed.] *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (347-365). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.

Ögüt, F., Kilic, M.A., Engin, E.Z., and Midilli, R. (2006).Voice onset times for Turkish stop consonants. *Speech Communication, 48*, 1094-1099.

Ohde, R. N., & Ochs, M. T. (1996). The effect of segment duration on the perceptual integration of nasal consonants for adult and child speech. *Journal of the Acoustical Society of America, 100,* 2486-2499.

Ohde, R. N., & Sharf, D. J. (1981). Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study. *Journal of the Acoustical Society of America, 69*, 297-300.

Ooijen, B. van (1994). The processing of vowels and consonants. The Hague: Holland Academic Graphics.

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

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

Palmer, J., Verghese, P., & Pavel, M.(2000). The psychophysics of visual search. *Vision Research, 40*, 1227-1268.

Picheny M.A., Durlach N. I., & Braida L.D. (1989) Speaking clearly for the hard of hearing. III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *Journal of Speech & Hearing Research, 32*, 600-3.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*, 253-260.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In Goodman, J.C., & Nusbaum, H.C. [Eds.] *The development of speech perception: The transition from speech sounds to spoken words* (121-166). Cambridge, MA: MIT Press.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285-290.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Pols, L. C. W., & Schouten, M. E. H. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America, 64,* 1333-1337.

Repp, B. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics, 30*, 217-227.

Repp, B., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 71*, 1562 -1567.

Rietveld, A. C. M., & Van Heuven, V. J. (2001). Algemeine fonetiek. Bossum: Uitgeverij Coutinho.

Rochet, B. (1991). Perception of the high vowel continuum: A cross-language study. In *Proceedings of XIIth ICPhS in Aix en Provence, France* (1, 273-76).

Rosner, B.S., Lopez-Bascuas, L.E., Garcia-Albea, J.E., and Fahey, R.P. (2000). Voice-onset times for Castillan Spanish initial stops. *Journal of Phonetics, 28*, 217-224.

Rothstein, R. A. (1993). Polish. In Comrie, B., & Corbett, G. [Eds.] *The Slavonic Languages* (686-758). London, New York: Routledge.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics, 19*, 394-398.

Savin, H. B., & Bever, T. G. (1970). The non-perceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behaviour, 9*, 295-302.

Schwartz J. L., Boë L. J., Vallée N., & Abry C. (1997). Major trends in vowel system inventories. *Journal of Phonetics, 25*, 233-253.

Schweickert, R. (1993). Information, time, and the structure of mental events: a twenty-five-year review. In Meyer, D. E., & Kornblum, S. [Eds.] *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (535 – 566). Cambridge, MA: MIT Press.

Sebastián-Gallés, N., Cuetos, F., Carreiras, M., & Martí, M. A. (2000). LEXESP. Léxicon formatizado del español. Barcelona: Universidad de Barcelona.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory & Language, 52,* 240-255.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition 72*, 111-123.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423 & 623-656.

Sharf, D. J., & Hemeyer, T. (1972). Identification of place of articulation from vowel formant transitions. *Journal of the Acoustical Society of America, 51*, 652-658.

Sharp, D., Scott, S. K., Cutler, A., & Wise, R. J. S. (2005). Lexical retrieval constrained by sound structure: The role of the left inferior frontal gyrus. *Brain & Language, 92*, 309-319.

Smits, R. (2000). Temporal distribution of information for human consonant recognition in VCV utterances. *Journal of Phonetics 28*, 111-135.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception & Performance, 27,* 1145-1162.

Smits, R., ten Bosch, L., & Collier, R. (1996). Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment. *Journal of the Acoustical Society of America, 100,* 3852-3864.

Smits, R., Warner, N. L., McQueen, J. M., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *Journal of the Acoustical Society of America, 113*, 563-574.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In David Jr., E. E., & Denes, P. B. [Eds.] *Human communication: A unified view* (51–66). New York: McGraw-Hill.

Stevens, K. N. (1998). Acoustic phonetics. Cambridge, MA: MIT Press.

Stevens, K. N. (2002): Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111*, 1872-1891.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*, 1358-1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In Eimas, P., & Miller, J. [Eds.] *Perspectives on the study of speech.* (1-38). Hillsdale, NJ: Lawrence Erlbaum.

Stevens, K. N., and House, A. S. (1956). Studies of formant transitions using a vocal tract analog. *Journal of the Acoustical Society of America, 28,* 578-585.

Stockmal, V. & Bond, Z. S. (2002). Same talker, different language: A replication. In *Proceedings of VII ICSLP in Denver, Colorado,USA* (77-80).

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America 85,* 2135-2153.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language & Speech 3*, 32-49.

Studdert-Kennedy, M. (1976). Speech Perception. In Lass, N. J. [Ed.] *Contemporary Issues in Experimental Phonetics* (243-293). New York: Academic Press.

Sussman, J. E. (2001). Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions? *Journal of the Acoustical Society of America, 109*, 1173-1180.

Sussman, H. M., Fruchter, D., Hilbert, J., & Sirosh, J. (1998). Linear correlates in the speech signal: the orderly output constraint. *Behavioral & Brain Sciences, 21*, 241-299.

Sussman, H. M., & Shore, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics, 58*, 936-946.

Tabain, M. (2000). Coarticulation in CV syllables: a comparison of locus equation and EPG data. *Journal of Phonetics, 28,* 137-159.

Tabain, M. (2002). Voiceless consonants and locus equations: a comparison with electropalatographic data on coarticulation. *Phonetica, 59,* 30-37.

Tabain, M., and Butcher, A. (1999). Stop consonants in Yanyuwa and Yindjibarndi: a locus equation perspective. *Journal of Phonetics, 27*, 333-357.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics 51,* 599-606.

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S-Plus. New York: Springer.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language, 40*, 374-408.

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

Wagner, A., Ernestus, M., & Cutler, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America. 120*, 2267-2277.

Walley A. D., & Carrell T. D. (1983). Onset spectra and formant transition in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 73,* 1011-1022.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *Journal of the Acoustical Society of America, 54*, 1248-1266.

Warner, N. (1998). The role of dynamic cues in speech perception, spoken word recognition, and phonological universals. Doctoral Dissertation, University of California, Berkeley.

Warner, N., Smits, R., McQueen J. M., & Cutler, A. (2005). Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. *Speech Communication 46*, 53-72.

Warren, P., & Marslen-Wilson, W. D. (1987). Continuos uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics, 41*, 262-275.

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

Weber, A. (2002). Assimilation violation and spoken-language processing: A supplementary report. *Language & Speech, 45*, 37-46.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory & Language, 50*, 1-25.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*, 35-44.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology, 50,* 509-535.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America, 69*, 275-282.

Whalen, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics, 46*, 284-292.

Wieringen, van A. (1995). Perceiving dynamic speech like sounds: psychoacoustics and speech perception. Doctoral Dissertation, University of Amsterdam.

Wieringen, van A., & Pols, L. C. (1994). Frequency and duration discrimination of short first-formant speech like transitions. *Journal of the Acoustical Society of America, 95*, 501-511.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *PNAS, 104*, 7780-7785.

Wright, R., Frisch, S., & Pisoni, D. B. (1995). Perceptual and articulatory factors in place assimilation: an optimality theoretic approach. *UCLA Occasional papers in linguistics 16*, 1-80.

Zatorre RJ, & Belin P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex, 11*, 946-953.

Zygis, M., & Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In *Proceedings of the 15th ICPhS* in Barcelona, Spain (395-398).

Samenvatting

Als volwassenen naar spraak luisteren passen ze automatisch luisterstrategieën toe die zijn afgestemd op hun moedertaal. Hierdoor nemen volwassenen met verschillende taalachtergronden spraak op verschillende manieren waar. Dit maakt het leren van een vreemde taal lastig, maar het betekent ook dat de perceptie van de moedertaal optimaal is. Luisteraars die gesproken uitingen horen worden geconfronteerd met een overvloed aan akoestische informatie. Om deze overvloed aan informatie te structureren, laten luisteraars zich leiden door betekenisvolle akoestische patronen in hun moedertaal. Voor de identificatie van spraakklanken moeten luisteraars informatie die de ene klank van de andere onderscheidt extraheren. Maar de vraag is wat deze onderscheidende informatie precies is. Dit is afhankelijk van de andere klanken in de desbetreffende taal; het foneemrepertoire van een luisteraar bepaalt welke informatie precies het verschil maakt tussen spraakklanken.

In deze dissertatie onderzoek ik of luisteraars met een verschillende moedertaal andere akoestische patronen kiezen om spraakklanken te identificeren als ze naar dezelfde uitingen luisteren. Bijvoorbeeld, alle luisteraars die /pssssttt/ horen zullen hierin de klank /s/ herkennen. Echter, onbewust moeten verschillende moedertaalsprekers de /s/ klank onderscheiden van de andere fonemen in hun taal die erop lijken. In de experimenten beschreven in deze dissertatie kregen luisteraars van zeven verschillende taalachtergronden dezelfde onzinwoorden te horen. De doelklanken die ze moesten identificeren waren contrastief in al deze talen. Op deze manier kan worden bepaald hoe verschillen in foneemrepertoire de foneemperceptie beïnvloeden.

De experimenten van Hoofdstuk 2 onderzoeken of luisteraars met een verschillend aantal categorieën binnen een foneemklasse (fricatief, stop consonant en klinker) de fonemen van deze klassen anders identificeren. Het Spaans heeft

bijvoorbeeld maar vijf klinkers, terwijl het Engelse er ongeveer 20 onderscheidt; het Nederlands heeft zes fricatieven en het Pools maar liefst elf. Beïnvloedt het aantal soortgelijke foneemcategorieën hoe snel en hoe accuraat een foneem wordt herkend? Eerdere studies die de identificatie tussen verschillende foneemklassen vergeleken maakten voornamelijk gebruik van slechts één luisteraargroep, meestal Engelstaligen (e.g., Foss & Swinney, 1973; Healy & Repp, 1982). De gevonden verschillen werden toegeschreven aan de verschillende fonologische functies van klinkers en consonanten en aan de akoestische eigenschappen die de foneemklassen onderscheiden. Al deze studies rapporteerden dat de stop consonanten het snelst werden geïdentificeerd, terwijl klinkers de moeilijkste doelfonemen zijn. In Hoofdstuk 2 heb ik luisteraars met een verschillend aantal categorieën binnen de foneemklassen vergeleken, om zo te achterhalen of de verschillen in de verwerking van klinkers versus fricatieven of stop consonanten hetzelfde zijn voor alle luisteraars. Een andere mogelijkheid is dat de identificatie van fonemen wordt gemoduleerd door het aantal soortgelijke fonemen in het foneemrepertoire.

De resultaten van Hoofdstuk 2 ondersteunen tot op zekere hoogte de aanname dat er een verschil is tussen de foneemklassen. Er werden over het algemeen meer fouten gemaakt voor de klinkers, en er was meer tijd nodig om de fricatieven te identificeren. Deze studie laat echter ook zien dat het effect van foneemklasse grotendeels afhangt van de manier waarop de reactietijden worden gemeten. Stop consonanten werden het snelst geïdentificeerd wanneer de reactietijd vanaf de plof werd gemeten, zoals gebruikelijk was in de eerdere studies. Wanneer de reactietijd echter gemeten werd vanaf het begin van de sluiting, waren de responsies voor de stop consonanten langzamer dan voor de fricatieven. Het is geen eenduidige beslissing waar een foneem begint in gesproken taal. Bovendien kunnen de coarticulatorische cues die een bepaald foneem aankondigen sterk verschillen tussen de foneemklassen. Het is dus onduidelijk op welk moment de evidentie voor een bepaald foneem begint toe te nemen voor luisteraars.

Daarnaast geven de resultaten in Hoofdstuk 2 aan dat er een taalspecifieke rangorde is in de snelheid en correctheid waarmee een foneem wordt geïdentificeerd, die afhankelijk is van het aantal categorieën in een foneemklasse. Hoe meer categorieën er zijn in de moedertaal van de luisteraars, des te langzamer is hun reactie en des te slechter hun prestatie. De aanwezigheid van slechts drie extra klinkers in het Catalaans in vergelijking met het Castiliaans Spaans verandert al het gemak waarmee klinkers worden geïdentificeerd relatief tot stop-consonanten en fricatieven. Extra spraakklankcategorieën lijken dus te concurreren met het foneem dat moet worden geïdentificeerd.

Met andere woorden, de vergelijking tussen vijf talen laat zien dat er geen duidelijke rangorde is in de identificatiesnelheid en correctheid tussen foneemklassen, die door hun akoestische eigenschappen alleen kan worden verklaard. Hoofdstuk 2 laat zien dat het aantal spraakklankcategorieën in de moedertaal verschillen veroorzaakt in de manier waarop fonemen worden verwerkt.

In Hoofdstuk 3 wordt door te focussen op de perceptie van fricatieven, verder onderzocht of de aanwezigheid van perceptueel vergelijkbare categorieën resulteert in verschillen in de selectie van akoestische cues. In deze experimenten werden Nederlandse, Engelse, Duitse, Poolse en Spaanse luisteraars gevraagd om de twee fricatieven /f/ en /s/ te identificeren. De fricatiefrepertoires van deze talen verschillen in de hoeveelheid categorieën die perceptueel vergelijkbaar zijn met de doelfonemen. Wellicht identificeren Nederlandse en Duitse luisteraars deze fricatieven puur op basis van de primaire cues, die liggen opgeslagen in de ruis van de fricatieven, omdat al hun fricatieven spectraal te onderscheiden zijn. Engelse en Spaanse luisteraars maken onderscheid tussen de labio-dentale /f/ en de dentale /θ/, die in feite erg op elkaar lijken. Poolse luisteraars hebben vier sisklanken met een palatale plaats van articulatie; daarom is het mogelijk dat de perceptuele opvallendheid van /s/ is gereduceerd. Mijn hypothese was dat luisteraars met soortgelijke fricatieven meer vertrouwen op coarticulatorische informatie in de omliggende klinkers. Het effect van de

coarticulatorische informatie werd onderzocht door luisteraars materialen te laten horen met coherente dan wel conflicterende cues in de aangrenzende klinkers.

De resultaten lieten zien dat de coarticulatorische informatie waarop luisteraars vertrouwen per taal verschilt. Conflicterende cues in de klinker belemmerden de fricatiefidentificaties van Engelse, Poolse en Spaanse luisteraars, maar niet van Nederlandse en Duitse luisteraars. Voor de Engelse en Spaanse luisteraars werd met name de identificatie van /f/ verstoord, en voor de Poolse luisteraars vooral de identificatie van /s/. Een vervolgexperiment controleerde of luisteraars de verkeerde akoestische informatie tussen de klinkers en fricatieven van de aan elkaar geplakte materialen konden waarnemen. Hieruit bleek dat zowel Spaanse als Nederlandse luisteraars de conflicterende informatie wel konden waarnemen. Dus hoewel beide groepen luisteraars hiertoe in staat waren, werden alleen de Spaanse luisteraars hierdoor in verwarring gebracht in het identificatie-experiment.

Samengevat laten de resultaten zien dat perceptueel vergelijkbare fricatieven in het foneemrepertoire luisteraars aanmoedigt om aandacht te schenken aan meer subtiele cues, zoals formanttransities. Daarnaast suggereren de resultaten dat het alleen nodig is om op extra coarticulatorische cues te letten als de fricatieven die geïdentificeerd moeten worden soortgelijke concurrenten hebben. Eveneens blijkt dat aandacht voor formanttransities niet beperkt is tot fricatieven met akoestische zwakke kenmerken zoals /f/ en /θ/. Poolse luisteraars laten zich leiden door formanttransities, zelfs als ze de akoestisch opvallende /s/ identificeren. Het lijkt er dus op dat niet de akoestische eigenschappen van een fricatief, maar de kennis van de luisteraars over gelijkenissen in het foneemrepertoire, bepalen of luisteraars op een bepaalde coarticulatorische cues vertrouwen of niet. Tot slot geven deze resultaten aan dat Nederlandse en Duitse luisteraars geen gebruik maken van de systematische akoestische informatie die wel wordt gebruikt door Spaanse, Engelse en Poolse luisteraars.

In Hoofdstuk 4 wordt onderzocht of dergelijke taalspecifieke verschillen invloed hebben op het moment waarop akoestische informatie wordt opgenomen. De

volgende drie specifieke vragen werden gesteld: (1) Zijn er verschillen tussen talen in de temporele opname van cues voor de plaats van articulatie? (2) Als luisteraars van een bepaalde taal meer vertrouwen op coarticulatorische cues bij het identificeren van fricatieven, doen ze dit dan ook bij stopconsonanten? (3) Halen luisteraars die op coarticulatorische cues vertrouwen deze informatie eerder of later uit het zich ontvouwende spraaksignaal? In een aangroeiproef (gating experiment) werden Nederlandse en Italiaanse luisteraars, wiens fricatieven spectraal distinctief zijn, vergeleken met Poolse en Spaanse luisteraars, die makkelijk te verwarren fricatieven hebben. De doelfonemen /k p t f s/ werden geïdentificeerd in afgeknipte klinker-consonant en consonant-klinker syllaben.

De resultaten laten zien dat, in vergelijking met Nederlandse en Italiaanse luisteraars, Poolse en Spaanse luisteraars de informatie die aangeeft wat de plaats van articulatie van een fricatief is, eerder uit het signaal halen (dus uit kortere stukjes van de klinker-consonant syllaben). De identificatie van de stopconsonanten lieten geen verschil zien tussen luisteraars. De verhoogde gevoeligheid voor coarticulatorische informatie voor fricatieven generaliseert dus niet naar stopconsonanten, maar die verschillen ook niet in het aantal categorieën tussen de verschillende talen. Samen ondersteunen deze resultaten de hypothese dat luisteraars de opname van cues alleen optimaliseren als hun foneemrepertoire fijner onderscheid vereist tussen soortgelijke contrasten. Bovendien letten luisteraars met perceptueel vergelijkbare fricatieven op additionele bronnen van informatie zodra deze beschikbaar zijn in het spraaksignaal. Luisteraars zonder perceptueel vergelijkbare fricatieven in hun foneemrepertoire halen informatie over de plaats van articulatie van de fricatieven later uit het signaal, omdat de cues die voldoende zijn om een onderscheid te maken in de plaats van articulatie voor al de fricatieven in hun moedertaal, later in het signaal zitten. Optimaal gebruik van cues betekent dus dat luisteraars kiezen voor de meest betekenenisvolle en noodzakelijke cues gegeven het foneemrepertoire van hun moedertaal.

THE FALSE TRACK

Why would it be that people do not understand each other even if they speak the same language? Consider Europe, and let's assume that everybody would master a common language, Esperanto, or, if need be, English. Would that help understanding? If we inquire this issue with the science that draws necessary conclusions we see that: a=1 & z=26 in English; a=1 & z=6 in Greek; in Basque a=1 & z=28, just like in Esperanto; a=2 & z=43 in Hungarian; in Dutch a=1 & z=26; a=1 & z=30 in Polish, while in German a=1 & z =29; and in Lithuanian a=1 & z=32, whereas a=1 & z=33 in Icelandic. Screaming for help (SOS) thus means 24 20 24 in Polish, 19 15 19 in English, 21 16 21 in German, 18 24 18 in Greek, 32 25 32 in Hungarian, 21 17 21 in Basque, and in Esperanto 22 19 22. To parse this illustratively short utterance, several data transformations are needed, and they clearly depend on the language of the one who seeks to understand. These transformations are not impossible, but they claim their time, and help cannot be guaranteed immediately. This delay may become even longer for those who master several languages because calibration will certainly take some extra time. Longer and more complex utterances will, of course, complicate things even more. It is boundless. This demonstrates the "unreasonable effectiveness of mathematics", and insinuates the widespread illusion of having understood

Curriculum Vitae

Anita Wagner was born in Katowice, Poland, on February 28, 1975. In 1987 she moved to Hannover in Germany. She studied Phonetics, Psychology and Italian Philology in Trier, and received her Magistra Artium degree in 2002. In August 2002 she was granted a Ph.D. stipend from the NWO SPINOZA grant "Native and Non-Native listening" awarded to Prof. Anne Cutler. She carried out her dissertation research at the Max-Planck-Institute for Psycholinguistics in Nijmegen, where she joined the Comprehension Group. In 2005 she was granted a NWO travel grant which allowed her to join the Grup de Recerca Neurociència Cognitiva in Barcelona for a period of six months.

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 articulograph. *Niels O. Schiller*

3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*

4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*

5. The acquisition of phonetic categories in young infants: A self-organising 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*

17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical coding of postural semantics in Goemai. *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermin Moscoso del Prado Martin*

24. Contextual influences on spoken-word processing. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing. *Rachèl J.J.K. Kemps*

29. At the same time ...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita E. Wagner*