Neural correlates of fricative contrasts across language boundaries

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Philosophie (Dr. phil.) genehmigte Abhandlung

Vorgelegt von

Silvia C. Lipski

aus Backnang

Hauptberichter: Prof. Dr. Grzegorz Dogil

Mitberichter: PD Dr. Bernd Möbius

Prof. Dr. Dr. Klaus Mathiak

Tag der mündlichen Prüfung: 22.05. 2006

Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart
2006

Contents

Danke!

Mein herzlicher Dank geht an die vielen Menschen, die mich während der Entstehung dieser Dissertation unterstützt haben:

Grzegorz Dogil danke ich besonders. Er hat mein Interesse an der Phonetik, Phonologie und Neurolinguistik geweckt und mich während des Studiums und der Promotion jederzeit unterstützt.

Bernd Moebius danke ich für die Gespräche und Anregungen und für ein tolles Seminar über Sprachsynthese.

Klaus Mathiak ermöglichte mir die Durchführung des MEG-Experiments. Herzlichen Dank auch für die Diskussionen der Ergebnisse und die zahlreichen Hilfestellungen.

Ein großer Dank an **Wolfgang Wokurek** für Rat und Hilfe mit Kabeln, Aufnahmen, Matlab und Sprachanalyse.

Am MEG-Zentrum habe ich von **Maike Borutta** gelernt, wie man die Aufnahmen durchführt.

Marzena Zygis ermöglichte mir den Kontakt zu **Andrzej Lyda**, der mir Räume und Versuchspersonen in Katowice und Gliwice organisiert hat.

Konrad Jusczyk half bei Experimenten in Poznan.

Diese Arbeit entstand im Rahmen des **Graduiertenkollegs "Sprachliche Repräsentationen und ihre Interpretation"** an der Universität Stuttgart, das von der Deutschen Forschungsgemeinschaft finanziert wird. Den Mitgliedern des Graduiertenkollegs möchte ich sehr herzlich für die freundliche und lustige Atmosphäre bei unseren Treffen und Klausurtagungen danken - die doch auch ganz schön anstrengend werden konnten (ich meine damit nicht nur die Bergwanderungen). Besonders danke ich **Manuela Schönenberger** für die Hilfe bei der Vorbereitung einer Präsentation und für die gute Laune, die sie überall verbreitet. **Beate Dorow** war ein echter Lichtblick, als das Ende des Tunnels fast erreicht war: danke für die nette Stimmung als es dann richtig eng wurde.

Für die Unterstützung von der **Phonetikgruppe am IMS** möchte ich mich ebenso bedanken: bei **Martin Haase**, der mir in der ersten Zeit ein geduldiger und freundlicher Bürokollege war; bei **Britta Lintfert**, meiner zweiten Bürokollegin, der ich für Unterstützung, Spaß und Ratschläge sehr dankbar bin; bei **Antje Schweitzer** und **Katrin Schneider**, von deren Arbeiten ich viele Anregungen mitnehmen durfte. Besonders danke ich **Sabine Dieterle** für viel, viel Kaffee und die wunderbare Unterstützung in allen formellen Belangen. Die nette Mittagsrunde und die schönen Kaffeepausen den IMSlern vermisse ich schon.

Herzlich möchte ich auch **Susanne Reiterer** danken, mit der ich so schön diskutieren kann und deren Arbeit und Ideen so interessant sind.

Paola Escudero was a great inspiration in terms of theoretical approaches to second language learning. Her works and her immense personal enthusiasm for SLA motivated me many times.

Für die Unterstützung meiner Eltern **Ema** und **Erich Lipski**, auf die ich immer zählen konnte, möchte ich mich an dieser Stelle herzlich bedanken.

Ein großer auch Dank an **Karin** und **Rüdiger Stegmeyer**, **Bruno** und **Yana Lipski** für viele gute Zeiten und Spaß mit dem Nachwuchs.

Ein besonderer Dank geht an **Heike Schulz** und **Richard Waltner** für das Nicht-Linguistische im Leben.

Aber an erster Stelle danke ich **Ruwen Kaminski** - für mehr, als ich hier aufführen könnte.

Abbreviations

BA	Brodmann area
COG	Center of gravity
EEG	Electroencephalography
EMF	Evoked magnetic field
ERP	Event related potential
FFT	Fast fourier transform
fMRI	Functional magnetic resonance imaging
ISI	Inter stimulus interval
LH	Left hemisphere
LPC	Linear predictive coding
MEG	Magnetoencephalography
MMN	Mismatch negativity
MMNm	Magnetic counterpart of mismatch negativity
PET	Positron emission tomography
RH	Right hemisphere
SMG	Superior marginal gyrus
SOA	Stimulus onset asynchrony
STG	Superior temporal gyrus
STS	Superior temporal sulcus

Chapter 1

Introduction

The phonological system, the way that speech sounds contrast and combine in a language to create lexical contrast, determines, to a large part, how they are perceived. Everybody who tries to learn a new language, will find some sounds irritating and difficult to tell apart. However, a clear perception of sounds is the most important prerequisite to memorize new words and to articulate them correctly. A German learner of Polish, for example, may take quite a long time, before even being able to to start a conversation by saying: "cze!" - [tʃɛɕtɕ]. The differences between the fricatives and affricates in this word are clear to Polish listeners, since they distinguish three different sibilant phonemes, as for example in: "sali - szali - siali", [sali - ʃali - ɕali] (room - scale - sowed). A German listener, however, is only used to distinguish the two sibilants [ʃ] and [s], such as in "Tasche - Tasse", [taʃə] - [tasə], (bag - cup).

This thesis investigated how linguistic experience affects the neural response to fricative sound contrasts. It was examined in what way auditory processing of native phonemic fricative contrasts differed from that of nonnative, unfamiliar fricatives. Moreover, it was tested if the contrastive function of native, phonetically distinct fricatives affects early auditory processing.

Speech perception is a very fast process and it is, to the most part, not consciously accessible. Investigations of underlying neural mechanisms of speech perception, therefore, require the usage of methods that can record auditory responses to speech with high temporal precision, such as electroencephalography (EEG) and magnetoencephalography (MEG). The mismatch negativity (MMN) component of the auditory event-related potential (ERP) and its magnetic counterpart (MMNm) signal auditory discrimination and auditory sensory memory. Recent studies provided evidence that the MMN reflects memory of native language speech sound categories.

The present study used the MMNm as a probe for native language-specific representations of fricative sounds, that have not been covered in electrophysiological research on speech processing so far.

Overview

Chapter 2 gives an introduction to the influence of the native language phonological system on speech perception. Evidence from behavioral research for the prominent role of distinctive speech segments is shown and specific observations for the perception of fricative sounds are reviewed. These observations are complemented by an overview on theoretical accounts of speech perception and models of second language perception and learning.

EEG and MEG measurements allow to closely trace the processing of the incoming speech signal. **Chapter 3** explains these methods and introduces the components of the neural auditory response that are relevant for speech sound processing.

Chapter 4 presents findings from previous neurophysiological studies of speech perception. It is shown how the processing of speech and nonspeech sounds might differ. Then, studies on the extraction of phonetic cues and the influence of native phonemic categories on auditory processing are covered, and the influence of second-language learning on auditory processing is described. The overview on related works is complemented by a summary of brain studies that used neuroimaging methods which provide a precise localization of speech-related responses.

The findings and ideas that are compiled in Chapter 4 provide the foundation for the concept and discussion of the experimental work of this thesis. **Chapter 5** presents the MEG experiment on native phonemic, allophonic, and nonnative fricative contrasts and a closer behavioral inspection of the impact of formant transitions on fricative perception.

Chapter 6 summarizes the findings of this study and provides some suggestions for future research.

Chapter 2

Linguistic experience and speech perception

2.1 Linguistic experience in speech perception

A basic assumption of linguistic analysis states that speech sounds are represented in terms of their function in the language. Sounds that can be used to create lexical contrast are mentally represented by discrete phonological segments. The presence absence or exchange of these phonemes can modify the meaning of a word. Systematic alternations are derived from these underlying representations by language-specific phonological rules. Thus, a distinction is drawn between phonemes and allophones, which are phonetic variations of phonemes and appear regularly in certain phonological contexts (Chomsky and Halle, 1968; Kenstowicz, 1994; Halle, 2000). In a phonetic sense, allophones may vary very strongly. It is the linguistic function of a speech sound, not its phonetic appearance which renders it to be a phoneme or an allophone. This is illustrated, for example, in the phonemic function of the alveolar and velarized [l] in Catalan vs. their allophonic variation in English, (Wheeler, 2005).

In speech perception, listeners have to connect the acoustic speech signal with the phonological representations that form words in the mental lexicon. Speech sounds occur with high variation, due to individual speaker characteristics, influences from other sounds, the position of a sound in a phrase, to background noise or room acoustics, and many other factors (Pisoni, 1997). Nevertheless, words can be usually understood without difficulty. To account for this phenomenon, speech perception is thought of as a mapping-process from the continuous, variable acoustic form to discrete sound category representations by means of normalization and abstraction. As will be demonstrated in the following sections, this process is strongly influenced by the native language. This influence very often appears as a decline in sensitivity to phonetic contrasts that are not used as phonemes in the native language.

2.1.1 Perception of phoneme contrasts

Language-specific knowledge guides the auditory perception of speech sounds already at an early age. Werker and Tees (1984) reported that English infants of 5 to 8 months of age could discriminate nonnative Hindi plosive contrasts as well as native adult listeners, whereas the disciminatory ability for the nonnative sounds decreased between 10 to 12 months of age. Similarly, Tsushima et al. (1994) showed that Japanese infants 6 to 12 months can discriminate the English contrast between [r] and [l] that is not phonemic in Japanese and very difficult to discriminate and to learn for adult Japanese speakers (Miyawaki et al., 1975; Takagi, 2000).

Studies of categorical perception in adults reveal the same pattern: native language experience enhances the sensitivity for differences between phonemically relevant categories and decreases the sensitivity for variations within these categories (e.g. Liberman et al., 1957; 1967; for review see Repp, 1984).[1] In order to test for categorical perception, an acoustic stimuli continuum is created in which the sounds gradually change in one or a few acoustic parameters. First, the stimuli are presented one by one in random order and listeners are asked to label them according to specified phonetic categories. Then follows a discrimination test, in which listeners have to decide whether two or more stimuli are acoustically different or identical. Perception is considered to be categorical when there is a sharp boundary between categories and the stimuli within each category are consistently labelled. In addition, the identified category boundaries have to coincide with a steep rise of discrimination. Strictly speaking, discrimination should be predictable on the basis of the identification performance, showing that listeners are hardly able to distinguish the stimuli within the category boundaries.

An example for the influence of the native language on plosive categorization is the cross-linguistic study by Abramson and Lisker (1970). They compared the perception of plosive-vowel syllables that varied continuously in voice-onset-time (VOT) for Thai and English listeners. Crucially, there are three phonemic plosive categories in Thai that are distinguished by VOT: voiced, voiceless, and aspirated stops, as in the words [p^hà], [pà], and [bà], "to cut", "forest", and "shoulder". In English, only voiced and voiceless plosives are phonemically distinctive. Unaspirated and aspirated stops are allophonic variations, for example "top" is spoken with aspiration whereas "stop" contains the unaspirated variant. The results closely matched the phonological function of the stops in the two languages: Thai listeners perceptually divided the continuum into three categories and judged all members of these categories to sound alike. English listeners divided the VOT continuum into two categories and could only perceive an acoustic difference at the boundary.

At closer inspection, not all types of speech sounds elicit the same kind of perceptual patterns. As reviewed above, the strongest evidence for an influence of the native language on categorical perception has been found for plosives. For fricatives, approximants, and vowels, however, the ability to discriminate between variants of a phonetic category is much higher.[2]

When vowels are tested in the classic categorical perception paradigm, perception is found to be basically continuous (Fry et al., 1962; Pisoni, 1973). The influence of native phonetic categories becomes more evident, if listeners do not focus on single acoustic cues, for

---

[1] When the first studies on categorical speech perception came out, it was believed that this mechanism is unique to human speech perception (Liberman et al., 1957). Further research with non-speech stimuli, like Miller et al.'s study (1976) with a noise-buzz continuum, or successful attempts to train animals for categorical perception, e.g. by Kuhl and Miller (1978) suggested that categorical perception is rather a common cognitive principle (for reviews see Rosen and Howell, 1987; Kuhl, 1987; and Iverson and Kuhl, 2000).

[2] The differences in categorical perception for different classes of sounds raised criticism on the categorical perception test paradigm. The problem is that many experimental factors can significantly change the outcome of a categorical perception test, i.e. the duration and naturalness of the stimuli, the discrimination task, and the inter-stimulus-interval (Pisoni, 1973; Massaro, 1998; Van Hessen and Schouten, 1999).

example on variations in the first and second formants, but have to compare more complex stimuli, for example naturally spoken vowels in syllable context (Gottfried, 1984).

Kuhl and associated researchers observed that, instead of a categorical perception, the perception of vowels is guided by prototypical category members. Those vowel exemplars, that are judged to be the best members of the category, perceptually attract other vowel variations. Consequently, listeners are less likely to perceive acoustic differences between a prototype and vowels that are similar to this best category member. This effect appears to emerge earlier than language-specific categorization of consonants. Grieser and Kuhl (1989) tested the discrimination of 6 months old infants for variations of the vowel category [i]. They trained the children to turn their head when they heard a change in a repetitive sequence of vowel exemplars. When a prototypical [i] exemplar was alternated with a less prototypical variation, children could not discriminate the two as well as when a poor exemplar of [i] had to be compared to other [i] variations. Concurrent results were obtained by Kuhl (1991) for adult listeners and in a cross-linguistic study by Kuhl et al. (1992), in which they observed that Swedish 6 months old babies were better at discriminating prototypical English [i] from other [i] variants than English babies. Likewise, English babies could discriminate prototypical Swedish [y] better from other members of this category than Swedish infants. These observations lead to the development of the Native Language Magnet theory, which is described later in section 2.3.4.

2.1.2 Perception of allophone contrasts

Allophonic variations of sound categories are always present in the studies that were cited above. The variant stimuli in categorical perception tests, for example, may correspond to freely varying allophones and these studies show that listeners map the variations to one phonetic category during perception. Distributional allophones vary in a rule-based fashion and occur in a certain phonological context. They are phonetically distinct, but they cannot be used for lexical distinctions. There are only few studies that explicitly address the perception of allophonic contrasts, but they suggest that allophone contrasts are perceived differently from sounds that function as phonemes.

Evidence that the phonological status of a phonetic contrast indeed influences speech perception comes from Hume and Johnson (2003). Their study of Mandarin tone perception in native and nonnative listeners illustrates that phonological contrast influences perceptual distinctiveness. The Mandarin tones 214 (low-falling-rising) and 35 (mid-rising) do not contrast in the context following 214, and thus present a partial phonological contrast. Native and American English listeners discriminated these tones and additionally two other tones (51, 55) which form a complete phonological contrast in Mandarin. The reaction times were recorded as a measure of ease of discrimination and perceptual distance. It was shown that the partially contrastive tones were perceptually less distinct for the Mandarin Chinese listeners than the fully phonemic tones. No such effects were observed for the nonnative listeners for which only the acoustic difference affected discrimination.

An effect of phonological contrastiveness in perception was, furthermore, found for the French allophonic contrast between [ʁ] vs. [χ] by Peperkamp et al. (2003). The voiceless fricative occurs in the context of voiceless consonants. Peperkamp and collegues compared how native listeners discriminated these allophones and the phonemic contrast between [m] and [n]. Poorer discrimination was observed for the allophones than for the phonemes as they were embedded in nonword "minimal" pairs, even if the phonological context did not match the distributional rule for the allophones.

2.1.3 Perception of fricative sounds

Tests of categorical perception of fricatives yielded mixed results. In most studies, synthetic noise continua were used, which contain one or two noise-poles or formant peaks. A percept of [ʃ] is obtained when the pole centers are between 2000 and 3000 Hz, and when the spectral peaks are roughly between 3800 to 5000 Hz the noise resembles [s].

Perception tests that used such synthetic noise continua ranging from [s] to [ʃ], showed a correlation of category boundaries with discrimination peaks. However, within category distinction was moderate for isolated noises and increased when the noises were followed by a vowel in the studies by Fujisaki and Kawashima (1968) and Repp and Mann (1978). In contrast, in Healy and Repp's study (1982) of the [s] to [ʃ] continuum, the perception was completely continuous.

Stronger categorical perception was observed by May (1981) who investigated the perception of Egyptian Arabic voiceless [s] and [ʃ], voiced and voiceless velar and pharyngeal fricatives. Pharyngeal fricatives have less strong frication and a more pronounced formant structure than velar fricatives, they are acoustically similar to approximants (Butcher and Ahmad, 1987). May tested whether the presence of formants and voicing would lead to continuous perception. Synthetic continua of these sounds with very fine intermediate steps were presented in intervocalic context (ə - ə) to eight native listeners. Since the differences between the individual stimuli were smaller than in the studies by Fujisaki and Kawashima (1968), Repp and Mann (1978), and Healy and Repp (1982), the perception of the voiceless fricatives was rather categorical, but discrimination performance was still better than predicted. Voicing and pharyngeal articulation, on the other hand yielded high within category discrimination performance.

2.1.3.1 Phonetic and auditory modes of perception

Repp (1984) suggests that the patchy results of categorical perception tests for fricatives result from the fact that listeners are able to switch from a "phonetic" to an "auditory mode" of perception for these sounds:

> Fricative stimuli seem to be especially suited for the application of different strategies so that they may be perceived fairly categorical in one situation but continuously in another. (Repp, 1984, p. 284)

An auditory mode of perception refers to perception without reliance on linguistic category labels. Fine acoustic detail is perceived well and supports listeners' judgements. The reason why detailed acoustic information is more available for some sounds could lie in the duration of acoustic cues (Fujisaki and Kawashima, 1968; Pisoni, 1973). Acoustic features that last over a longer period of time, such as in fricatives and vowels, could be represented better in an auditory echoic memory buffer (Crowder, 1971; Pisoni, 1973, Mirman et al., 2004).

The exertion of an auditory or phonetic mode of perception in fricatives, furthermore, seems to depend on the integration of formant cues, as exemplified by Repp (1981). In this study, he employed a synthetic continuum noise that ranged from [s] to [ʃ] followed by the vowels [a] or [u]. For most listeners, moderately categorical perception was observed. Additionally, the vowel context shifted the category boundary significantly: higher noise formants were required to yield the percept of [s] for syllables that contained [a] than for syllables that contained [u].

A few listeners, however, distinguished the stimuli with high accuracy. To follow up on this observation, Repp carried out a second experiment, this time using a continuum with smaller intermediate steps so that discrimination would be harder. Again, the same few listeners showed highly accurate discrimination. Interestingly, it was observed that the vowel context significantly affected the discrimination for those listeners who showed poorer discrimination performance, but for the exceptional listeners, almost no influence of vowels was seen. Repp hypothesized that those listeners who could distinguish the sounds very well, segregated the noise part from the vowel, while the others integrated noise and vowel.

This idea was tested in a third experiment were new participants were first trained to differentiate the stimuli. The listeners were encouraged to transfer the discrimination performance for isolated noises to syllable stimuli. This type of training was successful for most listeners who could now distinguish within-category changes very well. Thus, Repp's suggestions were supported: detailed auditory analysis of fricatives requires a separation of frication from vowels.

2.1.3.2 Perception of nonnative fricatives

Language-specific perceptual patterns for fricatives become more evident when listeners do not focus on one absolute acoustic cue but have to find complex relations between stimuli, similar as for vowel stimuli (e.g. Gottfried, 1984). Furthermore, the usage of synthetic noise which, in isolation strongly resembles non-speech sounds may artificially enhance a non-phonetic listening strategy.

Lisker (2001) tested the identification of Polish [s], [ɕ], and [ʃ] by American English listeners with naturally spoken exemplars. The subjects had no knowledge of Polish and were only briefly introduced the sounds. Several recordings of fricative-[a] syllables, which had been spoken by one male native speaker, were presented as whole syllables, clipped off isolated fricatives, and isolated vowels, that included the formant transitions.

The sound [s] was well identified in the syllable and in isolation, only when the vowel part was presented, it was misidentified as belonging to [ʃ] with about 20% and less as belonging to [ɕ].

The other two fricatives were confounded to a high extent. In the syllable condition, [ʃa] was identified in ca. 59% of the trials correctly and [a] in about 56%. However, the performance raised significantly when the sounds were presented in isolation. Palato-alveolar [ɕ] was identified even better (ca. 75% correct) when the listeners only heard the vowel.

Subsequently, the stimuli had to be rated for loudness and pitch. The judgements were clear for the isolated fricatives: [ʃ] was clearly perceived as being louder and [ɕ] was perceived as having a higher pitch (both in approx. 85% of the responses). When the syllables were rated for the impression of loudness, ratings became random. The [ɕa] syllables were still perceived as having higher pitch but to a lesser extent (ca. 68%).

Lisker's study shows that the Polish contrast between the post-alveolar and palato-alveolar fricatives is difficult to distinguish and label for nonnative listeners. Furthermore, Lisker's study suggests that even natural fricatives may be perceived less phonetically when presented as isolated noises. This became most evident when the listeners were asked to rate the sounds according to the non-linguistic properties of loudness and pitch. Lisker suggests that nonnative listeners may use such auditory strategies when confronted with unfamiliar fricative contrasts, and he notes:

> [...] it might reasonably be claimed that fricative noises, particularly in isolation, are no longer unequivocally speech, i.e. immediately identified as outputs of a human vocal tract in particular articulatory configurations, and are therefore amenable to the same psychoauditory processing as any other nonspeech acoustic signals. (Lisker, 2001, p. 236)

2.1.3.3 Formant transitions

The study by Repp (1981) suggested that a phonetic mode of perception for fricatives is enhanced when transitions to an adjacent vowel are perceptually integrated. Along similar lines, Lisker (2001) showed that nonnative listeners can make more definite judgements on a fricative's acoustic properties (loudness and noise-pitch) when it occurs as isolated noise than when it occurs in syllable context, suggesting that the lack of the vowel enables an acoustic analysis rather than a phonetic one.

The interaction of several acoustic features and language-specific weighting are main properties of phonetic perception and phonetic knowledge (Repp, 1982). The way that individual acoustic cues are weighted in perception has been shown to be part of language-specific auditory perception in several studies (Morrongiello et al., 1984; Underbakke et al., 1988; Bohn, 1995; Escudero, 2002; Nittrouer, 2002).

Concerning the relation between transitional and more steady acoustic cues, Nittrouer et al. (1993) proposed a "developmental weighting shift". According to Nittrouer and collegues, children are initially geared to dynamic movements of the vocal tract and therefore pay more attention to transitional cues. During maturation, children acquire the ability to use acoustic properties that do not involve spectral change.

This hypothesis has been verified in a number of studies. Morrongiello et al. (1984) have shown that children receive more information from the transitions in a "say-stay" continuum while adults rely more on the duration of the silent gap when differentiating these words. Investigations by Parnell and Amerman (1978) and Ohde and Haley (1997) on the perception of place in plosives brought similar results. Adults and 3 to 5 year old children gained more information from the vocalic transitions, but adults were significantly better at identifying the plosive from the noise burst alone.

Especially for fricatives, Nittrouer and Miller (1997) found that with increasing age (they tested 3.5, 5, 7 year olds and adults), the fricative part in [sa], [su], [ʃa], and [ʃu] syllables becomes more important than the vocalic portion. Listeners had to identify synthetic fricative continua from [s] to [ʃ] which were combined to syllables with natural vowels that contained the formant transitions of either one of the fricatives or with synthetic vowel continua with gradually varying transitions. The subjects' age was the significant factor in all of these tests: the youngest children based their decisions on formant transitions, the adults labelled the stimuli according to properties of frication.

In a follow-up study by Nittrouer (2002), natural fricatives of [s], [ʃ], [f], and [θ] were followed by synthetic vowels that varied in their appropriateness to the respective fricatives. Nittrouer tested 4, 6, and 8 year old children and adults. The findings for [s] and [ʃ] met the previously obtained results: the youngest children's judgements were guided more strongly by the transitional information, 8 year olds and adults labelled the stimuli according to the fricatives, and the responses of children of 6 years lay in between.[3]

The perceptual shift away from transitional cues appears to concern mainly the distinction of those fricative contrasts that are acoustically relatively salient, such as [s] vs. [ʃ]. When the distinction of a fricative contrast is more difficult, native listeners rely more strongly on the formant transitions. For the syllables containing [f] and [θ] which are spectrally very similar, all age groups in Nittrouer's study (2002) were affected by the formant transitions in the same way. The relevance of formant transitions even increased slightly for the adult participants.

Additionally, the study by Wagner and Ernestus (2004) has shown that the relevance of formant transitions for fricative perception differs for speakers of different languages. They used nonwords containing [sa] and [fa] syllables. Some syllables were cross-spliced and switched, so that the fricatives were followed by the wrong vowel transition. In a phoneme-monitoring experiment, Dutch, German, English, and Spanish listeners had to detect [s] and [f] in these nonwords. The Dutch and German listeners' responses were not affected by the wrong transitions, whereas the number of correct responses decreased for English and Spanish subjects when fricative and transition did not match. Wagner and Ernestus suggest that the fact that Engish and Spanish contains a third frontal phonemic fricative (θ) could generally enhance the perceptual focus on transitional cues for speakers.

---

[3] Recalling the observations by Repp (1981) of a perceptual switch between an auditory and phonetic mode of perception, it could be speculated that the older listeners in these studies had more practice in switching perceptual strategies. However this question is not addressed in Nittrouer's articles.

2.1.3.4 Fricative perception - summary

To summarize, experience with the native language does not uniformely affect the perception of all sound classes in the same way. Whereas the acoustic salience of within-category differences is very low for plosives, the best category members excert a more gradient "magnet" effect for vowels. Concerning fricatives, variation within a category can be distinguished, possibly due to a switch to a non-phonetic mode of listening which is free of native language representations. Generally, less sharp category boundaries and better discriminability of variation within phonetic categories than for plosives was observed. Still, the native language excerts a strong influence which has been seen in Lisker's study where American English listeners mixed up Polish [ʃa] and [ɕa] to a high extent. Furthermore, formant transitions seem to play an important role for the distinction of some fricative contrasts and their perceptual importance may differ between languages.

2.2 Models of speech perception

The observations that the perception of speech sounds is strongly guided by the native language, are conceptualized in psycholinguistic models. Since electrophysiological research on phonetic perception allows to trace the process of speech perception during the earliest stages, in which the acoustic signal is mapped onto native speech sound representations, a more detailed account of theoretical explanations of this process is justified.

The models of speech perception have rather different and often opposed views on the mechanisms involved and the contents of the mental lexicon. The previous overview on the perception of native and nonnative segments of various types and function suggests that many factors are involved in auditory speech processing. The large number of different and often contradictory approaches further indicates that speech perception is a highly complex phenomenon. However, common to most theoretical accounts of speech perception is the assumption, that native linguistic representations affect the perceptual process at a very early stage. This motivates research on early cortical responses to speech sounds.

This presentation of speech perception models does not include accounts that are mainly concerned with word recognition and word selection. Overviews of such models are given by McQueen (2004) and Goldinger et al. (1996). Instead, those models are described that explain how information, which is used to access lexical representations, is extracted from the acoustic signal and how native language representations shape this process.

Many accounts of speech perception assume that the speech signal is mapped successively onto more abstract representations before the lexical level is accessed. During this process, the acoustic signal is transformed into a sequence of discrete elements. This transformation provides a step of normalization of the acoustic variation, the integration of several acoustic cues, as it was seen for fricative sounds, and the detection of phonetic units. However, the number of different proposals about the size and nature of these intermediate phonetic units is long (cf. Cutler and Clifton, 1999; McQueen, 2004). Among the suggestions is a mapping from phonetic features to phonemes (Elman and McClelland, 1986; Pisoni and Luce, 1987). Then there is the assumption that phonemes provide an intermediate step in perception (Norris et al., 1997). Other researchers found evidence for syllables as basic units of perception (Mehler et al., 1981). It was also observed that languages can differ in the type of intermediate phonetic unit that they use (Cutler and Norris, 1988).

This list indicates that there is no agreement on the size and the level of abstractness of intermediate representations. This has to do with the fact that evidence for a certain type of intermediate representation is often dependent on the particular experimental method or task that is used. Recent proposals, such as the one by Boersma (1998, 2005) that will be detailed below, are not concerned about the definite size of intermediate representations. Rather, it is assumed that they can vary according to the requirements of the situation. In the experiment that is described in this thesis, syllables were used for testing, but this is not meant to imply that these would be basic units of speech perception.

2.2.1 Motor Theory

A particular suggestion about the basic units of speech perception comes from the Motor Theory which assumes that articulatory gestures form the invariant elements in speech which are detected in perception (Liberman et al., 1967; Liberman and Mattingly, 1985). Generally, the relationship between articulatory and acoustic properties of speech perception has long been the foundation of phonological description (Jacobsen et al., 1952; Fant, 1960). The Motor Theory took this relationship to be the basic principle of speech perception. The listener reconstructs the articulatory gesture that underlies the acoustic signal by means of a specialized mental module. This module connects the speech signal with the neural representation of articulatory gestures which are associated with phonological representations. This mental device is assumed to be innate and thought to operate independently of general auditory perception. Speech is, therefore, perceived differently from other auditory events: it is processed in reference to production.

2.2.2 Direct Realism

Similarly, in the view of Direct Realism (Fowler, 1989; Gibson, 1991) articulatory movements are the cues to speech perception. Unlike Motor Theory, Direct Realism claims that no specialized mental module is involved. Instead, the articulatory gestures are directly perceived from the speech signal and do not have to be deduced. The signal contains all information that is needed to identify the perceived objects. Acoustic features in speech invariantly correspond to place, manner, and specific patterns of articulation.

The acoustic-to-gesture mapping is basically a non-linguistic ability, but turns into a language-specific device through experience with the native language. During first language aquisition articulatory constellations are associated with certain speech sounds. The increase in lexical representations and communicative skills turns the learned constellations into abstract phonological units of representation. These phonological units contain language-specific phonetic detail. It is assumed that phonetic and phonological representations are not separated, but are closely intertwined. Experience with the native language increases the perceptual focus on phonetic details that signal important phonological units. Other phonetic detail is then perceived less well:

> Attunement to the native language increases attention to the portions of the phonetic domain that the language has harnessed to serve linguistic functions; that is, detection of crucial information in these regions of the domain becomes both more sharply defined and more efficient. (Best, 1995, p. 186-187)

2.2.3 Acoustic landmarks and analysis-by-synthesis

The model of speech perception by Stevens (2002, 2005) assumes that words in the lexicon are stored as sequences of segments that are defined by distinctive features, which are related to articulatory movements. Thus, Stevens' account is associated with articulatory accounts of speech perception. The lexical distinctive features are assumed to be discrete and binary, and similar to the concept of Generative Phonology, there is just one representation for each word (Halle and Stevens, 1991). Speech perception extracts acoustic cues and provides an estimation of phonological features which is then evaluated by an internal synthesis mechanism.

Stevens argues that the relation between acoustic cues and articulatory patterns is defined by clear principles and that this relation is utilized in perception. Each feature has a corresponding articulatory gesture which has a corresponding acoustic form. Therefore, the discreteness of phonological features and phoneme segments is mirrored in the acoustic and articulatory patterns of speech, even though speech movements are continuous and variable (see also Halle, 2002; and Stevens, 1980, 1985, 1989).

The initial step in feature extraction is the detection of acoustic "landmarks". These are the crucial cues for speech perception because they signal articulator-free features which distinguish the main sound classes. Landmarks guide all subsequent perceptual analysis. Articulator-free features do not specify one particular articulator but rather a general articulatory setting. Landmarks are ares of high intensity in the lower spectral region which refers to the feature [vowel]; the areas with rapid and abrupt acoustic changes signal the feature [-continuant] and [consonantal]; areas without rapid changes and most energy in higher frequences are [+continuant], and [+strident] if there is strong high-frequency intensity; and finally overall low energy without rapid changes corresponds to glides.

In the next step, acoustic cues in the vincinity of the landmarks are extracted. Articulatory movements mainly take place between landmarks. Therefore, articulator-bound features that signal a specific articulatory organ can be found here. Stevens distinguishes between the articulators that form the constriction and gestures that shape and position these articulators. Thus, the dependency between articulatory gestures and features becomes apparent: a certain type of landmark limits the set of possible articulators that can be associated with it and when the primary articulator is found, the position or shape that it can have is reduced to a small set of possibilities. This dependency, which can be seen as a hierarchy, strongly supports speech perception.

The conversion of acoustic cues to features is carried out in the next step of the perceptual process. For example, as mentioned before, a landmark of continuous energy in the high frequencies is translated to [+continuous] and [+strident], and furthermore, the acoustic cues are estimated as [+anterior] and [+high].

Feature estimations from acoustic cues are strongly determined by the native language:

> Once the cues for an articulatory-bound feature have been extracted, these cues must be weighted or combined in some way to yield an estimate of the feature. These combinations of cues must be learned by a speaker of the language. (Stevens, 2002, 1888-1889)

Upon lexical access, the validity of the extracted features is tested by an internal synthesis. It works like an internal generator of corresponding articulatory movements which create internal synthetic representations of acoustic cues and landmarks that belong to these gestures. These are compared with the perceived acoustics from the speech signal. Additionally, this process takes context information into consideration, i.e. prosody, stress, the position of a segment in a syllable, and even background noise.

2.2.4 Linguistic perception

Boersma's model of speech perception and production is based on the functional principles[4] for production, communication, and perception, which Boersma incorporates in an Optimality Theory (OT) grammar (Boersma, 1998; 1999; 2005).

OT explains the relations between underlying representation and surface form as a transformation from input to output. A function GEN generates a set of possible surface forms from the input from which the function EVAL selects the best candidate. The process of evaluation is based on a set of constraints. These are supposed to be innate, universal, and violable. They are hierarchially ranked in a language-specific way so that the violation of highly ranked constraints is more culpable than that of low-ranked constraints. Constraint-ranking, thus, renders the particular surface forms of individual languages.[5]

Boersma departs from this theory in that he does not adopt innate phonologically specified constraints. Rather, he assumes that general articulatory and perceptual capabilities, learning abilities, and the need to communicate are innate to every human being. These general principles provide constraints for perception that are initially non-linguistic. This includes the ability to detect statistical distributions (see also Kuhl, 2000, 2004; section 2.3.4).

During language acquisition, the general constraints are defined by precise language-specific values and are ranked in a language-specific way. Every language, therefore, has specific constraint values that directly influence speech perception. This implies that speech perception is the realization of a phonological perception grammar.[6] Speech perception is thus

---

[4] This view goes back to Passy (1891) and Martinet (1955), both cited in Boersma (1998). More information can be found in Akamatsu (1992).

[5] For a detailed introduction to OT see Kager (1999).

[6] The opposite view is held by Hume and Johnson (2001) who summarize speech perception under non-linguistic, non-phonological factors. Speech perception may shape and be shaped by phonological representations but since perceptual principles work in non-linguistic domains as well, Hume and Johnson do not view speech perception as a linguistic device.

seen as a linguistic device that is active as soon as the speech signal is sensed. This way, speech perception is optimally tuned to the requirements of the native language.

Speech comprehension is seen as a two-step process that includes a mapping of the original acoustic signal (Auditory Form, AudF) to discrete phonological units (Surface Form, SF) - this is speech perception -, and second, a mapping of the SF to lexical forms (Underlying Form, UF) - this is speech recognition.[7]

The first step, perception, includes constraints that enable the abstraction from the acoustic input. In particular, one family of constraints limits the difference between the acoustic signal and the percept. Here it is determined how much the speech percept is warped in comparison to the original signal. Another type of constraints regulates the categorization of acoustic features into perceptual classes. Language-specific weighting and integration of several cues that was seen in the behavioral tests on fricative perception for example by Nittrouer (2002), is characterized by constraints that determine the combination of several acoustic cues to obtain a single percept. This provides a language-specific balance between perceptual integration and differentiation of acoustic details.

In the second step, during recognition, faithfulness-constraints require the perceptual similarity between SF and UF. Finally, lexical constraints evaluate the appropriateness of a lexical item in a semantic context.

The perception-step is usually not consciously accessible. Listeners can only report about the results perception which are then handled by recognition-constraints. Furthermore, perception is not influenced by word meaning or sentence context. Lexical information first becomes available at the step of recognition. The perception grammar that transforms the acoustic signal to phonological units that can be used to access the lexicon, has no connection to higher level structures. However, Boersma admits that it is not always easy to differentiate the two steps for the experimenter:

> [...] of course, the comprehension system is optimized for comprehension, not phoneme identification, so the recognition grammar cannot be turned off and will influence the outcome of many so-called percpetion tasks, simply because the respondents cannot separate the output of the perception grammar from the output of the recognition grammar. (Boersma, 2000, p. 4)

2.2.5 Featurally Underspecified Lexicon

The Featurally Underspecified Lexicon (FUL) model does not assume an intermediate step to mediate between the acoustic signal and the lexicon. It hypothesizes that the underlying lexical representations are directly accessed on the basis of the speech signal (Lahiri and Marslen-Wilson, 1991; 1992; Lahiri and Reetz, 2002).

---

[7] Articulatory aspects of sounds (Articulatory Form, ArtF) are accessed in speech production but not in perception. This is exemplified in Boersma (2005b): Listeners tend to chose an exemplar as a category-prototype that is much better than their normal productions, i.e. an overarticulation. Boersma claims this to follow from the fact that in perception, articulatory constraints that require a minimization of effort and, thus inhibit overarticulation, are not involved.

This model assumes that the lexical entries consist of underspecified distinctive features. Nondistinctive and predictable elements, that is features that would not produce a lexical contrast are not part of the underlying mental lexicon (Archangeli, 1988). During speech perception, features are extracted from the acoustic signal that correspond to coarse acoustic cues. These features are transformed into phonological features which form the access to the mental lexicon. This process of extraction and transformation is speech-specific but it is not reliant on the particular language: all phonetic features can be extracted. In contrast to the aforementioned model by Boersma, language-specific processes are transfered to the lexical level.

The features enter the lexicon directly and the search for a word candidate begins. As soon a possible word is found, interactions between different levels of grammar occur, including prosodic units, syntax, and semantics. Also, postlexical and morphophonological rules start to interact with word selection. Now, larger structures, such as segments and syllables can play a role as well.

Access to the lexicon and the choice of a lexical entry function via a matching process which has three possible outcomes: (a) match: the extracted feature (surface feature) and the lexical feature are the same, (b) mismatch: the surface feature and the lexical feature are not the same, they clash, (c) no-mismatch: the surface feature is not listed in the lexicon, i.e. the extracted feature is not distinctive.

The choice of a lexical candidate depends on the goodness of fit which is determined by the number of matching features that result from the ternary matching process. A no-mismatch does not necessarily lead to rejection but candidates that contain a mismatch are rejected. Variation in the signal, allophony, and neutralization are intercepted by this matching process.

More support for lexical selection comes from the assumption that the extraction of features and the matching process is going on continually. The listeners does not "wait" until larger phonetic units or segments or other intermediate representations can be formed. This immediate access to the lexicon enables a quick interaction with other lexical levels, such as knowledge about morphophonology, syntax or semantics.

2.2.6 Exemplar theory

So far, the accounts that were presented are based on the assumption that the lexical entries consist of discrete, abstract components, that are lexically distinctive. Exemplar models, in contrast, assume that the stored lexical representations are not completely abstract. Instead, they claim that acoustic and phonetic details of words are stored in the lexicon. Therefore the acoustic signal does not have to be normalized or changed during speech perception. As described by Pierrehumbert (2001), phonological categories are represented by clouds of experienced instances of sounds.

A word may appear in an unlimited number of different acoustic surface forms. However, it can be observed that the values of acoustic features form clusters. This becomes immediately apparent if one records many exemplars of different vowel categories and measures the first and second formants. Even if only exemplars by one speaker are considered, a great amount of variation appears. The individual values of vowel may partly overlap, but as a larger amount of vowels is collected, distributional peaks will appear (cf. Hillenbrand et al., 1995). Accordingly, memory clouds are organized in relation to the physical characteristics of the exemplars: similar instances are close to each other, dissimilar ones are stored further away in a multidimensional parameter space.

The frequency and recency of experiences determines a "resting activation level": recently and frequently encountered exemplars have a higher activation level than rare and old ones. The exemplar clouds can, thus, be seen as statistical distributions along all phonetic parameters and they carry phonological category labels. In speech perception, the incoming acoustic signal is compared to the items in an exemplar cloud by comparing of the distance between phonetic parameters and by taking the activation level into account.

The observation that languages differ systematically in fine phonetic details imply that speakers have learned these details. Furthermore, listeners have very detailed memories of utterances. Models that assume abstract lexical representations cannot explain such observations in a linguistic framework but have to assume an associated paralinguistic memory.

2.3 Models of second language perception and learning

The models of speech perception that were described so far, focus on mechanisms for the transformation of the acoustic speech signal into representations that can be used to recognize words. In contrast, models of second language perception and learning aim to predict which perceptual abilities a listener has left as she encounters unfamiliar sounds and has to handle sound variations that categorize differently than those in her native language. Predictions for learning are incorporated in most of the models and it is emphasized that the detection of differences between nonnative sounds or between native and nonnative sounds is the prerequisite to learning.

2.3.1 Speech Learning Model

The Speech Learning Model (SLM) developed by J.E. Flege (e.g. 1981, 1987, 1991, 1993, 1995) attempts to predict the ability to learn sounds of a foreign language over the life-time. The studies that SLM is based on, mostly quantified learning success in terms of production. However, the ability to perceive nonnative sounds and sound contrasts is seen as the basic requirement for learning which makes SLM interesting for studies on second language perception.

Perceptual interaction between native and nonnative speech sounds are assumed to take place at a phonetic level, in a common phonological space in which all sound representations are represented with phonetic details. All predictions focus on this intermediate perceptual level: "Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level." (Flege, 1995, p. 239). This implies that the perception of detailed phonetic features of nonnative sounds is the starting point for the learning process. This assumption is based on observations that perceptual difficulties often arise within certain phonetic contexts and not absolutely for all instances of a phoneme (Polka, 1991).

Generally, SLM postulates that nonnative speech sounds are perceived in terms of their similarity to native phonetic categories, a process that is termed as "equivalence classification".[8] The perceived difference to existing phonetic categories decides whether a new mental representation for a foreign sound can be formed. If the perceived difference to a native category is too small, the nonnative sound is perceived as a variant of this category and no new phonetic representation will be established. If the nonnative sound is perceived as rather dissimilar, a new phonetic category can be founded more easily and the learning success will be higher.

With increasing age the effect of the native language on the perceptual abilities to detect subtle phonetic differences between nonnative and native sounds decreases. According to Flege (1995) this is the case because less attention to phonetic detail is needed when the native categories are established and attentional capacities are applied in other domains. Further, effects of anticipation occur that blur a detailed analysis, and finally, fortified perceptual strategies exist concerning specific acoustic cues, such as they are described in the models of speech perception in section 2.2.

Since it is assumed that all speech sound representations co-exist in the same phonological perceptual space, they must interact (Flege et al., 1997). Native categories may change in order to maintain phoneme contrasts or they may open up to encompass new categories. Examples of such "interlingual identifications" were reported for VOT values in plosives for people with a long experience of the second language (Flege, 1987; Flege and Hillenbrand, 1984): the plosives were produced with VOTs that lay between those of the first and the second language.

The prediction of learnability as measured in successful production, of a "new" second language category was partly supported by Guion et al. (2000). Sounds that were rated similar to native categories were not produced as well as some "new" sounds. Another investigation of language learners' ability to perceived and produce nonnative sounds by Munro et al. (1996) could only explain some of the learning success from the distinction between "new" and "similar" sounds. Mainly, the rather loose definition of "new" and "similar" sound has been criticised (Markham, 1997). To better predict learnability, the relation of native and nonnative sounds should be specified further. Particularly with regard to the assumed context dependence for phonetic perception the distinction between new and similar sounds is rather coarse.

For EEG and MEG studies, that test whether processing of native and nonnative sounds can be differentiated at an early level of auditory processing, Flege's claim about a universal perceptual space is of interest. Chapter 4 reviews electrophysiological studies on second language learning, that support the SLM model in this respect and show that native and second language category representations interact.

---

[8] The classification system provided by the International Phonetic Alphabet can serve as an initial guidance to predict perception and learning (Flege, 1991). Subsequently, the perceived difference to native categories has to be tested individually (Flege, 1992; Guion et al., 2000).

2.3.2 Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) (Best, 1993; Best, 1995) is specifically concerned with the perception of nonnative sounds and sound contrasts. It attempts to explain the perceptual difficulties of nonnative listeners within the framework of Direct Realism, that was described earlier in section 2.2.2. The PAM is the most detailed model on second language perception. Its main claim is that nonnative sounds are assimilated to native categories. The predictions of PAM address an early state of second language learning, in fact, they apply to the initial encounter with a foreign language.

Based on the claim that articulatory gestures are directly detected in the speech signal, PAM proposes a distinction between perceived speech-like and nonspeech sounds. Articulatory regions, types of articulation, constriction, and dynamic dimensions that are not employed for native categories but are located inside the native articulatory region are perceived as potential phonological units. If the articulation falls outside the native range, the sounds are not perceived as speech and inference on their articulatory constellations cannot be drawn.

Along these guidelines, six assimilation patterns for nonnative speech sound contrasts are possible and the respective discrimination levels are predicted:

Two-Category Assimilation - the nonnative sounds are associated with two existing native categories. Discrimination is excellent.

Category-Goodness Difference - both nonnative sounds are assimilated to one native category. One sound is perceived a "bad" member of this category. Discrimination can be good, depending on the perceived quality deflection.

Single-Category Assimilation - both nonnative sounds are assimilated to one native category and they are both perceived either as good or both as bad category members. Discrimination will be very poor.

Both Uncategorizable - the nonnative sounds cannot be assimilated to native categories, but they are perceived as speech. Discrimination can be good when they are far away from native categories or poor when they are close to one native category.

Uncategorized versus Categorized - One nonnative sound is perceived as a native category exemplar, the other not. Discrimination will be very good.

Nonassimilable - the nonnative sounds are not perceived as speech sounds because their articulation is outside the native range. Discrimination will be very good.

In summary, PAM assumes that the better a nonnative sound contrast is mapped onto distinct native categories, the better the discrimination of this contrast will be. Subsequent studies have supported this idea (e.g. Guion et al., 2000; Best et al., 2001).

Conceivably, nonassimilable sounds that are perceived as nonspeech are rarely encountered. One example is the perception of Zulu-clicks by American English listeners that Best et al. (1988) tested. All nonnative listeners described the clicks as nonspeech sounds such as water drips, claps, percussion instruments, etc., but discrimination was like that of native listeners. In a dichotic listening test with click sounds, Best and Avery (1999) observed a right-ear advantage for native listeners which suggests that the left hemisphere was more active in the auditory processing of these sounds.[9] No significant advantage of one side of presentation was observed for the nonnative listeners. Therefore, it is possible that the experience with clicks as phonemes elicited a stronger involvement of the speech dominant hemisphere, whereas nonnative listeners used a general auditory analysis for distinction.

2.3.3 Phonological interference

Brown (2000) adheres to a very strong role of phonology in speech perception. She models the influence of the native language on second language sound perception and learning with reference to radically underspecified feature geometry (Archangeli, 1988; Clements and Hume, 1994; Clements, 2003). According to Radical Underspecification, only distinctive features are represented at the phonological level (compare 2.2.5). Brown assumes that speech is perceived in terms of distinctive categories so that linguistically relevant contrasts are robustly perceived. The rest of the signal is systematically ignored:

> Those variations in the acoustic signal that do not contribute to differences in meaning are simply not perceived by the listener. (Brown, 2000, p. 19)

This view is based on a theory of first language acquisition which claims that a child starts out with a full set of universal phonetic features so that every sound can be perceived. The recognition of phonemic contrasts in the native language leads to a "pruning" of the universal initial feature set. This means that features that are not used distinctively in a language dissapear from the mental representation. The individual feature geometry is reduced to the set that is needed in the native language (Brown, 2000; Brown and Matthews, 1997).

Thus, nonnative sound contrasts that are based on distinctive features that are not used in the native language are very hard to perceive and cannot be learned perfectly. In fact, even after long years of learning, such contrasts will remain difficult. However, if a feature is present in the native system, even if it is used to distinguish a different phoneme contrast, the nonnative contrast can be perceived and learned.

Brown has a very radical view on the influence of phonology on perception that does not leave room for auditory learning or acoustic factors, that SLM and PAM incorporate. Moreover, Brown does not take into consideration that different types of sounds may be more salient than others, as it was shown in the first part of this chapter. However, her account may become more convincing if it is seen as a claim about a limit of cognitive-linguistic extraction, not literally as a limit of perception. While perception may or may not detect acoustic details, the influence of information extraction by the phonological system is so strong that sensory perception does not get a chance to establish new phonological representations. In relation to the present study that is described in this thesis, Brown's account is relevant insofar, as it predicts a reduced perception of non-distinctive features. This can explain the reviewed findings on the perception of allophonic contrasts (section 2.1.2)

---

[9] In dichotic listening tests two different stimuli or messages are presented simultaneously to both ears and subjects are asked to report the information from one or both ears or they are asked to detect a target stimulus. For speech stimuli, subjects are usually more accurate in reporting the information that was presented to the right ear. Since auditory information is projected bilaterally, but with an advantage for the contralateral pathways, dichotic listening tests are used to test for lateralization in auditory perception. More information on dichotic listening tests can be found in Hugdahl (1998).

and, likewise, would predict a reduced perception of the German allophones [] and [x], which were tested in the present study. However, a verification of Brown's predictions would require a longitudinal study on acquisition.

2.3.4 Native Language Magnet effect and Native Language Neural Commitment

The Native Language Magnet (NLM) effect and the Native Language Neural Commitment hypothesis (NLNC) by P. Kuhl are concerned with native sound perception, but have direct consequences for the processing of foreign speech sounds. The NLM effect was proposed when Kuhl (1991) observed that the perceived distance between vowels exemplars of one category ([i]) varied in relation to their acoustic distance to a very good member of this category, the prototype, as it was mentioned in section 2.1.1 (see also Kuhl and Iverson, 1995; Iverson et al., 2003). Prototypes refer to people's judgements of the best exemplars of a category and they contain all characteristic features of a category (Rosch, 1975; Rosch and Lloyd, 1978). Discrimination between sounds in the physical vincinity of the prototype was lower than between non-prototypical category exemplars.[10] Figuratively speaking, the prototypes function like magnets that warp the perceptual space.

This has implications for the perception of nonnative sounds. When listening to nonnative speech sounds, the reduced perceptual sensitivity around the magnet can lead to an insensitivity to acoustic cues that are critical for a nonnative sound contrast (Iverson et al., 2003; Kuhl and Iverson, 1995). Inversely, acoustic differences that are irrelevant in a foreign language could be enhanced by the native perceptual system (Boersma and Escudero, 2004).

The NLNC hypothesis (Kuhl, 2000; 2004) was based on these findings. NLM assumes abstract representations of phoneme-prototypes. Along these lines, the NLNC proposes that during first language acquisition neural networks are formed that are particularly sensitive to native sound categories and prototypical characteristics of these categories (cf. Guenther and Gjaja, 1996). The development of these networks is based on statistical learning which babies are very good at. NLNC does not assume innate phonetic dispositions like a Universal Grammar, but innate biases and strategies to detect distributions and to organize neural structures accordingly. "Parentese" or "motherese", the speaking style that adults use when they talk to children, aids this process because it exaggerates distinctive acoustic cues. In fact, children have been found to prefer this style of speaking over normal speaking styles. When new phonetic experiences do not change the memorized distributions any more, the neural networks stabilize. Most of this learning appears to take place during the first year of life (compare section 2.1.1). Later experience with new phonetic patterns could potentially change the established networks, although this may take a lot of experience and exaggerated productions, similar to "parentese" could be of help.

By making assumptions on neural networks that underly language-specific responses, the NLNC bridges the gap between linguistic representations and neural responses. Learning and the establishment of mental representations for linguistically relevant units is seen as a

---

[10] The NLM theory has been challenged by later findings, for examply by Lotto et al. (1998), Thyer and Hickson (2000), and by an EEG study by Sharma and Dorman (1998) which is described in Chapter 4. Thyer and Hickson could not replicate the NLM effect. They suggested that the non-prototypes that were used in Kuhl's study (1991) may have belonged to different vowel categories ([e] and [I]).

modification of neuronal cell connections.[11] In these assemblies, the cell's firing preference is highest for prototypical examples of a phoneme. Processing native speech sounds activates established cell assemblies and yields strong and fast responses. The processing of unfamiliar speech sounds may involve more inefficient processing which leads to longer response latencies (Zhang et al., 2005). The detection of such language-specific neuronal tuning that results in synchronized cell responses, can be measured by EEG and MEG.

After this review of the effects of native language representations on the perception of native and nonnative sounds, the next chapters address the neural correlates of these observations. Chapter 3 provides basic principles on the origin of the neural responses to acoustic signals, their measurement, and the detection of language-specific responses, which allows an understanding of the electrophysiological research that is covered in this thesis.

---

[11] This notion is based on Hebb's postulate of learning as a modification of synaptic connections between neurons (Hebb, 1949).

Chapter 3

Auditory event-related potentials and electrophysiological methods

3.1 Introduction

This chapter explains the origin of auditory event-related potentials and details how these signals can be measured. The components of the recorded signals are introduced and discussed with an emphasis on the components that are relevant for the analysis of early auditory speech processing. Thus, this chapter provides the background for an understanding of the results, that studies on the influence of the native language on speech sound processing have collected. Especially, the mismatch negativity component of the auditory event-related potential is described, that served as a probe for native language influences in the present MEG study.

3.2 Auditory cortical areas involved in speech processing

The auditory cortical areas lie on the superior temporal plane, mostly inside the lateral sulcus. It is difficult to study the neuroanatomy of the human cortical auditory areas and the function of specific parts. Therefore, findings from animal studies, especially primates, are used as a reference (Scott and Johnsrude, 2003; Semple and Scott, 2003). This points to the fact that many details about the human auditory system are not known yet. However, the comparison with auditory processing in primates has proven to be very useful. The most recent models of human neuroanatomy of speech perception that will be described in Chapter 4 are based on principles of auditory processing in animals.

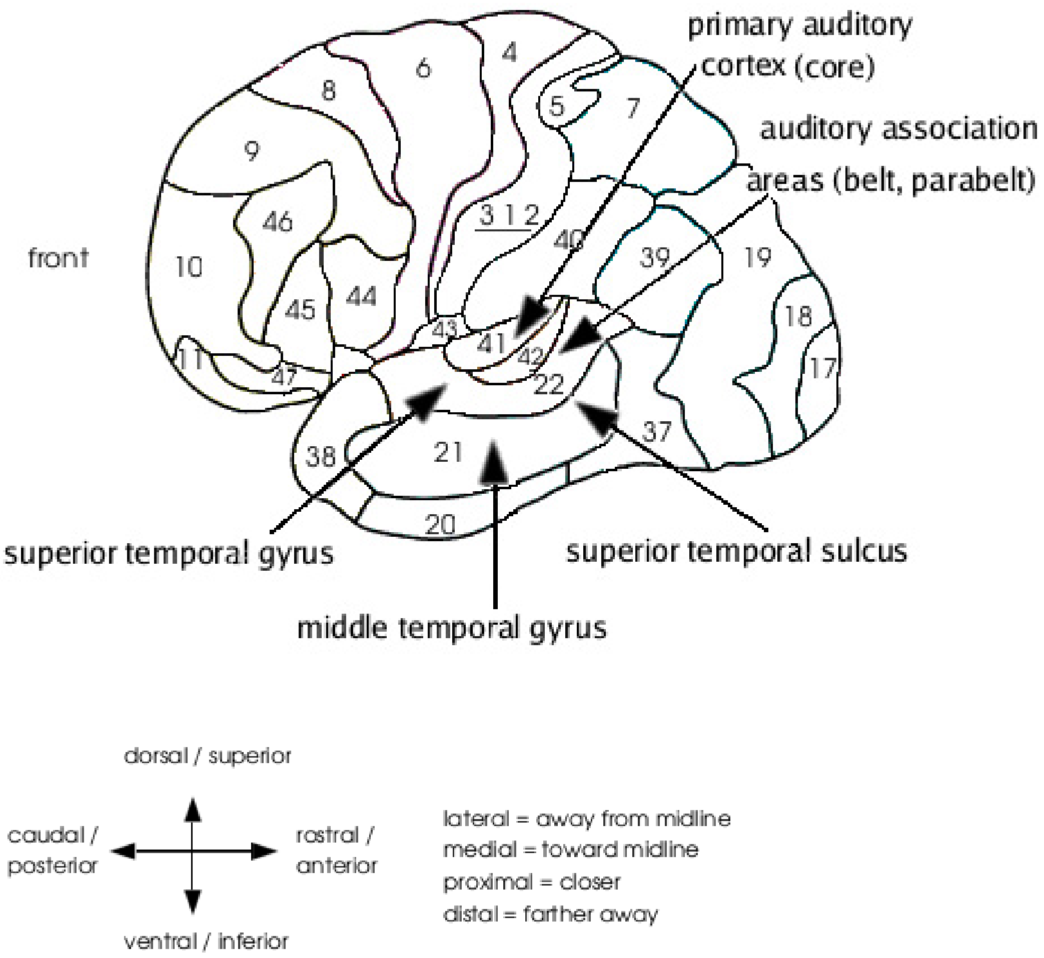


Figure 3.1: Lateral view on the left hemisphere including Brodmann areas (BAs). Areas of main interest for auditory speech processing are labeled. The figure was adapted from Demonet et al. (2005). Anatomical terms of location are explained below the figure.

Acoustic events are projected from the cochlear nerve to the cochlear nuclei in the brainstem and thalamus. The output from the thalamic areas arrives at the primary auditory cortex, the "core" regions which are located in Brodmann's area 41 (BA 41) in Heschl's gyrus. The left and right hemisphere core areas are strongly connected via the corpus callosum (Semple and Scott, 2003). Up to this level, similar processing for all types of sounds, speech and nonspeech, has been observed (Scott and Johnsrude, 2003; Bernstein, 2005). Input to the core areas is passed on to the surrounding "belt" areas which are considered to be located in BA 42 (Bernstein, 2005).

Core and belt areas display tonotopic organization and show different responses to noises of varying bandwidths (e.g. Wessinger et al., 1997; also see reviews in Semple and Scott, 2003; and Scott and Johnsrude, 2003). Tonal stimuli evoke most activity in the core areas. Spectrally more complex stimuli activate the belt areas (Wessinger et al., 2001; Zatorre and Belin, 2001). Zatorre and Belin (2001) noted that the left anterior superior temporal areas were more engaged in temporal processing and the corresponding areas in the right hemisphere were more active in spectral processing.

The belt areas project to lateral parabelt fields. In human listeners these areas could refer to BA 22, extending to the superior temporal sulcus and the superior temporal gyrus. Parabelt areas also receive input from nuclei in the thalamus (Semple and Scott, 2003). Responses specific to speech sounds have been observed in these areas, especially in the left superior temporal sulcus (see reviews in Scott and Johnsrude, 2003; Bernstein, 2005).

The parabelt areas then further project to the temporal, parietal, and frontal lobes.

Generally, auditory processing runs basically sequential from core to belt and parabelt areas (Scott and Johnsrude, 2003; Semple and Scott, 2003). All of these regions consist of multiple histologically and probably functionally different subregions (Kaas and Hackett, 2000; Morosan et al., 2001; Velenovsky et al., 2003). The contributions of these different areas to auditory processing are still not fully understood (Scott and Johnsrude, 2003; Arnott et al., 2004).

The current discussion of general sound and speech processing in the auditory cortex focusses on the temporal course of processing, the contribution of the left and right auditory and adjacent areas.

3.3 Origin of electrophysiological signals

In order to understand which physiological processes correspond to EEG and MEG measurements, an explanation of the origin of electrical signals in the brain is given.

Nerve cells in the auditory areas are mainly aligned vertically to the cortex' surface. Neurons consist of the cell body or soma, dendrites, and the axon.[1] The surface of dendrites is covered with junctions that receive incoming signals, the postsynaptic potentials. The axon, deriving from the axon hillock, conducts impulses away from the soma to the synapses of other neurons. The structure of dendrites and axon varies according to neuronal function. Generally, dendrites are short, branching fibers extending from the cell body that enlarge the surface area available for receiving incoming signals. Axons are single fibers with a length of a few millimeters up to 2 meters. Axons of more than 1 µm diameter are surrounded by insulating myelin segments, that are separated by gaps of 1 to 4 µm, the Nodes of Ranvier. The end of an axon has a high number of branches that end in small knobs, the Terminal Buttons. The Terminal Buttons and the end of the dendrites of the postsynaptic cell form synapses which are the points of connection between two neurons. One neural cell has more than 10,000 synapses. At the synapses, changes occur constantly which form the basis for adaptation to new stimuli and tasks.[2] These processes are the prequisites for the formation of neural networks that NLNC assumes to underly language acquisition and phonological representations (see Chapter 2, 2.3.4).

Electrical potentials emerge at a neuron's membrane due to the differential concentration of ions within and outside the cell. The membrane is nonpermeable for most ions. The difference-potential between the inside and outside of a cell is also kept stable by the Na/K

---

[1] Graphic illustrations of neurons and the transmission of action potentials can be found in many related textbooks, see, e.g.Schandry (1998).

[2] Synaptic changes include strengthening of connections between neighboring cells, growth of new dendritic spines, or changes of the receptivity. For more information on these complex processes see Polnau and Kossel, (2002).

pump which drives Na+ ions out of the cell. Postsynaptic potentials at the membrane force the membrane's molecule channels to open and positive sodium ions (Na+) diffuse into the neuron. An action potential is generated when a threshold of about -50 to -60 mV is reached. This works according to the "all-or-none" principle: If the neuron does not reach this critical threshold level no action potential emerges. But when the threshold is attained, the process cannot be stopped.

As more positive ions enter the cell, the membrane becomes more and more permeable. In about 0.2 to 0.5 ms the cell's potential rises to ca. +40 to +50 mV. At a depolarization of about 30 mV, the sodium channels close and the potassium (K+) channels open. Now, K+ leaves the axon due to the greater positive potential inside the neuron and the reversed voltage levels.

This compensates for the Na+ influx and leads to a repolarization. Until repolarization, the membrane cannot be excited by new stimuli for a refractory period of about 2 ms. The refractory period further causes the unidirectional flow of electricity. Finally, the Na/K pump restores the original ion concentration.

The amplitude of one action potential is about 100 mV. The action potential starts at the axon hillock and propagates along the axon. The electrical impulse jumps from one Node of Ranvier to the next at a rate as fast as 120 meters/second because only at the nodes an interchange of potentials can take place. Action potentials lead to very brief (1 ms, max. 10 ms) local currents (Hämäläinen et al., 1993; Malmivuo and Plonsey, 1995; Rowan and Tolunsky, 2003).

The impulse causes chemical neuro-transmitters at the synapses to diffuse through the synaptic cleft and attach to the receptors on the postsynaptic cell. Excitatory neurotransmitters cause the ion channels at the postsynaptic cell to open, leading to an excitatory postsynaptic potential (EPSP). This prepares the postsynaptic neuron to generate an action potential. Inhibitory neurotransmitters prevent the generation of an action potential at the postsynaptic neuron. One postsynaptic potential has a duration of approx. 10 ms and an amplitude of ca. 10 mV. Longer durations of 50 to 200 ms have been observed as well (Hämäläinen et al., 1993; Rowan and Tolunsky, 2003).

Figure 3.2 displays the potential changes at one cortical cell and shows their dipolar nature. The EPSP leads to an influx of positive ions into the cell and creates a negative potential outside the cell. The intracellular current flows towards the soma while the extracellular current, the volumetric current, moves in the opposite direction towards the dendrites.

Because of their short duration, action potentials hardly contribute to the signals recorded by electroencephalography and magnetoencephalography. Mainly, the postsynaptic potentials are recorded by these methods.

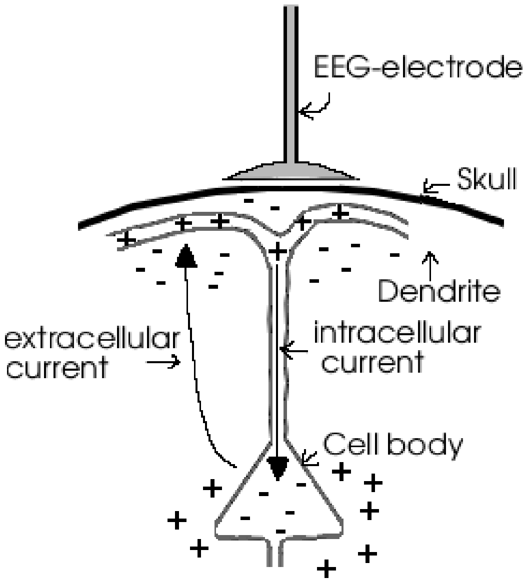


Figure 3.2: The intra- and extracellular currents at one pyramidal cell in vertical position to the skull are shown. At the skull's surface an EEG-electrode detects the negative potential of the volumetric currents (Figure adapted from Birbaumer and Schmidt, 1996, p. 492).

3.4 Recording of electrophysiological signals

Electroencephalography (EEG) and magnetoencephalography (MEG) mainly record the postsynaptic potentials. Both methods are non-invasive, provide a high temporal resolution, and can therefore record very early neural responses to external stimuli.

Figure 3.3 shows the relation of the electrical current and the surrounding magnetic field. The opposite electrical charges and the magnetic poles respectively form dipoles. According to the Biot-Savart law every electrical current produces a magnetic field (Williamson and Kaufman, 1990). Neural activity can be modelled by dipoles. A dipole (M) is characterized by its strength, called "moment", that is the product of pole strength (Q) and the distance between the poles (a): M = Q * a. The unit for M is ampere-meter (Am).

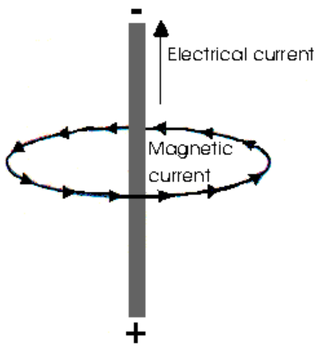


Figure 3.3: Electrical currents and the surrounding magnetic field.

As shown in Figure 3.2, every single cell may be regarded as a dipole. However, in MEG and EEG data analysis the activity of many thousands of neural cells are modelled as one or several dipoles. Detected signals in EEG and MEG studies, therefore, always reflect responses of whole cell populations over an area of at least 1 mm^2 (Hämäläinen et al., 1993; Hari et al., 2000). The more synchronous the cells respond, the stronger the detected signal will be.

Still, the amplitude of neural potentials is very low with about 75 µV in healthy subjects and the signal-to-noise ratio is poor (ca. 0.05:1 to 1:1; Regan, 1989). Eye-movements, heart-beat, muscular activity, pulsing arteries, neural activity e.g. the spontaneous alpha-waveAlpha activity has a frequency of ca.8 to 13Hz and is strongest when one is awake while the eyes are closed. Visual distractors block the alpha-wave,[3] and other internal influences produce artifacts. External sources of noise are, for example, the earth's magnetic field (10^{-4} Tesla), the laboratory, and traffic.

In order to measure neural responses to certain stimulus events, repeated measurements are carried out and their results are averaged. It is assumed that the repeated presentation of one stimulus event always causes the same response. Furthermore, it is assumed that brain activity that is not related to this event occurs randomly. Thus, averaging over multiple recordings with the same stimulus improves the signal-to-noise ratio and accentuates the response to a certain stimulus, the so-called event-related brain potential (ERP) and its magnetic correlate the evoked magnetic field (EMF). By definition, ERPs are all electrocortical potentials that occur before, during, and after a stimulation (Birbaumer and Schmidt, 1996, p. 500).

Artifacts, such as eye-blinks, are rejected,[4] the sources of the signals are estimated, and the components, the positive and negative peaks of the ERP are analyzed.

3.4.1 Electroencephalography

EEG-electrodes record extracellular currents at the surface of the skull. Neural activity, therefore, is recorded as a negative potential (compare Figure 3.2). Positive potentials, on the other hand, can signify activity from parts of the brain in which the neurons are aligned in the opposite direction, e.g. at the base of the skull.

The scalp electrodes are attached to the subject's head according to the international 10-20 system of electrode placement. Their locations are indicated by letters and number (see Figure 3.4). The electrodes are either placed individually by applying electrode paste or a cap with embedded electrodes.

---

[3] Thus, in EEG and MEG experiments with passive listening the subjects are usually instructed to read a book, watch a video, or solve visual tasks.

[4] Eye-blinks artefacts can be removed by using a filter that matches known characteristics of blink signals.

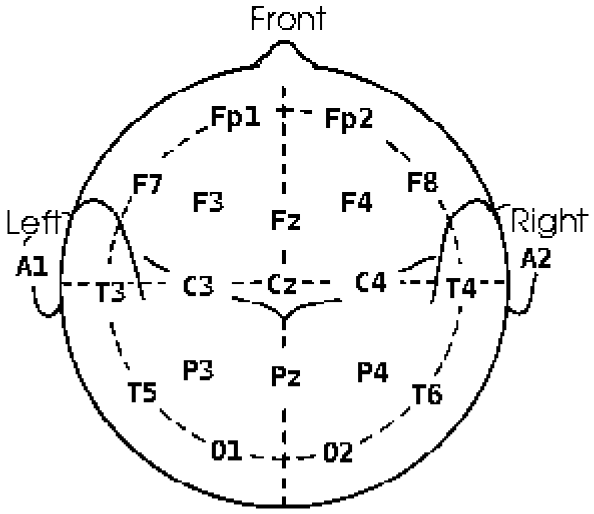


Figure 3.4: The letters F, T, C, P, and O stand for frontal, temporal, central, parietal and occipital cortical zones. Even numbers refer to the right hemisphere and odd numbers refer to the left hemisphere. The smaller the number, the closer the position to the midline. The letter z refers to an electrode placed on the midline. Fp stands for frontal pole. The reference points at the earlobes are labelled A1 and A2. The name of the system refers to the 10% or 20% interelectrode distance (the figure was adapted from Birbaumer and Schmidt, 1996, p. 490).

Generally, two electrodes are connected to an amplifier which measures the difference in potential. There are two types of recording, bipolar and referential recording. For bipolar recording, successive electrodes are linked, usually anterior-posterior or also transverse. The voltage at one electrode is compared to the voltage at the adjacent electrodes. Localization of activity is based on phase reversal. In referential recording the electrodes are not linked, but compared to a referential point of zero or small value (often the ears, see Figure 3.4). Localization of activity relies on amplitude (Birbaumer and Schmidt, 1996; Rowan and Tolunsky, 2003).

Electrical fields from vicinal areas can delete each other's detectable volumetric currents. Additionally, signals from neurons that are not oriented vertically to the skull's surface can extinguish the measurable fields. Therefore, a large number of vertical and parallel synapses must be active simultaneously to produce a detectable signal for EEG. Potentials measured at one electrode do not necessarily reflect near sources but can result from distant activity, as well as from subcortical sources.

3.4.2 Magnetoencephalography

Magnetoencephalography (MEG) records the magnetic fields that surround the neural electric PSPs, mainly of intracellular currents in cortical pyramidal cells.[5] As Figure 3.5 shows, mainly radial sources contribute to MEG measurements. Cortical pyramidal nerve cells are

---

[5] David Cohen was the first to record neural magnetic signals and the first one to measure magnetic brain activity with a SQUID system (Malmivuo and Plonsey, 1995, ch.14).

normally oriented orthogonally to the cortical surface. Thus, sources in the sulci are primarily recorded by MEG.

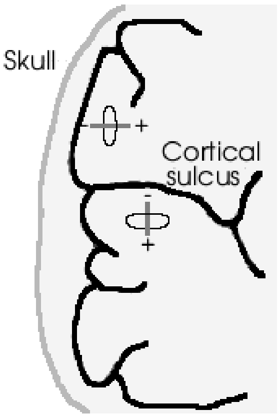


Figure 3.5: Schema of cell orientation in a part of the the cortex. Nerve cells are displayed as dipoles with electric potentials (plus and minus poles) and surrounding magnetic currents (ellipses). In the fissures, cells are oriented orthogonally to the surface of the cortex and their magnetic signals can be detected from the outside.

Neuromagnetic signals range from about 10^{-14}T to 10^{-13}T. In comparison, they are about 100 million times weaker than the earth's magnetic field. As mentioned above the signal-to-noise ratio is poor because the strength of the evoked magnetic fields is very low when compared to both external and internal potentials. To pick up the weak magnetic fields, Super Conducting Quantum Interference Devices (SQUIDs) are used. SQUIDs are the most sensitive detectors for magnetic flux.[6] They consist of ring-shaped supra conductors with one or two Josephson-contacts.[7] Supraconductors provide lossless current conduction. When cooled they suppress external magnetic fields (Meißner-Ochsenfeld-effect), and present a coherent quantum state of matter. Therefore, SQUIDS are placed in liquid helium at -269 degrees Celsius (Cohen and Halgren, 1993; Hämäläinen et al., 1993).[8]

---

[6] The foundation of SQUID technology was layed by the invention of the Josephson-contacts by Brian Josephson. The interested reader may find more on SQUIDS in Clarke and Braginski (2004). Apart from neurological research, SQUIDS are utilized for magnetocardiology. Other applications are material testing, e.g. of air-crafts or bridges, and geological research.

[7] Mostly, dc SQUIDS with two Josephson-contacts are used. A Josephson-contact consists of a supraconducting ring which is intermitted by one or two electric isolations. Supraconduction electrodes (Cooper-pairs) tunnel through these disjoints (Ernand Romani, 1990).

[8] There are other supra-conductors, that can be placed in liquid nitrogen at -196 degrees Celsius. However, these devices have a very poor signal-to-noise ratio.

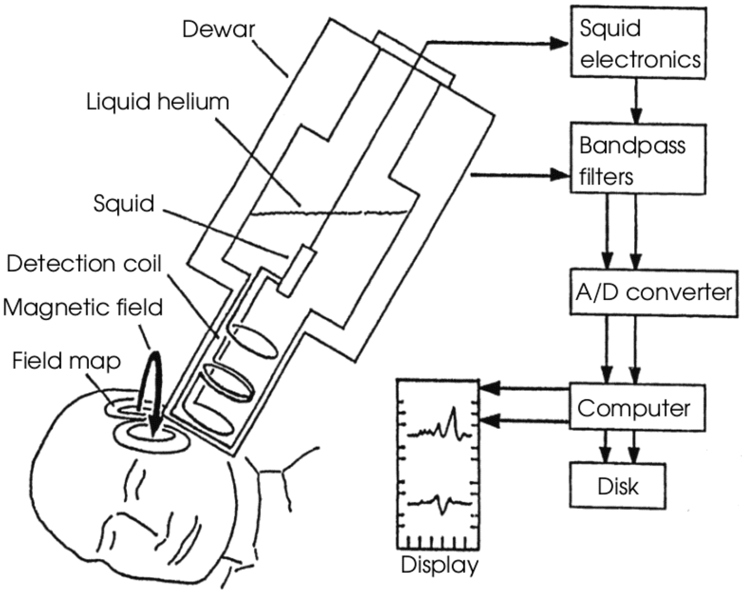


Figure 3.6: Recording with SQUIDS. The recorded data is bandpassed and digitized and transferred to a computer for control during the recordings and later stored on disk. For illustrations only one sensor is shown. Usually a helmet with multiple sensors is used. This figure was adapted from Williamson and Kaufman (1990).

External and internal magnetic fields are shielded by gradiometers, which distinguish between close and far sources and they reduce the influence of far-fields. Magnetically shielded rooms also add to undisturbed recording. These rooms consist of several layers of metal and one or more layers of copper or aluminum (Cohen and Halgren, 1993; Hämäläinen et al., 1993).

MEG has several advantages over EEG, although the SQUID technology makes it far more expensive. Magnetic fields are not affected by hair, skin, and skull tissues.[9] Thus, the signal is not distorted by artifacts of conduction. The sensors do not have to be attached to the subject's head. Instead, a helmet with sensors, called Dewar, is applied. MEG measures cortical signals from sources that are approximately 2 to 4 mm deep. Magnetic potentials from far areas and subcortical signals do not add to the magnetic measurements. As described above and shown in Figure 3.5, MEG mainly picks up signals from radial sources. Therefore, MEG is often preferred for the recording of auditory evoked potentials because the auditory cortex mainly lies in the Sylvian fissure. EEG and MEG data thus provide complementary information (Hari et al., 2000).

The differentiation of hemispheres can be difficult with EEG because auditory ERPs are best recorded at the vertex (for an example of source analysis with EEG data see Scherg et al., 1989). In comparison, MEG can separately record the activity in each hemisphere (Scherg, 1990; Hari, 1991; Hämäläinen et al., 1993).

---

[9] The conductivity of the skull, for example, is problematic for source location in EEG (Yao and Dewald, 2005).

3.4.3 Analysis

As mentioned before, the source of the recorded activity can be described by dipole models. One dipole represents many thousand neurons and often approximates more complex patterns of neural electrical sources (Hämäläinen et al., 1993; Tiitinen et al., 1999; Hertrich et al., 2002). The strength of the recorded signal varies with distance between source and sensor, the conductivity of the tissue, and the orientation of the dipole. The sources for MEG data are usually modelled by one or more, often two dipoles.

In order to determine the localization of the source of activity the forward solution and the inverse problem have to be considered. The forward solution calculates the distribution of an electric or magnetic field created by a source. The inverse problem is to estimate the sources of the recorded fields (Hämäläinen et al., 1993; Vroeijenstijn, 1998). No unique solution exists because every potential at the skull can result from an infinite number of sources. The inverse problem can only be overcome when the number of possible sources is reduced (Scherg, 1991).

Source modelling for EEG and MEG, therefore, is done by estimation. The analysis attempts to measure summed activity over a limited number of neural structures. In order to reduce the number of contingent sources as much as possible constraints from anatomy and physiology are considered, as well as data from imaging studies. Hereby, MEG has the advantage that mainly sources in the cortical sulci contribute to the detected signals.

One way to solve the problem of source localization is to iteratively fit one or few dipoles, e.g. bilateral position-symmetric dipoles (Hertrich et al., 2002). The position and moment of these hypothetical dipoles is varied until the difference between the forward solution created by the dipoles and the actual measured signals is minimal (Scherg, 1991).[10]

3.5 Auditory event-related brain potentials

The potentials past the brainstem and the thalamus in the primary auditory areas and beyond are of interest for psycholinguistic studies. They reflect stimulus characteristics, feature integration, auditory and phonetic/phonological memory, attentional factors, and subjects' sensory abilities. Of main interest for research on auditory speech perception are the potentials N100, MMN, and P300. Earlier auditory potentials also appear to play an important role in early sound processing, and therefore the responses up until N100 will be introduced as well.

A note on the terminology: the electric components are indicated by their positive or negative deflection with P or N and by their approximate latency. For example, the component N100 is the negative potential recorded by EEG about 100 ms after stimulus onset. The magnetic counterparts are indicated by adding the letter "m" for magnetic, e.g. N100m. Labels such as M50 specify that the component was observed by magnetic measurements. Small letters such as N100a, N100b, N200c, specify subcomponents of the main potential.

---

[10] This commonly applied method has also been used for the analysis of the MEG data from the present experiment that is described in Chapter 5. For other methods see Vroeijenstijn (1998).

In this text, the components are specified without "m", if no explicit reference is made to magnetic recordings. When referring to results of individual studies, the denotation is adopted from the respective publications.

3.5.1 Auditory Brainstem Response and Middle Latency Components

The Auditory Brainstem Response (ABR) is elicited 1.5 to 15 ms post stimulus and originates in the 8th cranial nerve and brainstem auditory structures, the lateral lemniscus and the inferior colliculus (Kraus and McGee, 1992). The middle latency response (MLR) is elicited 10 to 60 ms post stimulus. The MLR is generated by a complex system of neural structures within and outside the auditory pathway. It includes a series of negative and positive waves that originate in the primary auditory cortex.

Applications for ABR and MLR are tests of hearing abilities, e.g. in infants to assess the necessary function of a cochlear implant. So far, the contribution of ABR and MLR for the processing of speech sounds has not been extensively investigated. Imaging studies such as the one by Benson et al. (2001) that will be described in the next chapter do not indicate that speech sounds are processed differently than other sounds at these early levels. Also, in the present investigation on fricative perception these components do not play a role.

However, there is first evidence that ABR and MLR indicate language-based learning problems (Wible et al., 2005). Furthermore, the MLR component P50 apparently plays an important role in early sound encoding.

3.5.1.1 P50

The origin of P50 is roughly the primary auditory cortex, anterior to the location of the component N100 (see below) (Kanno et al., 2001). The function of P50 is yet unclear. A robust finding is that the amplitude of P50 in response to the second click in a click pair is reduced in healthy individuals whereas it is not reduced for schizophrenic patients (Adler et al., 1982). Therefore, P50 has been related to auditory sensory gating and the suppression of irrelevant information. Furthermore, the amplitude of P50 in relation to the later potential N100 changes during maturation (Cardy et al., 2004; Marshall et al., 2004). Also, the amplitude of P50 varies according to stimulus characteristics. Hertrich et al. (2002) reported an increase of the M50 amplitude, a decrease of the M100 amplitude for aperiodic signals, and the reverse effect for spectrally matched periodic signals (see also Kaukoranta et al., 1987; Mäkela et al., 1988; and Chapter 5, section 5.7.3.1).

A similar observation was reported by Chait et al. (2004) including a stronger left-hemispheric source for M50 in response to noise stimuli. Chait and collegues discuss the function of M50 in relation to M100 that is generally assumed to reflect the function of onset detection (see section 3.5.2.1). The paper raises the question: "why an onset detector would operate so late (100 ms post-onset) in the processing stream" (Chait et al., 2004, p. 2455). Chait and associates suggest that, instead, M50 might play that role whereas M100 in more specialized early cognitive processing, for example in figure-ground segregation. This illustrates that the function of early auditory potentials may be more complex than what is known so far. However, since these early components have only low relevance for phonetic processing and since the present study on fricatives did not concentrate on these components, this topic will not be pursued any further.

3.5.2 Late Auditory Evoked Potentials

3.5.2.1 N100

The component N100 is elicited by a relatively abrupt change in the auditory input signal. N100 is widely interpreted as the reflection of automatic onset detection (Näätänen and Picton, 1987).[11] Activity in the areas of the primary and secondary auditory cortex, and the parabelt area bilaterally is the source for N100 (Näätänen and Picton, 1987; Bernstein, 2005). N100 begins at ca. 60 to 80 ms and can last up to 160 ms (Woods, 1995). Functionally, N100 may be an attention-triggering process. It may also be the result of attention-triggering and a first sensory information pick-up which aids in the formation of a memory trace, so that stimulus changes can be processed (compare the suggestion made by Chait et al., 2004, in 3.5.1.1).[12] Furthermore, stimulus dependent changes in N100's amplitude, latency, and source location have been reported, suggesting that N100 reflects auditory feature encoding (Hari, 1991). This makes N100 very interesting for language-related studies, that for example investigate at which point in the process, speech signals are handled differently from nonspeech sounds.

According to Näätänen and Picton (1987), six or more subcomponents of the N100 potential can be recorded (see also Woods, 1995). They are distinguishable because they respond differently to certain stimulus characteristics and seem to differ in function (Näätänen, 1990; Näätänen, 1992; Woods, 1995). Thus, in their review of the N100 component, Näätänen and Picton (1987, p. 386) emphasize:

> [...] the auditory N100 wave does not reflect a single underlying cerebral process and therefore should not be considered as a unitary event.

The compositional nature of N100 has recently sparked a discussion about the memory-related component MMN (Jääskälainen et al., 2004). This issue will be taken up later in section 3.5.2.3.

The complexity of the neural mechanisms underlying N100 is demonstrated by the fact that N100 reflects a multitude of stimulus characteristics:

---

[11] As for an alternative interpretation, Näätänen and Picton mention that N100 could reflect synchronized neuronal reactions that generate a field potential and that cancel each other at longer latencies because they are not synchronized anymore. However, Näätänen and Picton dismiss this alternative in favor of the abovementioned interpretation because reactions to onsets seem to be typical for cortical neurons.

[12] Subjects' attention can affect N100, as well. This topic is reviewed in Rif et al. (1991).

Frequency: Spatial distinction of N100 sources could be related to stimulus pitch frequencies and, thus, a tonotopic organization has been suggested (Romani et al., 1982). Woods (1995) showed that high tones elicit a more frontal N100, mid-range tones elicit a central N100, and low tones lead to a posterior N100 signal.

In addition to tonotopy, the latency of N100 changes in relation to stimulus' pitch. Roberts and Poeppel (1996) reported the following tonochrony: low tones of 100 and 200 Hz and high tones of 4 and 5 kHz lead to long N100 latencies while mid-range tones elicit shorter N100 latencies.

Duration: The amplitude of N100 depends on stimulus duration, in that an increase in duration leads to an increase in N100 amplitude. At approximately 40 ms a saturation point is reached (Woods, 1995; Gage and Roberts, 2000) Rosburg et al. (2002), furthermore, reported differences in location of the N100 source in relation to stimulus duration.

Amplitude: Stimulus amplitude affects the amplitude and latency of N100 and perhaps also its source location. Vasama et al. (1995) and Stufflebeam et al. (1998) reported that latency decreases and amplitude increases with increasing intensity, while Stufflebeam and collegues demonstrated that the pitch-dependent latency is preserved.

Interstimulus interval (ISI): The N100 amplitude increases as the ISI is lengthened (Hari et al., 1982; Woods, 1995). This may suggest a habituation process. However, Budd et al. (1998) verified that a decrease in N100 is not a result of habituation to repeated stimulus exposure, but in fact follows from the refraction period of N100 generators. As ISI increases, more neurons can add to the N100 response.[13]

3.5.2.2 MMN

The mismatch negativity (MMN) reflects the response to changes in an auditory sequence (Näätänen, 1992). The source of MMN lies in the superior temporal auditory cortex (Hari, 1990; Alho, 1995; compare section 3.2).

The basic test arrangement to elicit MMN is the so-called oddball paradigm, in which a sequence of frequent standard stimuli is presented, that are physically equal or share a certain regularity. Infrequently, sounds are interspersed which deviate from the regularity of the standard events (Figure 3.7).

---

[13] The MEG-study by Tecchio et al. (2000) further shows the sensitivity of the M100 to the ISI. The behavioral ability of ten subjects was tested for finger tapping to the metronome of a 2 kHz tone while the clock pulse alternated between ISIs of 400, 490, 500, 510, and 600ms. At the same time, the auditory event-related fields were measured. The subjects were not able to consciously detect the 10ms changes, but still adapted their tapping rapidly. The amplitude of the M100 component increased or decreased with both the 100ms and the 10 ms ISI changes. Tecchio et al.suggest a link between the auditory cortex and the motor output that functions preconsciously. As many movements require very accurate timing and have to be exerted very fast after perceiving an auditory signal this conclusion seems very likely and with N100 may function as the connector. It would be very interesting to test whether such a similar link is responsible for the unconscious adaptation of VOT could be found between auditory speech processing and articulation, which may, for example, subserve conscious and unconscious imitation, for example for VOT.

The regular pattern of the standard events creates a sensory memory trace to which incoming deviant stimuli are compared. When a change between the representation of the regularities of the frequent sounds and the deviants is detected, the MMN emerges as a negative deflection peaking at about 100 to 250 ms after stimulus onset (Näätänen, 1992; Näätänen, 2001; Picton et al., 2000). This process of change detection does not require the listener to focus her attention on the sounds. On the contrary, MMN occurs non-volitionally and pre-attentively (Näätänen, 1992). Therefore, MMN is generally understood as a pre-attentive, automatic response (Picton et al., 2000; Näätänen et al., 2001), although its amplitude can be enhanced under attention (Aaltonen et al., 1987, cited in Näätänen, 1992, p. 147; see also Aulanko et al., 1993).

The sound changes yield a negative deflection over the frontal and/or temporal lobes at a latency of about 100 to 250 ms. MMN can be observed in the difference wave that is obtained as the standard-stimulus ERP is subtracted from the deviant-ERP. About 100 responses to a deviant stimulus have to be recorded to get a statistically valid result.[14]

---

[14] The need to record responses to such a large number of deviant stimuli leads to rather long recording times. Not only is this exhausting for the subjects, also the conductivity of electrodes in EEG experiments weakens over time. Thus, recently, modifications of the classic oddball paradigm were tested by Näätänen et al. (2004) in order to find a method that takes less time. Five different deviant types were presented (tones varying in frequency, intensity, duration, perceived sound location, and a deviant tone that included a 7 ms gap) in the context of standard tones in three paradigms: the traditional oddball paradigm with a standard-deviant ratio of 90:10 including one deviant type per block, a sequence in which three standards preceeded a deviants including all 5 deviant types, and a sequence in which every other sound was a deviant with all 5 deviants included. In this last sequence the first 15 events served as the standard. The MMNs in the last paradigm were almost identical to those obtained by the classical oddball sequence. Thus, this new design was proposed as a less time-consuming method to measure neural discriminative processing. However, until now this idea has not been copied.

A sequence of auditory stimuli containing frequent standard stimuli (black squares) and rare deviant stimuli (grey squares):

The stimuli elicit the N100 at about 100ms after stimulus onset. The response to the deviant stimuli (grey line) diverges from the response to the standards (black line) around 150ms. The difference (shaded area between standard and deviant curve) constitutes the MMN.

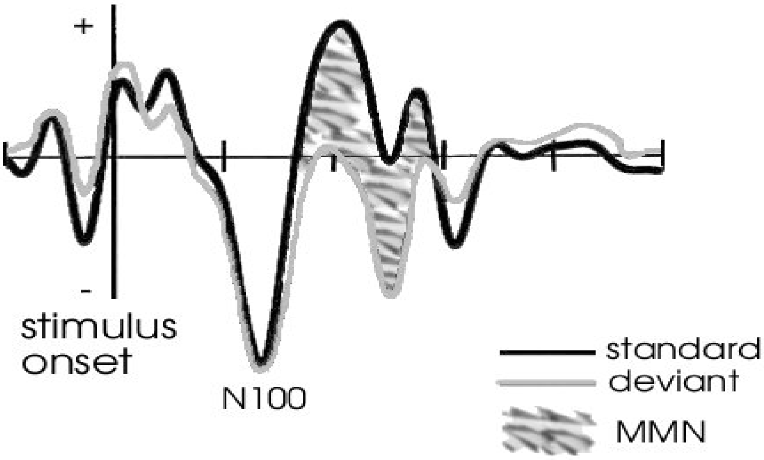


Figure 3.7: Schematic illustration of the oddball paradigm and the resulting MMN.

Sensory memory: The MMN reflects a mismatch between the sensory representation in auditory memory of the standard sounds and the new event (Ritter et al., 1995; Näätänen and Winkler, 1999). In other words, a MMN can only arise when a memory trace of the standard stimuli has been formed that differs from the characteristics of the deviant stimulus. Näätänen and Winkler (1999) and Näätänen (2001) proposed that the "central sound representation" (CSR) underlies MMN. CSR is understood as a percept of features that are integrated over a temporal window of about 150 to 200 ms:

> CSR, corresponding to the information content of the sound perception and sensory memory, emerges at the stage when the outputs of the central afferent processes are integrated and mapped on the neurophysiological substrate of sensory memory. (Näätänen, 2001, p. 3)

Alternatively, the "model adjustment hypothesis" by Winkler and collegues holds that the MMN reflects the adjustment of a neural model to the auditory background (Winkler et al., 1996; Winkler and Czigler, 1998). This idea emphasizes the detection of regular patterns instead of the detection of a mismatch. The representation of the standard must be readjusted when changes interrupt or stop the previous regular stream of events. Winkler et al. (1996) support their idea by showing that a new stimulus that initially yields MMN turns into the standard sound after only a few repetitions (see also Näätänen et al., 2004).

These two views on the MMN are compatible - Winkler and Czigler (1998), note that the detection of a deviant and the adjustment of memory may be interlinked and/or both be reflected by MMN.

What makes MMN applicable for the study of cognitive functions is the fact that it reflects processing of abstract regularities, long-term memory traces, and learning effects (Näätänen et al., 1989; Näätänen, 2001). Näätänen (1995) and Näätänen et al. (2001) describe the ability to detect abstract regularities and the incorporation of long-term memory as "primitive intelligence in the auditory cortex".

The incorporation of abstract information was demonstrated by Paavilainen et al. (1999). Several types of regular patterns were tested: first the standards were descending (or ascending) tone-pairs and the deviants were either ascending or descending pairs, respectively, while none of the standards were physically equal. Thus, the constant feature of the standards was abstract: the direction of frequency change. Secondly, the size of the interval of descend or ascend was varied, e.g. larger for the standard, smaller for the deviant. Finally, two tones were presented synchronously (like a musical chord) and their frequency intervall was either larger or small. In all of these tests significant MMNs emerged, indicating that the abstract rules inherent in the standard sounds had been extracted.

A related finding is reported by Paavilainen et al. (2001a). In their study, the regularities "the higher the louder" or "the higher the softer" elicited MMNs. Furthermore, Korzyukov et al. (2003) observed that memory for abstract patterns (descending or ascending tone sequences) could be reactivated after a relatively long ISI up to 10 seconds.

The illusion of continuity in a tone that is interrupted by noise also reflects in MMN, as Micheyl et al. (2003) demonstrated. Their tests contained continuous or interrupted standard tones which were contrasted with deviants that contained gaps filled with noise either bandpass-filtered in the same or a higher frequency region as the tone. Behavioral tests showed that the deviants were perceived as continuous when the noise had the same frequency range as the tone. The MMN amplitudes reflected this illusion for passively listening subjects.

Voice-characteristics are incorporated as well when the MMN is generated, as Titova and Näätänen (2001) have shown. In their study they used a female voice as standard, producing [e] and three different female and one male voice producing the same vowel as deviants. They conducted an additional behavioral experiment of dissimilarity rating. The results of both the behavioral and the electrophysiological experiment correlated and the largest MMN amplitude was recorded for the male deviants which were also rated most dissimilar.

Influence of long-term memory on MMN has mainly been observed in studies on speech sound processing (Näätänen, 2001). The next chapter will review how linguistic experience with phonetic categories, phonotactic patterns, and phonological rules affects the amplitude and latency of MMN. This is made clear especially when the responses of subjects with different native languages are compared. The observations that long-term memory representations are actively involved in neural responses signalled by the MMN have motivated the present research project that is described in this thesis.

Effects of stimuli and test-design: A greater magnitude of deviation often leads to a larger MMN amplitude with shorter latency and duration, at least for frequency (Näätänen, 1992). Näätänen (1995) illustrated this with unpublished data from Tiitinen et al. (1994) that shows an almost linear relation between frequency-deviation increase, enhancement of MMN amplitude and shortening of MMN latency (see also Sams et al. 1985). An increase in the difference of intensity or duration to the standard stimuli, however, did not yield a significant effect on MMN amplitude but shortened its latency in a study by Schröger and Winkler (1995).

The amplitude and latency of MMN is also influenced by the design of an experimental study. For example, a low probability of the deviant enhances the amplitude (Näätänen et al., 1983, cited in Näätänen, 1992, p. 149-150). No, or only a small, MMN is recorded for a 50:50 design (Sams et al., 1983, cited in Näätänen, 1992, p. 149; Näätänen et al., 2004). As a compromise between a minimum of recording time and obtaining a maximal result, the usage of a standard-deviant ratio of 85:15 or 90:10 is commonly applied.

Furthermore, ISI can change MMN latency (Schröger and Winkler, 1995). Commonly, ISIs of 500 to 700 ms have proven to be useful. The shortest ISIs that could be applied are 60 ms to 101 ms, the longest ones were about 10 ms (Näätänen, 1992). Winkler et al. (2002) found a reactivation of the standard memory trace even after 30 s when a "reminder" was presented.

Generally, the generators of MMN are the auditory areas in the left and right superior temporal gyri, but their location can vary for different types of deviations (Alho, 1995; Giard et al., 1995). For example, Alain and collegues (1999) showed that changes in tonal pitch elicited more lateral MMNs. Changes the pattern of a sequence were found to elicit more central or frontal MMNs (Korzyukov et al., 2003). Thus, changes in different stimulus features are detected by partly different neural populations.

Paavilainen et al. (2001b) tested whether the activity of the different generators would add up when a deviant included one, two or three feature changes. If the MMN amplitude would be twice as high for double deviants compared to single deviants and three times as high for triple deviants, the different MMN sources could be assumed to function independently. However, the study found a higher amplitude for the double deviants than for the single deviants, but no further increase for the triple deviants. This lead to the suggestion that the individual mismatch detectors interact for more complex changes.

Further applications: Further applications of MMN include tests on auditory capabilities in young children. This enables the detection of auditory problems at a very early age. It is already possible to record MMN in new-borns (Cheour-Luhtanen et al., 1996; Jansson et al., 2003; Jansson et al., 2004). For example, Alho et al. (1990), as well as Houtilanien et al. (2003) found an MMN-like negativity in newborns for tone-pitch changes that resembled adults' responses.

The development of language-specific neural representations in children has been studied under developmental and pathological aspects (Kraus et al., 1993). Concerning phonetic processing, MMNs in babies with about 10 months of age have been reported for phonetic category changes, even when acoustic variability had been added through the use of several talkers' utterances (Dehaene-Lambertz and Pena, 2001). Thus, young children appear to average over acoustic variability and show MMNs similar to adults (Dehaene-Lambertz and Gliga, 2004). However, several differences between adults' and childrens' MMNs have been recorded: longer MMN latencies (Cheour et al., 1997) and a more central and broader distribution of activity (Cheour-Luhtanen et al., 1996). Cheour et al. (2002) showed that new-born babies appear to have a shorter auditory memory span, since longer ISIs did not evoke a MMN for them. Moreowever, babies, unlike adults exhibit MMN during sleep, suggesting that neural activity and sensation functions differently in young children than in adults (for further review see Cheour et al., 2000).

MMN is also applied in adults' clinical studies. Since MMN is independent of attention and an objective measurement of auditory discrimination and memory it is well-suited for research on memory dysfunctions, such as dyslexia (Schulte et al., 1998; Kujala and Näätänen, 2001; Schulte et al., 2001; Maurer et al., 2003) and schizophrenia (Michie, 2001). MMN studies have also been conducted with autistic children, in order to determine whether their enhanced sensitivity to acoustic stimuli relates to pre-attentive neural abnormalities (Gomot et al., 2002). Furthermore, the presence of MMN in coma patients indicates their impending awakening (Kotchoubey et al., 2002; Näätänen, 2003).

3.5.2.3 N100 and MMN

The prevailing view holds, as described in section 3.5.2.1, that the N100 component reflects an afferent process, a general mechanism of onset detection, that encodes stimulus characteristics, whereas the MMN reflects a separate, memory-related process of change detection (Näätänen, 1992; Näätänen, 2001). This view has recently been challenged by Jääskeläinen et al. (2004). They proposed that the MMN is a by-product of an adaptation effect in N100-related neurons and does not reflect a separate component. For the realization of the present project on fricatives it was important to understand when neural responses could be expected that relate to language-specific memory. Therefore, the following paragraph addresses this discussion.

According to Jääskeläinen et al.'s "adaptation hypothesis" generators of the early N100 in the posterior auditory cortex with a latency of ca. 85 ms (the N1p component) are more broadly tuned on sound frequencies and adapt to external stimuli. When a new, deviant sound is presented, the response is suppressed, because of the broad tuning of the neurons underlying N1p, unless the difference between standard and deviant sound is very large.

Neurons in the anterior auditory cortex that produce the N1a component at approx. 150 ms are more finely tuned to sound frequency. Thus, the adaptation effect is not as strong and a response to deviant sounds is more likely. As the response to the standard is subtracted from the response to the deviant the resulting negativity reflects the activity of the unadapted N1a neurons.
Jääskeläinen et al.'s proposal is far-reaching: It dismisses a separate MMN component and it implies that the mismatch negativity is due only to afferent processes. Sensory memory and effects from long-term memory representations may not be involved. The function of attention switch to new auditory events, thus, would only be based on stimulus features and the refractoriness of the neural stimulus-response system but not on stored sensory percepts.

However, if no memory related mechanisms play a role, it is not possible to explain effects of learning on MMN, as e.g. found by Näätänen et al. (1993). In addition, studies showing the reactivation of sensory memory traces after several minutes as in Winkler et al. (2002) cannot be interpreted in terms of afferent responses only. Furthermore, Jääskelainen and collegues did not address the lateralization that has been observed for MMN in response to phonetic stimuli which may suggest a relation to language representations in long-term memory.

Näätänen, Jacobsen, and Winkler (2005) countered the ideas of Jääskeläinen et al. (2004) with a number of arguments which are comprehensible on the basis of the review of MMN in 3.5.2.2. They include the following:

1. Latency of MMN varies in relation to the magnitude of the mismatch. N1a, however, is time-locked to stimulus onset.

2. Changes induced by tone omission have been found to elicit MMN. Since the deviation consisted of no external stimulus, no N1 was present but the silent gap still presented a change from the standard background.

3. Many studies show MMN as a consequence of abstract concept formation.

4. Jääskeläinen et al.'s proposal is based on animal data (May et al., 1999; Ulanovsky et al., 2003) and human data by the authors, including fMRI recordings suggesting that the MMN is generated by the same sources that underly N1. However, Näätänen and collegues argue that the source analysis was incorrect because N1 was compared with N1 plus MMN activity. Therefore, the differences between the sources were blurred.

This discussion again points to the aspect of MMN that is crucial for the present study: its sensitivity to external auditory changes and its relation to recognition patterns in long-term memory. MMN incorporates both mechanisms and, hence, is relevant for research on mental language representations.

3.5.2.4 P300

Another component of the event-related potential that can be assessed by applying the oddball paradigm is P300, also called Late Positive Deflection (LPD). The generators of P300 have been localized slightly anterior to MMNm sources in the anterior superior temporal gyrus bilaterally (Alho et al., 1998b). P300 is related to the arousal or changes in subjects' attention, for example in response to stimuli that are very different from the standards, such as a telephone ring between simple tones (Sams et al., 1995; Alho et al., 1998b; Woods, 1992). Furthermore, for speech sound changes, LPD has been observed (e.g. Rivera-Gaxiola et al., 2000; Chapter 4, section 4.4.2). Studies on visual processing demonstrated that P300 is elicited by rare stimuli that are threatening, expected, surprising, or emotionally intense (see review in Gray et al., 2004). One example is the visual study by Gray et al. (2004). Printed words that were personally related to the subjects (e.g. the name of a subject's sibling) and other unrelated words were presented. The amplitude of P300 to the self-relevant stimuli was about three times higher than to other stimuli.

Concerning speech processing, P300 has been applied to study dyslexics' phonemic awareness. The study by Fosker and Thierry (2004) may serve as an example. In comparison to normal readers, the dyslexic subjects did not show a P300 effect in a lexical decision task with real and nonsensical words. Additionally, these words differed in the onset-sound. The lack of P300 in the dyslexic subjects, therefore, indicated a lack of attention to phonetic cues. Furthermore, the missing P300 could signify a general attentional deficit in dyslexics: as they focussed on the lexical decision they did not have any capacity left to attent to phonetic changes. However, whether this is related to a lack of speech-specific phonetic processing or to general attentional deficits is not clear yet.

Chapter 4

Neural correlates of speech sound processing

4.1 Introduction

The main question for research on speech perception is how information from the acoustic speech signal is used to access lexical representations. As reviewed in Chapter 2, a general consensus holds across theories of speech perception that features are extracted from the signal that can be mapped onto or related to abstract mental representations of linguistic categories. This process is mostly assumed to procede from features that are close to the actual signal to successively more abstract representations. Thus, the interplay between peripheral, general auditory, and language-specific mechanisms is a vital and recurrent issue in theories of speech perception. This chapter reviews previous studies that applied real-time recording with the EEG and MEG of auditory processing to reveal, at what point, language-specific representations are involved, what effect long-term representations of speech sound categories have on neural auditory processing, and also how abstract the representations may be, that are involved.

This literature review will show that not all classes of speech sounds have been investigated yet. In particular, fricative sounds have hardly been considered in electrophysiological research. This was a motivation to work with this sound class. One special issue concerning auditory processing of fricative sounds was elaborated in Chapter 3: the discrimination and identification of fricative sounds may evoke a processing strategy that is not guided by phonetic representations or that listeners can switch between the two modes of processing (Repp, 1981; Lisker, 2001; see also reviews by Repp, 1984 and Mirman et al., 2004). Thus, this point has to be considered in the present study on native language representations for fricatives. Therefore, the first part of this chapter reviews studies that investigated whether phonetic and nonphonetic perception can be differentiated in early auditory processing.

Related to this topic is the question, how early long-term memory representations of linguistic categories affect auditory processing. The previous chapter has shown that the component N100 is sensitive to acoustic features, such as pitch and temporal dynamics (Romani et al., 1982; Woods, 1995; Roberts and Poeppel, 1996). The second paragraph of this chapter shows that N100 is also affected by acoustic features that are relevant for phonetic processing. The questions whether speech-specific processes may be assumed at this point and whether long-term memory representations are involved will be considered.

A clear indicator of the influence of long-term memory and mental representations of linguistic categories is the MMN. The discovery that memory related processes underly the generation of the MMN has motivated many electrophysiological studies on the role linguistic experience on auditory speech processing. Therefore, this component of the ERP was used to investigate neural correlates of fricatives in the present study. The largest part of this chapter reviews studies that used MMN to assess aspects of phonetic grouping, the influence of phonological categories and rules, and the extraction of discrete phonological features. Studies that investigate the influence of native phonemic categories on the MMN component and trace responses to nonnative speech sounds are in the focus of this review. These studies were particularly important for the design, realization, and interpretation of the MEG experiment on fricative contrasts.

The results of EEG and MEG research should be considered, furthermore, in the context of imaging studies that enable a precise localization of neural processes. Therefore, the last part of this chapter describes results from positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies about phonetic and phonological processing in the brain.

4.2 Electrophysiological studies on speech versus nonspeech processing

One recurrent observation when correlates of speech and nonspeech processing are compared is hemisphere asymmetry. One problem in comparing the processing of phonetic vs. non-phonetic material is to guarantee that all stimulus material is of equal acoustic complexity. This problem persists in many studies on this subject.

Kayser et al. (1998) did not test specifically for MMN, but they observed higher left-lateralized N200 and P300 amplitudes for phonetic stimuli and higher amplitudes of these components in the right hemisphere in response to nonspeech sounds. As speech tokens Kayser and collegues used natural recordings of [ta], [da], and [ka] syllables. The nonspeech stimuli were musical chords consisting of six synthetic square wave tones differing in frequencies. The musical chords were acoustically less complex than the spoken syllables. Futhermore, the chords did not contain any rapid acoustic changes. Thus, it is questionable, whether the hemisphere asymmetry that Kayser et al. report is exclusively due to the distinction between phonetic and general auditory processing or whether it may be due to the differences in acoustics between the syllables and the musical chords.

Shtyrov et al. (2000) improved the nonspeech stimuli that were used in comparison to phonetic stimuli. They used complex wave stimuli consisting of a noise burst and a complex periodic signal in comparison to semi-synthetic [pa] syllables.[15] The sounds were presented separately in chains with 715 ms ISI. Additionally, a dichotic listening test was conducted in which natural spoken [pa], [ta], and [ka] syllables were presented simultaneously to the left and right ear in a forced-choice test. The dichotic listening task revealed a right-ear advantage in most subjects. The neuromagnetic results, however, showed a symmetric distribution of the

---

[15] The synthesis technique used for these stimuli takes natural vocal sounds as the basis and modifies them in an articulatory synthesis, for details see Shtyrov et al. (2000) or consult the PRAAT manual (Boersma, 2001).

components P100m, N100m, and P200m, and no effects of stimulus type. Shtyrov and colleagues assumed that speech specific processing would reflect in left hemisphere dominance. The lack of hemispheric asymmetry, thus, lead them to speculate that speech specific processes start later than P200m, probably at the onset of MMN.[16]

No information about hemisphere lateralization, but evidence that mismatch negativity mirrors phonetic normalization processes is reported by **Aulanko et al. (1993)**. They compared the magnetic mismatch response of 8 Finnish listeners to synthetic [bæ] and [gæ] syllables and to nonspeech tone stimuli. The syllables as well as the tones differered in pitch, altogether 16 different fundamental frequencies were used. Crucially, in order to detect the syllable change, the sensory memory trace would have to extract the invariant formant information and abstract from F0. EMFs were recorded over the left hemisphere while subjects either listened passively or attentively discriminated the syllables. A clear MMNm emerged for the phonemic changes but only a very small MMNm was detected for changes between tones. As a side-effect, subjects' attention enhanced the MMNm amplitude for the syllabes, but not for the tones. But would the MMNms necessarily be based on speech-specific memory formation? This is the criticism that Phillips et al. (2000) hold: the acoustics of the formant transitions in the plosive stimuli alone could be responsible for the MMNm without an influence from linguistic representations.

Differential lateralization of MMN was demonstrated in the study by **Rinne et al. (1999)**. They used 8-step continua from semi-synthetic [i] and [a] vowels to tonal correspondents The stimuli were gradually reduced in complexity until only sinusoidal sounds remained. Prior behavioral tests were used to evaluate which stimuli resembled speech and subsequently, one stimulus from each continuum was taken to create 8 oddball-paradigms. MMNs emerged for all contrasts around 139 ms. The stimuli that had been labeled as speech sounds elicited stronger MMN amplitudes in the left hemisphere and the tonal stimuli yielded stronger MMN amplitudes in the right hemisphere. Again, similar to the study by Kayser et al. (1998), the speech stimuli were generally more complex than the nonspeech stimuli. Thus, acoustic complexity may have confounded the results.

The MEG study by **Mathiak et al. (1999)** investigated whether hemisphere asymmetries would be driven by acoustic features, independently of the phonetic content of the stimuli. First, vowel [a] served as the standard against a vowel deviant [e], the syllable [ba], and a pseudo-syllable in which the F1 and F2 intersected in the transition phase, which cannot be produced by a human vocal tract.[17] Second, an [i] standard was combined with the syllable [bi] with a variation in the duration of the transition from 10 ms to 90 ms. Auditive evaluation of the stimuli revealed that the sounds with the shortest transitions were perceived as a click followed by a vowel. The intermediate transition length was perceived as the syllable [bi]. For the sounds with the longest transitions, listeners reported a diphtong-like impression.

No significant differences emerged between the syllable [ba] and the pseudo-syllable. For the vowel [e], the latency MMNm latency was slightly shorter than for the syllables, but the authors note that this may be an effect of a differential N1m response. No lateralization effects were observed for any of these stimuli. Overall, the syllables with varying length of transition yielded a bilateral MMNm response similar to that for [ba] and pseudo-syllable. The amplitude of the MMNm was found to increase with prolonged transition. Particularly,

---

[16] Relatedly, Hertrich and colleagues (2002) addressed the few recorded contrary findings and suggested that lateralization effects as early as M100 may be due to the experimental design. In oddball experiments MMN often overlaps with N100 which then may appear to be asymmetric.

[17] All stimuli were fully synthetic.

the largest MMNm amplitudes were found for stimuli with transitions of 30 ms and 40 ms. These sounds resembled the syllable [bi] the most. The MMNms for all stimuli were symmetrically distributed across the hemispheres, only for the stimulus with the shortest transition (10 ms), a stronger left hemispheric MMNm was found. Thus, a difference in MMNm amplitude was found for stimuli that most resembled speech sounds. Effects of lateralization, however, could not be associated with phonetic processing for these synthetic stimuli. However, perhaps the usage of synthetic stimulus material may not always automatically evoke phonetic processing. This issue will be resumed in section 4.4.5 of this chapter.

One way to solve the problem of finding adequate stimulus material is the usage of sine-wave speech. This was done in the study by **Dehaene-Lambertz et al. (2005)**. Sine-wave stimuli are not recognizable as language, unless listeners are told about their linguistic content and practice to recognize them as speech. Speech and nonspeech processing, thus, can be tested with the same stimulus material. Dehaene-Lambertz and associates tested two acoustically different analogues of [da] and of [ba] syllables. Heard as speech, one syllable pair presented an acoustic contrast within one phoneme category. The other syllable pair crossed the boundary between the alveolar and labial plosive. EEG and, additionally, fMRI recordings for oddball sequences were carried out in two parts: first subjects were not informed and perceived the stimuli as nonspeech noises. Then, the phonemic nature of the stimuli was revealed and the participants practiced the identification of the stimuli as syllables prior to the next recordings.

The EEG recording revealed an MMN that began about 60 ms earlier in the speech mode condition for the across-category contrast than for the within-category contrast and for both types of contrast in the non-speech mode. The subsequent fMRI experiment showed that the posterior part of the left superior temporal sulcus (STS) was most active during the across-category speech mode task. Additionally, the left-hemispheric supramarginal gyrus (SMG) was only activated in this condition. These results support the tentative suggestions of previous electrophysiological studies: left hemispheric auditory areas appear to be more strongly involved in speech sound processing when phoneme categories are discriminated.

4.3 Correlates of speech sound characteristics at N100

A number of studies reported that the sensitivity of the auditory areas that underly N100 to acoustic features of the stimulus could be utilized in speech perception. For example, **Gage et al. (1998)** observed that the latency of magnetic N100 reflects onset dynamics of speech sounds. They compared the N1m component plosives (rapid onsets) and [m] and [f] (slower onsets) and found shorter N1m latencies for stops than for [m] and [f].

More specifically, the N100 response has been shown to relate to VOT in plosives. **Sharma and Dorman (1999; 2000)** presented plosive-vowel syllables from a VOT-continuum in random order. The latency of N100 correlated with VOT: a single N100 peak emerged for stimuli with a VOT of less than 50 ms. In response to stimuli with 50 to 80 ms VOT two distinct N100 components appeared. A similar observation was made by **Phillips and colleagues (1995)**.

This finding correlates with results from direct single cell recordings in animals, showing "single-on" responses to short VOTs of < 20 ms and "double-on" responses to longer VOTs (Sinex et al., 1989; Steinschneider et al., 2003). The difference between N100 response to short and long VOT appears to reflect neuron-intrinsic characterics in the auditory cortex and may therefore display restrictions for onset detection in mammals (Steinschneider et al., 2003). These findings gave rise to the idea of a psychophysical predisposition for categorical perception in plosives (Phillips et al., 1995; Steinschneider et al., 2003). Therefore, categorical perception of voicing in plosives may be based on low-level neuronal constraints.

These findings relate to an early suggestion by Miller et al. (1976) that was made on the basis of findings of categorical perception in non-speech stimuli. Miller et al. proposed that language contrasts have evolved by taking advantage of innate perceptual constraints in the human or mammalian auditory system. The "double-on" response for voiceless stops vs. the "single-on" response for voiced stops could be responsible for the fact that voicing contrasts in stops are widely spread across languages of the world (compare Comrie, 1990; Ladefoged and Maddieson, 2002). However, if a physiological precondition for certain phonetic contrasts exists, the VOT categories could be expected to be very similar across the languages of the world. Yet, the study by Cho and Ladefoged (1999) of VOT contrasts in 18 languages has shown that this is not the case. Nevertheless, many behavioral studies report an auditory discontinuity at about 20 ms VOT (for discussion and recent findings see Holt et al., 2004).

On the other hand, the human listeners are able to detect 2 - 3 ms gaps in noise (Eggermont, 2000). Eggermont (2000) showed that neural responses to gaps depend on stimulus characteristics and can be carried out by different neuronal subsets with different gap-detection thresholds. Additionally, Elhilali (2004) and Elilhali et al. (2004) observed that the auditory cortical cells respond to very fine changes and to slower modulations at the same time. Thus, the human auditory system may show strong responses when certain thresholds are crossed, but is still sensitive to very fine acoustic modulations. It is not "deaf" to fine temporal differences.

For vowels, **Gage et al. (1998)** and **Poeppel et al. (1997)** observed that the latency of N100 and its source location varied for different types of vowels. Particularly, Poeppel et al. (1997) tested the vowels [i], [a], and [u] with two different pitches (100 and 200 Hz). They found that vowel type and not pitch lead to N100m latency changes with [a] leading to the shortest latency.

However, this finding stands in contrast to the results by **Roberts and Poeppel (1996)**. They report N100 latency changes in response to nonspeech tones of 100 and 200 Hz. Since vowel category, but not pitch lead to the latency effects in Poeppel et al. (1997), a coding of vowel type already at the level of N100 was suggested. This may not necessarily reflect an influence of linguistic representations, because the latency differences could reflect the differences in F1 (300 Hz for [u] and 700 Hz for [a]). But, as for VOT, the distinction of vowel categories and general auditory preconditions may be interrelated.

Support for vowel-category differentiation at the level of N100 comes from Roberts et al. (2004). In their MEG-study, Roberts and colleagues randomly presented an 11-step vowel continuum between [a] and [u]. First, behavioral responses were collected. Additionally, a control task was conducted using nonspeech tones that corresponded to the F1 values of the vowel stimuli. Three latency clusters or plateaus emerged for the vowel stimuli. Stimuli that unambiguously grouped with the category [u] evoked a long M100 latency (around 120 ms). The stimuli that grouped with [a] elicited a short M100 latency (around 95 ms). In contrast, the tone stimuli showed a linear latency distribution. Responses to ambiguous vowel stimuli followed the F1-related, linear latency, similar to that for the tones. Therefore, Roberts and colleagues propose:

> [...] the ultimate [i.e. speech-related] perception is parasitic on M100 encoding of physical stimulus properties when they unequivocally correspond to a native language category/neuronal cohort. For tokens that are not members of a distinct native language category, physical stimulus properties dominate M100 latency; in such cases, the role of context emerges as a biasing factor. (p. 1682)

Obleser et al. (2004) found that the N100m for the German vowels [i] and [e] had separate locations than for [u] and [o]. In reference to the FUL hypothesis (see Chapter 2, part 2.5.5), Obleser and colleagues suggest that the different N100m sources could relate to the contrasting phonological features of these vowels have, namely, their place of articulation, [coronal] vs. [dorsal]. Also, as observed by Poeppel et al. (1997) and by Roberts et al. (2004), Obleser et al. observed longer latencies in relation to F1 values: the back vowels yielded longer N100m latencies. Additionally, the N100m onsets of [i] and [e] were temporally separated by ca. 6 ms. This did not occur for [y] vs. [ø] and [u] vs. [o], even though the F1 differences are almost identical between these pairs. This indicates that not only F1 accounts for N100 latency differences for vowels but that several formants may interact. The compositional nature of N100 that Näätänen and Picton (1987) and Woods (1995) have noted (see Chapter 3, 3.5.2.1) supports this assumption.

To summarize the findings on N100 for speech sound processing, the reviewed studies show that the encoding of acoustic stimulus features as reflected by the N100 is very close to speech-specific processes and possibly interrelated with phonetic mechanisms. The studies indicate, furthermore, that N100 responses could be related to universal phonetic distinctions, such as voicing and place of articulation, or preconditions that are necessary for phonetic analysis. This close relation between afferent processes and units of speech shows how well language is adapted to the auditory system. Long-term memory representations of speech sound categories, however, do not seem to reflect when N100 is considered.

4.4 MMN and native speech sound representations

4.4.1 Voicing in plosives

The study by **Sharma and Dorman (1999)** shows clearly that N100 is more related to acoustic features of stimuli, regardless of their phonological function; whereas MMN correlates with behavioral categorical perception. The close relation of N100 to VOT in plosive-vowel syllables in Sharma and Dorman's study, that was described in 4.3, was accompanied by a mismatch response that reflected speech category boundaries. A within-category contrast and an across-category contrast taken from the synthetic continuum ranging from [da] to [ta] were presented in two oddball-sets. The resulting MMN for the across-category contrast was found to be more sustained between 100 to 310 ms than the MMN for the within-category contrast that peaked only shortly around 180 ms. The difference between the two conditions was significant, ($p < 0.05$). These results correlated with the behavioral ability of Sharma and Dorman's subjects: the across-category contrast could be differentiated better than the within-category contrast. Thus, categorical perception that has been observed for voicing in plosives very often (cf. Chapter 2, section 2.1.1) reflects in the mismatch response as well.

Phillips et al. (2000) extended the VOT continua tests. They applied VOT continua that ranged from the English syllable categories [dæ] to [tæ]. Additionally, the same stimuli were applied with an added 20 ms VOT. In this case, standards and deviants stimuli did not cross a phonetic boundary, as illustrated in Figure 4.1.

Eight American English subjects were tested for whom the category boundaries between the voiced and voiceless plosives had been individually assessed in prior behavioral tests. Four stimuli with different VOT values were selected for each plosive-category and presented as standard and deviant stimuli in a passive-listening oddball-sequence. The stimuli with the added 20 ms were then applied in a control experiment with 6 subjects. Magnetic neural signals were recorded over the left hemisphere while the stimuli were delivered to the right ear.

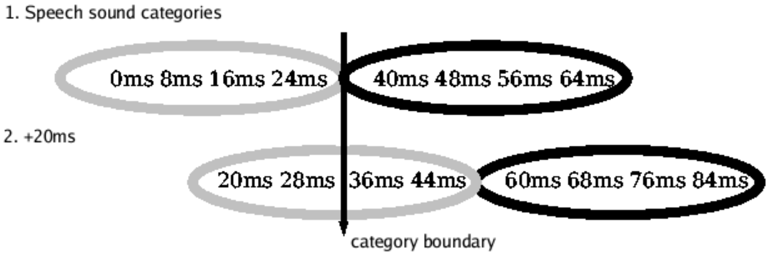


Figure 4.1: Oddball sequences with exemplary VOT values from the MEG study by Phillips et al. (2000). The arrow indicates the category boundary between [dæ] and [tæ]. Ellipses indicate the distinction between standard and deviant stimuli. In the first sequence, standard and deviant stimuli presented an across-category contrast.

A clear effect of linguistic memory was shown when significant mismatch response emerged between 100 and 210 ms ($p = 0.028$) emerged only for the [dæ] - [tæ] distinction but not for the stimuli with the prolonged VOT that did not correlate with speech sound categories.

Cross-linguistic comparisons provide strong demonstrations of native category effects. **Sharma and Dorman (2000)** compared Hindi listeners with American English listeners for plosive-vowel syllables from a VOT continuum from -90 to 0 ms. This continuum includes a voiced-voiceless phoneme contrast for Hindi listeners ([ba] vs. [pa]) but not for American English listeners, to which all stimuli categorize as [ba]. The syllables were natural productions in which silence was inserted to create a VOT continuum. Ten Hindi and 10 American English listeners were tested behaviorally in a discrimination/identification test. The full stimulus continuum was presented in random order and the N100 response was assessed. In an oddball-test to evoke the MMN response two stimuli were taken: VOT -10 ms and VOT -50 ms which crossed the category boundary for the Hindi, but not for the American English listeners.

The N100 responses were similar for both subject groups: N100 latency was prolonged as VOT increased. As it was discribed in section 4.3, N100 latency reflects the acoustic differences between stimuli. MMN responses, on the other hand, differentiated the two groups. A significant MMN emerged for Hindi subjects ($p < 0.05$) but no significant mismatch response was observed for the American English participants.

4.4.2 Place of articulation in plosives

Behavorial tests revealed strong categorical perception also for place of articulation in plosives (e.g. Repp, 1984). Is the MMN for this type of phonetic contrast similarly dependent on native language representations? To answer this question, three EEG-studies investigated the response nonnative listeners to Hindi plosives. Hindi contains a phonemic contrast between voiced alveolar [d] and voiced retroflex [ɖ] that is difficult to discriminate for nonnative listeners whose language does not utilize this contrast. For example, French and English listeners assimilate the retroflex sound to the native category of [d] in behavioral tests (Dehaene-Lambertz, 1997). **Dehaene-Lambertz (1997)** tested the behavioral and neural responses to this plosive contrast for 12 French subjects. Additionally, the French contrast between [d] and [b] was tested. The stimuli consisted of plosive -[a] syllables that were taken from a synthetic continuum. The acoustic difference between all sound pairs was kept equal.

The behavioral tests showed that the nonnative contrast was poorly discriminated by the French listeners. The contrast that is phonemic in French, however, yielded categorical perception. The EEG experiment yielded significant MMNs between 120 - 320 ms for the native contrast without significant hemisphere effects (all $p < 0.05$).[18] For the nonnative Hindi contrast, no significant mismatch responses were recorded. Therefore, the MMN seems to be highly dependent on long-term memory representations and, accordingly, the authors come to the conclusion that the ability to discriminate the nonnative contrast would be completely lost for the French listeners due to native language represenations.

However, other studies do not show such strong native-language influences. The study by **Rivera-Gaxiola et al. (2000)** on the same plosive contrasts obtained a different result. The stimulus material was similar, the subjects were 38 American English listeners. Both, the native and nonnative contrast yielded mismatch responses. Native and nonnative listeners' responses differed in that the latency of the MMN to nonnative change was longer than for the native contrast. Both contrasts, furthermore, yielded a late positive shift (LPD) that could signal an attention-orienting mechanism (see Chapter 3, section 3.5.2.4). Additionally, Rivera-Gaxiola and colleagues tested a sound-pair that presented a contrast within the category [ba]. This within-category contrast did not yield a mismatch response and no LPD, although the physical distance between the standards and deviants was the same as for the two other sound pairs.

Rather than a total loss of discriminability for the Hindi plosives, that Dehaene-Lambertz et al. had suggested, Rivera-Gaxiola and colleagues, therefore, claimed a "reorganization of information processing" (p. 22). Native within-category constrasts seem to be perceptually gated, whereas some sensitivity remains for differences in unfamiliar contrasts.

---

[18] For analysis of the electrophysiological data, four time windows were determined:120-176ms, 184-224ms, 248-320ms, and 328-400ms.

However, the difference between the French and American English listeners in Dehaene-Lambertz et al.'s and Rivera-Gaxiola et al.'s studies may result from the fact that American listeners are familiar with the phonetic features retroflex in their native [r]-sound.

The third EEG study that used the Hindi plosive-place contrasts by **Shafer et al. (2004)** compared the response of 13 native Hindi and 12 American English subjects to the dental-retroflex and the bilabial-dental plosive contrasts, again with synthetic stimuli, similar to the previous experiments. Here, the influence of the native language did not seem to be very strong and overall the results are not easily explained. Behaviorally, both subject groups exhibited categorical perception for the two contrasts and no group differences were found between the MMN responses. Unexpectedly, the dental-retroflex contrast did not yield a MMN response in either subject group. Subsequently, a recording of the response to the bilabial-retroflex contrast was carried out. Now, a MMN was seen for both Hindi and American English subjects and slight group differences emerged: deviant [ba] evoked an earlier and more left-lateralized MMN for Hindi than for American English subjects.

Categorical perception and MMN amplitude, thus, do not seem to correlate completely for place of articulation in plosives. Fine phonetic differences between sounds are detected, even if they do not signal a native sound category change.

4.4.3 Liquids [l] - [r]

The problems that Japanese speakers have when they learn try to learn the contrast between [l] and [r] have been extensively studied in behavioral tests (e.g. Takagi et al., 2002; Iverson et al., 2003). Theoretical considerations of nonnative language processing (cf. Chapter 2, 2.3) often use this case of strong native language influence as an argument for a restructuring of speech perception by language experience (see particularly Brown (2000), reviewed in Chapter 2, 2.3.3). Thus, if MMN is indeed dependent on native language representations, electrophysiological studies should observe strong differences in the mismatch response between English and Japanese listeners.

Buchwald et al. (1994) applied the word pair "rip - lip" in an oddball paradigm (80:20) and recorded the ERPs of 14 Japanese with good command of English and 14 American English subjects. The stimuli had been recorded from a female speaker and normalized for duration. The mismatch response occured rather late, but showed a significant effect of native language background: for the American English subjects a significant effect ($p < 0.05$) was recorded at about 385 - 525 ms for the "lip" deviant and at 415 - 480 ms for the "rip" deviant. No MMN emerged for the Japanese participants, strongly indicating MMN's relation to native phonetic represenations.

Zhang et al. (2005) compared the magnetic MMN responses to a syllable contrast that is phonemic in both Japanese and English, [ba] vs. [wa], and to the contrast [la] vs. [ra]. The stimuli were end-points from synthetic sound continua. A passive-listening oddball paradigm was applied and 10 Japanese with good knowledge of English and 10 American English were tested.

As could be expected, for the contrast [ba] - [wa], no group differences in amplitude and latency and no hemisphere effects emerged. Both groups responded with clear MMNms. A significant group difference emerged for the [la] - [ra] pair ($p < 0.01$). Here, a weak bilateral MMNm emerged for Japanese participants whereas the American English subjects had a left-lateralized MMNm with significantly stronger amplitude ($p < 0.05$). Surprisingly, though, the onset of MMNm occured significantly earlier for the Japanese group around 337 ms than for the English listeners, where the MMNm started around 413 ms.

The earlier MMNm onset stands in contrast to findings by other studies, for example, by Sharma and Dorman's plosive experiment (1999) or by Rivera-Gaxiola et al. (2000) where the Hindi plosives lead to a longer MMN latency in nonnative subjects (see 4.4.1 and 4.4.2). Moreover, studies with nonspeech stimuli have shown that contrasts that are more difficult to distinguish evoke longer MMN latencies (Näätänen, 1992; 1995; Sams et al. 1985; compare Chapter 3, section 3.5.2.2). To explain this paradox, Zhang and colleagues suggest that the earlier MMNm that emerged for the Japanese subjects may be the result of a different processing strategy. Normally, the transition of the third formant has to be integrated to differentiate English [la] and [ra]. This cue integration can be problematic for Japanese listeners (Underbakke et al., 1988). Therefore, the earlier MMNm for Japanese subjects could signal that the integration of the third formant did not take place, whereas the longer MMN latency for the native listeners reflected a longer window of integration for the discrmination of the syllables.

Zhang and associates carried out a second series of behavioral and MEG tests, this time with nonspeech sinusoidal tones that were acoustically matched to the syllables [la] and [ra]. Testing conditions were the same as in the previous tests. Here, no effects of language background emerged. Thus, the authors could surely conclude that the differences in the MMN responses to [la] and [ra] could be assigned to the influence of native language category representations.

4.4.4 Vowel categories

The study by Näätänen et al. (1997) tested the influence of language-specific memory on MMN for vowels. An EEG-experiment on 13 Finnish and 11 Estonian native speakers was conducted that applied an oddball-paradigm (85:15) in which the Finnish vowel [e] served as the standard and the Finnish [ø] or the Estonian close-mid back unrounded vowel [ɤ] were the deviant sounds. The stimuli were semi-synthetic prototypical versions of the vowels that differed only in F2.

A larger MMN amplitude emerged for the Finnish subjects in response to the native deviant [ø] than when the Estonian [ɤ] served as the deviant, although the acoustic difference between [e] and [ø] was smaller than between [e] and [ɤ]. For Estonian subjects, no such effect was observed. Additionally, the magnetic ERPs to these contrasts were recorded so that hemisphere effects could be traced. It was found that the MMNm activity was larger for the native vowel contrast over the left than over the right hemisphere. This lead Näätänen and colleagues to conclude:

> [...] the left auditory cortex is involved in phonemic discrimination, presumably because the neural phoneme traces are located there, whereas both the left and right auditory cortices serve acoustic discrimination. (p. 433)

The EEG-study by **Winkler et al. (1999a)**, likewise, found reduced MMN amplitudes in response to nonnative vowel contrasts. They tested 10 Finnish and 10 Hungarian listeners with no knowledge of the other language on a Finnish, a Hungarian and a bilingual vowel contrast, using synthetic stimuli. The Finnish contrast formed a within-category pair for Hungarians and the Hungarian vowels formed a within-category contrast for Finnish listeners. Subjects listened passively to an oddball-presentation with 82.5% standard stimuli, 15% exclusively Finnish or Hungarian deviants and 2% deviants that were phonemic in both languages.

This resulted in very similar neural responses to the deviant that existed in both languages. Effects of linguistic experience were seen most clearly in the Hungarian subjects which had a lower MMN amplitude with a longer latency than the Finnish for the native Hungarian vowel change. The MMN response of Finnish subjects to the nonnative Hungarian contrast was more similar to that of the Hungarians but compared to their response to the Finnish contrast the amplitude of the MMN was reduced.

Winkler and colleagues identified two MMN peaks and associated them with different perceptual mechanisms. The early peak may signal phonetic processing, that is fast and efficient because it relies on long-term memory traces. The later peak may indicate an analysis of physical differences which may not be as efficient but still possible for the nonnative listener. The results for the Hungarian subjects were taken as support for this assumption: the earlier MMN peak (100 - 200 ms) in Hungarians was larger for the native contrast. In contrast, the later peak (200 - 300 ms) was greater in response to the nonnative contrast.

Sharma and Dorman (1998) tested whether the concept of the Native Language Magnet (Kuhl, 1991; see Chapter 2, 2.3.4) has a correspondence at the early auditory processing level. In particular, they wanted to test if vowel exemplars in the vicinity of the prototype evoke less mismatch negativity than exemplars that are further away from the prototype. Prototypical synthetic [i] -stimuli and non-prototypical exemplars were identified by 10 English subjects. Behavioral discrimination tests and EEG tests that applied the oddball-paradigm were carried out between prototype and the vowel closest to it, between a vowel-pair that was less prototypical, and between the determined non-prototype and the vowel closest to it while the acoustic difference between all pairs was kept equal. The results did not affirm the proposed magnet effect. The discrimination and the MMN amplitude were highest for the vowel pair that contained the prototype and its nearest neighbor. However, the opposite would confirm the magnet effect theory that predicts lowest discriminability in the vincinity of the prototype. This study shows that behavioral findings and electrophysiological responses do not always correlate and that it is important to test theories of speech perception with methods that record the neural auditory activity.

The study by **Shestakova et al. (2002)** followed a different approach. Their aim was to test whether natural stimuli with very large acoustic variation would elicit MMNms related to the categories of the vowels. 150 different natural productions of male speakers for each of the three Russian categories [i], [a], and [u] were used. The MEG study applied a roving-standard paradigm, in which every deviant successively becomes the next standard stimulus. Nine Russian subjects were tested in a passive-listening condition. Results revealed clear MMNms that were significantly stronger in the left hemisphere in the 8 right-handed subjects and lateralized to the right hemisphere in one left-handed subject ($p < 0.0005$). The vowels that were used in this study could also be grouped according to formant frequencies without any reliance on linguistic representations. However, Shestakova et al. were able to show that auditory grouping of vowels is possible despite high acoustic variability.

4.4.5 Natural stimulus material

The usage of natural, spoken speech material provides a more realistic listening situation. Shestakova et al. (2002) demonstrated that natural stimulus material can be successfully applied in a MEG study on sound category processing. Differences between the processing of natural and synthetic stimuli are noted in several behavioral, electrophysiological, and other imaging studies, which are detailed in section 4.5.1 and in Chapter 5, 5.2.1. The study by Hertrich et al. (2002) was one of the first MEG experiments that focussed on potential differences between stimuli quality.

Hertrich et al. (2002) compared the magnetic auditory response to synthetic and natural CV syllables. Binaurally presented [ba] served as the standard stimulus, [da] as the deviant, either presented binaurally or to one ear, while [ba] was played to the other ear (80:10 left ear, :10 right ear). Synthetic and natural stimuli were presented in separate runs in a passive listening test with visual distraction. An additional session under attention was carried out for synthetic stimuli only.

In the passive-listening condition differential lateralization effects were observed for synthetic and natural stimuli between 80 to 120 ms. Here, the MMNm to natural stimuli was stronger in the right hemisphere. Hertrich and colleagues suggest that this could reflect an early analysis of non-linguistic features in natural stimuli, such as speaker characteristics.

For the attention task that was carried out with synthetic events, similar lateralization effects were recorded. The natural stimuli elicited a slightly stronger right-hemispheric MMNm while the synthetic events produced a bilateral MMNm and the MMNm latency for synthetic stimuli was shorter. Additionally, for the natural deviants that were presented to the right ear, the MMNm was strongest and lateralized to the left. Such an effect for the synthetic stimuli was only recorded under the attentive-listening condition. The authors, therefore, suggest that natural stimuli may evoke phonetic processing automatically and thus lead to a left hemispheric response even when the stimuli are not attended to.

The late onset of mismatch negativity for the natural tokens was explained by the fact that natural stimuli are acoustically more complex and require complex integrative processing. Between 300 to 400 ms a sustained MMNm was recorded for the natural but no MMNm for the synthetic syllables. Only in the attention-condition a similar duration of the MMNm for synthetic stimuli was observed. Hertrich et al. noted that synthetic stimuli probably have a "less attention-grasping structure" (p. 1913) and, therefore, do not maintain a persistent mismatch response.

Synthetic stimuli do not perceptually require as complex processes of normalization and cue integration as natural stimuli do and, thus, may not provide authentic testing conditions (see also Brugge, 1992; Eggermont, 2001). The issue of natural vs. stimuli will be resumed in the description of the MEG study on fricatives in Chapter 5, for which it was decided to use spoken material.

4.4.6 Phonotactics

Phonological "deafness" (Sebastian-Galles, 2005) in nonnative listeners for certain unfamiliar sound contrasts is perhaps most striking when the second language speaker is not aware of the fact that she uses epenthetic vowels. The study by **Dehaene-Lambertz et al. (2000)** leads one to suggest that the generation of the MMN can access native phonotactic rules. Neural responses of 12 French and 12 Japanese subjects were compared while listening to nonwords of the type: "igumo - igmo". Japanese phonotactics require epenthetic vowels in consonant clusters, therefore, such words could not appear as Japanese minimal pairs. In French, on the other hand, this difference could signal a phonemic change.

Behavioral discrimination tests yielded highly correct performance for the French subjects and low scores for the Japanese listeners. The EEG recording that applied an oddball-procedure (4:1) showed clear MMNs for French listeners between 139 - 238 ms, whereas no MMNs emerged for the Japanese group in that time period. Only very late, after 291 ms and after 523 ms, weak and short mismatch responses emerged for the Japanese subjects.

The Japanese participants in this study knew French well. Their late MMN shows that the acoustic contrast was processed to some extent. It seems, however, as if the fact that this sound change cannot occur phonemically in Japanese affected the auditory response. This result shows very clearly, how much MMN is influenced by native language representations.

4.4.7 Phonological features

Eulitz and Lahiri (2004) based their EEG experiment on the predictions of the FUL model of speech perception (e.g. Lahiri and Reetz, 2002). As described in Chapter 2, FUL assumes that underlying lexical representations consist of maximally underspecified phonological features. Phonetic surface features are extracted from the acoustic signal which are then mapped onto lexical representations. If the surface features conflict with lexically represented ones, the mapping is rejected.

Three German vowel pairs were tested in oddball-paradigms: [e] vs. [ø], [ø] vs. [o], and [e] vs. [o]. Each vowel in a pair served as the standard and in a second block as the deviant stimulus.

It was assumed that the standard stimuli would activate underlying abstract representations that correspond to the underspecified underlying representations of the vowels. This means that the transformation from acoustic to phonological features and lexical mapping would take place in the formation of the standard memory trace. In comparison, all surface features would be extracted from the deviant stimulus. Thus, depending on which vowel serves as standard or as deviant, conflict situations may occur. For a standard [o] and deviant [e] a conflict would be given because the standard activates the underlying feature [dorsal] and the feature [coronal] is extracted from the deviant. It was expected that feature mismatches would lead to higher MMN amplitudes than non-conflicting situations.

	e	ø	o
features from deviant	[COR]	[COR] [LAB]	[DORS] [LAB]
underlying features activated by standard		[LAB]	[DORS]

Figure 4.2: Rationale of the test paradigm in Eulitz and Lahiri (2004). A conflict situation occurs when [coronal] is extracted from the deviant and mapped to the representation [dorsal] that has been activated by the standard stimuli.

Three natural recordings of each vowel category by a male speaker were applied and presented in oddball-sequences. ERPs from 12 passively listening German subjects were recorded.

The results confirmed the predictions in that no significant difference in the mismatch amplitude emerged between the responses to the [e] - [ø] pair for which surface and underlying features do not conflict. For the other two vowel pairs, however, MMNs in the conflict situation were significantly higher. Thus, the authors concluded that phonologically underspecified mental representations of speech sounds are effective during vowel perception.

4.4.8 Context-dependent variation

Phonemic contrasts can be neutralized in certain phonetic environments. Does the early auditory response reflect the knowledge about such rule-driven neutralizations? To answer this question, the effect of the Dutch phonological rule of nasal-place-assimilation on perception was tested by **Mitterer and Blomert (2003)**. They applied a word pair in which assimilation is possible: "tuinbank" and "tuimbank" (meaning "garden-bench"), and one in which it is impossible: "tuinstoel" cannot become "tuimstoel" ("garden-chair"). If perception is affected by the language-specific assimilation rule, listeners could be expected to have more difficulties in discriminating [n] and [m] when it is a viable alternation than when it could not be exchanged. This was confirmed in a preparatory behavioral discrimination test in which "tuimbank" was often misperceived as "tuinbank", which shows that listeners tend to perceive the phonemic form of the word.

A subsequent EEG-experiment with 8 Dutch subjects presented these word-pairs in a passive-listening oddball-paradigm. The neutralization rule indeed reflected in the mismatch response: the MMN amplitude for the contrast "tuinstoel" vs. "tuimstoel" was significantly higher than for the pair "tuinbank" vs. "tuimbank" ($p < 0.025$). Thus, the study presents evidence for a very strong influence of the phonological context on change detection both behaviorally and on the auditory mismatch response.

However the authors remark that context dependence is a general phenomemon in perception.[19] Therefore, it is possible that place assimilation rules in language have emerged according to general characteristics of context dependence.

---

[19] See, for example, Webster et al. (in press) for adaption to context in visual processing.

4.4.9 Effects of second language learning

Many of the reviewed studies showed that linguistic experience interferes with MMN responses. The question arises whether learning a new language would influence the neurophysiological processes.

The distinction between [l] and [r] by Japanese listeners is rarely learned perfectly. However, changes in the event-related potential due to learning are reported by **Zhang et al. (2001)**. One Japanese adult subject was trained to distinguish English [l] and [r]. Subsequent MEG measures assessed the training effects. The training in Zhang et al.'s study consisted of identification tasks with incremental levels of difficulty by applying increasing talker variability and different phonetic contexts. Training stimuli were resynthesized from natural productions of several American English speakers. The training phase consisted of 12 x 60 minutes. In the behavioral and MEG pre- and posttest the original natural exemplars were used.

Behavioral tests revealed an improvement after training. The neuromagnetic data attested a change in lateralization of the MMNm. Before training, MMNm was lateralized to the right hemisphere, whereas after the training phase the left hemisphere was more active. The authors suggest that "training appeared to lead to more linguistic analysis of the speech stimuli in the left hemisphere". So far, this is the only study showing lateralization effects due to second language training.

Other studies found an increase in MMN amplitude as an effect of training, as for example, **Cheour et al. (2002)**. They worked with 3 to 6 year old Finnish children that took part in a French immersion program at a day-care center. After two months of exposure to French, these children exhibited a stronger MMN response to a French vowel contrast that does not exist in Finnish than before they had started to enroll in the immersion program. This effect was somewhat stronger in the 3 and 4 year olds than in the 5 to 6 year old children but the difference was not significant. However, it may be suggested that earlier exposure leads to a faster development of neural representations for nonnative sounds. In comparison, a control-group of age-matched children that had no experience with French did not show the recorded MMNs.

Winkler et al. (1999b) demonstrated learning effects as reflected by the MMN in adults. The phonemic Finnish vowels [æ] and [e] are perceive as allophones of a native category [ɛ] by Hungarian listeners and the contrast is very hard to distinguish for Hungarians without any experience with Finnish. Winkler et al. presented Finnish [æ] - [e] and the Hungarian/Finnish vowel contrast [y] - [e] to Hungarian and Finnish subjects in an oddball paradigm (standard: 82.5%, exclusively Hungarian deviant: 15%, y: 2%).

Ten of the Hungarian subjects had started to learn Finnish after childhood and lived in Finnland. The other 10 Hungarians did not know any Finnish. The stimuli were synthesized and one exemplar per category was applied. The contrast [y] - [e] that exists in both languages yielded very similar results for all three groups. For the Finnish contrast, no MMN emerged for the nave Hungarians. However, the MMNs were similar for the advanced learners and the Finnish group. Thus, this experiment showed that adult second language learners can exhibit native-like neural responses to an aquired phoneme contrast.

The EEG-study by **Winkler et al. (2003)** attempted to test whether learned nonnative sound representations are applied in the context of the second language but omitted in the context of the native language.

Ten Hungarian subjects that were fluent in Finnish and had lived in Finland for an average of ca. 11 years were tested in a Finnish and in a Hungarian context with an oddball-sequence. Natural productions of "pæti" were semi-synthesized to obtain to words differing in [æ] and [e] only. Furthermore, Finnish and Hungarian words were recorded from the same bilingual speaker. These words were used for an active semantic decision test and occurred with a probability of 1.3% whereas the standard and deviant words containing the critical vowels occurred with a ratio of 88.7:10.

In the first part of the experiments, subjects were only spoken to in Hungarian and then had to detect Hungarian target words while the ERPs were recorded. After a break, the scene changed to Finnish only. Now, Finnish words served as targets, while standard and deviants remained the same.

No differences appeared between the mismatch responses and other components (N2b and P3a) to the Finnish and Hungarian context.

A follow-up control study with 5 Finnish subjects (with no knowledge of Hungarian) applied the same test with Finnish target words. Again, no difference to responses that had been recorded earlier were found. Therefore, the authors suggest that "in listeners who have learned a phonetic distinction, auditory analysis uses this distinction, irrespective of its relevance to the current language context" (p. 842).

In the study by **Peltola et al. (2003)** learned nonnative sound categories seemed to blend in with native sound categories as well. They tested three groups of listeners for learning effects: 11 Finnish that knew very little English, 10 advanced Finnish learners of English, and 9 native English listeners. The following vowel contrasts were presented: Finnish [i] - [e], English [i] - [e], for which the Finnish contrast is physically not as distinct as the English one; English [i] - [ɪ], and English [e] - [ɪ]. All contrasts were presented in an oddball-paradigm with passive listening.

Overall, MMN responses were larger for the native than for the nonnative contrast and advanced learners of English showed higher MMN amplitudes for English contrasts. However, they still differed significantly from those of the native listeners.

Thus, in comparison to the results by Winkler et al. (1999b) no native-like performance was recorded. Peltola and colleagues reason that the experience of their advanced Finnish subjects was mainly based on classroom instruction. The language learners in the study by Winkler et al. (1999b), on the other hand, had been immersed in the environment of the second language. Thus, a comparison between the two studies indicates the importance of an authentic experience of the new language.

Furthermore, the advanced Finnish subjects in Peltola et al.'s study did not respond as strongly to the native contrast between [i] and [e] (which was acoustically less distinct than English [i] - [e]), as the Finnish subjects who did not know much English. This corresponds to the conclusions drawn by Winkler et al. (2003) that was mentioned earlier: a phonetic nonnative contrast that has been learned appears to activate pre-attentive change detection, whether or not this is required by the language environment.

The claims by Flege's SLM (1995; see Chapter 2, section 2.3.1) seem to be confirmed by these results. Flege hypothesized that nonnative categories that are similar to native categories share their phonetic representations in the same perceptual space.

The study by **Nenonen et al. (2005)** was also motivated by the predictions in Flege's SLM. They compared the responses of Russian listeners with good command of Finnish to those of Finnish subjects for vowel-quality and vowel-duration changes. Long and short natural productions of [ka] and [kæ] were used. The vowel [a] is common to both languages, but [æ] is an unfamiliar or new sound for Russians. Further, duration change is not phonemic in Russian, but in Finnish.

It was found that the MMN for the Russian listeners was larger and native-like in response to the duration change of [k] than for duration changes in [ka]. Thus, the duration change was processed better for the vowel that the Russians had to learned as a new vowel category. Nenonen et al. suggest that native representations of [a] may have interfered with the perception of the similar vowel in the second language.

4.4.10 MMN and linguistic experience - summary

As a general summary of the findings on the influence of linguistic experience on the mismatch component of the auditory neural response it can be stated that native sound contrasts evoke larger MMN with shorter latencies and often left lateralized dominance (Näätänen et al., 2001).

A closer look shows that the interplay of long-term representations and auditory sensitivity to phonetic or acoustic detail seems to differ between sound classes, similar to the differences that were found for behavioral categorization and discrimination (Repp, 1984). The strongest effect of language-based phonetic grouping appears for voicing in plosives, most remarkably shown by Phillips et al. (2000), whereas some auditory sensitivity to nonnative plosive place variation seems to be reflected in the mismatch response (Dehaene-Lambertz, 1997; Rivera-Gaxiola et al., 2000; Shafer et al., 2004).

Nonnative vowel contrasts always yielded a neural mismatch response, however lower in amplitude and with a longer latency. Winkler et al.'s study (1999) suggests that an auditory analysis of physical differences is carried out by nonnative listeners. However, no evidence for within-category grading effects of the native language could be traced (Sharma and Dorman, 1998).

A clear effect of native language experience has also been found for the liquid contrast. Interestingly, the findings by Zhang et al. (2005) meet the observations on cue-integration of behavioral tests that were mentioned in Chapter 2. As incorporated in the constraints of Boersma's perception model (1998; 2005, see Chapter 2, 2.2.4) the integration of several acoustic features seems to be a function of native language experience.

Left-hemispheric dominance for native sound processing has been observed by several studies (Näätänen et al., 1997; Hertrich et al., 2002; Shestakova et al., 2002; Zhang et al., 2005). Most notably, the normalization of high acoustic variation in Shestakova et al. (2002) and the response to natural stimulus material (Hertrich et al., 2002) seemed to amplify this effect. First results of a shift to the left hemisphere due to second language learning were observed as well for the [l] - [r] contrast, which thus may include the learning of correct cue-integration (Zhang et al., 2001).

A common finding of studies on second language learning is the interrelation of memory for native and new sounds (Winkler et al., 2003; Peltola et al., 2003; Nenonen et al., 2005). The issue of learning will not be discussed any further in this thesis, but it should be noted that phoneme representations of the native language exert a strong influence on contrast detection even in a well-learned second language (Nenonen et al., 2005).

The experiments on phonotactic regularities (Dehaene-Lambertz et al., 2000), context-dependence (Mitterer and Blomert, 2003), and feature extraction (Eulitz and Lahiri, 2004) suggest that more abstract representations and rules of neutralization are effective in pre-attentive auditory processing. Categorization according to underlying phonological representations, not only peripheral phonetic grouping, seems to be carried out during MMN generation.

4.5 Functional neuroimaging

Functional imaging[20] adresses very similar questions as electrophysiological research: the division between speech and nonspeech processing, the delineation of subprocesses involved in speech sound processing, and the assessment of the influence of abstract phonological knowledge. Any discussion of neural responses gains from insights of other brain imaging techniques. However, this overview does not attempt to cover all results of imaging research exhaustively. Rather, the aim is to highlight findings on topics of auditory speech sound processing that are relevant for the present MEG experiment.

Since speech sounds are tested in the present study that are different in phonological function and phonetic form, results and ideas of previous imaging studies will be reviewed that cover the differentiation of acoustic, phonetic, and phonological processes in speech perception. Secondly, since the MEG method that is used for the current experiment allows to differentiate activation in the two cerebral hemispheres, this section also present studies that further elaborate the contribution of the two hemispheres.

Tomographic imaging methods of functional magnetic resonance imaging (fMRI)[21] and positron emission tomography (PET)[22] enable the localization of mechanisms in auditory

---

[20] The term functional imaging here refers to methods such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies and excludes EEG and MEG. This classification is taken from Demonet et al.(2005).

[21] fMRI detects changes in metabolism or blood flow in the brain while the subject solves cognitive tasks. This method is based on the fact that active brain areas receive an enhanced blood flow. In the blocked design accumulated neural responses are recorded. The stimuli in one block all correspond to one testing condition. In

speech processing. fMRI provides the highest spatial resolution of about 1 to 2 micrometer in diameter or less (Demonet et al., 2005). In contrast, the temporal resolution is much lower than that of electrophysiological methods. With PET the delay between the neural process and the recording is about 30 seconds, with fMRI a minimum of several hundred milliseconds are possible, but usually a time delay up to 12 seconds is necessary (Demonet et al., 2005).

This shows that EEG/MEG and fMRI/PET studies do not provide the same kind of information. Caution is therefore warranted when comparing findings obtained by the different methods (cf. Horwitz and Poeppel, 2002). Still, results from electrophysiological and imaging studies are often considered as converging evidence (e.g. Dehaene-Lambertz et al., 2005). Likewise, the findings of this MEG study on fricatives are discussed in relation to results from imaging studies in the next chapter.

However, one has to keep in mind that these comparisons only point towards possible explanations and help to form a hypothesis, but cannot serve as a proof (Horwitz and Poeppel, 2002). A direct combination of the different methods and fully complementary results would require that the same neural sources are recorded when responding to the same task. Replications of measures of event-related potential with single-trial fMRI and simultaneous EEG/fMRI recordings are the most recent and promising advancements in this direction (Dogil et al., 2004; Demonet et al., 2005).

4.5.1 Localization of acoustic, phonetic, and phonological processes

A fMRI study that provides a detailed account on common and separate processes for speech and nonspeech stimuli was carried out by **Benson et al. (2001)**. In blocked fMRI recordings they distinguished responsive areas for speech of varying complexity (V, CV, CVC) and quality (natural, synthetic) and for spectrally similar nonspeech sounds that varied in complexity (tones, complex chords and chord progressions). The nonspeech sounds were also either synthetic or natural piano sounds.

For the processing of speech stimuli and the processing of complexity in speech stimuli, seven active regions were identified: left and right posterior STG, left ventral occipitotemoral cortex (BA 21 and 37), left inferior superior marginal gyrus (SMG) (BA 40), left middle frontal gyrus (BA 6/7), right superior SMG, and right anterior insula. The bilateral posterior STG and especially the left posterior STG were differentially active for all types of speech sounds. These areas appeared to be most selective for phonetic processing.

Past the subcortical level, Benson et al. found no common regions for the processing of complexity in speech and nonspeech stimuli. Thus, the authors conclude that speech and nonspeech sounds are encoded differently at an early level. Benson and associates disagree with the notion that all sounds receive the same acoustic analysis regardless of their linguistic nature. This meets the theoretical accounts that were described in Chapter 2, which, despite

---

the event-related design responses are sampled after each stimulus. Mixing of stimuli is possible (for a detailed description see Frese, 2004).

[22] Like fMRI, PET also utilizes the enhanced blood flow in active neural regions. Short-lived radionuclids (oxygen, fluorine, carbon, or nitrogen) are injected or inhaled and enter the bloodstream. This material breaks down quickly and releases positrons during its decay. The positrons annihilate when hitting electrons and releases photons or gamma rays. These rays are detected by the scanner and their origin can be determined.

their often contradictory views on mechanism of speech perception, agree that linguistic representations are effective at a very early stage of perception.

In relation to this issue, the results by **Belin et al. (2002)** are interesting. They observed an overall greater activity for words than for nonspeech verbal sounds (e.g. coughs or sighs) in all auditory areas, already at the primary auditory cortex. This supports Benson et al.'s suggestions that speech is treated differently from early on. Belin et al. (2002), who used natural material, further suggest that the higher activation they found for speech may be due to the fact that speech carries more acoustic information and requires more complex transformations than other sounds. The previously reviewed related electrophysiological studies lead to similar conclusions (e.g. Hertrich et al., 2002; see section 4.4.5).

Concerning the processing of natural versus synthetic speech stimuli, Benson et al. found that the left dorsolateral precentral sulcus was more active for natural speech sounds. Additionally, Benson and colleagues observed a tendency for stronger activity in the right secondary auditory cortex for natural speech stimuli.

This is in accordance with the findings by **Binder et al. (2000)** for nonspeech vocalizations, such as coughs or sighs as compared to nonvocal sounds. Binder et al. observed that the bilateral anterior STS region was selectively more active for the vocal sounds. Furthermore, Belin et al. (2002) also observed more activation in the bilateral STS when comparing the response to vocal sounds, including speech and nonspeech, to sounds that were not produced by a human vocal tract.

Benson et al.'s observation of bilateral activation of the STG with a dominance of the left posterior and superior STG is congruent with other imaging studies, such as Binder et al. (2000), Dehaene-Lambertz et al. (2005) which was described in section 4.2, and Vouloumanos et al. (2001).

The study by **Vouloumanos et al. (2001)** may serve as an example. By using sine-wave replications versus nonword speech sounds, Vouloumanos et al. provided physically matched nonspeech sounds. They found that speech sounds evoked stronger activation in the bilateral STG and the middle temporal gyrus and the left hemispheric activity in the posterior STG was significantly stronger for the speech sounds.

In summary of their results for speech sounds, Benson et al. (2001) recall a hypothesis by Boatman et al. (1998) that was made on the the basis of electrical stimulation[23] in a single subject. Boatman and colleagues suggested that bilateral auditory areas process phonetic units but that the left side may be preferred or of special importance for auditory speech processing.

The study by **Scott et al. (2000)** also compared the processing of speech and nonspeech stimuli and investigated where in the brain a mapping of speech sounds to semantic representations takes place.

Their stimuli included speech (short sentences) and rotated speech which is, basically, an upside-down spectrogram. Rotated speech is not intelligible but contains pitch information

---

[23] Electrical stimulation of brain areas enables the localization of functional areas in the conscious patient before neurosurgery. The patient has to fulfil several tasks while brain areas are electrically stimulated. When language-areas are stimulated, the patient becomes aphasic. The well-known map of the human motor and somatosensory system along the cortex, the homunculus, is a result of electrical stimulation (for an introduction to the method and findings, see Gazzaniga et al., 1998).

and some phonetic features. Further, they took noise-vocoded speech, which is intelligible but has very weak pitch. They also used noise-vocoded rotated speech, which is neither comprehensible nor does it have any resemblance to speech.[24]

Phonetic features which were present in all types of stimuli except the noise-vocoded rotated speech evoked activity in the posterior left STS and the left ventrolateral STG. A specific area for pitch or intonation that responded to both speech and rotated speech was found in the right lateral STG. The intelligible sentences (speech and noise-vocoded speech) yielded activity in the left mid and anterior STS. This area connects further to the association areas in the mediotemporal lobe. Thus, an interface to semantic access seems plausible in this region.

In reference to these findings, Binder (2000) remarks that a semantic mapping would not necessarily only be carried out for linguistic material. He takes into consideration that "many nonspeech vocalizations can be written down ('hmmm', 'ha ha') or named, implying a recoding process" (p. 2371). The STS region was selective for verbal utterances in Benson et al. (2001) and Binder et al. (2000). Binder's concerns point out that a definite distinction between a phonological-semantic interface and a general mapping of sound to referential meaning may be hard to draw.

However, regarding the results by Belin et al. (2002), a difference in processing of linguistic and non-linguistic material at the STS can be seen. Comparing spoken words to nonspeech vocalizations, Belin et al. found more activity for words in the left anterior STS and more activity in the right STS for the nonspeech items. Therefore, Belin and colleagues state that linguistic and paralinguistic information processing can be distinguished at the STS by hemisphere asymmetry.

The study by **Jacquemot et al. (2003)** was able to differentiate neural structures that are involved in acoustic/phonetic and phonemic discrimination. They applied a crosslinguistic ABX-discrimination design and recorded neural responses with fMRI. Japanese and French listeners were presented with nonword pairs that were either acoustically/phonetically or phonemically distinct. The stimuli were of the type "ebza - ebuza" which could signal a phonological change for the French but only an acoustic/phonetic change for Japanese listeners. The opposite was the case for pairs like "ebuza - ebuuza". This would be an appropriate phonemic change in Japanese but for French listeners this would only be an acoustic/phonetic difference.

Unlike in Dehaene-Lambertz et al.'s study (2000) that was previously described in section 4.4.6, Japanese and French listeners could distinguish all stimuli almost perfectly in Jacquemot et al.'s behavioral test. Their performance increased slightly, when the differences were phonemic. The fMRI recording revealed that the left STG and the left anterior supramarginal gyrus were significantly more active for the phonological distinction than for the acoustic/phonetic contrasts. A small part of the right STG was involved in processing the phonemic changes, as well. A dissociation between the two processes could, thus, be found. The findings by Jacquemot and colleagues, furthermore show that phonetic and phonological processing activated very similar, possibly overlapping regions.

---

[24] Except perhaps in the temporal-phrasal domain.

4.5.2 Contributions of the left and right hemisphere

Most studies that were reviewed above as well as several electrophysiological studies (e.g. Näätänen et al., 1997, and Shestakova et al., 2002) suggest that the left temporal cortex is dominant in phonetic/phonological processing.

At the same time bilateral activation in the temporal lobes for auditory speech processing is VOT continuatly found. Reviews of EEG/MEG and imaging studies, therefore, lead several authors to conclude that absolute hemisphere dichotomies in speech sound processing are rarely observed. Rather, a network of areas in the auditory regions seems to support speech sound processing up until semantic access, including both hemispheres (Brown and Kosslyn, 1993; Scott et al., 2000; Benson et al., 2001; Bernstein, 2005; Demonet et al., 2005).

The prominent role of the left hemisphere in speech perception has been accounted for by its advantage in temporal processing. This is illustrated in the PET study by Belin et al. (1998). They showed that for slow transitions of 200 ms of a nonspeech sound that resembled a syllable, the right auditory areas were more involved than for a short transition of 40 ms. Similarly, Zatorre and Belin (2001) observed more activity in the left auditory region for temporal changes in tones and stronger right hemispheric activation for changes in frequency. The advantage for temporal resolution in the left auditory cortex has been associated with anatomical features (Scott and Wise, 2004). Hemispheric asymmetries in terms of anatomy consist of larger myelin volumes in the left auditory areas (Penhune et al., 1996; Anderson et al., 1999) and also larger pyramidal cells (Hutsler, 2003). These features supposedly enable fast signal transmission (Hutsler and Galuske, 2003). Thus, the left hemisphere could serve finer temporal resolution, which is necessary to process rapid formant transitions and short burst signals.

Several ideas on the contribution of the left and right hemisphere for auditory processing are based on these computational differences. Zatorre et al. (2002) suggest that the advantage of the left hemisphere for temporal analysis is complemented by an advantage of the right hemisphere in spectral processing. These differences led to a specialization of the hemispheres: processing of music and similar auditory events is mainly supported by right hemispheric areas, linguistic processing is supported by the left hemisphere.

The "double-filtering-by-frequency" (DFF) model by Ivry and Robertson (1998) was proposed for auditory and visual processing. DFF states that the hemispheres differ in their filtering properties: the left one functions like a high-pass filter, the right one like a low-pass filter. Therefore, the left hemisphere is able to process fine details whereas the right processes global aspects of stimuli. According to DFF, hemispheric asymmetries arise when either high- or low-pass filtering is better suited to detect the information that is critical for a task. In terms of speech sound identification and discrimination, the left hemisphere can analyse formant transitions and other rapidly changing elements and the right hemisphere is important for prosodic elements.

The model of "asymmetric sampling in time" (AST) by Poeppel (2001, 2003) based on Allard and Scott (1975) also follows the idea of a difference in temporal resolution between hemispheres. AST states that the left hemisphere is better suited to integrate over short time-windows (ca. 20 - 40 ms). The right hemisphere is better at integrating over longer time-windows (ca. 150 - 250 ms). This implies that the analysis of formant transitions and other rapid aspects of segments are carried out by the left hemisphere. The processing of prosodic features and other suprasegmental elements, in turn, evoke right hemisphere processes.

4.5.3 Models of the neuroanatomy of speech perception

Finally, two recent models of the neuroanatomy of speech perception are described that sum up many of the previously mentioned findings and hypotheses. Common to both models is the idea of a division between sub-lexical processes that involve detailed articulatory and acoustic analysis and the process of speech comprehension, i.e. the mapping of sound to concepts and meaning.

Such a division of processing streams in auditory perception, particularly into ventral and dorsal streams, goes back to studies on visual (Milner and Goodale, 1995) and auditory processing in primates (Rauschecker, 1998). The ventral pathway generally subserves object identification, the dorsal pathway is involved in location and associated coordination of motions in visual perception. The role of the dorsal pathway in auditory processing in animals is often associated with locational analysis. Such a dual division of the auditory pathways can also be found in human subjects (Arnott et al., 2004; Scott and Johnsrude, 2003).

4.5.3.1 Ventral and dorsal pathways

Hickok and Poeppel (2000; 2004; Hickok, 2001) propose a model of speech perception where initial sublexical processing at the level of single speech sounds and syllables is carried out by the bilateral superior temporal gyrus. This posits a primary level in the speech perception process that could be described as a phonetic analysis and enables simple discrimination and identification tasks. Thus, the authors do not support the wide-spread assumption that the left superior temporal gyrus is mainly associated with phonetic-phonemic processing and the right side is involved in general auditory processing, that for example Näätänen et al. (1997) hold (see section 4.4.4). In support of this argument, Poeppel (2001) reviewed documented cases of word deafness, an impairment of speech understanding at the auditory-phonetic level. He found that in most cases, bilateral lesions were responsible, and that unilateral lesions did not lead to word deafness in many cases.

Hickok and Poeppel emphasize the involvement of both hemispheres in such a first phonetic processing stage but propose, further, that each hemisphere contributes in a different way. The differences lie in the temporal resolution and in the analytic strategies, which can be broken down to a difference between fast vs. slow processing (e.g. Belin et al., 1998), high-pass vs. low-pass filtering (Ivry and Robertson, 1998), temporal vs. spectral analysis (e.g. Zatorre et al., 2002), and short vs. long windows of integration (e.g. Poeppel, 2003). Subsequently, the processing system diverges into two processing streams:

(1) A ventral stream that is responsible for the mapping of sound to meaning, i.e. speech comprehension. The ventral stream projects towards the left superior temporal sulcus and the left posterior inferior temporal lobe.

(2) A dorsal stream which maps sounds to articulatory gestures. The auditory-motor interface is located at the left posterior inferior frontal areas. The authors propose that normal speech

comprehension does not necessarily involve this structure, if no phonemic segmentation, exact repetition, or other analyses of phonetic details are required.[25]

Data from aphasic patients, reviewed in Hickok and Poeppel (2004), supports this distinction between semantic mapping and access to articulatory representations. Lesions in Broca's area and adjacent regions yield poor performance on phoneme discrimination tasks but word comprehension was still possible (Caplan et al., 1995, cited in Hickok and Poeppel, 2004).

In consequence, one could expect a stronger acoustic-to-articulatory mapping for the processing of difficult nonnative sound contrasts. This was indeed found in a fMRI study by Callan and colleagues (2004) for Japanese subjects perceiving English [l] and [r]: neural areas associated with language articulation were more active in nonnative subjects. Moreover, those Japanese participants that performed best in an identification task showed most activation in these areas.

4.5.3.2 "What" and "how" pathways

Scott and Wise (2004), similar to Hickock and Poeppel, separate two distinct processing pathways:

(1) A left anterior pathway, that runs lateral and anterior to the primary auditory cortex that maps sound onto meaning. This "what" - pathway transfers to the prefrontal cortex and to medial temporal lobes.

(2) A "how"- pathway that maps sounds onto motor representations at the temporo-parietal junction and then passes on to the left posterior superior temporal sulcus to the premotor cortex. Mental representations of phonetic sound features could be stored here.

Scott and Wise emphasize that a distinction between speech-specific processing and processing of other sounds remains a challenge for future research. The authors point out that nonspeech sounds could be processed along the comprehension pathway, because they may also evoke a semantic analysis (compare Binder, 2000). Secondly, any sound that is reminiscent of speech may be processed as if it were speech, even if it is not intelligible.[26] Therefore, the question what is unique to speech sounds that would require specialized processing remains difficult to answer.

In relation to this, Scott and Wise remark that the contribution of the left and right auditory areas to prelexical speech processing is not easily distinguished. They disagree with a simple dichotomy between "fast" and "slow" processing on which most ideas of hemisphere differences are based. The left and the right hemisphere extract and process different parts of information. This may well be connected with differences in temporal resolution, but Scott and Wise have reservations against rash interpretations of lateralization effects. They demur that the underlying acoustic basis for hemisphere asymmetries may be more complex than

---

[25] Baddeley's "phonological "loop applies here (Baddeley, 1992; Baddeley et al., 1998). It refers to a working memory for speech sounds which allows for short-term maintenance of sound representations through articulatory rehearsal.

[26] This precedence of a speech mode of perception was also emphasized by Whalen and Liberman (1987).

what has been proposed so far. These cautions will be considered for the interpretation of the results of the present MEG study on fricatives that is described in the next chapter.

Chapter 5

Perception tests and MEG experiment

5.1 Introduction

The reviewed studies in Chapter 4 have shown that native phonological memory traces affect the generation of the mismatch negativity. The cross-linguistic studies that examined MMN suggest that this component reflects a relatively abstract level of processing in which linguistic experience plays a decisive role. Especially in the study by Phillips et al., (2000), MMN responses reflected a grouping of sounds only on the basis of sound category representations (compare Chapter 2, 2.1.1). Studies that used acoustically very diverse exemplars of speech sounds, showed that variation within a phonetic category is integrated into a neural sensory representation. The influence of phonotactic regularities and context assimilation on MMN amplitude further suggests the auditory areas have access to language-specific rules.

As shown in Chapter 2, the influence of the native language is not uniform. Speech sounds that function as phonemes warp the perceptual space, so that nonnative sounds are perceived as more or less well exemplars of a native phoneme category and, in the worst case, are grouped under one native category. The models PAM and SLM, that were described in the second chapter, detail these effects. The phonological function of sound contrasts also affects their perception. Additonally, the acoustics modify auditory processing, for example, whether a sound change is signalled by rapid or slow spectral changes or by voiced or voiceless cues.

The fact that acoustic properties of speech sounds as well as their phonological relevance influence auditory processing at the level of the MMN, motivated this investigation of pre-attentive processing of fricative sounds.

The neural responses of Polish and German participants to the contrast post-alveolar [ʃ] vs. alveolar [s] that is phonemic in both languages and to the Polish phoneme contrast between post-alveolar [ʃ] vs. palato-alveolar [ɕ] were compared.[1] Further, German and Polish listeners'

---

[1] The description of Polish [ʃ] as a post-alveolar (e.g. Rubach, 1984; Stone, 1987; Dogil, 1990) has recently been questioned. Hamann and Zygis (2003) argue that the articulatory features apicality, sub-lingual cavity, and the flat, retracted tongue suggest that this sound is a retroflex [ʂ]. In comparison, however, the retroflex articulation of a sibilant of Toda that is described by Ladefoged and Maddieson (2002) is much stronger and

neural responses to the German allophone pair [ç] and [x] were recorded. This sound contrast is a clear example of a contextually predictable sound variation with no relevance for lexical distinctions in German (Wiese, 1996). Unlike [ʃ] and [s], the German palatal fricative [ç] and the velar [x] cannot be used to distinguish words, because they do not occur in the same sound context, their distribution is complementary. Back vowels are followed by [x],[2] while [ç] occurs after front vowels and approximants, after [n], [l], and [r], and at the onset of words, as shown in the following examples:

"Tuch - Tücher", [tu:x] - [ty:çɐ] (cloth - cloths)
"Dach - Dächer", [dax] - [dɛçɐ] (roof, roofs)
"Chemie", [çɛmi:] (chemistry)

It is assumed that an phonologically underspecified segment underlies the sounds [ç] and [x]. Only as they are produced in phonological context, their place of articulation is first specified (Jessen, 1988; Wiese, 1996; Hall, 1992). Testing these three sound contrasts for the two subject groups allows to compare the auditory processing of native phonemic vs. nonnative, and native phonemic vs. native allophonic fricative sound changes. Electrophysiological studies on native sound processing, as described in Chapter 4, mostly work with phonemic contrasts. The present experiment with a native sound contrast that does not have a phonemic function can show, whether the generation of the auditory mismatch response incorporates the relevance of a native sound contrast for lexical distinctions. Evidence from behavioral research, that this might indeed be the case, is sparse. However, tests on the perception of partically contrastive tones (Hume and Johnson, 2003) and contextually distributed allophones (Peperkamp et al., 2003), showed that speech perception is sensitive to phonological function beyond the level of phonetic distinctiveness, even before word meaning is accessed. Furthermore, the fMRI study by Jacquemot et al. (2003) observed distinct activations between acoustic/phonetic change detection and the discrimination of phonemic changes.

For speakers of Polish, this contrast contains the native phonemic segment [x] and the unfamiliar palatal fricative.[3]

One challenge for the study of fricative perception is the lack of an adequate model of acoustics and articulation, which complicates the preparation of stimulus material. This added to the decision to use naturally spoken sounds. However, more importantly, there is evidence that natural and synthetic sounds are not processed in the same way. This issue is addressed in section 5.2 and the selection and editing of the stimulus material are described.

---

traceable than for the Polish sibilant. Furthermore, a retroflex sound could not be palatalized (Zygis, 2003; Zygis and Hamann, 2003). In loanwords, however, and at word boundaries, [ʃ] is palatized before [i]: e.g. “żigolak” [ʒigolak], Engl. "gigolo” and "oddarz dzisiaj” [odaʒʲ dʑiɕaj], Engl. "you will give it back today” (more examples can be found in Dogil, 1990). Consequently, an allophone with a different place of articulation would have to be introduced: [ʃʲ]. In this thesis, the traditional classification of Polish as a post-alveolar sound is adopted (Dogil, 1990). However, since the perceptual processing of the Polish and German sibilants was compared, a preparatory behavioral test (see section 5.6) was carried out to test whether Polish listeners assimilate German [ʃ] and [s] to the corresponding native categories.

[2] Velar [x] can alternate with uvular voiceless [χ] (Kohler, 1977; Wiese, 1996).

[3] A palatal variation of [x] is, however, found as an allophonic variation, dependent on the context (cf. Nawrocki and Gonet, 2004). Velar [x] occurs in all contexts, only before [i] or [j] it changes to [ç], (see also Wisniewski, 1998; and Wrobel, 1995, cited in Nawrocki and Gonet, 2004).

Section 5.3 presents the analysis of the acoustic properties of the fricatives, that were used for testing. The acoustic situation in the setting of the MEG recording was considered as well.

Before the MEG experiment was conducted, the behavioral discrimination performance of German and Polish listeners was assessed for the contrasts [ʃ] vs. [ɕ] and [ç] and [x]. These tests are described in sections 5.4 and 5.5.

A subsequent behavioral test, concerning the identification and rating of nonnative fricatives, is detailed in section 5.6. This test had two goals: first, to confirm the stimulus material for the MEG recording. Therefore, it was tested whether the nonnative sounds [ʃa] and [sa] were perceived as good exemplars of native correspondends. Second, the mapping of the Polish [ɕa] by German listeners and the mapping of the German [ça] by Polish listeners was quantified.

Subsequently, the MEG experiment is described, in which the magnetic auditory neural responses of German and Polish subjects were recorded for the three sound contrasts German [ʃa] - [sa], Polish [ʃa] - [ɕa], and German [ça] - [xa] (section 5.7).

Finally, section 5.8 describes a behavioral experiment that tested the impact of formant transitions by means of cross-spliced stimuli with conflicting cues in frication and transition.

5.2 Preparation of stimuli

5.2.1 Natural material

Naturally spoken stimuli were chosen. Many of the previous behavioral and brain studies have used synthetic materials. However, comparisons of natural and synthetic materials, such as the studies by Hertrich et al. (2002) and Benson et al. (2001) that were discussed in the previous chapter, found that the two types of sounds evoke different neural processes. Hertrich and collegues (2002), thus, suggested that synthetic sounds may only evoke phonetic processing if they are closely attended to.

Behavioral tests provide concordant results (Saffran et al., 1976; Gow and Caplan, 1996; Lacerda, 2001; Blomert and Mitterer, 2004). In particular, Lacerda (2001) shows that the perception of synthetic stimuli is much less immune against manipulations. Small acoustic changes in synthetic stimuli to lead to disproportionate perceptual confusion while natural material is processed more robustly.

In the case of fricatives, this may be problematic because the precise acoustic parameters that contribute to phonetic classification are not well known (Ladefoged and Maddieson, 2002; see section 5.3). Without an accurate model of fricative acoustics, it would be impossible to say whether perceptual effects are due to linguistic representations or problems with the stimulus material. Previous perception tests applied filtered noise, e.g. Healy and Repp (1982; see Chapter 2, section 2.1.3) or Manrique and Massone (1981), but in these tests, the fricational differences were not as subtle as they are in the case of the Polish post-alveolar vs. alveolo-palatal fricative contrast.

5.2.2 Syllables

Voiceless fricatives resemble nonspeech noise and this impression is reinforced by the repetitive presentations that are necessary in MEG experiments. This, and the fact that transitions to subsequent vowels may provide important perceptual cues for the identification of Polish fricatives (Nittrouer, 2002), motivated the usage of CV syllables in these experiments.

It was decided to apply the vowel [a] because in contrast to [u] and [o], [a] does not lead to anticipatory lip rounding during the production of the fricative. Compared to [i] and [u], there is no coarticulatory influence of a palatal glide for [a]. The vowel [a], furthermore, is common to both Polish and German and it is acoustically similar in the two languages.

For confirmation, a small random sample of 18 exemplars of [a] in the word "tasa" in carrier sentences (see section 5.2.3) was recorded by 5 male and 4 female German and by 3 male and 5 female Polish speakers. F1 and F2 values were found to be similar for the two language groups as shown in Table 5.1.

Table 5.1: Averaged F1 and F2 values of the vowel [a] over productions of German and Polish male and female speakers. Values were measured with PRAAT 4.2.18 for 40 ms over the vowel center.[4] F1 and F2 means are displayed in Hertz.

[a] exemplars of	*F1 mean*	*F1 median*	*F2 mean*	*F2 median*
Ger. males	723 ± 67	697	1235 ± 52	1235
Pl. males	714 ± 78	674	1251 ± 109	1217
Ger. females	644 ±135	647	1398 ± 100	1395
Pl. females	686 ± 202	686	1363 ± 103	1325

5.2.3 Recording

All stimuli were digitally recorded in the anechoic chamber of the Institute for Natural Language Processing, Stuttgart, at 16 bit quantization and 48 kHz sampling rate, downsampled to 22050 Hz.[5] Native speakers articulated the German syllables [sa], [ʃa], [a], and [xa], and the Polish syllables [sa], [ʃa], and [ɕa] in non-words in the following carrier sentences:

German: "Er hat ta_ gesagt" ;
for [a]: "Er hat te_ gesagt." (He said ta_ / te_.)

---

[4] Data is not normalized for speaker characteristics, such as F0 or vocal tract length.

[5] Down-sampling was necessary because the acoustic transmission system at the MEG center in Tübingen only supported a sampling rate of 22050Hz.

Polish: "On ukrad ta_ dzisiaj rano."[6] (He stole ta_ this morning.)
The German speakers were 5 females (25 - 29 years) and 5 males (23 - 31 years) who came from Stuttgart, Bonn, and Kln. The Polish speakers were 4 females (21 - 24 years) and 5 males (20 - 25 years) from Poznan and Gdansk. Several speakers were professionals (teacher, speech therapist, radio host, rap-singer, and actor) and spoke little or no dialect.

In a first recording, the speakers were asked to produce the sentences as naturally as possible to avoid overarticulation. The mean duration of the fricatives were calculated by analyzing three exemplars of each speaker and by determining the duration of the unvoiced part between the adjacent vowels. Voicing was assessed by the pitch analysis routine in PRAAT 4.2.18 (cross-correlation, measurement interval 3.33 ms for male exemplars, 2.5 ms for female exemplars).

Mean durations of the German fricatives:
s: 89.7 ± 12.8 ms
ʃ: 86.4 ± 12.6 ms
ç: 76.0 ±17.5 ms
x: 71.6 ±24.8 ms

Mean durations of the Polish fricatives:
s: 137.0 ± 21.1 ms
ʃ: 136.7 ±36.3 ms
ɕ: 137.8 ±31.9 ms

The measurements suggest that [ç] and [x] are generally shorter than the sibilants. This observation has been previously made by Mbius and van Santen (1996) for German fricatives and by You (1979) for English fricatives. You suggested that fricative duration decreases as fricatives are produced further back. However, Gordon et al. (2002) observed in their extensive survey of the acoustics of voiceless fricatives in eight languages that duration is only a weak distinctive parameter. If at all, duration only appears to be a distinctive parameter for voicing: voiced fricatives are shorter than voiceless ones (see also Behrens and Blumstein, 1988).

As a compromise for the differences between the duration of fricatives in Polish and German it was decided to set the duration of the full frication to 120 ms. The duration of the full vowel-part was also fixed at 120 ms. To obtain fricatives of this length, a second recording was made and the speakers were asked to talk slowly and clearly, but to avoid overarticulation.

5.2.4 Selection and editing

The stimuli were selected from the second recordings according to the following criteria:

- minimal aspiration after the fricative,

---

[6] Polish speakers were advised not to pronounce [ɫ] which is common in casual speech.

- duration for fricative and vowel about 150 ms each,
- even, constant fundamental frequency (F0) trend,
- exemplars of [x] with strong uvular production and predominantly vibrant characteristics were rejected.

Editing included:

- Duration was normalized by clipping the fricative onset and vowel offset (onset of the vowel was defined as the onset of periodicity), so that each fricative and each vowel had a duration of 150 ms.[7]
- Overall intensity was set to 70 dB SPL. This did not change the intensity relation between fricatives and vowels in the individual syllables.[8]
- At the onset and at the end of the syllables a smooth rising/falling ramp with a duration of 30 ms was added (Gaussian filter).
- F0 was set to a constant value throughout the vowel with respect to initial F0 value (adjustments < 30 Hz) by applying the Pitch-Synchronous-Overlap-and-Add algorithm (PSOLA), (Valbret et al., 1991; Boersma, 2001). Syllables that were judged to sound unnatural after the PSOLA resynthesis were rejected.

Finally, a set of 12 different tokens for each fricative category was obtained. This included three exemplars for each syllable category that were spoken by 2 male and 2 female German speakers and by 2 male and 2 female Polish speakers.

F0 did not distinguish the syllable categories (ANOVA: F = 1.74, sign. = 0.146).

---

[7] It has not been reported in the literature that the initial part of fricatives could be an important cue for fricative identification. Furthermore, the edited syllables were readily identified by native listeners. The initial part, specifically the duration of the rise of the noise amplitude appears to be important only for a distinction of fricatives and affricates (van Heuven, 1983).

[8] The intensity of frication depends on the size of the constriction and the distance to the mouth opening: the narrower the constriction the higher the intensity and the further back in the vocal tract the more noise is absorbed (Neppert, 1999). The intensity of a fricative, however, is generally not used to distinguish place of articulation. Rather, the relation between amplitude of the fricative and the following vowel can influence fricative identification (Stevens, 1980; Behrens and Blumstein, 1988). However, it has been shown that listeners only rely on this parameter when no other acoustic cue is available (Hedrik and Ohde, 1993).

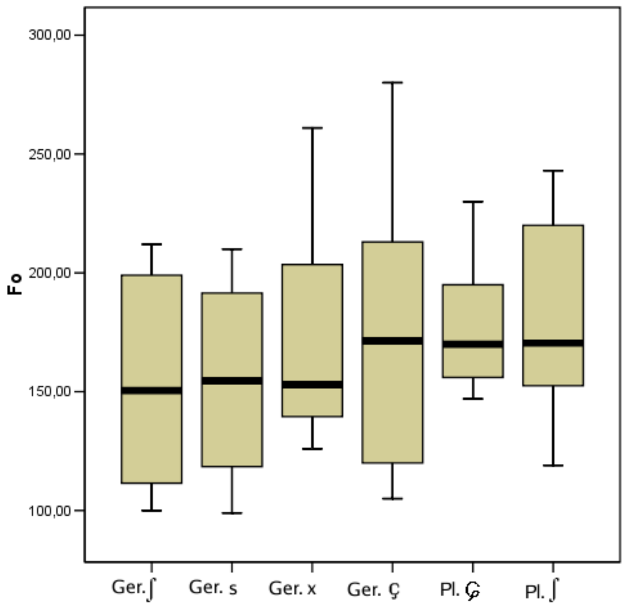


Figure 5.1: Variation of fundamental frequency F0 in the vowel parts of the stimuli.

The point of transition between fricative and vowel was at a mean of 150 ms ± 4 ms and each vowel had exactly 150 ms, including offset. An example is shown in Figure 5.2.

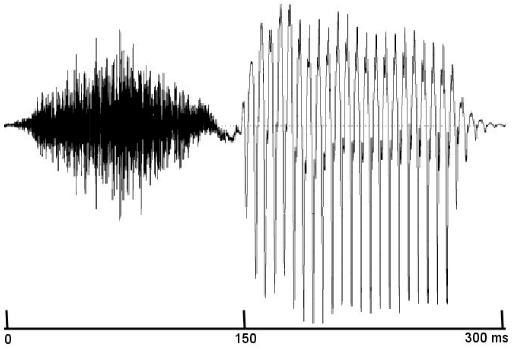


Figure 5.2: Representative stimulus waveform.

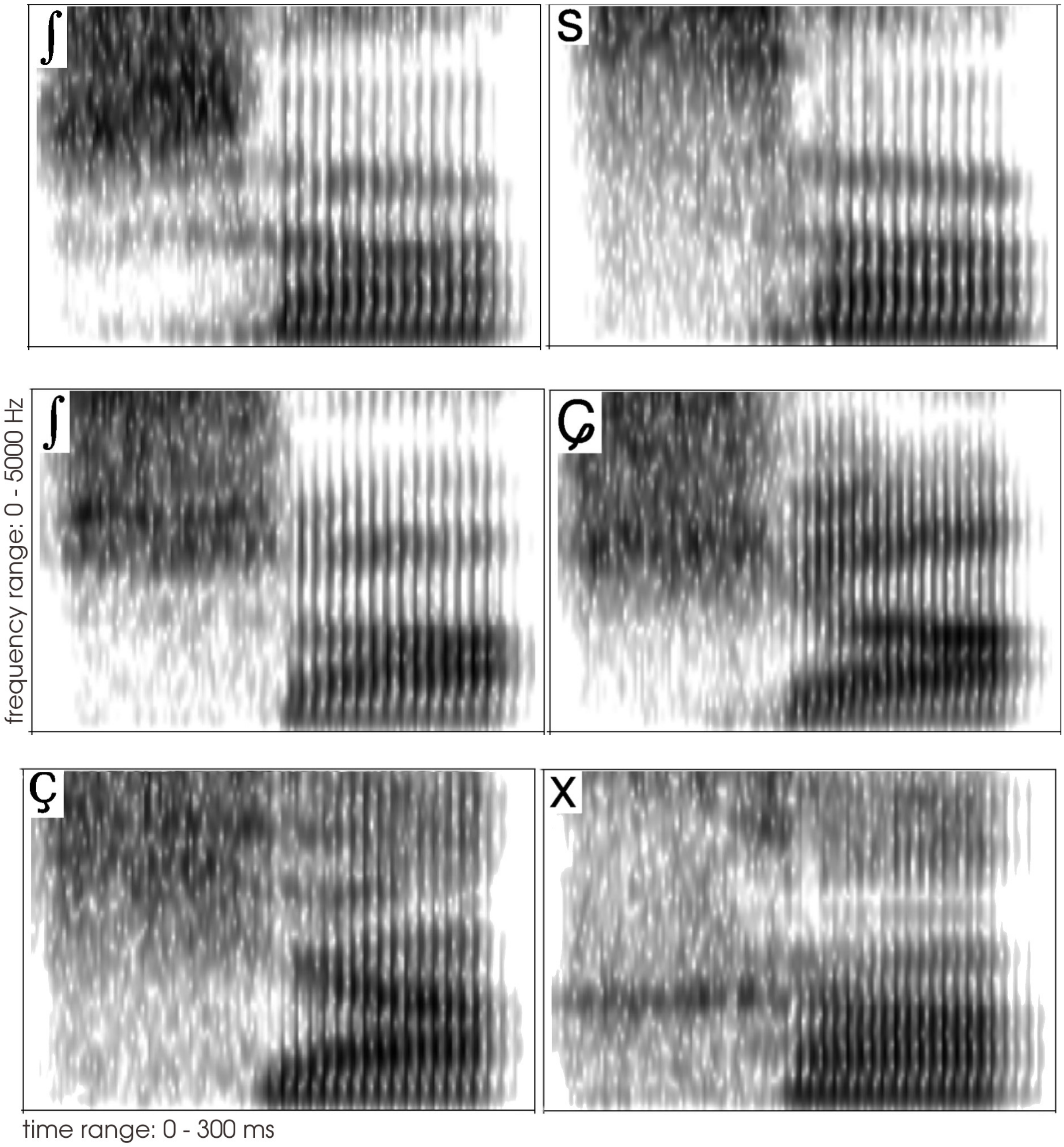


Figure 5.3: Representative spectrograms of stimuli. Top row: German [ʃa] and [sa], mid row: Polish [ʃa] and [ɕa], bottom row: German [ça] and [xa].

5.3 Acoustic analysis

Fricatives are characterized by a turbulent air flow that is generated as the air is forced through a narrow constriction and/or over a sharp edge (teeth). Turbulent signals are complex and chaotic random pressure fluctuations.[9] Fricative spectra result from the resonance of the vocal tract which shapes the noise spectrum. Up to now, no satisfactory model of the acoustics and articulation of fricatives has been proposed (Shadle and Mair, 1996; Ladefoged and Maddieson, 2002).[10]
Ladefoged and Maddieson (2002, p. 139) describe the situation as follows:

> The acoustic structure of fricatives seems to vary widely from individual to individual, but this really reflects only the unfortunate fact that we do not yet know what it is that we ought to be describing. We do not know how to sum up what is constant, and what is linguistically and perceptually most relevant in acoustic terms. As we do not yet have an adequate model for the acoustics of fricatives, we are in a position comparable to having to describe vowels without having the notion of formants, or at least peaks in the spectrum. Our best guess is that what matters for fricatives (more especially for sibilant fricatives) is the overall intensity, the frequency of the lower cut-off point in the spectrum, and something corresponding to the center of gravity and dispersion of the spectral components above a certain threshold.

These preliminary remarks suggest that acoustic analyses of fricatives only present approximations to a description of acoustic features that are relevant for identification and discrimination.

A number of parameters have turned out to be useful to distinguish fricatives: the spectral shape as characterized by Fast Fourier Transform (FFT) and Linear Predictive Coding (LPC) spectra (Ladefoged et al., 1994; Evers et al., 1998; Gordon et al., 2002), spectral moments (Jassem, 1979; Forrest et al., 1988; Matthies et al., 1996; Jongman et al., 2000), and formant transitions (Gordon et al., 2002). These methods of analysis were applied in the present study. The acoustic analyses were made for the syllables that served as stimuli for the MEG experiment.

5.3.1 Frication

FFT spectra were computed for each fricative over the mid-third part (40 ms). The individual exemplars of each fricative were averaged separately for the utterances by male and female

---

[9] The degree of turbulence can be described by the Reynolds Number (Re): Re = Vd/v. Vd = velocity of flow, v = kinematic viscosity of air (0.15cm^2/second). The degree of turbulence in relation to the distance between constriction and obstacle was incorporated into C.H. Shadle's differentiation between “obstacle source” and “wall source” fricatives (Shadle, 1985; Shadle, 1990). These types refer to sibilant and non-sibilant fricatives which can be differentiated acoustically by amplitude relations in the spectrum (Shadle and Mair, 1996). A definite set of parameters to model and distinguish place of articulation in fricatives, however, has not been developed yet.

[10] But see Nguyen et al. (1994) for a model that creates [ʃ] and [s] spectra with high accuracy by applying articulatory parameters.

speakers. The spectral moments center of gravity (COG first moment), skewness (third moment), and kurtosis (fourth moment) were calculated over the 40 ms FFT spectra.

As mentioned before, acoustic measurements for fricatives can vary highly between individual speakers. Thus, to provide a better insight into the data, individual values are displayed in Tables 5.2 to 5.4. Data of male and female speakers are averaged separately, as gender exerts an influence on frication (Jongman et al., 2000; Fox and Nissen, 2002).

To calculate the spectral moments the FFT spectrum is treated as a random probability distribution. Based on the obtained power spectra, the center of gravity, skewness, and kurtosis were computed using the routine provided in PRAAT, (see also Matthies, 1996).

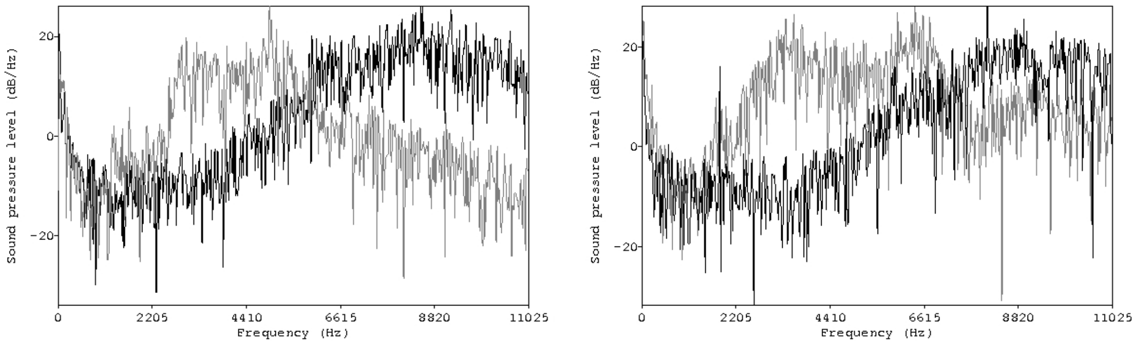
COG correlates with front cavity length: the more anterior the constriction, the higher is COG (Pols and Van Son, 1996).

Skewness indicates if the energy below and above COG are symmetrically distributed. A value of zero indicates a perfectly symmetric spectrum. For positive values, the right side of the distribution extends further than the left side, for negative values the opposite is the case.

Kurtosis states how much the spectral shape differs from a normal distribution. A zero value indicates a normal distribution. Higher, positive values indicate a more peaked spectrum. For negative values the spectrum has a flatter shape.

5.3.1.1 German [ʃ] - [s]

The spectrum of [ʃ] contains most energy in the range between 3 to 4.5 kHz. For [s] the frequencies between 6 and 9 kHz are prominent. This reflects in the skewness values. The kurtosis measurements indicate a much higher peakedness for [s] than for [ʃ].



male utterances

female utterances

Figure 5.4: Averaged acoustic spectra for German [ʃ] (grey) and German [s] (black).

Table 5.2: COG values for German [ʃ] and [s]. GM and GF stands for German male and female speakers, respectively. The numbers indicate the individual speaker and a, b, and c indicate the exemplar. ′

	German [ʃ]			*German [s]*		
	COG	*Skew.*	*Kurt.*	*COG*	*Skew.*	*Kurt.*
GM1a	4369	-0.09	1.64	8064	-0.41	1.4
GM1b	3665	0.29	1.03	8066	-0.8	3.36
GM1c	4227	0.06	2.42	7972	-0.76	2.56
GM2a	4022	0.25	3.35	8956	-1.62	7.98
GM2b	4629	-0.2	2.79	8729	-2.57	13.84
GM2c	4185	0.57	2.94	7817	-2.15	4.89
Mean	**4183**	**0.15**	**2.36**	**8267**	**-1.38**	**5.67**
GF1a	5466	0.09	0.16	8690	-1.27	3.08
GF1b	4759	0.78	0.72	8925	-2.41	8.98
GF1c	5845	-0.57	0.7	8706	-1.24	3.23
GF2a	4409	0.84	2.16	8064	-3.92	26.34
GF2b	4218	0.14	0.82	8066	-3.07	24.35
GF2c	3861	0.21	1.24	7972	-1.94	16.72
Mean	**4760**	**0.25**	**0.97**	**8404**	**-2.31**	**13.8**

The statistical difference between the spectral moments of [ʃ] and [s] was calculated in t-tests (all df = 10). COG and skewness were significantly different for the two sounds:

COG: male speakers: T = -17.7, p < 0.0001; female speakers: T = -10.3, p < 0.0001;

skewness: male speakers: T = 4.13, p = 0.002; female speakers: T = 5.34, p = 0.0003.

No significant difference was seen for kurtosis in the productions of male speakers: T = -1.73, p = 0.11; but those of female speakers differed significantly in kurtosis: T = -3.05, p = 0.012.

5.3.1.2 Polish [ʃ] - [ɕ]

Polish is one of the very few languages that contrasts three fricatives, with a place of articulation in the front of the mouth: alveolar [s], post-alveolar [ʃ], and alveolo-palatal [ɕ] (Rubach, 1984; Stone, 1987; Dogil, 1990). Polish [ɕ] is produced with the tongue blade and a doomed tongue behind the alveolar ridge towards the palate. Polish [ʃ] is produced slightly more fronted, towards the alveolar ridge with a flat tongue. The difference between the articulation of the two sounds lies the longer area of constriction for Polish [ɕ] and the protruded lips for [ʃ], whereas the lips are spread for [ɕ] (Dogil, 1990; Flemming, 2002; Ladefoged and Maddieson, 2002; Zygis, 2003; Zygis and Hamann, 2003; Padgett and Zygis, 2003).

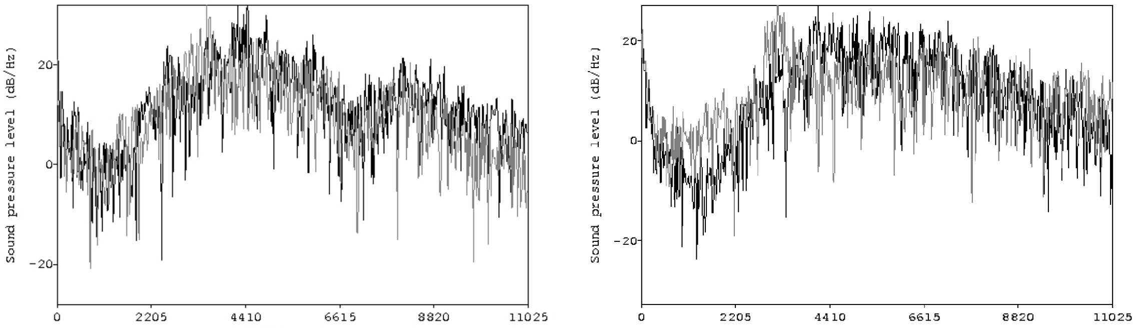
The similarity in the place of articulation for these fricatives reflects in their similar fricative spectra. Figure 5.6 shows mean LPC spectra of the two sounds which indicate a lower spectral peak for the post-alveolar fricative while [ɕ] contains more energy in a wide band starting at about 3400 Hz.

For exemplars by male speakers the main peak as approximated by the LPC spectra for [ʃ] is roughly at 3480 Hz and for [ɕ] at ca. 4370 Hz. Furthermore, [ʃ] contains more energy below 2000 Hz than [ɕ].

For the exemplars by female speakers the main spectral peak for [ʃ] is at approx. 3140 Hz, for [ɕ] at ca. 4140 Hz (both values were taken from the LPC spectra). Both spectral peaks are slightly higher but, overall, the energy falls off earlier for the exemplars by male speakers.

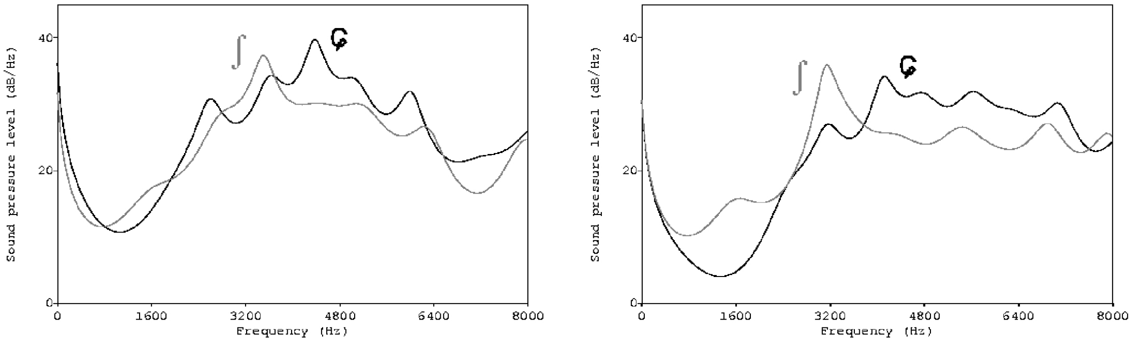
Similarly, Dogil (1990) as well as Kudela (1968) reported that Polish [ʃ] frication contains a spectral peak around F2 between 2 to 4 kHz, whereas Polish [ɕ] shows more energy slightly above this area. As Dogil argued, the higher energy in the F2 range for the post-alveolar results of the closure of the sublingual cavity and lower incisors cavity which is closed by the protruded lips for Polish [ʃ].

Thus, the perceptual discrimination of Polish [ʃ] and [ɕ] requires a finer analysis of frication differences than the German sound pair [ʃ] and [s].



male utterances female utterances

Figure 5.5: Averaged acoustic spectra for Polish [ʃ] (grey) and Polish [ɕ] (black).



male utterances female utterances

Figure 5.6: 24-order LPC spectra of Polish [ʃ] (grey) and [ɕ] (black), averaged over the productions of male and female speakers obtained over a time-window of 40 ms over the center of the fricative.

Table 5.3: Spectal moments for Polish [ʃ] and [ɕ]. PM and PF stands for Polish male and female speakers, respectively, the numbers indicate the individual speaker, a, b, and c indicate the exemplar.

	Polish [ʃ]			*Polish[ɕ]*		
	COG	*Skew.*	*Kurt.*	*COG*	*Skew.*	*Kurt.*
PM1a	3978	0.92	1.75	3719	0.94	1.74
PM1b	3994	0.56	1.85	3673	1.02	2.52
PM1c	4284	0.86	2.08	3479	0.79	1.17
PM2a	4535	1.38	1.68	4627	1.66	5.28
PM2b	3999	1.95	6.4	5306	0.45	0.04
PM2c	4083	1.69	4.49	4629	2.16	7.04
Mean	**4146**	**1.23**	**3.04**	**4239**	**1.17**	**2.96**
PF1a	4057	1.15	2.02	4954	-0.45	2.86
PF1b	5224	0.51	0.93	5677	-0.36	2.32
PF1c	4830	0.3	1.54	5715	-0.65	1.03
PF2a	5857	0.07	-0.65	6152	-0.13	0.15
PF2b	4799	0.9	-0.38	6086	-0.29	0.82
PF2c	4836	0.37	-0.57	6214	-0.2	0.18
Mean	**4934**	**0.55**	**0.48**	**5479**	**-0.35**	**1.23**

The statistical difference between the spectral moments of Polish [ʃ] and [ɕ] was calculated in t-tests (all df = 10).

The fricatives produced by male speakers were not reliably distinguished by the spectral moments of the two fricatives: COG: T = -0.3, p = 0.77; skewness: T = 0.17, p = 0.87; kurtosis: T = 0.57, p = 0.96.

The productions of the female speakers showed distinct values for COG: T = -2.8, p = 0.019; and for skewness: T = 4.95, p = 0.0006; but not for kurtosis: T = -1.13, p = 0.29.

Dogil (1990) argues that the protrusion of the lips for Polish [ʃ] is the major contrasting feature between the Polish post-alveolar and the palato-alveolar fricatives because their places of articulation are very close. The lower peak that was observed in the spectrum of [ʃ] may result from the involvement of the lips. In comparison, the German [ʃ] spectra do not show such a pronounced, but a broader peak in the frequencies around 3500 Hz.

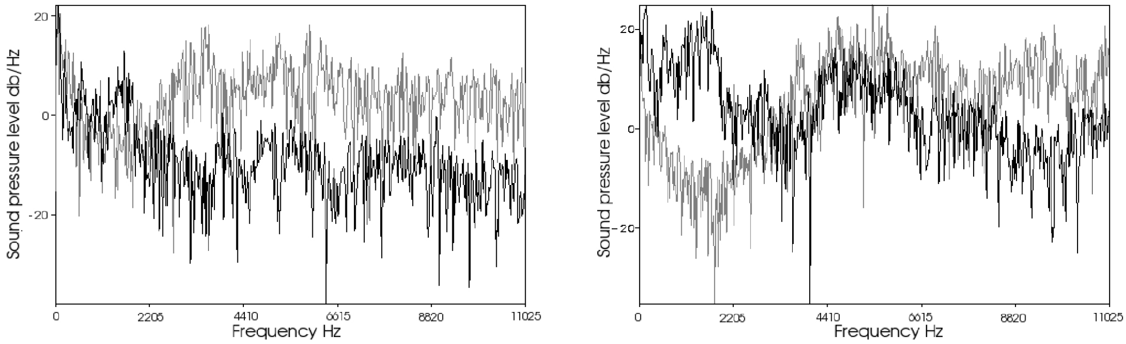
Since there is no need to distinguish another sibilant that is similar in the place of constriction, German speakers may not involve the lips as much as Polish speakers. When the spectral moments of the 12 Polish and German exemplars of [ʃ] were compared by t-tests, however, only the parameter of skewness in the productions of the male speakers differed significantly (T = 4.37, p = 0.0014). Otherwise not significant difference in the spectral moments were found (all p > 0.2). Thus, the need for a stronger lip protrusion in Polish [ʃ] yields only slight acoustic differences between this sound in German and Polish.[11]

---

[11] Furthermore, this finding is critical for the claims by Zygis and Hamann (e.g. 2003) about the retroflex articulation of the Polish sound, that were mentioned in 5.1.

5.3.1.3 German [ç] - [x]

The German allophones are distinct in all measured spectral moments with large differences between male and female productions in COG and kurtosis. The averaged spectra show that the velar fricative was produced with lower amplitude by the male speakers. The values in Table 5.4 reveal high variation between individual productions by male speakers for [x]. The exemplar by GM1a was probably produced further back at the uvular position.



male utterances female utterances

Figure 5.7: Averaged acoustic spectra for German [ç] (black) and [x] (grey).

Table 5.4: Spectral moments for German [ç] and [x]. GM and GF stands for German male and female speakers, respectively, the numbers indicate the individual speaker, a, b, and c indicate the exemplar.

	German [ç]			*German [x]*		
	COG	*Skew.*	*Kurt.*	*COG*	*Skew.*	*Kurt.*
GM1a	2566	0.83	-0.29	368	5.36	30.38
GM1b	2868	0.57	-0.49	890	3.10	10.31
GM1c	3251	0.34	-0.62	833	3.23	11.3
GM2a	6256	-0.36	0.53	1306	2.63	6.16
GM2b	5432	-0.21	-0.15	1661	2.49	5.99
GM2c	5650	0.05	-0.4	1491	3.03	11.08
Mean	**4337**	**0.2**	**-0.24**	**1092**	**3.31**	**12.5**
GF1a	6395	0.43	0.19	2206	1.22	0.56
GF1b	7866	-0.42	-0.78	2259	1.29	0.63
GF1c	6275	0.6	0.63	2010	1.0	-0.01
GF2a	5793	-0.42	0.37	2323	1.63	2.71
GF2b	5518	-0.07	0.69	2301	1.94	4.03
GF2c	5515	-0.15	0.18	2532	1.73	2.47
Mean	**6227**	**-0.005**	**0.21**	**2272**	**1.47**	**1.73**

The spectral moments were highly distinctive between the two fricatives, as determined by t-tests (df = 10):

COG: male speakers: T = 4.71, $p < 0.0001$; female speakers: T = 10.8, $p < 0.0001$;

skewness: male speakers: T = -6.65, $p < 0.0001$; female speakers: T = -6.47, $p < 0.0001$;

kurtosis: male speakers: T = -3.45, $p = 0.006$; female speakers: T = -2.24, $p = 0.05$.

5.3.2 Formant transitions

Differences in the transition from a fricative to a vowel are most pronounced in the second formant (F2) as reported by Kudela (1968) in an acoustic analysis on Polish fricatives. Data from Wilde (1993) on English fricatives further shows that the second formant systematically relates to the place of articulation (see also the discussion in Jongman, 2000). For plosive obstruents, F2 has long been listed as an important indicator of place of articulation (Delattre et al., 1955; Sussman and Shore, 1996; for review see Neppert, 1999).

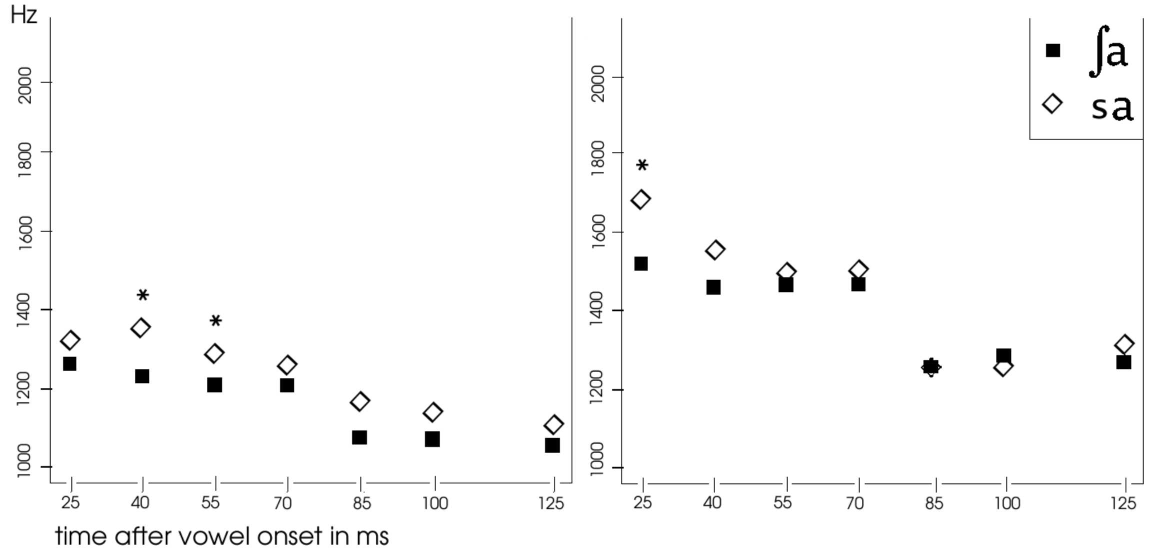
Therefore, the F2 transitions were measured in seven consecutive time-frames: from the onset of periodicity to 25 ms into the vowel, five consecutive frames of 15 ms each, and a seventh frame from 100 to 125 ms after vowel onset. The obtained formant values represent the means in these time-frames. The F2 values were averaged separately for utterances of male and female speakers.

Unpaired t-tests were applied to compare the F2 values between the German and Polish contrasts in every time-frame.

5.3.2.1 German [ʃa] - [sa]

The transitions of F2 are overall similar for German [ʃa] and [sa]. Slight differences are noticable at the onset of the vowel.

One may conclude that the transitional information is not very critical for the differentiation of these fricatives. A perception experiment by Heinz and Stevens (1961) supports this notion. Heinz and Stevens tested synthetic versions English fricatives ([ʃ], [s], [f], and [θ]) and found that the cues for the differentiation of [ʃ] and [s] mainly lie in the frication, whereas for [f] and [θ] the transitions to the following vowel are important. Similar observations by Nittrouer (2002) for English fricatives were mentioned in Chapter 2, section 2.1.3.3.



male utterances female utterances

Figure 5.8: Mean values of F2 after German [ʃ] - [s], measured in 7 time-frames after vowel onset (full squares = following German [ʃa], open diamonds = following [sa], stars indicate significant differences, $p < 0.05$).

Table 5.5: Mean values of F2 after [ʃ] and [s] spoken by male German speakers.

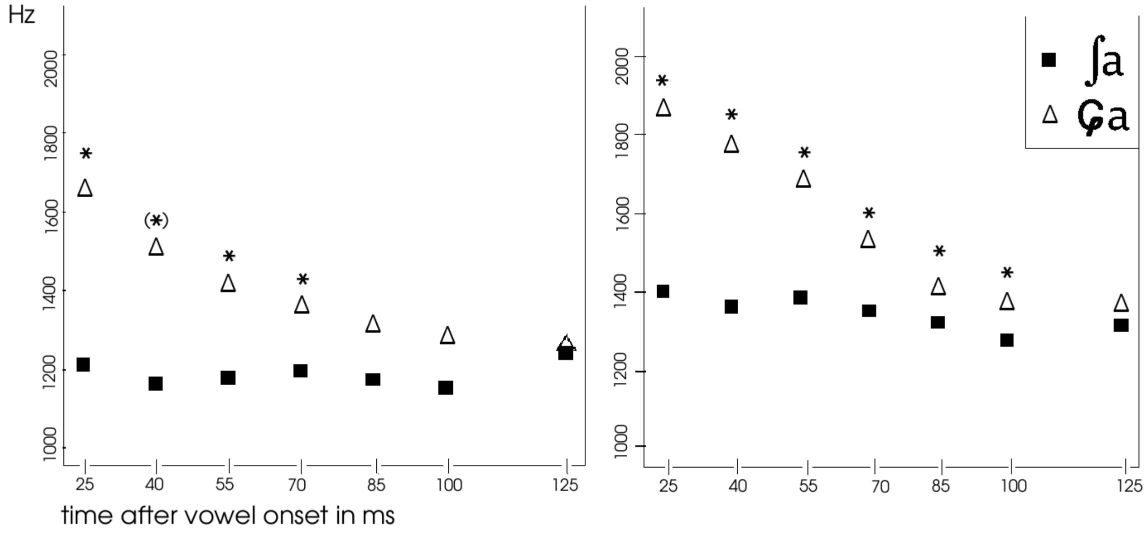
	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Ger. [ʃ]	1268 ±109	1227 ±61	1204 ±41	1196 ±47	1069 ±76	1059 ±58	1050 ±90
Ger. [s]	1336 ±62	1340 ±71	1304 ±71	1239 ±51	1141 ±140	1292 ±381	1103 ±72
p	0.21	0.0023	0.014	0.16	0.24	0.17	0.28

Table 5.6: Mean values of F2 after [ʃ] and [s] spoken by female German speakers.

	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Ger. [ʃ]	1528 ±142	1444 ±120	1424 ±98	1421 ±118	1248 ±73	1289 ±57	1263 ±43
Ger. [s]	1684 ±28	1552 ±60	1477 ±51	1499 ±70	1248 ±38	1241 ±34	1318 ±87
p	0.025	0.076	0.27	0.19	1.0	0.11	0.19

5.3.2.2 Polish [ʃa] - [ɕa]

The formant transitions of F2 are significantly different for the Polish sibilant contrast over a period of about 70 ms after vowel onset for the male speakers' exemplars and up to about 100 ms for the female utterances.



male utterances

female utterances

Figure 5.9: Mean values of F2 after Polish [ʃ] - [ɕ], measured in 7 time-frames after vowel onset (full squares = Polish [ʃa], open triangles = [ɕa], stars indicate significant differences, $p < 0.05$). The difference between formant values in the second measurement of the male speakers' productions approached significance ($p = 0.06$).

Table 5.7: Mean values of F2 after [ʃ] and [ɕ] for productions by male Polish speakers.

	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Pl. [ʃa]	1165 ±192	1161 ±86	1170 ±71	1188 ±68	1934 ±81	1125 ±170	1239 ±71
Pl. [ɕa]	1655 ±71	1564 ±489	1528 ±363	1381 ±171	1391 ±242	1302 ±110	1256 ±124
p	0.0002	0.06	0.04	0.028	0.087	0.058	0.77

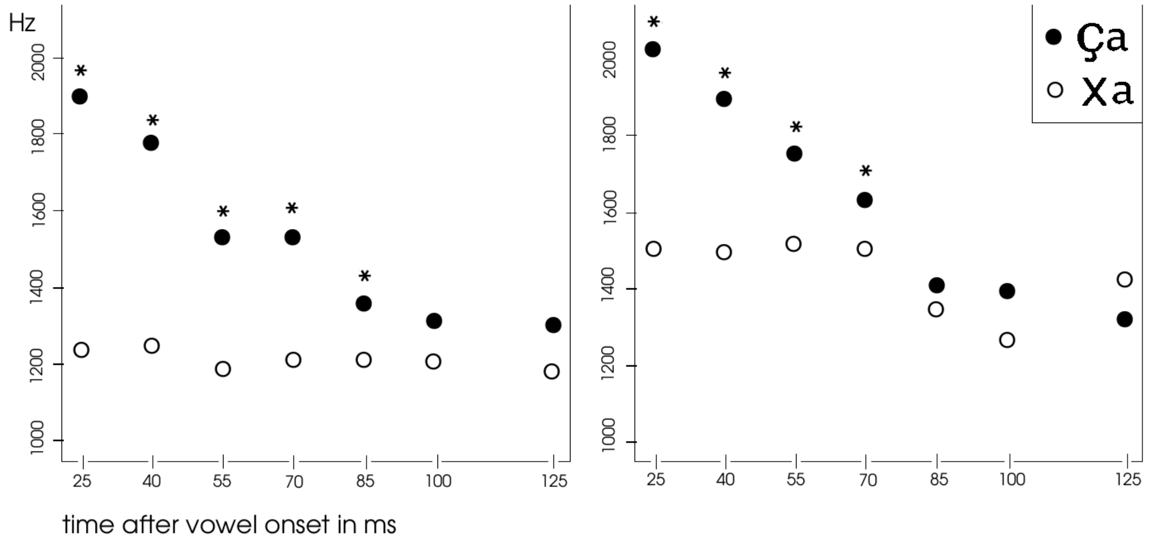
Table 5.8: Mean values of F2 after [ʃ] and [ɕ] for exemplars by female Polish speakers.

	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Pl. [ʃa]	1354 ±218	1281 ±269	1306 ±236	1301 ±202	1295 ±103	1271 ±108	1267 ±134
Pl. [ɕa]	1897 ±70	1813 ±24	1696 ±97	1523 ±76	1418 ±23	1384 ±23	1368 ±28
p	0.0002	0.0007	0.0038	0.024	0.017	0.032	0.1

5.3.2.3 German [ça] - [xa]

The F2-transitions clearly differentate the two sounds. The downward movement of the tongue in the palatal German fricative results in clear differences in F2 in contrast to [x]. F2 differs between the two sounds up to about 70 ms after vowel onset for male and female speakers' productions.

Considering also the differences in frication between these two sounds, the fact that they are allophones does not seem to lead to a blurred acoustic contrast.[12]



male utterances

female utterances

Figure 5.10: Means of F2 after German [ç] - [x], measured in 7 time-frames after vowel onset (full circles = a, open circles = xa, time is shown in seconds after vowel onset, stars indicate significant differences, p < 0.05).

Table 5.9: Mean values of F2 after [ç] and [x] produced by male German speakers.

	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Ger. [ça]	1869 ±342	1880 ±440	1510 ±157	1583 ±357	1363 ±204	1329 ±161	1307 ±146
Ger. [xa]	1227 ±211	1224 ±86	1211 ±75	1269 ±67	1257 ±218	1276 ±264	1168 ±186
p	0.0029	0.0049	0.0018	0.032	0.40	0.68	0.69

---

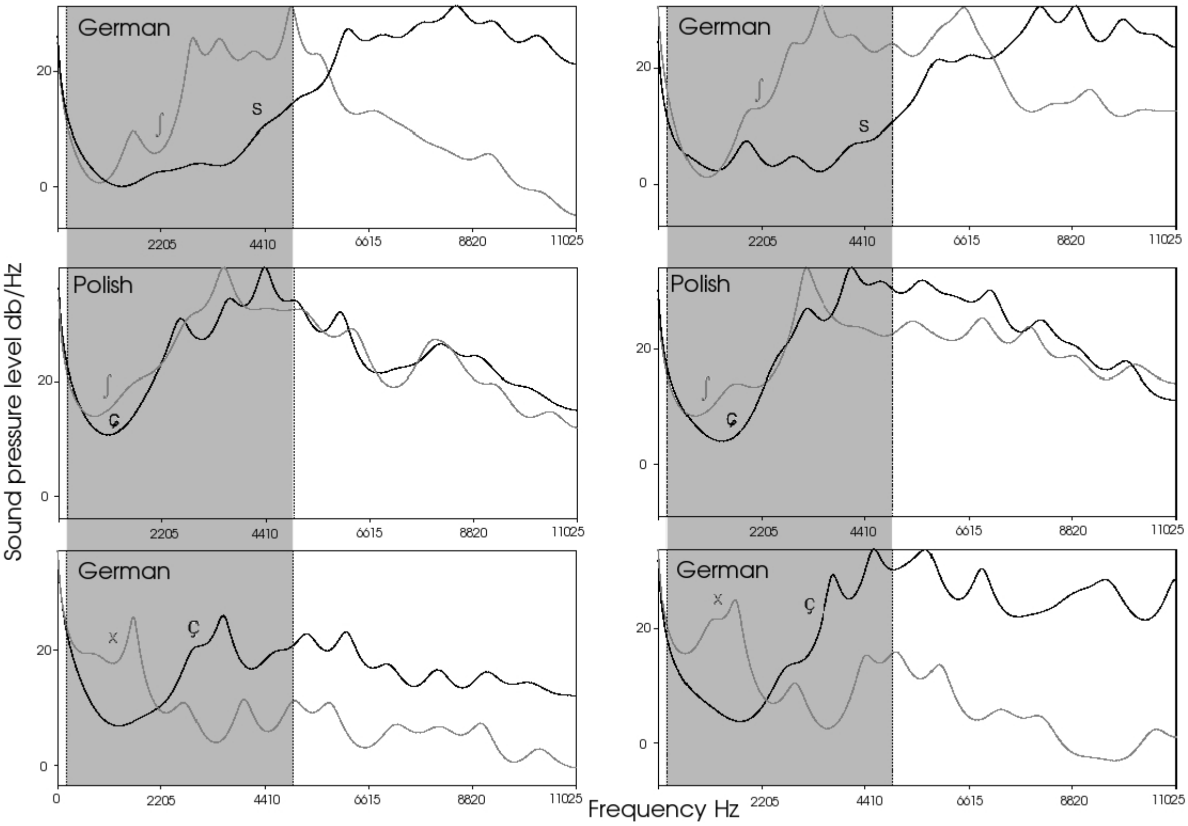
[12] That phonological status does not necessarily affect the acoustic contrastiveness was also found by Evers et al. (1998). They investigated whether the contrast between [s] and [ʃ] was acoustically reduced when the sounds were allophones (as in Bengali and Dutch) as compared to when they served as phonemes (as in American English). Evers et al. observed that acoustic variation could not be associated with the phonological status of the sounds.

Table 5.10: Mean values of F2 after [ç] and [x] spoken by female German speakers.

	25ms	*40ms*	*55ms*	*70ms*	*85ms*	*100ms*	*125ms*
Ger. [ça]	2025 ±90	1887 ±95	1750 ±127	1645 ±72	1372 ±161	1350 ±163	1321 ±175
Ger. [xa]	1493 ±124	1507 ±121	1516 ±105	1511 ±114	1321 ±117	1291 ±135	1386 ±137
p	<0.0001	0.0001	0.0059	0.036	0.54	0.51	0.49

5.3.3 Fricative acoustics in the MEG setting

No electrical devices, like headphones or earphones, can be operated close to the MEG sensor. Accordingly, plastic tubes transmitted the sounds during the MEG-recordings. These tubes act like a bandpass filter with a passband of about 200 - 5000 Hz, dropping off steeply at the higher frequencies until the signal is completely suppressed above 8 kHz. Below 200 Hz transmission is damped by about 10 dB.[13] The parts that were transmitted clearly are visualized in the LPC spectra in Figure 5.11.



male / female utterances

Figure 5.11: 24th-order LPC spectra of frication, averaged over 40 ms over the center of the fricatives. Exemplars of male are shown in the left panels, exemplars by female speakers in right panels. Shaded areas denote the frequencies that were transmitted undamped in the MEG setting.

The spectral moments were measured for the stimuli that had been filtered according to conditions in MEG setting (pass-band between 200 - 5000 Hz). T-tests were used to test whether frication provided distinctive cues in the filtered conditions (all df = 22).

Filtered German [ʃ] and [s] differed significantly in COG (T = 2.52, p = 0.02), but not in skewness (T=1.61, p = 0.14), and kurtosis (T = 1.83, p = 0.08).

---

[13] Many thanks to Wolfgang Wokurek for help with analyzing the transmission of the MEG air tube system.

Polish [ʃ] and [ɕ] were close to a significant difference in COG (T = -2.08, p = 0.05) and differed in skewness (T = 2.66, p = 0.01), but not in kurtosis (T = -0.14, p = 0.89).

COG values for German [ç] and [x] differed significantly in COG (T = 17.4, p < 0.0001), skewness (T = -8.22, p < 0.0001), but not in kurtosis (T = 0.8, p = 0.41).

Table 5.11: Average values of spectral moments for filtered fricatives (passband = 200 to 5000 Hz)

	COG	*Skew.*	*Kurt.*
Ger. [ʃ]	3703 ± 158	-0.2 ± 0.57	3.55 ± 2.66
Ger. [s]	3196 ± 684	-0.75 ± 1	0.75 ± 4.6
Pl. [ʃ]	3547 ± 204	-0.14 ± 0.48	5.41 ± 5.6
Pl. [ɕ]	3873 ± 501	-1.06 ± 1.09	5.66 ± 2.88
Ger. [ç]	4057 ± 356	-1.32 ± 1.14	6.69 ± 6.76
Ger. [x]	1790 ± 277	1.84 ± 0.69	4.62 ± 5.25

The spectral moments and spectra indicate that the main acoustic differences between Polish [ʃ] and [ɕ] and between German [ç] and [x] are transmitted in the acoustics of the MEG setting. Only German [s] is strongly affected by the limited acoustics.

Therefore, the comprehensability of the stimuli was verified behaviorally. Four German and 4 Polish listeners (students from Stuttgart and Polish exchange students[14]) were asked to discriminate the stimuli in the setting of the MEG-recording. Stimuli were presented as chains of standard and deviants as in the later MEG recording. German listeners heard the contrasts [ʃa] - [sa] and [ça] - [xa]. Polish participants listened to [ʃa] - [sa] and [ʃa] - [ɕa]. Ten alternations of each contrast were presented. The listeners had to report verbally to the experimenter which contrast they had heard. All contrasts were correctly identified. Then, each stimulus exemplar was presented once in isolation. Again, German speakers listened to the [ʃa], [sa], [ça], [xa] tokens, and Polish speakers heard German and Polish [ʃa], German [sa], and Polish [ɕa]. The order of stimulus presentation varied randomly. The participants were asked to identify each sound upon presentation by verbalizing it to the experimenter. All 12 exemplars were correctly identified by all listeners. Therefore, it was assumed that the fricatives could be applied in the MEG setting.

---

[14] Three of these Polish participants took part in the later MEG experiment.

5.4 Behavioral discrimination of Polish [ʃ] - [ɕ]

At informal testing, German listeners have difficulties distinguishing the Polish contrast between [ʃ] and [ɕ]. This limitation has also been documented in American English listeners in a phonetic labelling test by Lisker (2001; see Chapter 2, section 2.1.3.2).

5.4.1 Materials and methods

5.4.1.1 Subjects

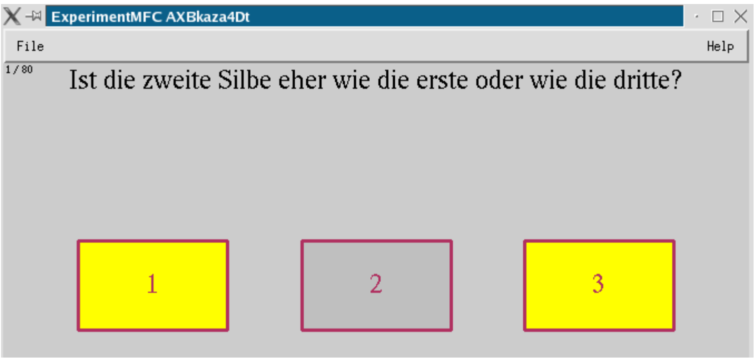
Twenty German subjects (university students and first-year students of speech therapy, mean age = 27.7, range = 20 - 40) without any knowledge of Polish and 6 Polish listeners (mean age = 23, range = 20 - 26) participated. The Polish volunteers were on a study-exchange in Germany and were fluent in German and English. All subjects reported undisturbed speech and hearing capacities.

5.4.1.2 Stimuli and procedure

The Polish syllables [ʃa] and [ɕa] that were to serve as stimuli in the MEG recording (see section 5.2 for recording and editing) were used in this test. In order to assess the contribution of formant transition for discrimination performance, the fricative-noise parts were cut from the CV-syllables at approx. 20 ms before onset of voicing.

An AXB procedure was applied in which subjects heard three sounds in a row (ISI = 1 s). The isolated fricatives and the syllables were tested in separate trials. Stimuli were presented randomly in an AXB scheme and none of the three sounds in any one trial were identical. Eighty trials per subject and condition (isolated or syllables) were applied. The order of testing was balanced across the groups. The listeners had to decide whether the second sound had more resemblance to the first or to the third sound. The experimental setup provided by PRAAT was applied in these tests. The subjects were tested in a quiet office with a DELL Latitude D600 laptop and AKG K501 headphones.

Before the test started, the subjects passed 3 to 5 trials without receiving feedback in order to familiarize them with stimuli and procedure.



Ist die zweite Silbe eher wie die erste oder wie die dritte?

Figure 5.12: User-interface for the AXB discrimination experiment.

AXB instead of AX was chosen, because in AX tests very difficult discriminations may lead to the answer "same" (Macmillan et al., 1977). The advantage over the ABX paradigm is that the time interval between A and X and between B and X is the same. Repp (1984) notes that in ABX-tests subjects may do only a BX comparison. A similar paradigm is the 4IAX test, in which two AX pairs are presented and subjects have to decide which of them was different. Yet another method is the so-called "oddity test". Hereby, subjects have to decide which one of three stimuli is different from the others: it can be the first, second, or the third (Miyawaki et al., 1975). In comparison to these tests, the AXB test was preferred because it requires lower memory effort, (for further reviews of test paradigms see Beddor and Gottfried, 1995).

5.4.2 Results

The order of testing did not significantly affect the results (German subjects: CVs: $T = 0.05$, $p > 0.2$ for CV condition; isolated fricatives: $T = 1.15$, $p > 0.2$; Polish subjects: the number of correct responses of the Polish controls was identical in both sequences).

The syllables were assigned correctly in the German group with 69.7 ± 13.0%. Isolated fricatives yielded 61.5 ± 10.4% for German listeners. Polish listeners identified the syllables and the isolated fricatives almost perfectly: CVs: 99 ± 1.2%, and isolated fricatives: 93.7 ± 3.5%.

These results show that German listeners correctly discriminated the Polish contrast above chance, however, their performance was significantly lower than that of the native listeners. One reason for the moderate discrimination abilities of German listeners may be the fact that Polish [ɕ] resembles German [ç] acoustically. German subjects may have assimilated the nonnative Polish sound to this native category and, thus, were able to differentiate [ʃ] and [ɕ]. The experiment in section 5.6 tests this possibility.

Furthermore, the results suggest a high relevance of formant transitions for differentiation of this contrast for both Polish and German listeners because the number of correct responses was significantly higher in the CV condition as compared to the isolated-noise condition (German subjects: $T = 2.49$, $p < 0.05$; Polish subjects: $T = 3.33$, $p < 0.05$).

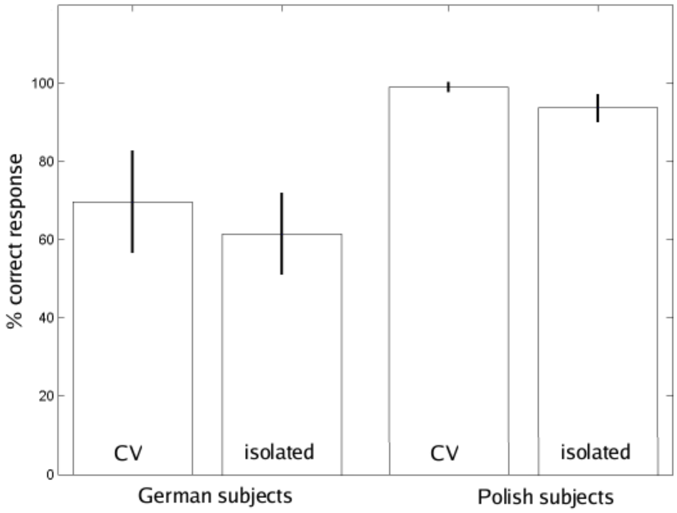


Figure 5.13: Discrimination of Polish [ʃ(a)] - [ɕ(a)]. Correct responses for German and Polish subjects for syllables and isolated fricatives.

In comparison to the present results, the American English subjects in the aforementioned phonetic labelling test by Lisker (2001) for the Polish [ʃ], [ɕ], and [s] in fricative-[a] syllables and in isolation, could identify the isolated fricatives better than the syllables. Lisker's subjects seemed to assimilate [ɕa] strongly to the native syllable [ʃa], but were able to distinguish the isolated sounds on the basis of acoustic comparisons, that Lisker describes as a detection of noise-"pitch" and intensity. Lisker, therefore, suggests that a nonphonetic mode of perception was used.

5.5 Behavioral discrimination of German [ç] - [x]

5.5.1 Materials and methods

5.5.1.1 Subjects

The same 20 German subjects that had taken part in the AXB discrimination of the Polish phoneme contrast[15] and 13 new Polish subjects who were students at the University at Poznan, Poland (mean age: 22.8, range: 20 - 29 years) with no knowledge of German participated.

---

[15] The German subjects participated in both test in the same session.

5.5.1.2 Stimuli and procedure

CV-syllables and isolated fricatives of each sound category, as described in section 5.2, were presented in an AXB test in the same way as in the discrimination test for Polish [ʃ] and [ɕ]. Eighty trials per subject and condition were applied. Testing was done by means of the experimental set-up provided by the software package PRAAT with a DELL Latitude D600 laptop and AKG K501 headphones. German subjects were tested in Stuttgart, Polish subjects in Poznan, each time in a quiet office. Before the test, subjects passed 3 to 5 trials without receiving feedback so that they became familiar with stimuli and task.

5.5.2 Results

For the CV condition the average score of correct responses for German listeners was 88.1 ± 16.8%. For fricatives presented in isolation, German participants attained an average score of 92.5 ± 10.7%.

The average score for Polish listeners in the CV condition was 96 ± 6%, and in the isolated fricative condition Polish listeners had 99.0 ± 1.6% correct answers.

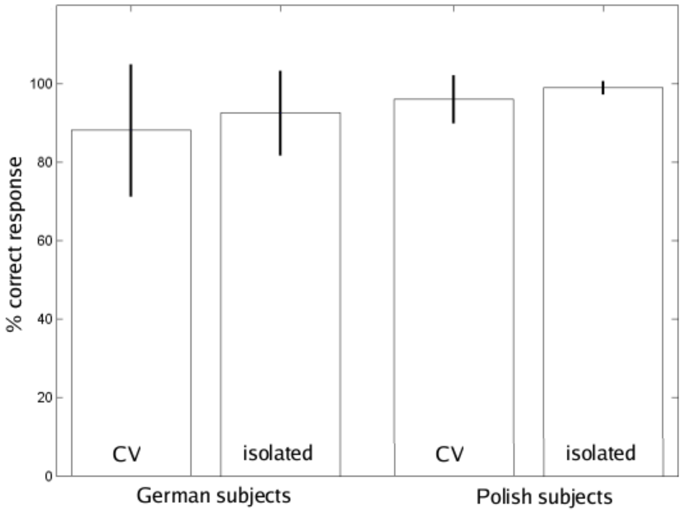


Figure 5.14: Discrimination of German [ç(a)] - [x(a)]. Correct responses for German and Polish subjects for syllables and isolated fricatives.

Although the Polish group showed a slight tendency for better discrimination, the performance of Polish and German listeners did not differ significantly in the syllable condition (t-test: $p = 0.14$). In the discrimination of the isolated noises, the Polish subjects attained slightly better results and the difference to the German group was significant ($p = 0.05$). The fact that the two sounds are not used contrastively in German did not seem to severely diminish difference detection for German listeners. For Polish listeners, this contrast appears to be equally salient.

Responses to the CV stimuli and isolated fricatives did not differ significantly for either group (t-test: Germans: $p = 0.34$; Polish: $p = 0.11$). The formant transitions, therefore, did not seem to influence the discrimination of [ç] and [x].

In addition, the informal reports that the majority of German subjects made after they had finished the discrmination tests. They remarked that they had never thought about the acoustic difference between [ç] - [x] before because both sounds are written in the same way (which is "ch"). There is evidence for a connection between orthographic knowledge and awareness of sound segments which, in turn, influences speech perception (Jusczyk, 1985; Read et al., 1986; Morais et al., 1985; Derwing, 1992; Morais and Kolinsky, 1995).[16] It could be speculated that slightly, although not significantly reduced discrimination performance of the German group may reflect an influence of orthographic knowledge.

5.6 Identification and rating of nonnative fricatives

5.6.1 Materials and methods

5.6.1.1 Subjects

The same 20 native German speakers that had participated in the discrimination test[17] and 19 new Polish subjects (mean age = 22.8, range = 20 - 29) participated in the rating test. The Polish subjects knew English and French but had no knowledge of German. Thirteen Polish subjects were students of the University of Poznan and 6 participants were students at the Teacher's college in Katowice.[18] None of the subjects had any speech or hearing problems.

5.6.1.2 Stimuli and procedure

Ten exemplars each of German [ʃa], [sa], and [ça]; and Polish [ʃa], [sa][19], and [ɕa] were used. In a random sequence, each stimulus was presented four times resulting in 240 trials per

---

[16] Further thoughts come from Aronoff (1992) who provides an interesting discussion on the notion that orthography may reveal the psychological reality in many cases, i.e. the phonological representations of sounds, as proposed by Sapir (1933). A similar claim comes from Chomsky and Halle: "conventional orthography is a near optimal system for the lexical representation of English "words (Chomsky and Halle, 1968, p. 49; cited in Aronoff, 1992, p. 77).

[17] The two tests took place two weeks apart.

[18] I am very thankful for the help by Andrzej Łyda from the Institute of English, University of Silesia at Sosnowiec for providing subjects and facilities for the tests.

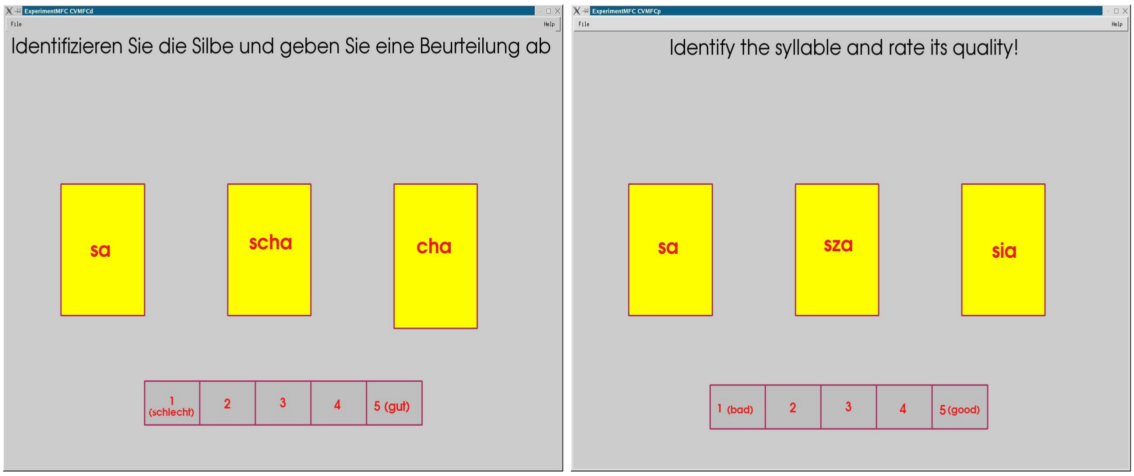
[19] Polish [sa] exemplars were recorded and edited in the same way as the other syllables (see section 5.2).

subject. The stimuli were repeated twice in each trial with an ISI of 1 second. Subjects were asked to identify the stimuli and rate them for intelligibility.[20]

They had to choose from three categories labeled according to the respective native orthography: scha, sa , and cha - German for ʃa, sa, and ça, and sza, sa, and sia - Polish for ʃa, sa, and ɕa. Intelligibility was rated on a scale from 1 (bad) to 5 (good).

Testing was performed by means of the experimental set-up provided by the software package PRAAT with a DELL Latitude D600 laptop and AKG K501 headphones. The stimuli were played at comfortable loudness of about 70 dB. German subjects were tested in Stuttgart, and Polish volunteers in Poznan, in a quiet office, each.

To familiarize the subjects with the stimuli and labels, an introduction was presented in which two examples of each syllable category, spoken by one male and one female speaker were played while the category labels were presented. Then followed 3 to 5 training trials without feedback to familiarize the subjects with the procedure.



Identifizieren Sie die Silbe und geben Sie eine Beurteilung ab

Identify the syllable and rate its quality!

Figure 5.15: User-interface for the identification-rating test. Left panel for German subjects, right panel for Polish subjects.

---

[20] Thanks to Paola Escudero for suggesting this rating measure. Goodness rating tasks without an explicit criterium are difficult to interpret, especially for natural stimuli, because listeners may choose any measure of goodness, e.g. personal preferences for a particular voice.

5.6.2 Results

The Polish subjects categorized and rated the German sounds [ʃa] and [sa] in a similar way as the Polish correspondends. Comparing the responses of the Polish listeners between the German and Polish [ʃa] and [sa] syllables, no significant differences in categorization and rating were found (all $p \geq 0.2$).

German subjects categorized Polish [ʃa] as scha, but rated it significantly lower in intelligibility as compared to the corresponding German sound ($p < 0.001$). Thus, German listeners did not identify Polish [ʃa] equally to native [ʃa]. All Polish syllables had been spoken by the same four speakers. One explanation could be that the fact that [ɕa] syllables were spoken by the same voices as [ʃa] lead the German listeners to associate some [ʃa] exemplars with the unfamiliar sounds.

The German [sa] syllables were rated lower than the Polish [sa] syllables by German listeners ($T = 3.6$, $p < 0.001$).

The syllables [ça] and [ɕa] were given low scores in intelligibility by the nonnative listeners. Germans did not uniformly map Polish [ɕa] onto German [ʃa]. Polish [ɕa] was categorized a palatal and as a post-alveolar fricative by German listeners while the intelligibility ratings were significantly lower than for corresponding native sounds (both $p < 0.001$).

The Polish listeners categorized all exemplars of German [ça] as sia, but gave it significantly lower intelligibility scores than for Polish [ɕa] ($T = 26.2$, $p < 0.001$).

In summary, the Polish listeners appear to assimilate German post-alveolar and alveolar fricative syllables with their native correspondents.

Table 5.12: Identification and rating of the syllables [ʃa], [sa], [ɕa], and [ça]. Means and standard deviations across groups are shown. Main perceived categories in bold print.

Stimuli	Polish group		German group	
	Identification	Rating	Identification	Rating
Ger. [ʃa]	**99.8 ± 0.8 % sza**	**4.6 ± 0.6**	**98.8 ± 2.2% scha**	**4.3 ± 1.6**
	0.1 ± 0.6 % sia	3 ± 0	2.2 ± 1.4% cha	3 ± 1.4
	0.2 ± 0.04% sa	3 ± 0	0% sa	-
Pl. [ʃa]	**99.3 ± 0.6% sza**	**4.6 ± 0.9**	**94.3 ± 7.1% scha**	**3.7 ± 1.5**
	1.4 ± 0.4% sia	1.5 ± 0.7	4.4 ± 5.1% cha	2.8 ± 1.6
	0.4 ± 1.3% sa	3.8 ± 1.5	1.4 ± 3.3% sa	3.7 ± 1.5
Ger. [sa]	0.1 ± 0.6% sza	3 ± 0	0.4 ± 0.8% scha	1 ± 0
	1.0 ± 2.8% sia	4.4 ± 0.9	0.4 ± 0.8% cha	1.3 ± 0.4
	98.8 ± 2.8% sa	**4.4 ± 0.9**	**99.4 ± 1.1% sa**	**4.1 ± 0.9**
Pl. [sa]	0.3 ± 0.8% sza	2.5 ± 2.1	0.2 ± 1% scha	1.8 ± 1.2
	1.5 ± 5.2% sia	3.3 ± 1.4	0.5 ± 0.6% cha	3 ± 0
	98.3 ± 5.2% sa	**4.4 ± 0.9**	**99.4 ± 1.1% sa**	**4.3 ± 0.8**
Ger. [ça]	0.3 ± 1.2% sza	2 ± 0	0.3 ± 0.1% scha	3 ± 0.7
	99.5 ± 1.6% sia	**2.3 ± 1.4**	**99.6 ± 1% cha**	**3.7 ± 1.5**
	0.3 ± 1.6% sa	1 ± 0	0.3 ± 1% sa	2 ± 0.7
Pl. [ɕa]	3 ± 5.1% sza	3.2 ± 1.4	**35.4 ± 17.8% scha**	**3.0 ± 1.4**
	96.8 ± 5% sia	**4.2 ± 1.1**	**63.3 ± 18.8% cha**	**3.1 ± 1.4**
	0.1 ± 0.6% sa	5 ± 0	1.6 ± 2.5% sa	2.0 ± 1.6

5.7 MEG experiment

5.7.1 Hypotheses

The memory-related component of the auditory event-related potential, the mismatch negativity was used to probe for an influence of linguistic experience on the processing of the three fricative contrasts. A group of German subjects and a group of Polish subjects were tested for the German/Polish phoneme contrast [ʃa] - [sa], the Polish phoneme contrast [ʃa] - [ɕa], and the German allophone contrast [ça] - [xa].

The contrast between [ʃa] and [sa] was expected to evoke very similar responses in both groups since this contrast is phonemic in both languages. The stimuli that were used to test this contrast had been spoken by German speakers. Since the results of the identification and rating test had suggested that Polish listeners assimilate these German sounds to the corresponding native categories, it was assumed that the German exemplars are acceptable for Polish listeners.

The discrimination performance of German listeners for the Polish contrast between [ʃa] and [ɕa] was significantly below that of native listeners. However, the result of 69.7 ± 13.0% correct responses for the German listeners shows that the Polish sounds are difficult but not completely impossible to distinguish. The results of the identification and rating experiment suggest that German listeners assimilate the unfamiliar Polish fricative [ɕ] partly to native [ʃ] but as poor realizations of this category. Polish [ɕ] was also identified as a poor example of the German palatal fricative. The identification and rating tests indicated that this contrast may be classified as a differentiation of category goodness or even as a two-category assimilation, according to the predictions of the Perceptual Assimilation model by Best (1995; see Chapter 2, 2.3.2). The acoustic difference between [ʃ] and [ɕ], thus, appears to be moderately discernible for German listeners.

Therefore, a mismatch response to the unfamiliar Polish contrast was expected for the German subjects. However, in accordance with findings of previous EEG and MEG studies (Chapter 4), the amplitude was expected to be lower. Additionally the latency of the German group's MMNm could be longer than that of native listeners.

On the basis of behavioral tests (see Chapter 2, 2.1.3.1), it was suggested that fricatives can be perceived in a non-phonetic way, i.e. without reference to sound category representations (Pisoni, 1973; Repp, 1981; Lisker, 2001). The fact that voiceless fricatives, resemble nonspeech noise may facilitate this effect. Lisker's study (2001) suggests that this strategy could be used especially by nonnative listeners. Since German subjects are be able to detect acoustic differences, a non-phonetic auditory analysis of the distinctions may be reflected in later MMN onsets, similar to the findings by Winkler et al. (1999) for vowels and Dehaene-Lambertz et al. (2000) for epenthetic vowels.

In reference to findings on the phonetic versus non-specific auditory distinction (Rinne et al., 1999; Näätänen, 2001; Dehaene-Lambertz et al., 2005; Vouloumanos et al., 2003), a stronger reliance on a general auditory analysis of the difference can lead to a more right-lateralized mismatch field for the German group in response to the unfamiliar Polish contrast.

The German allophones are assumed to be lexically underspecified. If the phonological function of the sound contrasts affect early change detection, a difference between the responses to the German phoneme contrast [ʃa] - [sa] and the German allophone contrast [ça] - [xa] is possible: the amplitude of the MMNm for German subjects was expected to be lower for the allophonic than for the phonemic German contrast.

In comparison, the mismatch response of the Polish group to [ça] vs. [xa] for whom this contrast is acoustically salient and not represented by an underspecified phonological category, was expected to be of higher amplitude than the MMNm for the German participants.

5.7.2 Materials and methods

5.7.2.1 Stimuli

For each syllable category, several naturally spoken exemplars were used. In order to detect sound changes, invariant characteristics of the variable standard stimuli have to be extracted and compared to the variable deviant events. The use of spoken material simulates a more realistic situation for speech perception by facilitating speaker normalization and the distinction of phonetically relevant and irrelevant signal characteristics (Shestakova et al., 2002; Hertrich et al., 2002; see also Eggermont, 2001). Twelve different exemplars per fricative-category, spoken by two male and two female native speakers, were used, as described in section 5.2.

5.7.2.2 Subjects

Fourteen right-handed Polish and 14 right-handed German listeners participated (mean age = 25.7, range = 21 - 44) (see appendix for details on participants). The volunteers had been recruited among students, exchange students, and employees of the University of Stuttgart, the Academy of Arts in Stuttgart, and the University of Tbingen. None of the German subjects had taken part in the previous experiments and in the verification tests of the stimulus material. Three of the Polish subjects were familiar with the stimuli because they had taken part in the evaluation of the material in the MEG setting (see section 5.3.3, this chapter). Most of the German subjects had grown up in the Swabian area and spoke with a slight to moderate Swabian accent. None of them had any knowledge of or experience with the Polish language.

Five of the Polish subjects had started to learn German in school at the age of 12. Six had started at the age of 15 to 18, and three had started to learn German within the last year. One subject had grown up with Polish until the age of 6 and then moved to Germany, growing up bilingually. All Polish participants reported daily usage of their native language. None of the

Polish subjects spoke any dialect.[1] Subjects gave informed consent, confirmed unimpaired hearing sensitivity and no history of neurological disorders. Handedness was assessed by a test based on the criteria by Sattler (1998). All participants showed superiority of their right hand. No differences between Polish and German participants in respect to handedness were found (see appendix). Subjects were naïve to the aim of the experiment. They were paid for their participation.

5.7.2.3 Experimental setup

Stimuli were presented in the common oddball design, as explained in Chapter 3, 5.2.2. A sequence of frequent standard stimuli (85%) and deviant stimuli (15%) was pseudo-randomized such that at least 3 standards preceded each deviant sound. Each subject underwent six sessions with 550 sweeps each. The stimulus-onset-asynchrony varied randomly between 800 to 820 ms.

Stimuli were presented binaurally by means of air-tube headphones at a comfortable loudness (about 70dB SPL). During the entire recording session, subjects watched a silent movie by Charlie Chaplin. They were advised to watch the film attentively and ignore the auditory stimuli.[2]

The three contrasts were presented to half of the subjects in the order: 1. [ça] (standard), [xa] (deviant), 2. Polish [ʃa] (standard), [ɕa] (deviant), 3. [sa] (standard), German [ʃa] (deviant), 4. [xa] (standard), [ça] (deviant), 5. [ɕa] (standard), Polish [ʃa] (deviant), 6. German [ʃa] (standard), [sa] (deviant). The other half of the subjects had the reverse order.

5.7.2.4 MEG measurements and data processing

MEG recordings took place in a sound-attenuated, electromagnetically shielded chamber at the MEG Center of the University of Tübingen. Subjects lay supine during the recordings.

A whole-head device (CTF System Inc., Vancouver, Canada; 151 sensors; sampling rate = 250 Hz, low-pass anti-aliasing filter with a cutoff at 80 Hz, 600 ms sweeps, 148 ms pre-stimulus baseline) recorded the auditory evoked fields.

The entire data analysis was fully scripted and automatized to avoid observer errors. The scripts for artifact reduction, averaging, dipole modelling, and statistical testing were provided by Prof. Klaus Mathiak (Department of Psychiatry and Psychotherapy, RWTH Aachen) (Mathiak, 2002; Mathiak et al., 2002).

---

[1] One of the initially 15 Polish subject was excluded for her Mazurian dialect, spoken in Ostroda, which does not contain the phoneme contrast between the post-alveolar and alveolo-palatal fricatives.

[2] This is a standard testing procedure used at the MEG-center in Tübingen and it has been applied in many other EEG and MEG studies.

In order to remove the consistent offset, the mean across the pre-stimulus baseline was subtracted from the signals of each of the 151 channels.

Equivalent eye dipoles, i.e. a symmetric two dipole model, served to control for artifacts. Trials that were contaminated by eye blinks or movements and other noise exceeding moments above 60 nAm were rejected. The remaining sweeps obtained from each subject were averaged.

Source analysis relied on a simple-sphere head model with two tangential dipole currents.[3] Residual variance was reduced to a minimum. The dipoles were fitted around P50m (50 to 96 ms after stimulus onset) at the point of maximum global field power.
Calculation of the individual lead-field functions relied on dipole position and direction. These two linear combinations of the 151 channels maximize the signal-to-noise ratio of current estimation at the level of the left and right auditory areas (Mathiak et al., 1999; 2000; Hertrich et al. 2000). Auditory activity as revealed by dipole strength could, thus, be measured separately for the left and the right auditory cortex.

To account for the distance to the skull, the dipoles were normalized to their global field power. The pre-stimulus variance was subtracted from the data in order to minimize noise. Quantitative estimates of mismatch fields were obtained within a time window derived from the average mismatch time course.

MEG fields represent the sum of multiple signal and noise sources at separate locations. Consequently, intra-subject variability was considered an additive error term allowing for robust linear statistics. In each subject, the probability of the null-hypothesis H0 was derived by bootstrap procedures that were based on the corresponding regression model and transformed to a z-score. This method has been suggested by Efron and Tibishirani (1993). The z-scores were averaged across the groups to test whether H0 for all subjects could be rejected. The advantage of this non-parametric ANOVA is that identical distributions across subjects are not required (see also Mathiak et al., 2002).

Responses to the deviant stimuli and to the preceeding standard stimuli, that is the last standards, were averaged. To measure the MMNm the responses to the standard stimuli were subtracted from the responses to the deviant stimuli. The mismatch responses for each sound contrast were combined. This means that, for example, the responses to the standard in the contrast [ʃa] - [sa] comprised the responses to standard [ʃa] and to standard [sa]. This prevents a comparison of mere stimulus effects.

Ten analysis windows were determined on the basis of expectations about typical components of the neural auditory event-related potentials and visual inspection of the recorded MEG trajectories. This rigid determination of windows of analysis avoids the possibility of observer bias. In addition, if the recorded difference waves are small, an assessment of individual peaks is not reliable and the averaging over consecutive time-frames is advantageous. The following time domains were chosen: 80 - 100 ms, 100 - 120 ms, 120 - 140 ms, 140 - 170 ms, 170 - 210 ms, 210 - 250 ms, 250 - 290 ms, 290 - 330 ms, 330 - 390 ms, and 390 - 470 ms. The onset of the MMNm can be expected by 100 to 120 ms after the sound change between standard and deviant (Näätänen, 1992). The earlier time frame was included to check for possible early differences. It was decided to use rather late time-windows, because the visual inspection of

---

[3] This rather rough approximation of head anatomy has been shown to work very well in comparison to more realistic models (Hämäläinen and Sarvas, 1989; Lütkenhöner et al., 1990).

the data indicated long lasting differences between standard and deviant responses. The successive windows of analysis allow to approximate the latency of mismatch responses.

Since both the frication and the vowel parts of the stimuli provide distinctive cues, two additional larger integration windows were considered for analysis: 100 - 250 ms and 250 - 390 ms. Approximately, the first window shows the mismatch responses to frication, the second time window includes mismatch responses to changes in transition of the syllables.

Repeated measures ANOVA was performed to test for the effects of native language, sound contrast, and hemisphere side. The dipole strengths in the selected time windows were taken as the dependent parameters of the statistical analysis. All data analysis was performed with Matlab 6 (Mathworks Inc., USA), using a significance level at $p < 0.05$.

5.7.3 Results

5.7.3.1 General observations

Averaged over all conditions and subjects, the location of the left dipoles in relation to the center of the sphere was $x = 6.8 \pm 12.3$ mm, $y = 46.0 \pm 6.9$ mm, $z = 56.7 \pm 8.8$ ms. The mean location of the right dipoles was at $x = 13.9 \pm 8.5$ mm, $y = -47.9 \pm 9.5$ mm, $z = 54.4 \pm 7.7$ mm. Thus, the dipole sources lay approximately in the left and right auditory areas. Figure 5.16 shows the magnetic field distribution.

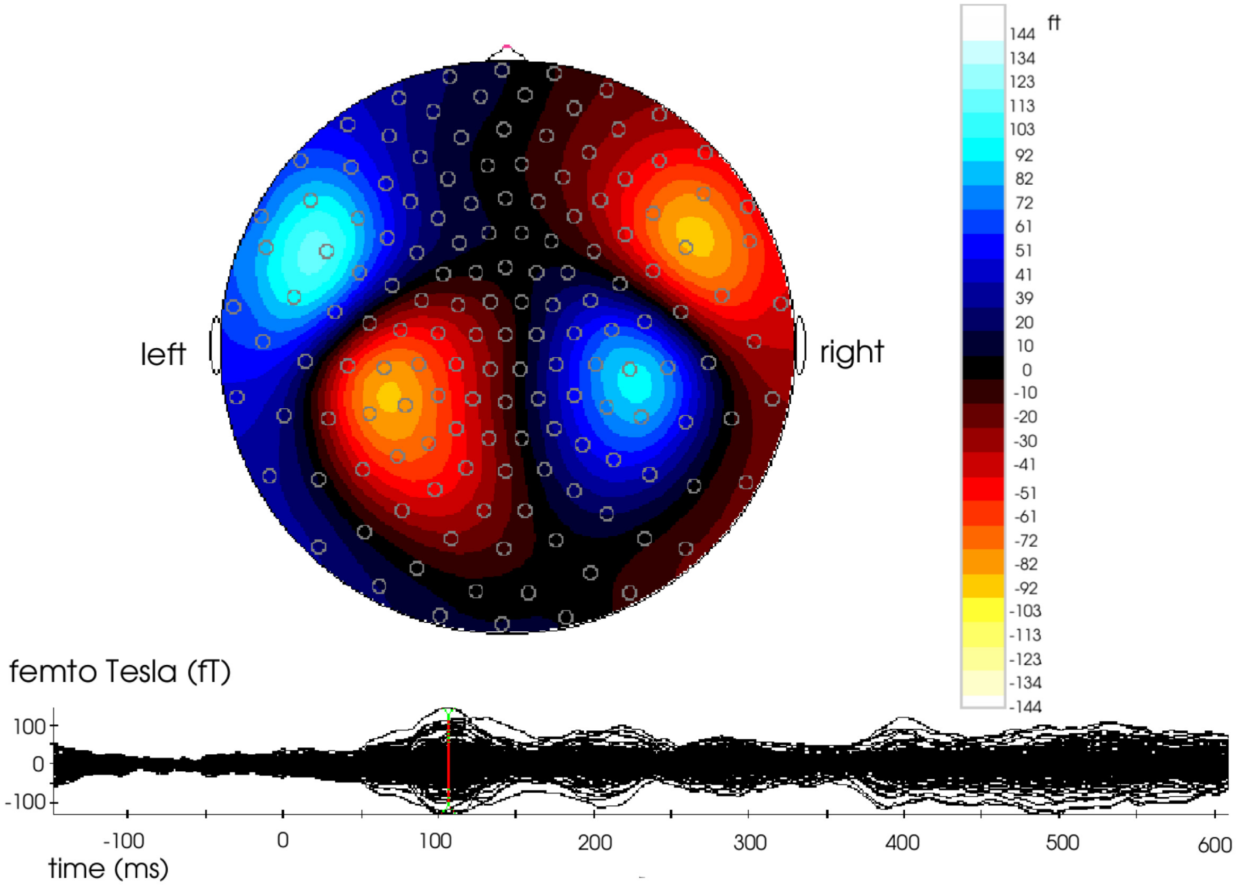


Figure 5.16: Neuromagnetic response of one representative subject. Upper graphic displays the magnetic field distribution averaged across all sweeps after eye-blink rejection at 106 ms after stimulus onset. The inward (red) and the blue (outward) field lines correspond to dipole locations at the left and right auditory areas (scale at the right in ft). Grey circles indicate the position of the SQUID sensors. The lower graph shows the time course of all 151 sensors throughout the whole recording time.

A characteristic waveform was present in all responses with two prominent deflections about 50 to 100 ms after the onset of the fricative and the onset of the vowel and two large peaks at about 100 to 120 after fricative and vowel onsets (compare Figures 5.17, 5.18, and 5.20).

Previous studies that have used stimuli containing a noise part followed by a periodic sound made similar observations. Kaukoranta et al. (1987) tested the response to fricative-vowel syllables such as "sei" in order to investigate the N100 response to fricatives (MMN was not tested). They found that both the fricative and the transition to the vowel elicited a distinct N100 component each.

Mäkelä et al. (1988) used nonspeech noises followed by square waves to further test whether this could be found for nonspeech sounds. Similar to Kaukoranta et al.'s findings they observed two P50m and two N100m components in response to the onset of the noise and the periodic wave. The second N100 can be considered a response to an onset, rather than to the offset, because in Mäkelä et al.'s experiment, the duration of the noise did not affect the amplitude of the second N100. In addition, when they presented a noise stimulus in isolation, the offset response was much lower in amplitude than for the noise-wave combinations.

As reviewed in Chapter 3 for the P50 component (section 3.5.1.1), Hertrich et al. (2000) compared the M50 and M100 components between unvoiced, fricative like segments and voiced vowel-like segments when the sounds were presented separately. They observed that the strength of the M50 was larger and the amplitude of the M100 was lower for unvoiced sounds. For periodic signals a higher M100 peak and a slightly low M50 peak was found. Visual inspection of the present findings shows a similar pattern. The amplitude difference in the study by Hertrich and collegues and dipole localizations by Mäkelä and associates, further, suggest that the responses to noisy and periodic signals are generated by different neural populations. The waveform of the event-related fields that were obtained in this study resemble these previous findings.

Generally, significant mismatch responses were present for all fricative contrasts. Thus, despite the large acoustic variability of the stimulus material, which was introduced by the usage of spoken sounds of male and female speakers, invariant regularities could be detected. However, the great acoustic variation of the stimulus material may partly account for the individual differences of MMNm amplitudes, that emerged for all sound contrasts (see standard deviations from the means in Tables 5.14, 5.16, and 5.18). In some subjects, the responses were weak and a significant difference between responses to standard and deviants was not found in all cases.

Considering the time-span between 100 to 390 ms after stimulus onset, significant MMNms in the left hemisphere were obtained for 9 German and 8 Polish subjects, and in the right hemisphere for 11 German and 8 Polish subjects for the German contrast between [ʃa] and [sa].

For the Polish contrast [ʃa] vs. [ɕa], significant difference waves in the left hemisphere were observed in 3 German and 9 Polish subjects, and in the right hemisphere in 9 German and in 5 Polish participants between 100 to 390 ms.

In that time window, the contrast [ça] vs. [xa] elicited significant mismatch responses in the left hemisphere in 11 German and 7 Polish subjects, and in the right hemisphere for 6 German and 9 Polish participants.

The following description of the results for the three different sound pairs includes data from all recorded subjects.

A further observation common to most responses was that the MMNms lasted for a long period of time. Since syllables were used that differed in the initial portion, the frication and in the second part, the vowel, this finding could be expected. Additionally, in accordance with previous findings on spoken stimuli (see section 4.4.5), the naturalness of the material could have further contributed to a longer duration of the mismatch response.

5.7.3.2 Contrast [ʃa] - [sa]

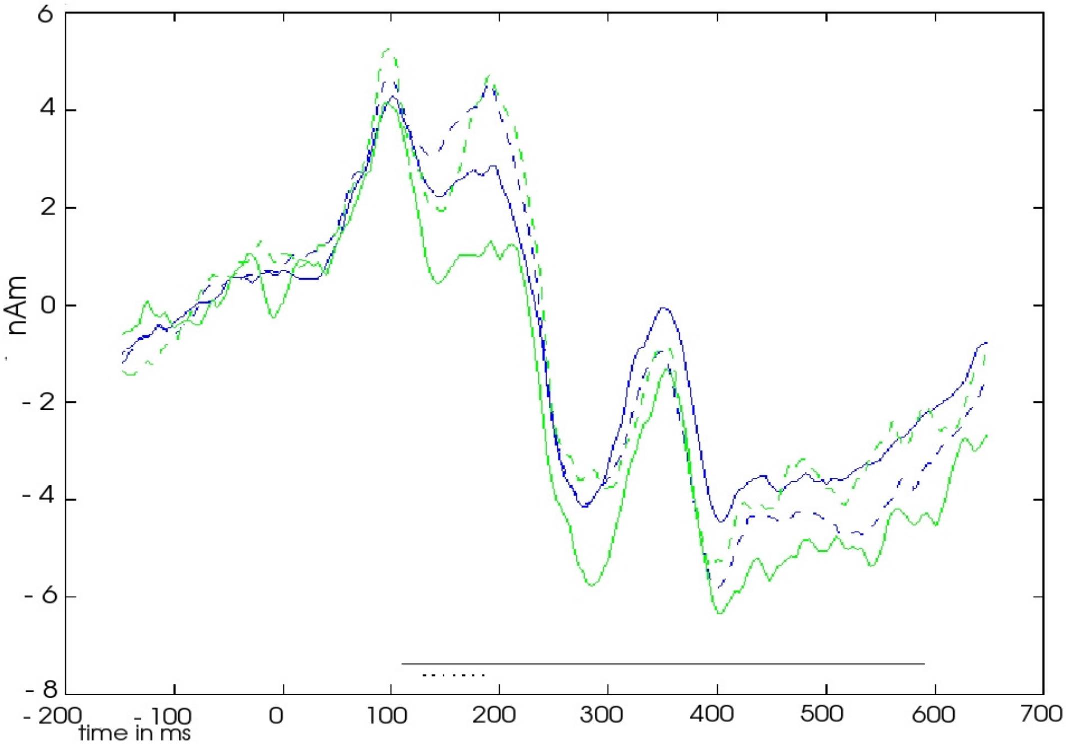
In the earliest two time-windows until 120 ms no significant effect of the sound contrast was observed (see Figure 5.17 and Table 5.14).

The mismatch response first appeared in the time window 120 to 140 ms for the Polish subjects in both hemispheres and for German subjects the first significant MMNm emerged in the right hemisphere dipole. The left hemisphere difference wave just failed to reach significance between 120 to 140 ms for German listeners (z = 1.61).

The mismatch amplitude in both groups raised slightly until about 210 ms. No significant group differences emerged (p was above 0.07 in every time window; 100 - 250 ms: z(LH) = 1.02, z(RH) = 0.3; 250 - 390 ms: z(LH) = 1.2, z(RH) = 0.1). However, the right hemisphere mismatch component was weak in the Polish group. This yielded a significant difference between hemispheres for the Polish subjects between 210 - 490 ms where the mismatch response in the left hemispheric response was stronger (z = 3.6, $p < 0.001$). As it can be seen in Table 5.14, the difference between the responses to standards and deviants remained significant for the Polish listeners throughout the recording, although between 210 and 250 ms, the mismatch response was very weak (compare also Figure 5.17).

Table 5.15 displays the results for the larger integration windows. This broader average points out that the MMNm activity was larger in the left hemisphere for the Polish subjects. Although the weakness of the dipole moments did not yield significant group differences, the Polish subjects showed a tendency for a more sustained left hemispheric mismatch response. Significant differences between the response to standard and deviant lasted until 470 ms. For the German group, a late left-sided difference deflection emerged between 390 to 470 ms.

POLISH SUBJECTS



GERMAN SUBJECTS

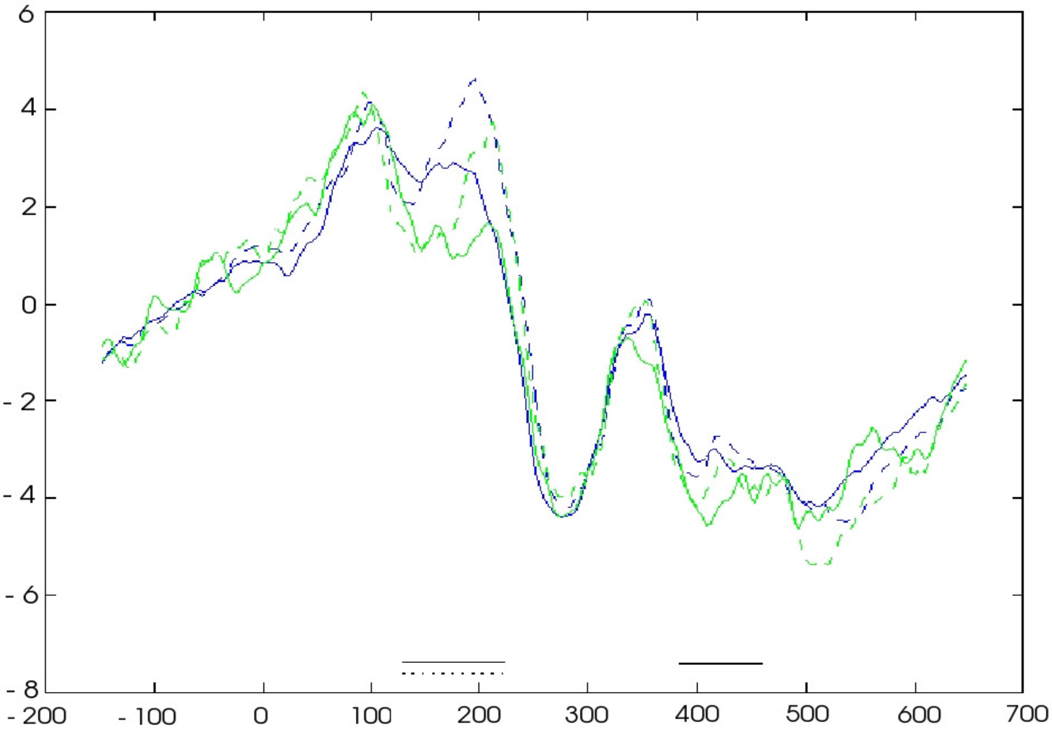


Figure 5.17: Phoneme contrast [ʃa] - [sa]: Group average of dipole moments in response to standards (blue) and deviants (green). Left-hemispheric responses are marked by a solid line; right hemisphere responses are marked by a dashed line. Significance of mismatch at the left hemisphere is indicated by a solid line and at the right hemisphere by a dashed line at the bottom of each graph.

Table 5.14: Amplitudes in nAm with standard deviations and z-scores for the difference in the response to standards and deviants in Polish and German subjects for [ʃa] vs. [sa] for each window of analysis. Significant values are printed bold. For exemplification, a z-value of 1.96 equals a probability of 0.05.

		Polish subjects		*German subjects*	
		dipole moment	*z*	*dipole moment*	*z*
80-100	*LH*	-0.6 ± 3.19	0.51	1.0 ± 4.2	-1.4
	RH	-0.49 ± 2.94	-0.21	1.63 ± 4.93	-1.16
100-120	*LH*	-0.95 ± 3.27	0.57	-0.67 ± 4.34	-0.15
	RH	-1.66 ± 3.0	0.29	0.56 ± 4.89	0.88
120-140	*LH*	-0.5 ± 3.3	**2.53**	-2.84 ± 4.42	1.61
	RH	**-1.21** ± 3.21	**1.97**	**-0.95** ± 5.13	**2.0**
140-170	*LH*	**-2.24** ± 3.27	**3.54**	**-2.64** ± 4.4	**3.61**
	RH	**-1.6** ± 3.05	**2.96**	**-3.27** ± 4.68	**2.95**
170-210	*LH*	**-2.8** ± 3.23	**4.1**	**-4.64** ± 4.27	**4.05**
	RH	-0.35 ± 3.04	0.58	**-5.48** ± 4.64	**3.86**
210-250	*LH*	**-0.87** ± 3.49	**2.58**	0.67 ± 4.55	0.65
	RH	-0.96 ± 3.05	0.23	-2.56 ± 4.76	0.88
250-290	*LH*	**-1.97** ± 3.7	**3.65**	-0.38 ± 4.6	-0.3
	RH	-0.48 ± 3.11	0.03	-1.07 ± 5.02	0.14
290-330	*LH*	**-2.92** ± 3.78	**4.21**	-0.33 ± 4.67	0.44
	RH	0.58 ± 3.28	0.22	0.55 ± 5.21	0.38
330-390	*LH*	**-1.95** ± 3.82	**3.72**	-0.97 ± 4.61	0.7
	RH	-0.04 ± 3.3	0.88	-2.14 ± 5.16	0.43
390-470	*LH*	**-2.42** ± 3.74	**3.76**	**-2.31** ± 4.57	**2.02**
	RH	-0.99 ± 3.21	0.35	-2.73 ± 5.19	1.28

Table 5.15: Significant MMNm amplitudes in nAm ($p < 0.05$) for [ʃa] - [sa] for two integration windows.

	100-250ms		*250-390ms*	
	LH	*RH*	*LH*	*RH*
Polish	-2.0 ± 1.7	n.s.	-2.0 ± -2.7	n.s.
German	-2.9 ± 2.2	-2.9 ± 2.2	n.s.	n.s.

5.7.3.3 Contrast **[ʃa] - [ɕa]**

As in the German phoneme contrast, no significant differences between responses to standards and deviants emerged in the early time-windows around the M100 component (see Table 5.16 and Figure 5.18). Later than for the previously described results for [ʃa] vs. [sa], the difference waves for the Polish syllable pair [ʃa] - [ɕa] first emerged in time window of 210 to 250 ms, in the left hemisphere for the native group and in the right hemisphere for the German subjects. For the Polish group the left hemispheric mismatch activity approached significance ($z = 1.78$) in the preceding time window.

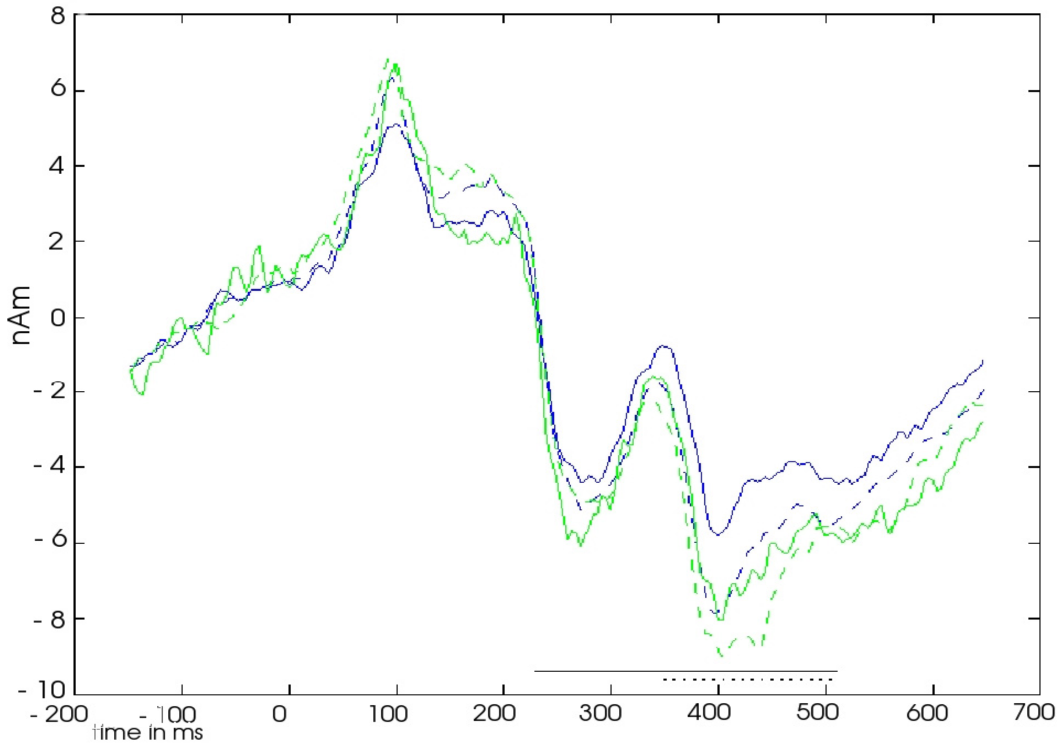
The mismatch response in the left hemisphere was observed in the subsequent time-windows for the Polish group and a late right hemispheric difference wave emerged between 390 to 470 ms. The left-sided MMNm was significantly stronger for the Polish group than for the German group between 210 to 330 ms ($z = 2.6$, $p < 0.05$).

The MMNm for the German group reemerged in the right hemisphere between 290 to 330 ms and was present in all subsequent sections of analysis with increasing amplitude. The German subjects' earlier right hemisphere MMNm yielded a significant hemisphere effect between 140 to 250 ms ($z = 2.6$, $p < 0.05$). Later, no differences between the hemispheres were observed.

The stronger right hemispheric difference response of the German group yielded a significant group difference between 100 and 250 ms ($z = 3.1$, $p < 0.002$) and between 250 and 390 ms ($z = 2.9$, $p < 0.005$). For the Polish group, significantly stronger left MMNm lead to a group difference between 210 and 330 ms ($z = 2.6$, $p < 0.05$).

The differences between the two groups can be clearly seen when the two larger windows of analysis are considered (see Table 5.17). An average of the mismatch responses over 100 to 250 ms and 250 to 390 ms shows the stronger involvement of the right hemisphere in German subjects versus the earlier and stronger left-hemispheric MMNm in Polish subjects.

POLISH SUBJECTS



GERMAN SUBJECTS

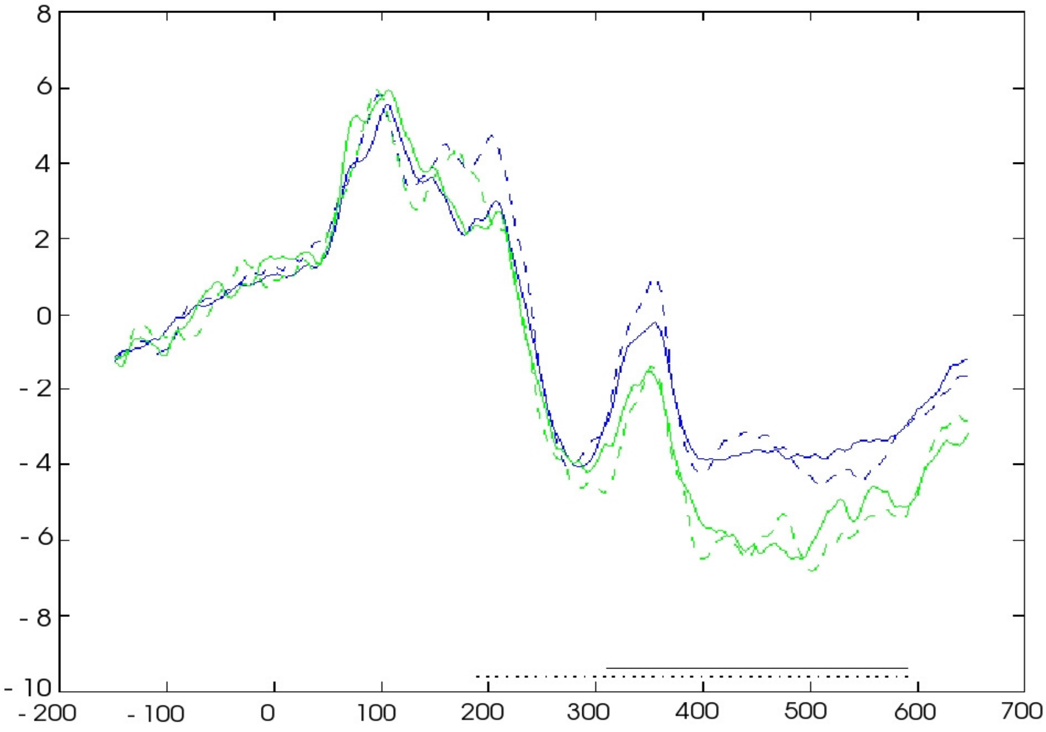


Figure 5.18: Sound contrast [ʃa] - [ɕa]: Dipole moments in response to standards (blue) and deviants (green) were averaged across groups. Left hemisphere responses are marked with solid lines; right hemisphere responses with dashed lines. Significance of mismatch at the left hemisphere is indicated by a solid line, and at the right hemisphere by a dashed line at the bottom of each graph.

Table 5.16: Amplitudes in nAm with standard deviations and z-scores in each window of analysis for the difference in the response to standards and deviants in Polish and German subjects for [ʃa] - [ɕa]. Significant values are in bold print.

		Polish subjects		*German subjects*	
		dipole moment	*z*	*dipole moment*	*z*
80-100	*LH*	-1.34 3.4	0.06	1.24 ± 4.19	-2.61
	RH	0.66 ± 2.89	-1.17	0.19 ± 4.61	-0.73
100-120	*LH*	-0.49 ± 3.65	-0.28	1.35 ± 4.44	-2.12
	RH	0.21 ± 3.05	0.164	2.3 ± 4.64	-0.67
120-140	*LH*	-1.19 ± 3.46	-0.07	0.62 ± 4.62	-1.73
	RH	0.25 ± 3.13	-0.53	-0.63 ± 4.89	0.52
140-170	*LH*	-0.49 ± 3.48	0.95	0.84 ± 4.43	-1.77
	RH	0.4 ± 3.04	-0.92	0.04 ± 4.79	0.85
170-210	*LH*	-3.02 ± 3.46	1.78	0.42 ± 4.27	-1.29
	RH	0.31 ± 2.97	0.22	-0.89 ± 4.68	1.91
210-250	*LH*	**-1.77** ± 3.46	**2.4**	-0.38 ± 4.44	0.13
	RH	-0.02 ± 3.06	0.74	**-2.17** ± 4.89	**3.27**
250-290	*LH*	**-4.8** ± 3.7	**3.58**	1.07 ± 4.82	0.1
	RH	-0.51 ± 3.17	0.82	-3.07 ± 5.39	1.66
290-330	*LH*	**-3.17** ± 3.66	**3.24**	0.37 ± 4.77	0.44
	RH	-1.41 ± 3.27	1.16	**-3.0** ± 5.66	**2.97**
330-390	*LH*	**-3.01** ± 3.8	**3.1**	**-2.52** ± 4.77	**2.58**
	RH	-0.96 ± 3.31	1.55	**-3.91** ± 5.44	**4.6**
390-470	*LH*	**-4.03** ± 3.86	**3.92**	**-4.83** ± 4.75	**3.6**
	RH	**-2.03** ± 3.4	**2.93**	**-5.07** ± 5.46	**5.04**

Table 5.17: Significant MMNm amplitudes in nAm ($p < 0.05$) for [ʃa] - [ɕa] for two integration windows.

	100-250ms		*250-390ms*	
	LH	*RH*	*LH*	*RH*
Polish	-1.6 ± 1.6	n.s.	-3.0 ± 2.7	n.s.
German	n.s.	n.s.	n.s.	-3.5 ± 3.8

The hemisphere differences emerged also when German subjects' responses to the German phoneme contrasts were compared to the responses to the unfamiliar Polish contrast. The difference wave for [ʃa] vs. [sa] was significantly stronger in the left hemisphere between 80 to 250 ms ($z = 3.1$, $p < 0.002$). A stronger right hemisphere mismatch response in the German group emerged for the unfamiliar Polish contrast between 250 to 470 ms ($z = 3.1$, $p < 0.002$). The difference between the dipole moments in the two hemispheres were strong for both subject groups in the time-window between 210 to 250 ms.

Figure 5.19 illustrates the differences in the onset of MMNm for [ʃa] vs. [sa] and [ʃa] vs. [ɕa] between Polish and German participants. The MMNm first emerged for both groups of listeners in the left and right hemisphere for the sound change between [ʃa] vs. [sa] stimuli. For the Polish contrasts between [ʃa] and [ɕa], the first significant mismatch deflection was measured in the left hemisphere, whereas for German listeners it started in the right hemisphere.

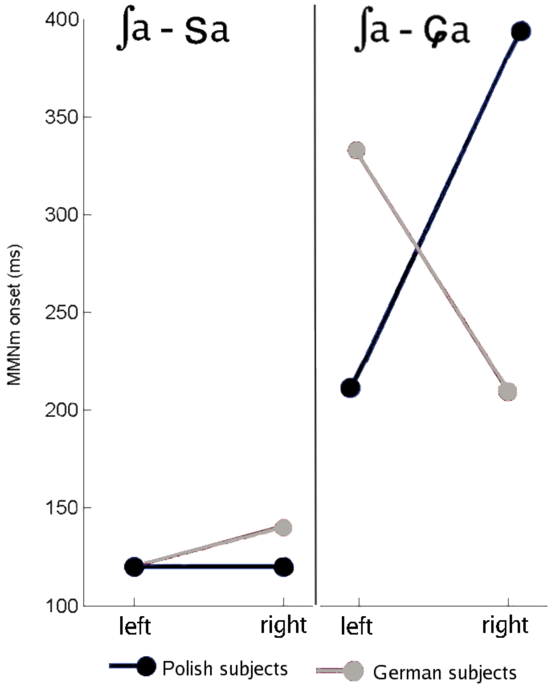


Figure 5.19: Comparison of MMNm onset for German and Polish listeners for the standard-deviant pairs [ʃa] vs. [sa] (left panel) and [ʃa] vs. [ɕa] (right panel).

No significant differences in the activity of the left hemisphere were found for this comparison of the Polish participants' responses (all time-windows $p > 0.7$). In the right hemisphere auditory areas the difference wave was stronger to the German phoneme contrast than to [ʃa] vs. [ɕa] between 100 - 170 ms ($z = 3.1$, $p < 0.002$).

5.7.3.4 Contrast [ça] - [xa*]*

As for the responses to the other fricative pairs, the early time-windows did not show effects of sound contrast (Figure 5.20, Table 5.18).

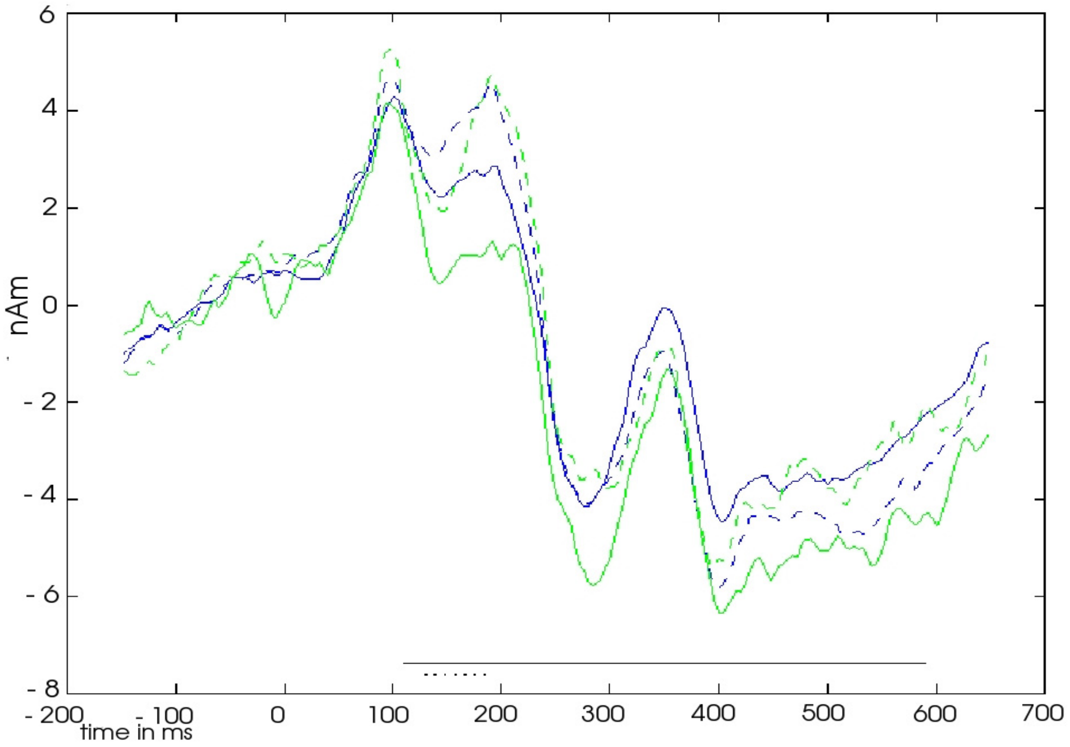
A significant difference between the response to the standards vs. deviants emerged for the German group at 120 ms in the left hemisphere and continued until the time-window of 210 to 250 ms. Between 250 and 290 ms the mismatch response of the German group almost reached significance ($z = 1.92$) and after 290 ms it was significant again. The right hemisphere MMNm was observed for the German group between 140 to 210 ms. Between 390 to 470 ms the response of the right hemisphere just failed to reach significance ($z = 1.91$). The responses of the German group to the sound difference tended to be dominant in the left side but this effect was not significant.

The MMNm for the Polish group started bilaterally in the time-window of 140 to 170 ms and lasted in both hemisphere until 210 ms. The right hemispheric MMNm for the Polish group was significantly stronger in this period ($z = 4.5$, $p < 0.001$). In the earlier time-window of 120 to 140 ms the right hemisphere difference wave of the Polish group approached significance ($z = 1.88$).

Group differences were significant at the left hemispheric dipole between 100 to 250 ms ($z = 3.1$, $p < 0.002$). The left hemispheric mismatch response were significantly stronger for the German group than for the Polish group (compare Table 5.19). MMNms in the right hemisphere were not significantly different between Polish and German listeners. Overall, the MMNm amplitude was larger for German subjects in each hemisphere.

To test whether the native phoneme contrast lead to different responses than the native allophone contrast, German participants' responses to the [ʃa] - [sa] contrast and were compared with [ça] - [xa]. No significant differences emerged ($p < 0.08$ in all time-windows).

POLISH SUBJECTS



GERMAN SUBJECTS

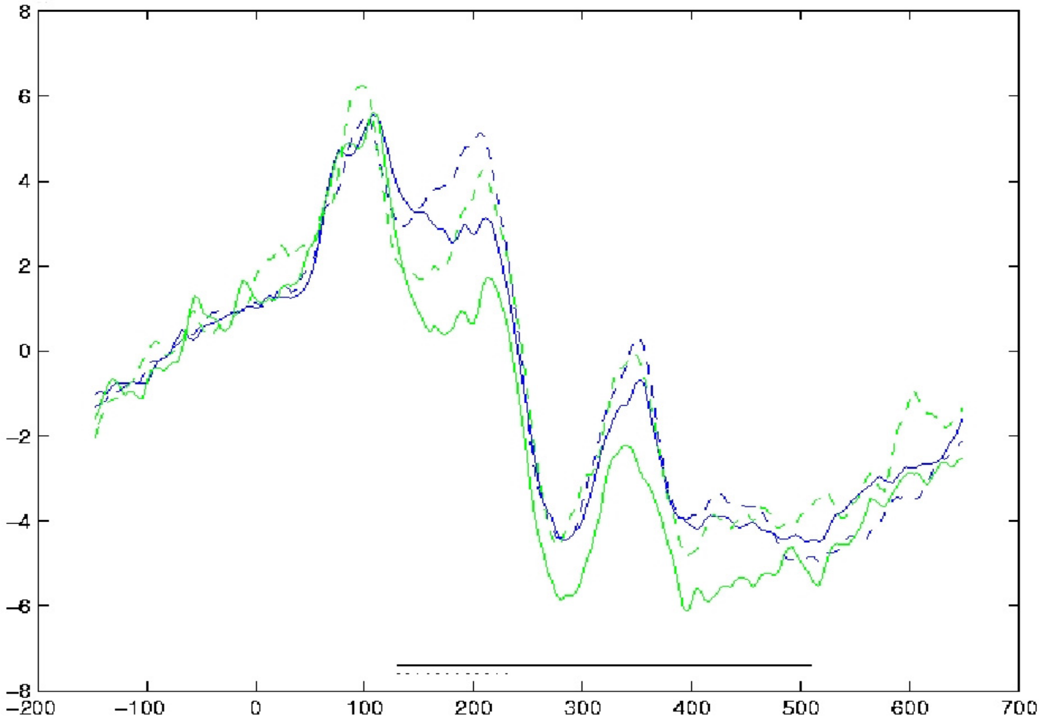


Figure 5.20: Group average of dipole moments in response to standards (blue) and deviants (green) for the non-native, respectively allophone contrast [xa] - [ça]. Left hemispheric responses are marked by solid lines; right hemisphere responses are marked by dashed lines. Significance of mismatch at the left hemisphere is indicated by a solid line, and at the right hemisphere by a dashed line at the bottom of each graph.

Table 5.18: Amplitudes in nAm with standard deviations and z-scores for the difference in the response to standards and deviants in Polish and German subjects for [ça] - [xa]. Significant values in bold print.

		Polish subjects		*German subjects*	
		dipole moment	*z*	*dipole moment*	*z*
80-100	*LH*	1.61 ± 3.3	-2.3	-0.92 ± 4.32	-0.12
	RH	-0.52 ± 2.8	1.2	0.18 ± 4.4	-0.87
100-120	*LH*	1.23 ± 3.45	-1.77	-1.35 ± 4.22	0.44
	RH	-1.11 ± 2.9	1.67	0.34 ± 4.68	-0.67
120-140	*LH*	0.77 ± 3.38	-0.56	**-3.26** ± 4.43	**2.46**
	RH	-1.19 ± 2.83	1.88	-2.5 ± 4.78	1.3
140-170	*LH*	**-1.51** ± 3.26	**2.48**	**-6.18** ± 4.32	**4.76**
	RH	**-2.31** ± 2.8	**3.97**	**-5.28** ± 4.56	**3.5**
170-210	*LH*	**-2.0** ± 3.29	**3.73**	**-6.3** ± 4.26	**5.3**
	RH	**-2.62** ± 2.97	**3.45**	**-5.14** ± 4.48	**3.46**
210-250	*LH*	0.02 ± 3.41	0.63	**-1.2** ± 4.45	**2.84**
	RH	-0.49 ± 3.0	0.21	-2.14 ± 4.83	1.8
250-290	*LH*	0.03 ± 3.53	1.01	-3.14 ± 4.51	1.92
	RH	-0.02 ± 3.18	0.23	-2.2 ± 5.17	0.36
290-330	*LH*	0.49 ± 3.63	0.74	**-1.15** ± 4.87	**2.61**
	RH	1.08 ± 3.20	0.43	-2.98 ± 5.21	0.84
330-390	*LH*	-0.23 ± 3.83	1.11	**-4.23** ± 4.63	**3.22**
	RH	-0.18 ± 3.25	0.71	-3.49 ± 5.25	1.41
390-470	*LH*	0.37 ± 3.83	0.53	**-3.67** ± 4.86	**3.58**
	RH	-0.7 ± 3.25	0.66	-2.11 ± 5.24	1.91

Table 5.19: Significant MMNm amplitudes in nAm ($p < 0.05$) for [ɕa] - [xa] for two integration windows.

	100-250ms		*250-390ms*	
	LH	*RH*	*LH*	*RH*
Polish	-0.4 ± 1.7	-1.9 ± 1.3	n.s.	n.s.
German	-4.3 ± 2.2	-3.3 ± 2.3	-2.6 ± 3.4	n.s.

5.7.4 Discussion

To summarize the results of the MEG experiment on the influence of the native language on auditory processing of fricatives, the main finding is that the mismatch response in the left and right hemispheres differed between native and nonnative listeners.

In response to the Polish phoneme contrast [ʃa] vs. [ɕa], the MMNm of Polish subjects started earlier and was predominant in the left hemisphere. In contrast, for German subjects, the right hemispheric mismatch response was initially stronger (compare Figure 5.19). This is most evident when the two windows of integration are considered: Polish listeners show significant MMNm between 100 and 250 ms and between 250 to 390 ms in the left hemisphere, German subjects show a significant MMNm only in the right hemisphere between 250 to 390 ms (see Table 5.17).

Therefore, the results of the recording of responses to native and nonnative fricative contrasts indicate that the native and nonnative listeners' responses were based on different neural processes.

The phonological contrastiveness - allophone vs. phoneme - did not observably influence the neural mismatch detection and no differences in response to the phoneme and allophone contrast were observed in German subjects. Differently than expected, the MMNm responses of the German listeners were not lower in amplitude than those of the Polish group, but rather significantly higher.

5.7.4.1 Phonetic versus auditory processing

Left and right hemispheric activity has often been associated with a division between phonetic, speech-related processing and general, non-phonetic auditory processing. This idea is evident in the studies of Kayser et al. (1998), Rinne et al. (1999), and Dehaene-Lambertz et al. (2005) that compared the neural response to speech sounds or syllables and non-speech sounds of a similar or of the same complexity (cf. Chapter 4, section 4.2). The training study by Zhang et al. (2001), that was described in Chapter 4, 4.4.9, furthermore, showed a change towards more left-lateralized responses after phonetic training.

The assumption of a left hemisphere dominance for phonetic processing is also verified by a number of imaging studies that localized activity connected with phonetic/phonological tasks at the left superior temporal gyrus (Benson et al., 2001; Vouloumanos et al., 2001; Jacquemot et al., 2003; Dehaene-Lambertz et al., 2005). Further support for left-hemispheric phonetic traces comes from studies that investigated the magnetic MMN for native and nonnative sound changes and checked for hemisphere participation (Näätänen et al., 1997; Alho et al., 2000; Näätänen, 2001; Zhang et al., 2005).

According to these findings, the dominance of the left hemispheric activity for the native Polish listeners as they processed the difference between [ʃa] and [ɕa] could be associated with a phonetic mode of processing. The earlier onset of right hemispheric activity for the

German group for this nonnative contrast could reflect auditory sensory analysis of the differences that did not rely on linguistic representations.

This interpretation of the findings is in accord with the observations and claims by Lisker (2001) and Repp (1981) on the basis of behavioral perception tests with fricative sounds. Lisker noted that nonnative listeners could benefit from a nonphonetic mode of perception for the subtle difference between the frication in the Polish post-alveolar and alveolo-palatal sibilants. Repp had observed that some listeners can divert from a speech-mode of perception for voiceless fricatives in order to detect fine differences. The fact that voiceless fricatives resemble nonspeech noise was mentioned as a possible reason for this effect by Lisker (2001; compare Chapter 2, section 2.1.3.2).

Furthermore, the steady-state nature of the sounds has been associated with a robust formation of a sensory memory trace, which is not given for rapidly changing stimuli (Mirman et al., 2004). Relatedly, the nonnative listeners in the dichotic perception experiment by Best and Avery (1999), for click consonants, that are most likely perceived as nonspeech sounds, displayed a left-ear advantage, which points to a stronger involvement of the right hemisphere (compare Chapter 2, section 2.3.2).

Additionally, a late left-lateralized MMNm was found for German subjects for the unfamiliar Polish contrast between 100 to 130 ms after the full transition from fricative to vowel. If left hemispheric responses indicate an involvement of phonetic mapping, it may be speculated that this indicates a phonetic processing of the transitional differences. According to informal accounts of German participants after the behavioral discrimination test, the Polish [ɕa] sounded to them as if an additional "j" was inserted. Thus, the vowel part of the Polish contrast may have evoked phonetic assimilation to a native contrast. This could have been [a] versus [ja], extracted separately from frication.

Overall, the MMNm to [ʃa] and [ɕa] started later for both subject groups as compared to the contrast [ʃa] - [sa]. The more subtle spectral differences between these fricatives must have required more complex processing which delayed the onset of the MMNm. As it was shown in the acoustic analysis of these sounds in section 5.3.1.2, male speakers' exemplars of the palato-alveolar sound could not be distinguished in spectral moments from the post-alveolar fricative, although the palato-alveolar sound tendentially has a higher COG. The fricational difference was mainly seen as a broader noise distribution for [ɕ] and a peak of energy in the region of F2 for [ʃ].

Smaller acoustic differences between standard and deviant generally lead to longer latencies of the mismatch response (Näätänen, 1992; Schröger and Winkler, 1995; see Chapter 3, section 5.2.2). Apparently, the inherent acoustic saliency affects the time-course of mismatch detection independent of linguistic experience. Similarly, Shafer and collegues (2004) reported that the difficulty of plosive discriminiation was reflected in the brain activity of both native and nonnative subjects.

Considering the Polish subjects' bilateral mismatch fields in response to the contrast [ʃa] - [sa] it could be speculated that the comparatively smaller acoustic difference between [ʃ] and [ɕ] enhanced the need to rely on long-term memory representations for these sound categories. Therefore, Polish subjects' response to [ʃa] vs. [ɕa] was lateralized more strongly than their response to the post-alveolar vs. alveolar fricative contrast. Support for this speculation comes from a PET study by Demonet et al. (1994) which found an increase in activity in the left

STG as native speech sounds became harder to identify. Alternatively or in addition, a reliance on corresponding articulatory gestures may apply here. This could involve the "how" pathway that Scott and Wise (2004) proposed, which runs through the left posterior superior temporal sulcus to premotor areas (see Chapter 4, 5.3.2).

Different from the asymmetric lateralizations for the Polish contrast [ʃa] vs. [ɕa], the sound change between [ʃa] and [sa] elicited an mismatch response in both hemispheres between 120 to 140 ms for both subject groups. For German subjects, the right hemisphere MMNm commenced earlier, although a tendency for a left lateralized response was seen as well. The left hemispheric MMNm in the Polish subjects persisted throughout the measured time-period until 470 ms. For the German listeners a second late left hemispheric MMNm peaked again between 390 to 470 ms.

Since natural stimulus material was used for this MEG experiment, the initially bilateral neural activity could be assigned to a parallel analysis of non-linguistic differences, i.e. speaker characteristics. Relatedly, imaging studies found stronger activations of right hemispheric structures to various paralinguistic aspects. Voice characteristics and nonspeech vocal sounds were found to activate specifically the right superior temporal gyrus (Belin et al., 2002; Belin and Zatorre, 2003) and patients with right hemispheric lesions were reported to have difficulties with fine pitch analysis (Samson and Zatorre, 1994; Johnsrude et al., 2000; for review see Zatorre, 2001). Previous imaging studies that applied spoken items, such as the present study did, have also observed stronger activity in the right areas, specifically in the right anterior superior temporal sulcus (Belin and Zatorre, 2003; Von Kriegstein et al., 2003). The right-hemispheric activity even increased when more than one speaker's voice was presented (Belin and Zatorre, 2003). For a task of voice differentiation in comparison to word differentiation, activity was also found to increase in the right hemispheric temporal area by Von Kriegstein et al. (2003). These results strongly support that the right auditory areas are involved in human voice recognition or distinction.

Generally, however, native fricative contrasts evoked longer lasting MMNms in the left hemisphere. This was seen for the contrast [ʃa] vs. [sa] for both groups of participants, for Polish [ʃa] vs. [ɕa] for the native listeners, and for [ça] vs. [xa] for the German listeners. This finding is in accordance with Näätänen et al. (1997) and with findings from several imaging studies that recorded neural activity in response to single speech sounds and syllables (Benson et al., 2001; Vouloumanos et al., 2001; Dehaene-Lambertz et al., 2005) which showed that the processing of native sound changes is based on memory traces in the left hemispheric auditory regions.

5.7.4.2 Phonetic memory and transitional features

The interpretation of general auditory versus phonetic processing, however, may not be the only explanation of the hemisphere asymmetry that was observed in this MEG experiment. In their outlines of speech sound processing, Hickok and Poeppel (2000) and Scott and Wise (2004) do not adopt a simple dichotomy between general, non-phonetic processing in the right and speech-related processing in the left auditory areas. Along the same lines, Demonet et al. (2005) point out that, although numerous studies find left-lateralized responses to speech sounds:

> Attempts to localize neural responses that are specific to the human voice or to speech components do not seem to point to a single, homogeneous, and clearly left-sided area [...]. (Demonet et al., 2005, p. 64)

Differential hemisphere participation has been associated with the acoustic nature of stimuli and not only with phonetic/phonological content. As elaborated in Chapter 4, 4.5.2., Scott and Wise (2004) emphasize that there may be a correlation between cytoanatomic characteristics of the left hemisphere and the left hemisphere's advantage in temporal processing. This is also the suggestion that Hutsler and Galuske (2003) make. Relatedly, Poeppel's model on temporal integration in the hemispheres (2001, 2003) is based on differences in temporal resolution.

Fricative spectra display rather steady acoustic cues compared to, for example, plosives. But in the transitions to the adjacent vowel, important cues to fricative place can be found (Kudela, 1968; Lisker, 2001). Especially for the contrast between [ʃa] and [ɕa] with its smaller differences in frication, formant transitions appear to provide salient cues for distinction. This became evident in the acoustic analysis (section 5.3.2) as the progression of the second formant was traced. Furthermore, the lack of formant transitions in the isolated fricative sounds lead to a significant decrease in subjects' performance in the behavioral discrimination test (section 5.4). Furthermore, a behavioral study by Nowak (2003) suggested that the formant transitions were highly relevant in the discrimination of the contrast [ʃa] - [ɕa] for Polish listeners.

The relevance of transitional cues appears to vary for different languages. Recall the observations from behavioral tests that were reviewed in Chapter 2. When more subtle sound distinctions have to be carried out, transitional cues appear to play a more important role in perception. Nittrouer (2002) has shown this to be the case for English [f] vs. [θ]. Wagner and Ernestus (2004) suggested, that the relevance of transitional cues for fricative identification and discrimination is enhanced when fine fricative differences have to be distinguished. Furthermore, Hume et al. (1999) observed in a behavioral perception test, that Korean listeners relied more on transitions in plosive-vowel syllables than English listeners. The authors related this finding to the fact that Korean distinguishes one more plosive category along the VOT dimension than English.

Thus, the later onset of the mismatch field in both groups for the Polish contrast [ʃa] vs. [ɕa] contrast, in the present MEG experiment, could reflect that the transitions were incorporated for difference detection. German and Polish listeners, however, could have processed the auditory distinctive cues in frication and transition differently. The integration of both frication and transitional cues may be part of the mental representation of native Polish listeners for the fricatives [ʃ] and [ɕ]. For German subjects, on the other hand, difference detection could initially be carried out by general sensory mismatch detection for the steady cues in the frication and the difference in the vowel transition could have been processed separately from the differences in frication.

A closer analysis of the transition requires temporal resolution, which could rely on the left hemisphere aptitude for transients, although the spectral changes in the fricative-vowel syllables are relatively slow as compared to spectral changes in plosives. The knowledge that cues in the frication and in the transition have to be combined to obtain a phonetic percept, requires learned language-specific representations (Stevens, 2002; 2005; Boersma, 1998). Weighting and combination of several acoustic cues is part of native linguistic knowledge (Repp, 1982; Morrongiello et al., 1984; Underbakke et al., 1988; Escudero, 2002; Nittrouer, 2002; Stevens, 2002). The MEG experiment by Zhang and collegues (2005) on Japanese listeners' processing of English [l] and [r] comes to a similar conclusion (Chapter 4, 4.4.3).

Additional evidence for an involvement of the left hemisphere in cue integration during syllable perception comes from a MEG study on duplex perception by Mathiak and collegues (2001).[4] They tested synthesized [ba], [ga], and [da] syllables that were split into a base consisting of the first and second formant and the transition of the third formant. The syllables [ga] and [da] could only be differentiated when the F3 transition was integrated. They observed that the perceptual integration of the two cues syllable base and F3 transition yielded enhanced left hemisphere activity.

Furthermore, considering that for the contrast [ʃa] vs. [sa] a slightly, albeit non-significantly, stronger left hemispheric MMNm response was found for Polish subjects, it could be speculated that Polish listeners generally focus on transients in fricative contrasts and integrate them because they have learned that these cues can be distinctive for fricative categories.

The effect of formant transitions on the perception of the Polish and German fricatives for Polish and German listeners were tested further in a behavioral test that is detailed in section 5.8.

A final comment on the analysis of responses to frication and vowel transitions should be added. As shown in the study by Mäkelä and collegues (1988) for nonspeech noise signals followed by periodic waves, the neural generators of the ERPs to both types of signals are located at different positions. Other studies on nonspeech sounds by Alain et al. (1999) and by Korzyukov et al. (2003) showed that the source location of MMN differs for the detection of change in pitch and changes in abstract regularities (see also Chapter 3, 3.5.2.2). In addition, a study by Diesch and Luce (1997) suggested that MMNm locations for contrast detection between vowel stimuli and between plosive-vowel syllables differ. Future analysis of auditory processing fricative-vowel syllables could consider to fit dipoles individually for the response to frication and vowel. Cross-linguistic comparisons could investigate, whether the sources of the mismatch responses converge for listeners, who may integrate transitional cues when distinguishing subtle fricative differences, and are rather separate for subjects, who are not used to rely on transitions for fricative processing. Functional imaging with fMRI could furthermore be applied to investigate this question.

5.7.4.3 Phoneme versus allophone contrasts

The second aim of the study was to test whether the functional status of a speech sound contrast influences pre-attentive neural processing. The comparison of the auditory processing of German [ʃ] and [s] to [ç] and [x] in this study was based on the assumption about a differential mental representation of phonemic and allophonic contrasts. The German sounds [ç] and [x] do not occur in phonemic contrast and do not support lexical distinctions, but

---

[4] Duplex perception experiments show that dichotically presented parts of a syllable, i.e. an ambiguous base and a formant transition that differentiates [da] from [ga], are automatically fused into a complete sound percept. At the same time, the isolated transition perceived as a non-speech chirp. Thus, both types of processing, speech (i.e. fusion) and non-speech (isolated transition) are simultaneously present (see e.g. Bentin and Mann, 1990).

alternate according to the phonological context. It was hypothesized that the influence of a common phonological representation of both sounds at an abstract level could reduce the mismatch response to the allophone contrast as compared to the phonemic sound pair for German subjects. This approach was based on to the common, fundamental assumption, that phonemic speech sounds are represented by discrete phonological segments and that any systematic variation is derived from these underlying representations (Chomsky and Halle, 1968; Kenstowicz, 1994; Halle, 2000; cf. Chapter 2, section 2.1).

As a result, however, for both the German phoneme and allophone condition, German subjects displayed clear MMNms, and no significant differences between the responses in the two conditions were observed. Compared to the responses of the Polish participants, the MMNm amplitude of the German subjects was significantly stronger. Thus, the findings of the present MEG experiment do not indicate that the processes underlying MMNm are affected by representations of phonological function.

However, since mismatch negativity is elicited by any discriminable change in an acoustic sequence, the acoustic contrast between the two sounds may have elicited a strong mismatch response so that no influence of the phonological function emerged. The behavioral discrimination test, that is described in section 5.5, furthermore, showed good discriminability for both subject groups. The acoustic analysis indicated clear differences between the palatal and velar fricatives' frication and transitional parts (compare sections 5.3.1.3 and 5.3.2.3). The present results, therefore, could show a ceiling effect of acoustic change detection.

Differential responses between the two subject groups were observed. The responses of the left hemisphere differed already very early around 80 ms after stimulus onset with stronger amplitudes in German listeners, whereas brain responses of Polish subjects were stronger at the right hemisphere. In accordance with suggestions that hemisphere differences indicate phonetic vs. auditory processing, the results could reflect that [ça] - [xa] signals a familiar phonetic change for German listeners, but may evoke stronger general auditory processing in Polish listeners, for whom [x] is familiar and [ç] is new or at best a poor version of native [ ], as suggested by the results of the rating experiment (5.6). An alternative explanation of the stronger mismatch response for the German group is, that the presence of the velar fricative at the onset of a syllable was noticed as highly unusual, since [x] does not appear at the onset of words.

A further point that can be relevant to the present findings is, that the lexical representation of the German allophones is not uncontroversially considered to consist of a unifying or underspecified representation that both sounds are mapped onto. The discussion about the representation of the allophones is motivated by the fact, that allophone alternation is the sensitivity to morpheme boundaries, as it appears in the words "tauchen vs. Tauchen", tauxən - tauçən, (to dive vs. small rope; "-chen" is a diminuitive ending). Morpheme boundaries block the spreading of the feature [back] from the vowel onto a following consonant (Jessen, 1988). This phenomenon is taken up in two alternative views on the lexical representation of the German palatal and velar fricatives.

T. Borowsky (1993) discussed this case in the framework of Lexical Phonology. In his theory of Lexical Phonology, Kiparsky (1982) proposed a distinction between lexical and post-lexical phonological rules. According to the principle of Structure Preservation, lexical rules preserve the underlying lexical structure and cannot create non-distinctive features. Allophones are always derived post-lexically. However, post-lexical rules apply after morphological rules so that the consideration of morpheme boundaries is not possible. Therefore, the creation of the lexically redundant elements [ç] and [x] and their dependence on morpheme boundaries would violate the principle of Structure Preservation. This problem has been considered by Hall (1992) as well, who argued that Structure Preservation is a strong universal tendency, but that it is not absolute and may be violated in individual languages.

However, Borowsky (1993) uses this case to exemplify a different view on the architecture of the mental lexicon, according to which phonological rules can precede morphological rules. Borowsky takes palatal to be the starting point for the process of assimilation: ç -> x /[+back] _ . She claims that this rule applies at the level of word derivation within the lexicon before morphology creates morpheme sequences. The phonological rule that is responsible for the fricative alternation, therefore, resides within the lexicon. This implies that both sounds, ç and x belong to the lexical level, and are not represented by an underspecified representation, as claimed, for example, by Jessen (1988).

A different approach that questions the representation of the two sounds is taken up by Bybee (1999). Bybee does not equate lexical status with distinctivity, but instead emphasizes the influence of usage on the mental lexicon. Bybee's position is that the phonetic form and the input to the lexicon are not separated by a phonemic level of distinctive elements. Therefore, non-distinctive alternations can achieve lexical status. Since the distribution of the German voiceless palatal and velar fricatives is sensitive to morpheme boundaries, Bybee suggested they could well be lexically represented like other "real" phonemes. Bybee's suggestion is consistent with exemplar based models of speech perception that claim that the lexicon contains surface features, for example Pierrehumbert (2001; see Chapter 2, 2.2.6).

The present experiment did not show evidence that the phonological function of the sound contrasts was of influence. Usage-based phonology, as proposed by Bybee or a different organization of the lexicon as suggested by Borowsky could be considered in further investigations of the auditory processing of the non-distinctive German fricatives.

Moreover, future studies on auditory processing of complementary distributed allophones may present the sounds in the appropriate phonetic context. Complementary distributed allophones may be treated differently than phoneme contrasts, if the phonetic context is provided. In the behavioral test on complementary distributed French allophones by Peperkamp et al. (2003), significant differences between the perception of the phonemic and allophonic segments emerged when the sound occurred in the context of nonwords. This was observed even if the allophones occured in the wrong phonological context.

In his model on speech perception, Boersma (1998, 2005, Chapter 2, section 2.2.4) suggests a two-step process, in which the acoustic signal is first mapped onto native sound categories, which, in the second step, are mapped onto lexical representations of words. Boersma does not assume any feedback from the lexicon during the initial step of perception. The fact that German [ç] and [x] do not create lexical contrast may only become important at the second step, in which word forms are accessed. Testing isolated syllables, as it was done in the present experiment, could involve only processes that are responsible for the formation of a sound category percept, an intermediate representation. Thus, information about the lexical relevance of the sound contrast may not be present. Placing the the sounds in the context of words or longer nonword sequences, could enforce lexical mapping and thus, reveal whether the functional status of the sound contrast reflects in the generation of the auditory mismatch deflection.

5.7.4.4 Usage of naturally spoken syllables

The long duration of the mismatch deflections and, in some cases the reemergence of a late mismatch response in the present study could result from the usage of natural stimulus material. Conceivably, this is due to the complex acoustic structure of natural sounds and the amount of information that they carry, which could occupy the automatic detection mechanisms for a longer period of time. Similarly, Hertrich et al. (2002) reported that synthetic stimuli generally evoked a shorter mismatch response than spoken syllables (cf. Chapter 4, section 4.4.5).

Furthermore, large individual differences in the MMNm amplitudes were observed and not all subjects reliably showed a mismatch response to all contrasts. A study by Winkler and collegues (1990) for tone stimuli showed that the MMN amplitude is reduced if the standard stimuli vary, and that it is larger if the standard stimuli are all the same. They interpreted these findings as a sign of the generation of multiple "substandards" in sensory memory, even if the general regularity that all standard event share is represented as well. The amplitude of the MMN, thus, could be inversely related to the number of substandard memory traces. In addition, Titova and Näätänen (2001) have shown, the amplitude of the MMN in response to vowel changes is affected by voice characteristics, in that larger voice differences generate larger MMN amplitudes (see Chapter 3, section 3.5.2.2). The large acoustic variation in the stimuli that were used for the present experiment, thus, could have reduced the mismatch response. Just considering the values for F0 in the present stimuli, shows how large the variation between stimuli was (compare Figure 5.1 in section 5.2.4). Although the individual stimuli were played in random order in the present MEG experiment, difference detection of voice characteristics could have added to the detection of differences between syllable types.

5.8 Impact of formant transitions

5.8.1 Introduction

The results of the MEG-experiment suggest that Polish listeners integrate the transitions from fricative to vowel, especially when distinguishing the post-alveolar and the palato-alveolar fricatives. Thus, their perception of these sounds may be strongly affected when wrong formant transitions are present. This was tested in an identification test in which the fricative - [a] syllables were cross-spliced and fricative and vowel portions were switched. Polish and German subjects' identification performance for these stimuli were tested, to examine whether Polish listeners may be guided even more strongly than German listeners by the transitional cues.

If transitions are generally more imporant for Polish speakers' perception of fricatives, their performance may differ also from that of German listeners when post-alveolar and alveolar fricatives are preceded by the wrong vocalic parts.

The formant transitions were not as distinct between the syllables [ʃa] and [sa] as they were for [ça] and [xa] (compare section 5.3.2). The comparison between the perception of palatal and velar fricatives with conflicting formant cues and the perception of post-alveolar and alveolar fricatives can show, whether the presence of distinctive formant transitions generally affects the identification of fricatives.

5.8.2 Materials and Methods

5.8.2.1 Stimuli

The stimuli that had been applied in the MEG experiment were used to create conflicting cue stimuli. This was done by separating the frication from the vowels just before the onset of periodicity. Frication-parts and vowels were then pasted to other fricatives and vowels. Frication and vowel were always taken from productions of the same speaker.

Additionally, non-conflicting stimuli were created by joining the frication of one syllable type with a vowel of a different speaker of the same syllable type. Male and female utterances were not mixed.

In this description, the origin of frication and vowel will be indicated by symbols given in parentheses, e.g.: F(ç) V(x) stands for: frication taken from [a], vowel taken from [xa]. Twelve exemplars of each of the following stimulus categories were used which are shown in Figures 5.21 and 5.22.

Polish stimuli:

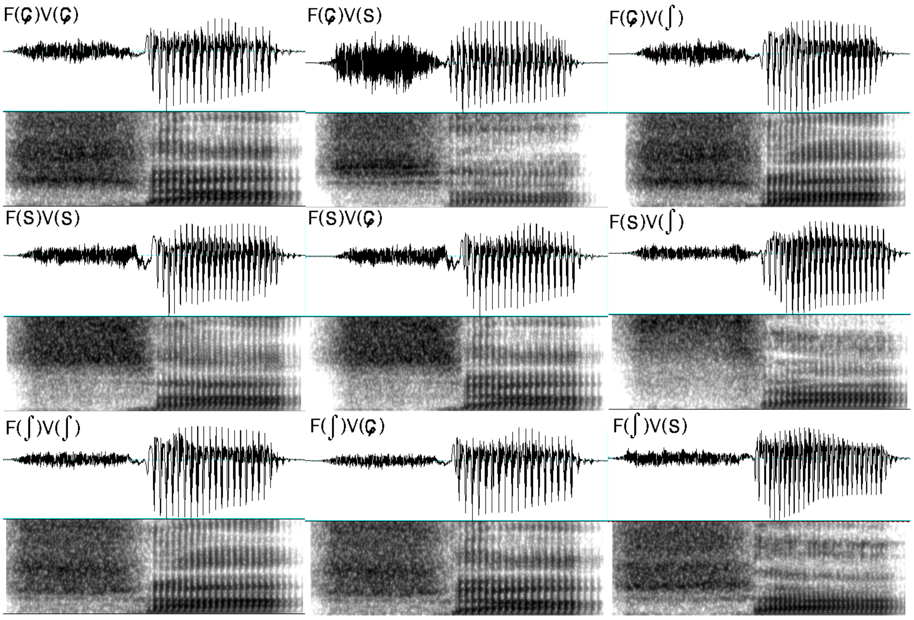


Figure 5.21: Exemplars of Polish stimuli, waveform, and spectrograms (0 - 10 kHz, 300 ms).

German stimuli:

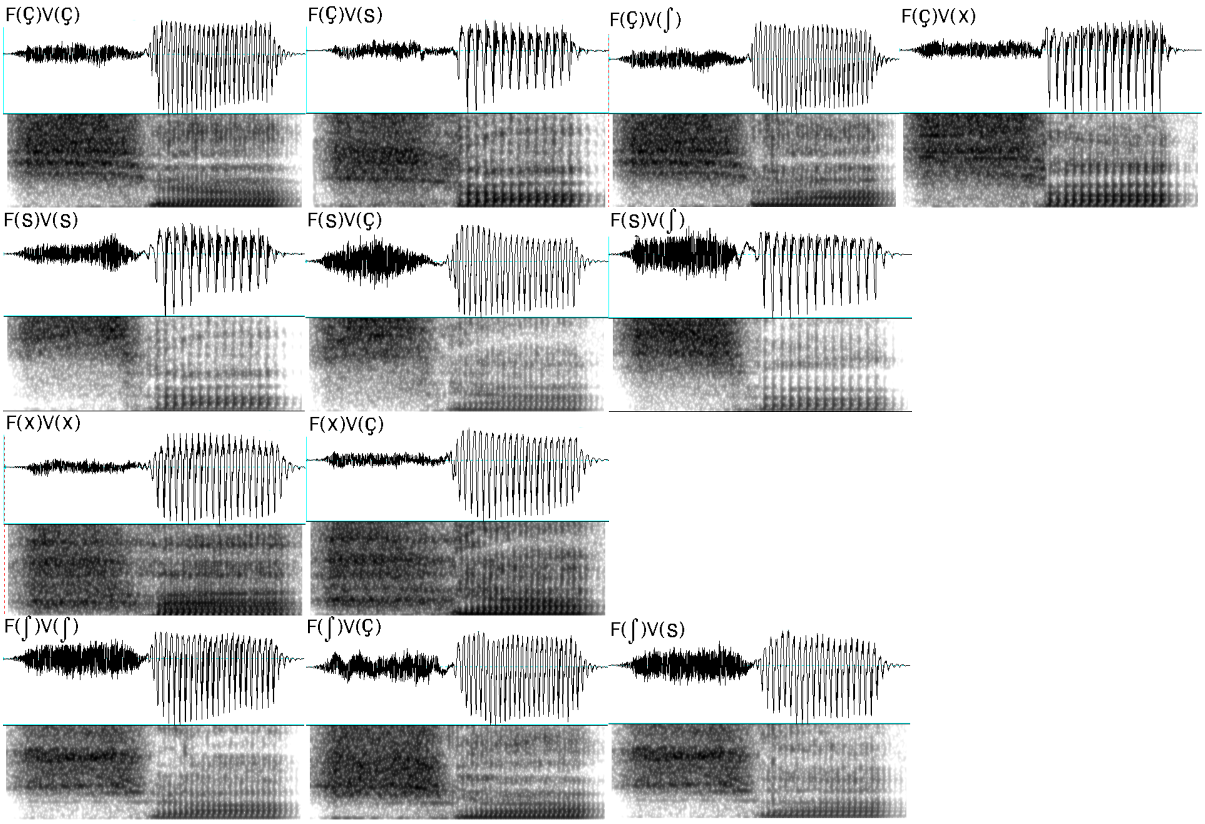


Figure 5.22: Exemplars of German stimuli, waveform, and spectrograms (0 - 10 kHz, the duration is 300 ms).

5.8.2.2 Subjects

20 German subjects (10 females, mean age = 27.7, range = 20 - 40) and 20 Polish subjects (13 females, (mean age = 22.7, range = 20 - 29) participated. The German subjects had already taken part in the discrimination test for the Polish phoneme contrast and were already familiar with the Polish sounds.

Nineteen of the Polish participants had taken part in the identification and goodness rating test (section 5.6). The Polish subjects knew English and French but had no experience with German. Thirteen Polish subjects were students of the University of Poznan and 7 participants were students at the Teacher's college in Katowice. All subjects reported normal hearing.

5.8.2.3 Procedure

PRAAT was used to run the experiment. The Polish and German fricatives were presented separately in two parts, the order of which was balanced between subjects. Between the two parts, participants took a break of 20 minutes.

Every stimulus was repeated three times during the test. This resulted in 324 trials per subject for the Polish part and 432 trials per subject in the German part. For each stimulus category a total of 36 exemplars per subject was presented. The stimuli were randomly presented via headphones one by one at a comfortable loudness. Two seconds after a response was given, the next stimulus started. After every 130 stimuli participants could take a small break. The subjects had to choose the sound category by selecting buttons which were labelled as follows:

Polish stimuli: [ʃa]: "Sa", [ɕa]: "Ca", [sa]: "sa"; German stimuli: [ʃa]: "Sa", [sa]: "sa", [ça]: "ca", [xa]: "xa".

The Polish and the German part of the test started with an introduction to the procedure and the symbols. This included a presentation of four exemplars of each stimulus category, two spoken by a male speaker and two spoken by a female speaker. The labels were presented along with the sounds. The stimuli that were used in this presentation had not been spliced.

5.8.3 Results

Tables 5.20 to 5.25 display the responses of Polish and German subjects in % and show the results of a Kolmogorov-Smirnov test that served for a comparison between the groups.[5]

5.8.3.1 Polish syllables

Not affected by splicing was the identification of the stimuli F(s)V(s). They were almost exclusively identified as "sa" by both groups.

F(ʃ)V(ʃ) and F(ɕ)V(ɕ) lead to more confusions, especially for the German listeners. With a rate of 78.89 ±11.19% German subjects identified F(ʃ)V(ʃ) syllables as "Ca". With this result they lay significantly below the identification rate of these stimuli as [ɕa] by the Polish group.

Furthermore, German subjects confused F(ʃ)V(ʃ) more often with "Ca" than Polish listeners did. Thus, German subjects seemed to be more affected than native listeners by slight acoustic irregularities as they occur when syllables are cross-spliced.

---

[5] The KS test was chosen because it is non-parametric and it can handle small datasets with non-normal distributions.

F(ʃ)V(s) and F(s)V(ʃ) were identified by both groups mainly according to the frication. The groups did not differ significantly in their responses when the post-alveolar and the alveolar fricative parts were switched. Frication may be salient enough and/or the similarity of the transitions for these two sibilants prevents confusions.

A significant difference between the responses of the two groups arose only for the F(ʃ)V(s) stimuli for which German more often than Polish listeners chose "Ca". Conceivably, this results from the lack of experience with [ɕ] and the lack of a detailed knowledge of this sound's acoustics of German listeners. This could have lead them to classify any syllable with irregular acoustics as the unfamiliar Polish sound. Since the German listeners assumed that some of the syllables they heard began with the alveolo-palatal fricative, they may have labelled any strange sounding items as "Ca". In a similar way, a tendency for the German listeners to classify F(ʃ)V(s) syllables as [ɕa] was observed as well.

Despite the lack of the alveolo-palatal transition to [a], Polish listeners were still guided in their responses by the frication part in the F(ɕ)V(ʃ) and F(ɕ)V(s) syllables. In contrast, German subjects labelled these sounds more often as "Sa". An explanation for the German subjects' behavior could be that they followed the transitional cues more "blindly" than the native listeners.

The stimuli that contained [ʃ] frication and [ɕ] vowels yielded similar results for both groups: the majority of the responses was "Ca" (i.e. [ɕa]). Therefore, German listeners seem to take the transitions after the alveolo-palatal fricative as an indicator of this unfamiliar sound. That German listeners partly rely on the transitional cues has also been seen in the discrimination test described in section 5.4. There, isolated fricatives lead to lower correct identification rates than fricative-[a] syllables.

For the F(ʃ)V(ɕ) stimuli, both groups were strongly guided by the formant transitions. The glide in the transition from [ɕ] to the vowel that was clearly seen in the second formant (section 5.3.2.2) is a strong cue for Polish listeners and it is also a salient cue for German listeners.

[ɕ]-transitions can even dominate Polish listeners' responses. This was especially seen for the stimuli F(s)V(ɕ) where differences between the two listener groups were most striking: Polish subjects classified the stimuli F(ɕ)V(ʃ) and F(ɕ)V(s) significantly more according to the frication than German subjects who labelled the stimuli more often as "Sa". It is surprising that here the transitional cues had such a strong effect in comparison to the F(ʃ)V(ɕ) syllables where vowel information did not affect Polish listeners' choices as much. This finding supports the notion that the transition is an important cue for Polish listeners. It appears that this representation comprises constraints that channel the perception of sibilants with a COG higher than that of [ʃ] and a palatal glide transition towards the percept of [ɕ] for Polish listeners.

The results of this behavioral test support the notion that was discussed in section 5.7.4.2, that the earlier and significantly stronger left hemispheric mismatch response to [ʃa] vs. [ɕa] for Polish subjects in the MEG experiment is based on learned integration of the fricational and transitional cues to perceive the alveolo-palatal fricative.

Table 5.20: Polish subjects: Means and standard deviations of responses in %.

	response of Polish subjects		
	sa	*Ca*	*Sa*
F(ɕ)V(ɕ)	0.42 ± 1.36	94.2 ± 9.92	5.56 ± 9.28
F(ɕ)V(s)	2.78 ± 4.32	66.1 ± 20.89	31.1 ± 17.53
F(ɕ)V(ʃ)	3.19 ± 4.61	58.61 ± 18.86	38.32 ± 16.35
F(s)V(s)	99.43 ± 1.7	0.28 ± 1.24	0.28 ± 1.24
F(s)V(ɕ)	24.17 ± 23.43	75.28 ± 23.1	0.56 ± 1.92
F(s)V(ʃ)	98.06 ± 2.55	1.25 ± 1.67	1.56 ± 1.44
F(ʃ)V(ʃ)	2.78 ± 3.35	0.68 ± 1.99	96.67 ± 4.75
F(ʃ)V(ɕ)	0.96 ± 2.73	62.78 ± 4.25	36.39 ± 23.31
F(ʃ)V(s	0.42 ± 1.02	0.82 ± 2.04	98.61 ± 2.28

Table 5.21: German subjects: Means and standard deviations of responses in %.

	response of German subjects		
	sa	*Ca*	*Sa*
F(ɕ)V(ɕ)	0.56 ± 1.45	78.89 ± 1.19	21.25 ± 1.67
F(ɕ)V(s)	1.25 ± 3.42	43.89 ± 0.89	55.83 ± 9.75
F(ɕ)V(ʃ)	2.78 ± 6.06	43.06 ± 0.39	54.43 ± 6.67
F(s)V(s)	97.78 ± 4.64	1.39 ± 4.64	0.83 ± 1.59
F(s)V(ɕ)	76.11 ± 37.78	23.47 ±37.78	0.28 ± 1.24
F(s)V(ʃ)	98.06 ± 5.69	1.25 ± 4.47	0.83 ± 1.83
F(ʃ)V(ʃ)	2.22 ± 4	7.64 ± 13.94	85.83 ± 24.94
F(ʃ)V(ɕ)	0.56 ± 1.45	62.78 ± 0.56	36.67 ± 0.28
F(ʃ)V(s	1.94 ± 4.14	8.06 ± 16.47	90 ± 16.86

Table 5.22: Results of comparison between responses of Polish and German listeners for Polish stimuli (Kolmogorov-Smirnov, all df = 38, D and p values).

	comparison of responses		
	sa	*Ca*	*Sa*
F(ɕ)V(ɕ)	0.05/1	**0.65/<0.000**	**0.65/<0.000**
F(ɕ)V(s)	0.3/0.275	**0.55/0.003**	**0.55/0.003**
F(ɕ)V(ʃ)	0.2/0.771	**0.45/0.023**	**0.45/0.023**
F(s)V(s)	0.25/0.497	0.1/1	0.2/0.771
F(s)V(ɕ)	**0.7/<0.000**	**0.7/<0.000**	0.05/1
F(s)V(ʃ)	0.25/0.497	0.3/0.275	0.05/1
F(ʃ)V(ʃ)	0.2/0.771	**0.6/0.001**	0.4/0.059
F(ʃ)V(ɕ)	0.05/1	0.2/0.771	0.2/0.771
F(ʃ)V(s	0.15/0.965	**0.5/0.008**	**0.55/0.003**

5.8.3.2 German syllables

As for the spliced Polish syllables with frication and vowel parts taken from the same places of articulation, the responses for F(s)V(s) and F(ʃ)V(ʃ) were similar for German and Polish listeners. The identification rate for both groups and both stimulus types was close to 100% correct.

When the frication and vowel parts were switched for German [ʃ] and [s], both groups responded according to the frication.

German F(ç)V(ç) stimuli were misclassified as "xa", indicating the velar fricative, by several participants in both groups. The F(x)V(x) stimuli were identified more correctly by Polish subjects, but labelled more often as "ca" (palatal fricative) by the German subjects. Here, the differences between the groups were significant. The German subjects seem to be more affected by small irregularities at the vowel transition. This could indicates that they have a less precise concept of the allophone segments. Velar [x] is a Polish phoneme, which could have enhanced Polish subjects' ability to identify even spliced F(x)V(x) stimuli.

Germans and Polish showed equally reduced recognition of the fricative when [ɕ] appeared with transitions from other syllable types. For F(ç)V(x) both groups responded mainly according to the frication, for F(x)V(ç) both groups were similarly guided by the transitional cues, but also responded mainly according to the fricative.

However, when the transitions were taken from the palatal fricative syllables, Polish participants were significantly more influenced by the vowel information. Group differences emerged for the stimuli F(s)V(ç) and F(ʃ)V(ç). Polish subjects classified F(s)V(ç) mostly according to the fricative but also to the transition, whereas the responses of German subjects were almost exclusively guided by the fricative. Similar results were obtained for the F(ʃ)V(ç) stimuli: the Polish group was more inclined to label the stimuli according to the vowel cues than the German listeners.

This is reminiscent of their responses to the Polish F(s)V(ɕ) stimuli and suggests that Polish listeners applied similar recognition strategies to German [ç] than to native [ɕ]. This idea is supported by the fact that Polish listeners identified and rated the German palatal fricative as a poor exemplar of [ɕ] in the behavioral experiment described in section 5.6.

Table 5.23 Polish subjects: Means and standard deviations of responses in %.

	response of Polish subjects			
	ca	*sa*	*Sa*	*xa*
F(ç)V(ç)	76.11 ± 22.19	1.81 ± 6.92	1.08 ± 2.04	16.39 ± 20.32
F(ç)V(s)	60.28 ± 23.31	13.06 ± 11.11	6.11 ± 10.14	15.56 ±15.47
F(ç)V(ʃ)	55 ± 24.92	16.81 ± 14.69	11.81 ± 11.72	12.22 ± 10.97
F(ç)V(x)	68.61 ± 23.42	3.89 ± 9.11	6.53 ± 10.14	15.97 ± 17.33
F(s)V(s)	5.56 ± 12.36	94.44 ± 12.36	-	-
F(s)V(ç)	35.28 ± 22.19	61.39 ± 23.03	0.42 ± 1.36	1.39 ± 3.97
F(s)V(ʃ)	4.31 ± 10.44	95.56 ± 10.42	-	-
F(ʃ)V(ʃ)	0.42 ± 1.02	0.42 ± 1.36	99.17 ± 1.59	-
F(ʃ)V(ç)	37.22 ± 23.81	0.56 ± 1.93	53.06 ± 24.94	4.72 ± 6.31
F(ʃ)V(s)	0.56 ± 1.93	0.28 ± 0.86	98.89 ± 2.26	0.14 ± 0.86
F(x)V(x)	1.25 ± 3.42	0.14 ± 0.63	-	98.6 ± 13.44
F(x)V(ç)	19.72 ± 25.97	0.14 ± 0.63	0.28 ± 0.447	68.61 ± 25.69

Table 5.24: German subjects: Means and standard deviations of responses in %.

	response of German subjects			
	ca	*sa*	*Sa*	*xa*
F(ç)V(ç)	84.44±18.06	0.42 ± 1.02	2.78 ± 4.68	12.5 ± 18.94
F(ç)V(s)	62.78 ± 20.94	21.39 ± 12.75	6.11 ± 7.25	9.57 ± 12.72
F(ç)V(ʃ)	58.06 ± 23.42	25 ± 18.06	10.69 ± 15.06	10.96 ± 17.89
F(ç)V(x)	73.61 ± 22.08	-	8.19 ± 9.22	10.56 ± 16
F(s)V(s)	0.28 ± 0.86	99.72 ± 1.02	-	0.14 ± 0.62
F(s)V(ç)	6.39 ± 15.17	91.94 ± 17.08	1.53 ± 4.89	0.28 ± 0.86
F(s)V(ʃ)	0.56 ± 1.44	98.89 ± 2.75	0.14 ± 0.63	0.42 ± 1.85
F(ʃ)V(ʃ)	0.56 ± 0.696	0.14 ± 0.62	98.89 ± 2.07	0.28 ± 0.86
F(ʃ)V(ç)	19.03 ± 23.25	0.14 ± 0.62	74.72 ± 24.39	5.82 ± 12.67
F(ʃ)V(s)	1.53 ± 2.92	0.42 ± 1.02	97.22 ± 3.67	0.82 ± 2.04
F(x)V(x)	14.58 ± 24.75	-	-	85.56 ± 24.75
F(x)V(ç)	31.11 ± 30.56	-	0.28 ± 0.86	68.61 ± 30.28

Table 5.25: Results of comparison between responses of Polish and German listeners for German stimuli (Kolmogorov-Smirnov, all df = 38, D and p values).

	comparison of responses			
	ca	*sa*	*Sa*	*xa*
F(ç)V(ç)	0.3/0.275	0.05/1	0.25/0.497	0.2/0.771
F(ç)V(s)	0.1/1	0.35/0.135	0.35/0.135	0.35/0.135
F(ç)V(ʃ)	0.2/0.771	0.3/0.275	0.25/0.497	**0.45/0.023**
F(ç)V(x)	0.15/0.965	0.3/0.275	0.2/0.771	0.3/0.275
F(s)V(s)	0.2/0.771	0.2/0.771	<0.000/1	0.05/1
F(s)V(ç)	**0.7/<0.000**	**0.7/<0.000**	0.1/1	0.15/0.965
F(s)V(ʃ)	0.15/0.965	0.150/0.965	0.05/1	0.05/1
F(ʃ)V(ʃ)	0.05/1	0.05/1	0.05/1	0.1/1
F(ʃ)V(ç)	**0.45/0.023**	0.05/1	**0.45/0.023**	0.1/1
F(ʃ)V(s)	0.2/0.771	<0.0000/1	0.25/0.497	0.15/0.965
F(x)V(x)	**0.55/0.003**	0.0476/1	<0.000/1	**0.5/0.008**
F(x)V(ç)	0.25/0.497	<0.000/1	0.05/1	0.25/0.497

5.9 Summary of experimental results

Behavioral discrimination tests revealed that discrimination performance of German listeners for the unfamiliar Polish contrast was high. Results of an identification and rating test provided evidence that German listeners' mapping of the Polish fricatives to German categories may be characterized as a category goodness difference assimilation (Best, 1995). Furthermore, it was observed, that both Polish and German listeners could discriminate this contrast significantly better when the transitions to the following vowel were present, than when the fricatives were presented as isolated sounds.

As expected, the sound contrast between [ʃa] and [sa] yielded very similar neural responses in both subject groups. An early MMNm response emerged bilaterally around 120 ms poststimulus and a long duration of the left hemispheric MMNm was seen for Polish subjects, while the MMNm reemerged in the left hemisphere for the German group. The sustained left hemispheric response could result from the phonetic content of the stimulus material, in accordance with previous studies, especially those that applied natural stimulus material (Shestakova et al., 2002; Hertrich et al., 2002).

The unfamiliar Polish contrast between [ʃa] and [ɕa] lead to an earlier onset of the MMNm in the left hemisphere for Polish subjects and to a predominant and earlier MMNm in the German group. This finding suggests that the discrimination of relied on different neural processes in native and nonnative listeners. One explanation for the left hemispheric dominance of the Polish group for the syllable pair [ʃa] - [ɕa] is that Polish listeners rely more strongly on the transitional information than German listeners. Their native category representation of these sounds apparently include the integration of transitional cues.

The subsequent behavioral identification test, that used syllables with conflicting formant transitions produced by cross-splicing, confirmed that Polish listeners' identification of [ʃa] and [ɕa] is strongly affected if the syllables contain wrong transitions. In addition, the behavioral experiment with cross-spliced stimuli suggests that Polish listeners are generally more inclined to use a palatal transition to identify fricatives: the transition from the alveolo-palatal fricative yielded the perception of [ɕa] for Polish listeners, even if the fricative [s] was present in a syllable. Furthermore, Polish subjects' responses to spliced German syllables with the fricatives [s] and [ʃ], that contained the transitions from the palatal fricative were guided significantly stronger by the vowel portions than German listeners' responses.

Thus, while the formant transitions provide also a salient cue for nonnative listeners when distinguishing the Polish fricatives, the integration of frication and transition seem to form part of Polish listeners' representation of these sounds. This cue integration, which has been learned as part of native language phonology, could underly the strong participation of the left hemisphere in prelexical phonetic processing for the Polish subjects.

A difference between the processing of phonemic and allophonic contrasts was not observed in the MEG recording for the German subjects. It is suggested that at the level of processing that is reflected in the MMNm response, the phoneme and allophone segments are not treated differently. However, a critical point for the present study design is, that both phoneme and allophone contrasts presented clear acoustic changes, as it was also shown in the acoustic analysis of spectral and transitional properties of these sounds (section 5.3). Thus, the lack of significant effect of phonological function could be ascribed to a ceiling effect, since the mismatch deflection emerges for any discriminable acoustic contrast. Additionally, behavioral discrimination of the allophones was high for German and Polish subjects (see section 5.5).

However, as the sounds were presented as isolated fricatives in the behavioral discrimination test, the results for the German group were slightly, but significantly lower than that for the Polish group. In the perception test that employed cross-spliced syllables, German subjects confused the [ça] and [xa] syllables, that were cross-spliced but did not contain misleading formant transitions, and showed significantly poorer identification performance than the Polish listeners. In contrast, when the wrong formant transitions were present in these syllables, German listeners were more inclined to categorize the sounds according to the frication. These results point out that the judgement of German listeners of these fricatives is more sensitive to slight interferences at the transition to vowels. Only when the transitions are completely misleading, they can rely on the information in the frication. These findings could suggest a slight effect of the phonological status of these sounds. Another possible aspect that affected German listeners' discrimination and identification performance would be an influence of orthographic knowledge (e.g. Morais and Kolinsky, 1995; see section 5.5.2).

A significant difference in the magnetic mismatch responses of the Polish and the German subjects emerged for the German allophone contrast, in that the German participants' mismatch response was of significantly higher amplitude than that of the Polish group and had a longer duration in the left hemisphere. This difference could be ascribed to the fact that [x] at the onset position of a stimulus is unusual for German listeners. A suggestion for further investigations of the influence of phonological function in pre-attentive auditory processing is to place these sounds in the context of words. This may reduce the potential effect of irritation for [x] at syllable onset and it could enhance the access to a level of representation, at which phonological function becomes relevant.

Furthermore, in terms of the methodological issue about the usage of natural stimulus material, the present investigation showed that the normalization and extraction of distinctive, invariant cues is possible at the early level of speech perception, even if acoustically very diverse material from male and female speakers is used. As discussed in section 5.2.1, the usage of authentic material had been considered to be especially useful for the testing of fricative sounds, because no complete models of the acoustics of these sounds yet exist (see also section 5.3). In confirmation of previous related research, the present study confirms that natural stimuli can be used successfully in MMN - studies. This motivates further research with natural stimuli.

Chapter 6

Conclusions

This last chapter will briefly summarize aims, hypotheses, and main results of the present investigation. Implications for an understanding of auditory speech processing will be discussed and impulses for future research will be given.

The aim of this study was to determine whether native language experience affected the processing of voiceless fricatives in the auditory cortex. For this purpose, three fricative contrasts were tested, one that is used phonemically in both Polish and German, one that is only phonemic in Polish, and one that is an allophonic contrast in German. All tests were carried out with natural stimulus material spoken by several native speakers, that was edited as little as possible. Therefore, listeners had to extract distinctive cues from highly variable acoustic signals.

The underlying neural mechanisms for the processing of these contrasts were investigated by recording the auditory event-related potentials with MEG. The neural signature of contrast detection is the MMN(m) which was used as a probe. In the MEG experiments, that are described in Chapter 5, the sound contrasts were presented in sequences in which one syllable occured regularly as the standard stimulus and the other appeared infrequently. Previous studies, reviewed in Chapter 4, that investigated the effect of linguistic experience on speech sound processing, report a lower amplitude and a longer latency of the MMN in response to nonnative speech sound contrasts and to sound contrasts that do not cross phoneme category boundaries in comparison to native across-category contrasts. The present study investigated whether this would also hold for fricative sounds - a sound class that has not been tested in this way before.

A reduction of the MMN amplitude could be expected for the German listeners' responses to the Polish contrast as compared to the responses of the native listeners. Such differences in amplitude have been subscribed to the presence or absence of long-term memory representations of speech sound categories (Näätänen, 2001). Furthermore, nonnative listeners' reduced ability to differentiate unfamiliar contrasts is assumed to be caused by magnetic or assimilatory effects of native sound categories (Best, 1995; Kuhl and Iverson, 1995; Flege, 1995).

Moreover, if the phonological function of a speech sound is incorporated in its memory representation, an amplitude difference between the response to a phoneme and an allophone contrast was expected. If a phoneme representation which can be understood as an abstract label, arbitrary to the phonetic qualities of the sounds, is accessed during early auditory processing, phonetically distinct would be mapped to this unique representation, not to two distinct phonological representations. As a consequence, the allophone contrast may evoke a lower amplitude of the MMNm than phoneme sounds.

Differences in MMN latency between native and nonnative listeners have been observed less frequently than amplitude differences. Since nonnative listeners can not rely on stored representations that may support a fast detection of sound contrasts, they may rely on a detailed analysis of the acoustic input. This could be expected to lead to later MMN responses (Winkler et al., 1999a). Fricatives can evoke auditory processing as opposed to processing that is based on memorized phonetic representations (Repp, 1981; Lisker, 2001; Chapter 2, 2.1.3.1). Nonnative listeners may resort to this mode of processing, thus showing longer MMN latencies. Furthermore, auditory processing that is not speech-specific has often been associated with enhanced right-hemispheric activity (cf. Chapter 4, 4.2). Thus, differences in lateralization between native and nonnative subjects' responses signal different underlying neural processes for contrast distinction.

The present results are partly meet the expectations based on previous cross-linguistic research on native language representations.

The neural responses of German and Polish listeners to the contrast between [ʃa] vs. [ɕa] confirm that native memory representations modulate pre-attentive auditory processing.

The main result is a differential hemisphere participation between native and nonnative listeners for the Polish contrast [ʃa] vs. [ɕa], which was unfamiliar to the German group. The pre-attentive change detection process to this contrast elicited an earlier and dominant MMNm response in the left hemisphere for native listeners and an earlier and dominant MMNm response in the right hemisphere for German subjects. These differences in lateralization of the auditory response suggest that different neural mechanisms support the processing and discrimination of the Polish contrast [ʃa] vs. [ɕa] for native and nonnative listeners.

However, the phonological status or function of the fricative sounds which are distinct for the German phonemes [ʃa] and [sa] and the German allophones [ça] and [xa] did not influence the response in the auditory areas during pre-attentive mismatch detection. A strong MMNm was found for German subjects which did not differ significantly from their response to the native phoneme contrast. Thus, the phonological function of fricative contrasts, whether they may create a lexical contrast or not, did not influence the measured neural mismatch detection.

The present findings have several implications for future work on speech perception using electrophysiological methods and they contribute to research on the influence of speech sound representations during auditory processing in several respects. These issues will be discussed in the following paragraphs.

First, an important methodological point has to be emphasized. This study applied natural material with high acoustic variation. Numerous previous studies have used synthetic materials only. This enables a controlled modulation of acoustic cues that are of interest (see e.g. Phillips et al., 2000). However, it is questionable to what extend the authentic function of perceptual systems can be studied. (Eggermont, 2001). Furthermore, it may not be safely assumed that listeners really adopt a phonetic mode of processing when presented with synthetic, unnatural sounds (see also Hertrich et al., 2001).

The clear MMNm responses to fricative contrasts that were observed here show that relevant phonetic cues can be extracted despite the great acoustic differences between individual stimuli within one sound category. This confirms previous findings with natural or acoustically diverse materials (Aulanko et al.,1993; Shestakova et al., 2002; Jacobsen et al., 2004). A new feature of the present study is that utterances by male and by female speakers were mixed within one sound category. Conceivably, the ability to extract invariant features in speech sounds subserves the outstanding ability of human listeners to readily comprehend spoken language in a natural environment. As a consequence for future research it can be stated that the usage of natural material is possible in electrophysiological research. Therefore, spoken material can be used to further understand how naturally occuring variation is mastered by the listener.

In future research, knowledge about the processing of variable stimulus material in the auditory cortex that is complex and rich in phonologically nondistinctive acoustic detail, could be extended beyond the observation that normalization and auditory grouping according to native categories takes place. Instead of investigating, how variation is normalized at the level of the MMN, it could be tested how variation is implemented as listeners adapt to idiosyncratic pronounciations. The speech perception model by Stevens (2002, 2005), that was described in Chapter 2, 2.2.3, includes an examination of context information in phonological encoding. Exemplary based models of speech perception (e.g. Pierrehumbert, 2001) assume that listeners utilizes the whole range of acoustic detail that speech has (Chapter 2, 2.2.6). In this view, confidence about phonemic units is partly based on sound characteristics that go beyond primary phonetic cues. Titova and Näätänen (2001; Chapter 3, 3.5.2.2) have already shown that speakers' voice characteristics contribute to MMN responses. In line with the notions of Stevens (2002, 2005) and Pierrehumbert (2001), recent behavioral research indeed suggested that context information, such as speaker characteristics, can be used to identify and discriminate speech sounds (Norris et al., 2003; Eisner and McQueen, 2005). A future approach in electrophysiological research could be based on these findings. By means of natural stimuli, it could be investigated, if speaker characteristics affect the detection of sound changes as signalled by the MMN. It could be tested, for example, whether pre-attentive auditory change detection adapts to foreign accented speech or dialects, so that listeners use idiosynchratic features of the signal for better discrimination.

Turning to the results for the German and Polish phoneme contrasts, it can be stated that they are in accordance with the majority of electrophysiological studies and that the claim of major models of native and nonnative speech perception is supported: linguistic experience does affect early speech sound processing, (cf. Chapter 4, 4.4.10). Thus, notions that early auditory processing of speech is not part of a process based on language-specific mental representations can be refuted.

Previous findings on neural correlates of auditory processing of non-speech sounds show that auditory processing mechanisms are strongly predisposed to be affected by long-term memory representations. Representations for certain regularities in the acoustic input can be created at once. Outside the research on speech perception itself, it has been shown that neural structures that underly MMN generation in the auditory areas adapt to regularities in sounds and to respond to derivations from these regularities (cf. Chapter 3, 3.5.2.2; Winkler et al., 1996; Winkler and Czigler, 1998; Näätänen et al., 2001; Paavilainen et al., 1999; 2001a; Micheyl et al., 2003). Recent observations on single neurons in the primary auditory cortex of animals by

Ulanovsky et al. (2003) point to the neuronal basis for the MMN generation such as the ones observed in the present study and show further, how deeply the mechanism of auditory concept formation is entrenched in the auditory cortex. Auditory concept formation relies on the ability to detect higher-order regularities in acoustically variable signals.

That this is crucial for speech perception and especially for the acquisition of speech sound representations in the first place is emphasized in Kuhl's theory of the language magnet (Kuhl, 2000; 2004). Neural "commitment", as Kuhl puts it, is nothing else than drawing a reference between an input signal and stored regularites that have been detected during language acquisition (cf. Chapter 2, 2.3.4). The multiple speech-related studies that have been reviewed in Chapter 4 that used the MMN as a probe have shown that the native language is active at an early, pre-lexical point in auditory processing. Some researchers even suggest an interaction between language representation and N100 generation (Roberts et al., 2004; Chapter 4, 4.3).

The specific contribution of the present study to this line of research is that the observation that language experience is crucially involved in auditory processing can now safely be extended to fricatives. Finding a native-language influence on fricative processing in the auditory cortex is also insofar interesting, since fricative sounds generally lead to less categorical perception and listeners show perceptual ambiguity for these sounds between a phonetic and an auditory mode of perception (Repp, 1981; Repp, 1984, Chapter 2, 2.1.3).

Fricative sounds cover most places of articulation across languages and are especially dense over the front part of the vocal tract (Ladefoged and Maddieson, 2002; Gordon et al., 2002). Thus, there is rich material for future investigations in language specific auditory processing of this sound class. The large amount of different fricative categories across languages even calls for advanced research in this specific area. Here, mainly coronal fricatives were investigated. The way that fricative sounds that are produced further back are processed and distinguished could be a future project.

Besides the main point of an involvement of the native language, the present findings point to further important issues of language specific phonetic representations.

The first point to be mentioned is a lack of complete assimilation for the nonnative fricative sounds. Both, the SLM by Flege (1995) and the PAM by Best (1995) emphasize the notion that nonnative sounds are automatically assimilated to the nearest native speech sound representation. While the SLM places more weight on perceptual representations, the PAM considers sound production: sounds that require an articulation that lies outside the native articulatory range, most poignantly seen in the case of clicks (Best and Avery, 1999), will not be assimilable to native sound representations. Within the native range of possible articulatory targets, assimilatory forces function. Behaviorally, German listeners can assimilate the nonnative Polish alveolo-palatal fricative to close native neighbors (Chapter 5, 5.6). Still, however, as their rating of Polish [ɕ] revealed, the differences to native sounds are perceivable. The MEG results that show a clear earlier right hemispheric MMNms for the German group suggest that these differences are detected in pre-conscious processes, as well. The assimilatory effects of close native sound categories, as proposed by the leading models of second language perception and learning, are not supported by the present data. Such cases of such partially successful assimilation are understood as "category goodness" classification in PAM's terminology. This means that the native language partly affects perception but that there is also room for the detection of the distance to native representations. In this way, PAM reflects what the findings of this study show: the process of mapping incoming auditory events onto long-term memory representations and dealing with the residual perceived mismatch between the two. The ability to process nonnative sounds in detail and without strong magnetic or assimilatory effects from native representations has to be considered for fricative sounds: native language influence exists but does not eliminate the ability to discriminate and categorize.

Secondly, the differences in lateralization between Polish and German groups that were observed in the MEG experiment, point to a special aspect of linguistic experience in speech perception.

In the case of Polish [ɕ] the present results specifically suggest that native language perception forces a special treatment of the transition after the fricative sound. On the basis of the present MEG experiment and behavioral testing, as well as previous research (Morrongiello et al., 1984; Underbakke et al., 1988; Bohn, 1995; Hume et al., 1999; Escudero, 2002; Nittrouer, 2002; cf. Chapter 2, 2.1.3.3), it is suggested that the differences between native and nonnative listeners may result from differential integration of the acoustic cues frication and transition. The present acoustic analysis of formant transitions, based on the second formant, revealed that this cue is significantly distinct between Polish [ʃ] and [ɕ] (Chapter 5, 5.3.2.2). The results of the behavioral experiment on formant transitions, described in Chapter 5.8, shows that Polish listeners are very sensitive to transitions that point to a palatal place of articulation. Transitions, thus, may be crucial for Polish listeners to identify the alveolar-palatal fricative and to distinguish it from it's acoustically close neighbor, the postalveolar fricative. The observed stronger left-hemispheric MMNm activity for the Polish group in the MEG experiment has been suggestively linked to the presence of acoustic integration according to native language rules.

Boersma's account of Linguistic Perception details constraints that favor or inhibit the perception of distinct cues as a unity (Boersma, 1998, ch. 18). Unification is possible for acoustic cues that appear simultaneously as well as for those cues that occur in a sequence. This is obvious for plosive sounds (Boersma, 1998; Hume et al., 1999) and may also be claimed for fricatives (Nittrouer, 2002). Thus, the present results suggest that Polish listeners' perception grammar likely contains constraints that require the integration of transition and frication for the alveolo-palatal fricatives. This distinguishes German from Polish listeners who do not have to distinguish fricative phonemes that are acoustically as close as [ʃ] and [ɕ] and for whom palatalization is not common.

Stevens' account of speech perception also details the way that the acoustic signal is mapped onto language-specific phonological representations (Stevens, 2002; 2005; Chapter 2, 2.2.3). Stevens assumes a close interplay of the outcome of general auditory perception and the knowledge about relations between acoustics and articulation that are specific for the listener's language. Peaks in low-frequency bands are detected and differentiated. The acoustic changes around these peaks or landmarks are incorporated as well. Together, landmarks and surrounding acoustic cues are used to derive estimates of phonological features. Thus, the points of transitions between segments are of high importance - a claim that has been previously made by Dogil and Braun (1988). Stevens (2002) emphasizes, furthermore, that the integration of all acoustic cues that have been extracted from the signal in order to find a possible phonological correspondence is language-specific. Such integration is mainly taking place at transitions between acoustic landmarks.

It has been proposed that neural phoneme memory traces are located in the left auditory areas (e.g. Näätänen et al., 1997, Chapter 4, 4.2 and 4.5.2). Furthermore, it has been suggested that the left hemisphere is specialized in the detection of temporal changes (e.g. Zatorre and Belin, 2001; Poeppel, 2003). In addition, there is evidence that processing of complex acoustic sounds, such as naturally spoken stimuli (e.g. Shestakova et al., 2002), and the fusion of several cues for speech sound discrimination (Mathiak et al., 2001) enhances left hemispheric activity. Therefore, the results of the present MEG recording for the Polish listeners could be explained by a language-specific process of combination of multiple stimulus features to form a category percept. The left-hemispheric MMNm for Polish listeners may signal that a perceptual integration of several acoustic cues is taking place.

In summary, the present findings emphasize the importance of individual acoustic features for sound category detection. This important part of language-specific memory and perceptual processing should be investigated further in studies on speech sound processing in the auditory cortex. In this way, the findings of this thesis could serve as the starting point for a closer consideration of neurophysiological research on early speech perception and theoretical accounts of language-specific perception. This could lead beyond a detection of native sound representations in auditory processing and towards the question what these representations consist of. The basis for such investigations is provided by detailed accounts on feature extraction and integration, such as those of Stevens and Boersma. Furthermore, integrative processes in speech perception such as suggested by the present work indicate that the focus of research on linguistic mental representations should not be narrowed down to single, isolated sounds, only. Rather, it should be kept in mind that surrounding acoustic cues can be important cues for the identification of sounds.

The last issue that this thesis raises is the question whether the functional significance of sound contrasts within the native language affects early auditory processing or not. In other words, it has been tested how abstract the involved language-specific mental representations are from the physical signal.

The observations on the processing of the allophone contrast for German listeners suggest that these representations are not completely abstract and reduced just to detect phonemic differences. The lack of an effect of the phonological function of a sound contrast on the MMNm which was observed for German [ça] and [xa] does not support accounts of speech perception which propose a direct access to underlying phonological representations. Furthermore, this result does not support the notion of a very strong effect of abstract underlying representations on processing of speech sounds or single syllables prior to word access. At least at the level of mismatch negativity, native representations of sound categories are involved, but the lexical relevance of sound contrasts does not seem to be effective. Therefore, a very close interrelation between native language phonological inventory and speech perception in the case of second language learning as proposed by Brown (2000; Chapter 2, 2.3.3) is not born out by the present results. Contrary to Brown's claims, speech perception is not completely dominated only be those elements that contribute to differences in meanings.

Instead, according to the present data, allophones in single syllables are treated in the same way as phonemic sounds. The auditory system abstracts from the acoustic variation in order to distinguish sound contrasts. However, this step of abstraction does not entail abstract phonological knowledge - this may only become active at a later step.

Electrophysiological mismatch negativity studies on native and nonnative speech sound processing often use the term "phonemic" to describe the sound categories that they apply. However, as this study shows, one must be careful in making such a claim. The term "phoneme" refers to an abstract, functional entity in the listener's mind. As seen in the present study, MMN(m) generation is based on abstraction over several acoustic events, but it is still not completely removed from the acoustic events.

Still, results of previous behavioral perception studies by Peperkamp et al. (2003) and Hume and Johnson (2003) (Chapter 2, 2.1.2) suggest that phonological function affects perception at some point during the process of perception. Thus, it is still be a worthwhile endeavor to prosecute the issue of phonological function of speech sounds and to determine when it begins to play a role in auditory processing of speech. The negative results of the present investigation on phonological function in early auditory processing may serve as a basis for future studies that attempt to answer the question at what point in the process abstract underlying functional representations are accessed.

One option is to apply larger units of speech than those that were used here. The present results imply that the phonological function of a speech sound does not affect auditory processing at the level of isolated syllables. It is possible that the fricatives have to be placed in the context of morphemes in order to activate an allophone-specific processing. According to Borowsky (1993, cf. Chapter 5, 5.7.4.3), phonological rules may apply before morphological rules and this may be given for the German velar and palatal fricatives.

Furthermore, Boersma's model of Linguistic Perception (Boersma, 1998, 2005; Chapter 2, 2.2.4) assumes two constraint-based steps of speech sound processing. The similarity of the auditory percept and the activated lexical entry is evaluated in the second step in which context begins to play a crucial role. According to Boersma, contextual information can be essential for chosing the correct lexical candidate and it may even override faithfulness to the percept gained by initial sound perception. Thus, this influence may be studied best by presenting the sounds in the phonetic context that determines their place of articulation. Phonological processes that may map both [ç] and [x] onto a unified mental representation may only be effective when, for example, [x] follows back vowels and [ç] is preceded by front vowels.

To conclude, the present study shows that the native language is involved in early auditory processing of fricative sounds. However, it is not the case that highly abstract mental representations dominate. Instead the influence of one's native language is already given at a low level of auditory processing, when feature of the acoustic signals are still being processed. Therefore, the present findings support accounts of speech perception that take native language influence seriously at a level that is close to the acoustic form. Models of second language perception which have to account for the obvious difficulties that nonnative listeners have with foreign speech sounds, strongly focus on such low-level processes. The Speech Learning Model (SLM) by Flege (e.g. Flege, 1995; Chapter 2, 2.3.1) claims, that only the perceived phonetic quality of a speech sound and the similarity of its phonetic quality to native sound categories determines a sound's classification. Mental representations of speech sounds in terms of SLM are understood as being closely linked to their phonetic properties. SLM does not even state the influence of a level of representations which are abstract from a sound's physical appearance. Still, the experience with sound of the native language determines, how acoustic signals are categorized and differentiated.

In a similar way, the Perceptual Assimilation Model (PAM) by Best (1995) strongly links the phonetic reality of speech sounds to mental representations and from that derives hypotheses for nonnative language perception (Chapter 2, 2.3.2). Like SLM, PAM claims that listeners will always try to assimilate any speech sound to native representations.

Boersma's model of Linguistic Perception details the shaping of peripheral perceptual processes by native language categories in more specifically as the PAM. Coming from a functional view of speech perception, Boersma (1998, 2005) emphasizes that the speech perception system is equipped with general abilities that speech perception can be based on. As mentioned above, this assumption is directly borne out by electrophysiological research on speech and non-speech perception: the detection of higher order regularities is a key feature of the auditory system, at least at the level of the auditory cortex (Paavilainen et al., 1999; 2001a; Ulanovsky et al., 2003). Crucially, Boersma sees native language representations as constraints and constraint orderings that are actively guiding speech perception from the first input until the access of abstract mental representations. In other words, perception itself is a process that is based on initially non-linguistic abilities and tendencies of the auditory system and includes rules created through native language experience.

Zusammenfassung

Neuronale Korrelate der Wahrnehmung von nativen und fremdsprachlichen Frikativen

Das Thema dieser Studie ist der Einfluss linguistischer Repräsentationen auf die auditorische Verarbeitung von Sprachlauten. Das Ziel war es, diese Vorgänge speziell für Frikativlaute zu testen. Mit Hilfe der Magnetenzephalographie (MEG) wurden die Reaktionen der auditorischen Hirnregionen auf muttersprachliche phonemische, allophonische und fremdsprachliche Frikativkontraste untersucht.

Wahrnehmung muttersprachlicher und fremdsprachlicher Laute

Die Phonologie der Erstsprache übt einen entscheidenden Einfluss auf die Wahrnehmung von Sprachlauten aus. Generell wird in Perzeptionstests festgestellt, dass Lautkontraste, die in der Erstsprache eine phonemische Funktion einnehmen, klar unterschieden werden können, während allophonische Variationen oft schwer zu diskriminieren sind (Strange, 1995). Dies wurde in besonderem Maße für Plosive festgestellt, die sich in Stimmhaftigkeit und Artikulationsort unterscheiden (e.g. Abramson und Lisker, 1970; Werker und Tees, 1984; Strange, 1995). Bekannt sind auch die Schwierigkeiten japanischer Hörer mit dem [l] vs. [r] Kontrast (e.g. Tsushima et al., 1994; Takagi, 2002; Iverson et al., 2003). Bei Vokalen zeigt sich der Einfluss der Erstsprache bei Vokalen im sogenannten "Magneteffekt" des Kategorieprototypen (Kuhl, 1991; Kuhl und Iverson, 1995).

Ebenfalls wurde festgestellt, dass auch die phonologische Funktion eines Lautkontrastes einen Einfluss auf die Wahrnehmung haben kann: allophonische oder nur partiell phonemisch kontrastive Laute werden weniger distinkt wahrgenommen, als Phonemkontraste (Hume und Johnson, 2003; Peperkamp et al., 2003).

Für Frikative wurde insbesondere gezeigt, dass Hörer zwischen einer rein auditiven, nicht von sprachlichen Kategorien bestimmten Wahrnehmung, und einer phonetischen Wahrnehmung wechseln können (Repp, 1981; Lisker, 2001). Die Integration von Transitionen zu angrenzenden Vokalen spielt bei der Wahrnehmung vor allem von feinen Unterschieden zwischen Frikativen eine wichtige Rolle und diese Integration findet nur bei einer phonetischen Wahrnehmung statt (Repp, 1982; Nittrouer, 2002; Nowak, 2003; Wagner und Ernestus, 2004).

Die Tatsache, dass die phonologische Funktion, aber auch akustische Charakteristika die Sprachwahr-nehmung beeinflussen, gab den Anlass, in dieser Studie eine Lautklasse zu untersuchen, die bislang in der neurophysiologischen Forschung zu sprachlichen Repräsentationen nicht berücksichtigt wurde.

Die Sprachperzeption, also die Dekodierung des akustischen Signals in phonologische Einheiten, die zur Worterkennung verwendet werden können, wird in verschiedenen theoretischen Ansätzen erklärt. Gerade neuere Ansätze, wie die von Boersma (1998), Stevens (2002) oder Lahiri und Reetz (2002), versuchen genau den Prozess von der Aufnahme des physikalischen Sprachreizes bis zum Erhalt von phonetischen Einheiten oder Merkmalen nachzuzeichnen. Der Unterschied dieser Theorien besteht hauptsächlich in der Auffassung über die Stufen des Perzeptionsprozesses. Gemeinsam ist ihnen die Annahme, dass sprachspezifische Repräsentationen und Regeln bereits sehr früh in den Wahrnehmungsprozess eingreifen. Besonders die Modelle von Boersma und Stevens gehen von einer sprachspezifischen Integration und Gewichtung von einzelnen akustischen Merkmalen aus. Die genannten Theorien teilen ebenfalls eine Auffassung über lexikalische Repräsentationen als relativ abstrakte, diskrete und rein distinktive Einheiten. Dagegen steht die Exemplartheorie (Pierrehumbert, 2001), die eine Speicherung von genauen phonetischen Details im Lexikon annimmt.

Theorien zur Wahrnehmung von fremdsprachlichen Lauten, wie zum Beispiel das Second Language Learning Modell von Flege (1995) und das Perceptual Assimilation Modell von Best (1995), sagen gute Erfolge für Wahrnehmung und Erwerb von Lauten voraus, die dem Hörer relativ fremd sind, d.h. die keine Entsprechung in der Erstsprache haben. Laute dagegen, die denen einer Erstsprache ähnlich sind, werden häufig nicht korrekt wahrgenommen, denn sie werden muttersprachlichen Kategorien zugeordnet. Die Theorie des Native Language Neural Commitments von Kuhl (2004) postuliert dazu eine Bildung von neuronalen Netzwerken, die optimal auf muttersprachliche, prototypische Varianten einer Lautkategorie reagieren, jedoch schwächer auf fremdsprachliche bzw. solche, die nicht einzuordnen sind.

Untersuchungen der Sprachwahrnehmung mit Elektroenzephalographie (EEG) und Magnetenzephalographie (MEG)

Mit Hilfe der EEG und MEG Technologie kann die Verarbeitung von Sprachlauten mit hoher zeitlicher Genauigkeit verfolgt werden. Um sprachspezifische Einflüsse auf die Lautunterscheidung zu messen, wird den Probanden eine Abfolge von häufigen, regelmäßigen Standardstimuli (ca. 85%) präsentiert, die von seltenen, abweichenden Reizen unterbrochen wird. Die Komponenten P50 und N100 des dabei entstehenden ereignis-korrelierten Potentials zeigen eher sprachunspezifische Reaktionen. N100 jedoch wird von akustischen Eigenschaften des Stimulus, z.B. der Grundfrequenz oder der Stimmeinsatzzeit beeinflusst.

Subtrahiert man die gemessenen Reaktionen auf den Standardreiz von denen auf den Devianten, erhält man eine weitere Komponente, die Mismatch Negativität (MMN) bzw. deren magnetische Entsprechung MMNm. Die MMN(m) tritt mit einer Latenz von ca. 100 bis

250 ms auf, wenn ein Unterschied zwischen Standard- und Deviantreizen festgestellt wird, selbst wenn die Hörer mit einer anderen Aufgabe beschäftigt sind und nicht auf die akustischen Reize achten. Die MMN(m) wird als ein Korrelat zeitlich früher, weitestgehend aufmerksamkeitsunabhängiger, automatischer, akustischer Informationsverarbeitung verstanden (Näätänen, 1992; Näätänen und Winkler, 1999; Näätänen, 2001; Näätänen et al., 2005). Sie entspricht einer Verarbeitungsstufe, in der Regelmäßigkeiten zwischen auditiven Reizen erkannt werden. Da dies auch abstrakte Regularitäten umfasst, prägte sich der Begriff der "Sensorischen Intelligenz" im auditorischen Kortex, die durch die MMN angezeigt wird (Näätänen et al., 2001). Aber auch erlernte akustische Zusammenhänge oder erworbene Sensitivität für feine akustische Unterschiede haben einen Einfluss auf diese Komponente (Näätänen, 1992).

Daher können sich die neuronalen Reaktionen auf erst- und fremdsprachliche Laute in der Amplitude und Latenz der MMN(m) unterscheiden (Näätänen et al., 2001). Dies wird besonders deutlich in Studien, die Plosivkategorien untersucht haben. Kontraste, die sich innerhalb einer muttersprachlichen Kategorie bewegen, lösen häufig keine oder nur sehr geringe signifikante Mismatchreaktionen aus (Dehaene-Lambertz, 1997; Phillips et al., 2000; Rivera-Gaxiola et al., 2000).

Studien, die sich mit der Verarbeitung des [l] - [r] Kontrastes von japanischen Hörern beschäftigt haben, kommen zu vergleichbaren Ergebnissen (Buchwald et al., 1994; Zhang et al., 2001; 2005). Darüberhinaus, deuten die Beobachtungen der Studie von Zhang und Mitarbeiter (2005) darauf hin, dass, im Gegensatz zu muttersprachlichen Hörern, bei der vorbewussten Verarbeitung der japanischen Probanden keine Merkmalsintegration vollzogen wird.

Für die Differenzierung von nicht-muttersprachlichen Vokalen wurden ebenfalls Unterschiede in der MMN Amplitude im Gegensatz zu muttersprachlichen Kontrasten gezeigt. Hierbei scheinen die nicht-muttersprachlichen Hörer jedoch noch eine Analyse der akustischen Vokalunterschiede zu vollziehen, die zu einer längeren Latenzzeit der Mismatchreaktion führt (Näätänen et al., 1997; Winkler et al., 1999).

Zusätzlich zum Einfluss von phonetischen Lautkategorien wurden Anzeichen für einen Einbezug abstrakter phonologischer Regeln und Kategorien festgestellt (Dehaene-Lambertz et al., 2000; Mitterer und Blomert, 2003; Eulitz und Lahiri, 2004).

Im Vergleich zur Methode der EEG, können durch die Anwendung der MEG-Technologie Reaktionen des primären und sekundären auditiven Kortex in der Sylvischen Furche gemessen werden, während die Reaktionen anderer Hirnstrukturen kaum mit aufgenommen werden. Darüber hinaus können die Potentiale der linken und rechten Hirnhälfte gesondert bestimmt werden. Lateralitätseffekte werden häufig sprachlicher Verarbeitung zugeschrieben.

Mehrere Studien beobachteten eine linkshemisphärische Dominanz der Mismatchreaktion auf muttersprachliche Lautkontraste, gerade bei der Verarbeitung von natürlichem Stimulusmaterial (Näätänen et al., 1997; Hertrich et al., 2002; Shestakova et al., 2002; Zhang et al., 2005). Bildgebende Studien zur Verarbeitung von Einzellauten und Silben deuten ebenfalls auf eine stärkere Beteiligung der linken Hemisphäre an phonetisch-phonologischen Vorgängen hin, bevor noch semantische und syntaktische Verarbeitungsschritte einsetzen (Benson et al., 2001; Vouloumanos et al., 2001; Jacquemot et al., 2003).

Dabei erklärt jedoch die Zuordnung von linkshemisphärischer Aktivität zu phonetischer Verarbeitung die Gründe für Lateralitätseffekte nicht ausreichend. Unterschiede in der zeitlichen und spektralen Auflösung scheinen zu Hemisphärendifferenzen beizutragen (Zatorre et al., 2002; Hickok und Poeppel, 2004; Scott and Wise, 2004). Eine MEG-Studie von Mathiak et al. (2001) zur Duplexwahrnehmung (e.g. Liberman, 1996), zeigte auch, dass eine Fusion akustischer Reize zur phonetischen Unterscheidung linkshemisphärische Aktivität im Bereich des Brodmann Areals 22 hervorruft.

Ergebnisse der vorliegenden Studie

Im vorliegenden MEG Experiment wurden deutschen und polnischen Probanden ein Frikativkontrast präsentiert, der in beiden Sprachen phonemische Funktion hat: [ʃa] vs. [sa], ein weiterer Kontrast, der im Polnischen phonemisch eingesetzt wird, für Deutsche jedoch unbekannt ist: [ʃa] vs. [ɕa], und der Kontrast zwischen [ça] und [xa], der im Deutschen nicht zur Unterscheidung von Wörtern eingesetzt werden kann und somit eine allophonische Alternation darstellt.

Ein Wahrnehmungstest zur Diskrimination und Assimilation der polnischen Laute bei deutschen Hörern zeigte, dass [ɕ] als schlechte Realisation der muttersprachlichen Kategorien [ʃ] oder [ç] wahrgenommen wird und die Fähigkeit zur Unterscheidung des polnischen Kontrastes zwischen [ʃ] und [ɕ] zwar nicht der von Muttersprachlern gleicht, mit einem Durchschnitt von ca. 70% korrekter Ergebnisse jedoch relativ hoch ist. Die Wahrnehmung des fremden polnischen Frikativs kann in Anlehnung an das Perceptual Assimilation Modell von Best (1995) als eine "category goodness difference assimilation" beschrieben werden. Darüberhinaus zeigte sich, dass sowohl polnische als auch deutsche Probanden signifikant schlechtere Ergebnisse erzielten, wenn die polnischen Frikative nicht als Silben, sondern isoliert dargeboten wurden.

Die bewusste Unterscheidung der Allophone [ç] und [x] war beiden Hörergruppen in einem Perzeptionstest problemlos möglich. Die Rate der korrekten Antworten deutscher Probanden lag nur knapp unter denen der polnischen Hörer.

Die MMNm wurde für die Verarbeitung dieser drei Kontraste ermittelt (vgl. Tab. A). Die Reaktion der Probandengruppen auf den Kontrast, der in beiden Sprachen dieselbe phonologische Funktion erfüllt, unterschied sich kaum voneinander. Es zeigten sich zunächst bilaterale MMNms, die im späteren Verlauf zu linkslateralisierten MMNm wurden. Eine signifikante Differenz zwischen deutschen und polnischen Hörern wurde für den polnischen Phonemkontrast festgestellt. Hier war die MMNm Reaktion bei polnischen Hörern in der linken auditorischen Region dominant, während für deutsche eine rechtshemisphärisch stärkere MMNm gemessen wurde.

Tabelle A: Signifikante MMNm Amplituden in nAm ($p < 0.05$) für die Frikativkontraste [ʃa] - [sa], [ʃa] - [ɕa] und [ça] - [xa]. MMNm Amplituden wurden gemittelt über die Zeitbereiche 100 - 250 ms und 250 - 390 ms nach Beginn des Stimulus.

		100-250ms		*250-390ms*	
		LH	*RH*	*LH*	*RH*
[ʃa] - [sa]	*Pl.*	-2.0 ± 1.7	n.s.	-2.0 ± -2.7	n.s.
	Dt.	-2.9 ± 2.2	-2.9 ± 2.2	n.s.	n.s.
[ʃa] - [ɕa]	*Pl.*	-1.6 ± 1.6	n.s.	-3.0 ± 2.7	n.s.
	Dt.	n.s.	n.s.	n.s.	-3.5 ± 3.8
[ça] - [xa]	*Pl.*	-0.4 ± 1.7	-1.9 ± 1.3	n.s.	n.s.
	Dt.	-4.3 ± 2.2	-3.3 ± 2.3	-2.6 ± 3.4	n.s.

Hinweise auf einen Einbezug der phonologischen Funktion der Kontraste konnten nicht gefunden werden. Die Reaktionen der deutschen Teilnehmer auf den phonemischen und allophonischen Kontrast unterschieden sich nicht signifikant voneinander. Allerdings zeigte sich ein Unterschied zwischen den Reaktionen der polnischen und deutschen Teilnehmer. Die MMNm Amplitude war signifikant höher für die deutsche Gruppe und hielt in der linken Hemisphäre länger an. Dies entsprach nicht den Erwartungen, denn es wurde zunächst vermutet, dass die Mismatch Amplitude der deutschen Probanden unter denen der polnischen Teilnehmer liegen könnte. Der Grund für diese Annahme lag darin, da diese Laute nur im Deutschen als Allophone einer zugrundeliegenden mentalen Repräsentation angesehen werden können, für polnische Hörer jedoch einen deutlichen akustisch-phonetischen Kontrast darstellen. Mögliche Gründe für diese Differenzen könnten darin liegen, dass Deutschen beide Laute bekannt sind, während das palatale [ç] für polnische Hörer nur eine relativ schlechte Realisation den nativen [ɕ] darstellt, wie es im Wahrnehmungstest zur Assimilation der Laute gezeigt wurde. Eine alternative Erklärung ist, dass das Auftreten von [x] am Anfang einer Lautsequenz für deutsche Hörer ungewöhnlich und daher irritierend ist, so dass sie stärker auf diesen Kontrast reagieren.

Die Unterschiede in der MMNm zwischen polnischen und deutschen Hörern in Reaktion auf den polnischen und für Deutsche fremden Frikativkontrast, zeigen, dass der Diskriminierung der Laute unterschiedliche neuronale Prozessen zugrunde liegen. Die Rolle der Transitionen für die Wahrnehmung der polnischen Frikative wurde in einem Perzeptionstest überprüft, und es zeigte sich, das polnische Hörer die Transition nach [ɕ] insbesondere gerade zur Identifikation des alveolo-palatalen Frikativs einsetzen. Auch in Bezug auf Ergebnisse von Verhaltenstests (Nittrouer, 2002; Wagner und Ernestus, 2003) und MEG Studien (Mathiak et al., 2001; Zhang et al., 2005) wird vermutet, dass die stärkere und frühere linkshemisphärische Aktivität für [ʃa] vs. [ɕa] bei polnischen Probanden auf eine Integration von Friktion und Transition zurückzuführen ist, die Teil der nativen phonetischen Repräsentation ist (vgl. auch Repp, 1982; Bohn, 1995). Im Zusammenhang damit, schlagen Modelle zur Sprachwahrnehmung von Stevens (2002, 2005) und Boersma (1998) vor, dass die Bildung von phonetischen Perzepten auch in der Anwendung sprachspezifischer Regeln zur Merkmalsgewichtung und -kombination liegt. Die vorliegenden Ergebnisse zur Verarbeitung des polnischen Silbenkontrastes können daher die Grundlage für weiterführende Studien bilden, die sich mit der Frage nach der Integration und Gewichtung verschiedener akustischer Merkmale im Sprachwahrnehmungsprozess beschäftigen.

Zusammenfassend kann man festhalten, dass in dieser Studie erstmals ein Einfluss der Erstsprache auf die auditorische Verarbeitung von Frikativlauten festgestellt wurde. Dieser Einfluss wurde auf sprachspezifische Prozesse bei der Integration phonetischer Merkmale zurückgeführt.

Die Ergebnisse lassen darüberhinaus vermuten, dass es sich bei der Verarbeitungsstufe, wie sie von der MMN(m) reflektiert wird, kein Unterschied zwischen der Verarbeitung von phonemischen und allophonischen, phonetisch distinkten Lautkontrasten handelt (vgl. das Modell der Sprachwahrnehmung von Boersma, 2005; Kapitel 2, 2.4). Jedoch sollte diese Annahme in zukünftigen Untersuchungen überprüft werden, in denen die Allophone in phonetische Umgebungen eingebettet werden.

Zusätzlich bestätigt und erweitert diese Studie Befunde zur Extraktion invarianter Merkmale in akustisch stark variierenden sprachlichen Stimuli (e.g. Aulanko et al., 1993; Shestakova et al., 2002; Jacobsen et al., 2004). Trotz der Verwendung von natürlichen Silben, die von mehreren weiblichen und männlichen Sprechern produziert wurden, zeigten sich MMNm Reaktionen auf phonetische Kategorien. Dies unterstützt die Verwendung von authentischem Stimulusmaterial und kann auch weiterführende Untersuchungen zur sensorischen Repräsentation von verschiedenen Lautexemplaren motivieren (siehe auch Titova und Näätänen, 2001), z.B. zum Einbezug von Kontextinformationen in der Sprachwahrnehmung, wie es von Stevens (2002, 2005) vorgeschlagen wird.

Aufbau der Arbeit

Das **Kapitel 2** führt anhand von Ergebnissen der behavioralen Forschung zur Wahrnehmung von erst- und fremdsprachlichen Lauten in die Thematik ein. Dabei gilt den Ergebnissen zur Frikativwahrnehmung besonderes Interesse. Die Untersuchung mit der zeitgenauen Methode der Magnetenzephalographie erlaubt es, sehr frühe Verarbeitungsschritte zu überprüfen. Daher stellt das Kapitel 2 theoretische Erläuterungen der Sprachwahrnehmung vor, die diese frühen Verarbeitungsschritte im Sprachwahrnehmungsprozess modellieren und erklären.

Im **Kapitel 3** wird erklärt, wie auditorische neuronale Potentialen zustande kommen und wie sie gemessen werden. Diese Einführung erlaubt das Verständnis der EEG und MEG Experimente, auf die in den folgenden Kapiteln Bezug genommen wird. Desweiteren werden die Komponenten des auditorischen ereignis-korrelierten Signals erklärt, die bei der Untersuchung der Sprachwahrnehmung bzw. dem Einfluss sprachspezifischer Prozesse eine Rolle spielen.

Damit ist die Grundlage für die Besprechung von Experimenten geschaffen, die mit der hier durchgeführten Untersuchung in Beziehung stehen. Diese werden im **Kapitel 4** vorgestellt, wobei eine Einteilung vorgenommen wird in Studien, die sprachliche von nicht-sprachlicher Verarbeitung abgrenzen, die phonetische Kategorisierungen an muttersprachlichen und fremdsprachlichen Hörern testen, und drittens Studien, die den Einfluss von erlernten Lautkategorien nachweisen. Ein weiterer Teil behandelt Ergebnisse und Theorien aus der bildgebenden Hirnforschung mit den Methoden fMRI und PET. Die Arbeiten, die in Kapitel 4 vorgestellt werden, bilden die Grundlage für die Konzeption des eigenen MEG Experiments, für die Hypothesenbildung und für die Interpretation der Ergebnisse.

Dieses wird in **Kapitel 5** beschrieben und diskutiert. Zur Vorbereitung auf das MEG Experiment wurden Wahrnehmungstests der getesteten Frikativkontraste durchgeführt. Im Anschluss an das MEG Experiment erfolgte ein weiterer Wahrnehmungstest zur Rolle der Formanttransitionen für die Identifikation von von Frikativen. In diesem Kapitel werden somit die Fragestellungen dieser Arbeit getestet und ausgewertet.

Im **Kapitel 6** werden die Resultate dieser Untersuchung zusammengefasst. Dabei wird erläutert, in welcher Weise die Befunde der vorliegenden Studie richtungsweisend für zukünftige Forschungsvorhaben werden können.

Appendix

Subjects in MEG recording

Subject ID	L1	Age	Gender	Experience with Polish/German
T1	G	40	f	-
T2	G	30	m	-
T3	G	28	f	-
T4	G	24	m	-
T5	G	30	f	-
T6	G	27	m	-
T7	G	23	f	-
T8	G	43	m	-
T9	G	27	m	-
S0	G	22	m	-
S1	G	24	f	-
S2	G	23	f	-
S3	G	26	m	-
R6	G	44	m	-
S3	P	23	m	-
S5	P	43	m	1 year in Germany
S6	P	20	m	since age 29 in Germany
S7	P	24	m	since age 16 in school, 1/2 year in Germany
S8	P	21	f	bilingual P/G since age 7, before P
S9	P	22	f	since age 12 in school, 1/2 year in Germany
R0	P	23	m	since age 12 in school, 1/2 year in Germany
R1	P	22	m	since age 17 in school, 1/2 year in Germany
R2	P	23	f	since age 18 in school, 1/2 year in Germany
R5	P	24	f	since age 12 sporadically, 2 months in Germany
R7	P	21	f	since age 14 in school, 1/2 year in Germany
R8	P	23	f	since age 15 in school, 1/2 year in Germany
R9	P	23	f	1 year in Germany
Q0	P	23	m	1/2 year in Germany

Handedness test

German:

Bitte geben Sie an, welche Hand Sie für gewöhnlich bei den untenstehenden Aktivitäten benutzen. Tragen Sie 2 Kreuze ein, wenn Sie diese Aktivität IMMER mit dieser Hand ausführen. Wenn Sie sich nicht sicher sind, welche Hand Sie bei einer Aktivität benutzen oder wenn Sie die Tätigkeit einmal mit der einen, ein andresmal mit der anderen Hand ausführen, tragen Sie bitte EIN Kreuz in BEIDE Spalten ein.

AKTIVITÄT	*LINKE HAND*	*RECHTE HAND*
Schreiben		
Zeichnen		
Werfen		
Mit einer Schere schneiden		
Zahnbürste halten		
Messer halten (ohne Gabel)		
Löffel halten		
Obere Hand am Besen beim Fegen		
Streichholz anzünden		
Deckel einer Schachtel öffnen		
Bevorzugter Fuß beim Kicken		
Bevorzugtes Auge, wenn nur eines benützt wird		

English:

Please indicate the hand you preferably use for the following activities. Enter two crosses in one field if this is the hand you always use for an activity. If you are not sure which hand you use or if you switch hands from time to time for an activity, enter one cross in EACH of the fields.

ACTIVITY	*LEFT HAND*	*RIGHT HAND*
Writing		
Drawing		
Throwing		
Cutting with scissors		
Holding a toothbrush		
Holding a knife (without fork)		
Holding a spoon		
Holding a broom, upper hand		
Lighting a match		
Opening the lid of a box		
Preferred foot for kicking		
Preferred eye if using only one		

Results of Handedness tests:

1 German and 2 Polish subjects had a right-handedness of 66.7%, all the other participants reported usage of the right hand with 76 - 100%.

References

Abramson, A.S., Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague, 569-573.

Adler L.E., Pachtman E., Franks R.D., Pecevich M., Waldo M.C., Freedman R. (1982). Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biological Psychiatry, 17,* 639-654.

Akamatsu, T. (1992). *Essentials of Functional Phonology*. Louvain-la-Neuve: Peeters.

Alain, C., Achim, A., Woods, D.L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology, 36,* 737-744.

Alho, K., Connolly, J.F., Cheour, M., Lethokoski, A., Huotilainen, M., Virtanen, J., et al. (2000). Hemispheric lateralization in pre-attentive processing of speech sounds. *Neuroscience Letters, 20,* 791-797.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16,* 38-50.

Alho, K., Sainio, D., Sajaniemi, N., Reinikainen, K., Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology, 77,* 151-155.

Allard, F., Scott, B.L. (1975). Burst cues, transition cues, and hemispheric specialization with real speech sounds. *Quarterly Journal of Experimental Psychology, 27,* 487-497.

Anderson, B., Southern, B.D., Powers, R.E. (1999). Anatomic asymmetries of the posterior superior temporal lobes: A postmortem study. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 12,* 247-254.

Archangeli, D.B. (1988). *Underspecification in Yawelmani Phonology and Morphology*. New York: Garland.

Arnott, S.R., Binns, M.A., Grady, C.L., Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage, 22*, 401-408.

Aulanko, R., Hari, O.V., Lounasmaa, R., Näätänen, R., Sams, M. (1993). Phonetic invariance in the human auditory cortex. *Neuroreport, 4,* 1356-1358.

Baddeley, A.D. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A.D., Gathercole, S.E., Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Beddor, P.S., Gottfried, T.L (1995). Methodological issues in cross-linguistic speech perception research with adults. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press, 207-232.

Behrens, S., Blumstein, S.E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America, 84,* 861-867.

Belin, P., Zatorre, R.J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14,* 2105-2109.

Belin, P., Zatorre, R.J., Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17-26.

Benson, R.R., Whalen, D.H., Richardson, M., Swainson, B., Clark, V.P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language, 78,* 364-396.

Bentin S., Mann V. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America, 88*, 64-74.

Bernstein, L.E. (2005). Phonetic processing by the speech perceiving brain. In: D.B. Pisoni, R.E. Remez (eds.), *The Handbook of Speech Perception.* Blackwell, 79-98.

Best, C.T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In: B. de Boysson-Bardies, S. de Schoenen, P. Jusczyk, P. MacNeilage, J. Morton (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer Academic Publishers, 289-304.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 171-206.

Best, C.T., McRoberts, G., Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 354-360.

Best, C.T., Avery, R.A. (1999). Left-hemisphere advantage for click consonants is determined by linguistic significance and experience. *Psychological Science, 10,* 65-70.

Best, C.T., McRoberts, G.W., Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America, 109,* 775-794.

Binder, J.R. (2000). The new neuroanatomy of speech perception. *Brain, 123,* 2371-2372.

Birbaumer, N., Schmidt, R.F. (1996). *Biologische Psychologie*. Springer.

Blomert, L., Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language, 89,* 21-26.

Boatman, D., Hart, J., Lesser, R.P., Honeycutt, N., Anderson, N., Miglioretti, D., Gordon, B. (1998). Right hemisphere speech perception revealed by amobarbital injection and electrical interference. *Neurology, 51,* 458-464.

Boersma, P. (1998). *Functional Phonology.* The Hague: Holland Academic Graphics.

Boersma, P. (1999). On the need for a separate perception grammar. *ROA-361,* http://roa.rutgers.edu/.

Boersma, P. (2000). *Phonetically-driven acquisition of phonology*. Unpublished manuscript, available at: http://www.fon.hum.uva.nl/paul/chrono.html.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5* (9/10), 341-345.

Boersma, P. (2005). Some listener-oriented accounts of hache aspir in French. *ROA-730,* http://roa.rutgers.edu/.

Boersma, P. (2005). Prototypicality judgments as inverted perception. *ROA-742,* http://roa.rutgers.edu/.

Boersma, P., Escudero, P. (2004). Learning to perceive a smaller L2 vowel inventory: An Optimality Theory account. *ROA-684,* http://roa.rutgers.edu/.

Bohn, O.-S. (1995). Cross-language speech perception in adults, first language transfer doesn't tell it all. In: W. Strange (ed.), Speech *Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 279-304.

Brière, E. (1966). An investigation of phonological interference. *Language, 42,* 768-800.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In: J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory.* Malden, MA: Blackwell, 4-63.

Brown, C., Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In: S.J. Hannahs, M. Young-Scholten (eds.), *Focus on Phonological Acquisition.* Amsterdam, Philadelphia: John Benjamins, 67-112.

Brown, H.D., Kosslyn, S.M. (1993). Cerebral lateralization. *Current Opinion in Neurobiology, 3,* 183-186.

Buchwald, J.S., Guthrie, D., Schwafel, J. (1994). Influence of language structure on brain-behavior development. *Brain and Language, 46,* 607-619.

Budd, T.W., Barry, R.J., Gordon, E., Rennie, C., Michie, P.T. (1998). Decrement of N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31,* 51-68.

Butcher, A., Ahmad, K. (1987). Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic. *Phonetica, 44,* 156-172.

Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differetially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage, 22,* 1182-1194.

Cardy, J.E.O., Ferrari, P., Flagg, E.J., Roberts, W., Roberts, T.P.L. (2004). Prominence of M50 auditory evoked response over M100 in childhood and autism. *Neuroreport, 15,* 1867-1870.

Chait, M., Simon, J.Z., Poeppel, D. (2004). Auditory M50 and M100 response to broadband noise: Functional implications. *Neuroreport, 15,* 2455-2458.

Cheour, M., Alho, K., Sainio, K., Reinikainen, K., Renlund, M., Eerola, O. Aaltonen, O., Näätänen, R. (1997). The mismatch negativity to speech sounds at the age of three months. *Developmental Neurophysiology, 14,* 167-174.

Cheour, M., Leppänen, P.H., Kraus, N., (2000). Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. Clinical *Neurophysiology, 111*, 4-16.

Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., Näätänen, R. (2002). Mismatch negativity shows that 3-6-year-old children can learn to discrmininate non-native speech sounds within two months. *Neuroscience Letters, 325,* 187-190.

Cheour-Luhtanen, M. Alho, K., Sainio, K. (1996). The ontogenetically earliest discriminative response of the human brain. *Psychophysiology, 33,* 478-481.

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

Chomsky, N., Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Clarke, J., Braginski, A.I. (eds.) (2004). *The SQUID Handbook. Fundamentals and Technology of SQUIDS and SQUID Systems, vol. 1.* Wiley-VCH.

Clements, G.N. (2003). Feature economy in sound systems. *Phonology, 20,* 287-333.

Clements, G.N., Hume, E. (1994). The internal organization of speech sounds. In: J. Goldsmith (ed.), *A Handbook in Phonological Theory*. Oxford: Blackwell, 245-306.

Cohen, D., Halgren, E. (1993). Magnetoencephalography *(Neuromagnetism).* In: *Encyclopedia of Neuroscience*. Elsevier, 3rd edition.

Comrie, B. (1990). *The World's Major Languages.* New York: Oxford University Press.

Crowder, R.G. (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 587-596.

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

Cutler, A., Clifton, C., Jr. (1999). Comprehending spoken language: A blueprint of the listener. In: C. Brown, P. Hagoort (eds.), *Neurocognition of Language Processing.* Oxford: Oxford University Press, 123-166.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8,* 919-924.

Dehaene-Lambertz, G., Dupoux, E., Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12,* 635-647.

Dehaene-Lambertz, G., Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12,* 3155-3158.

Dehaene-Lambertz, G., Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience, 16,* 1375-1387.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage, 24,* 21-33.

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

Demonet, J.F., Price, C., Wise, R., Frackowiak, R.S. (1994). A PET study of cognitive strategies in normal subjects during language tasks. Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain, 117,* 671-682.

Demonet, J.F., Thierry, G., Cardebat, D. (2005). Renewal of the neurophysiology of language: Functional neuroimaging. *Physiological Reviews, 85,* 49-95.

Derwing, B.L. (1992). Orthographic aspects of linguistic competence. In: P. Downing, S. Lima, M. Noonan (eds.), *The Linguistics of Literacy.* Amsterdam, Philadelphia: John Benjamins, 193-210.

Diesch, E., Luce, T. (1997). Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research, 116,* 139-152.

Dogil, G. (1990). *Hissing and hushing fricatives: A comment on non-anterior spirants in Polish.* Unpublished manuscript, University of Bielefeld.

Dogil, G., Braun, G. (1988). The PIVOT Model of Speech Parsing. *Veröffentlichungen der Kommission für Linguistik und Kommunikationsforschung, 19,* Wien: Österreichische Akademie der Wissenschaften.

Dogil, G., Frese, I., Haider, H., Rhm, D., Wokurek, W. (2004). Where and how does grammatical processing take place - and why is Broca's area often involved. A coordinated fMRI/ERBP study of language processing. *Brain and Language, 89,* 337-345.

Efron, B., Tibishirani, R.J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability.* New York, London: Chapman and Hall.

Eggermont, J.J. (2000). Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds. *Journal of Neurophysiology, 84,* 1453-1463.

Eggermont, J.J. (2001). Between sound and perception: Reviewing the search for a neural code. *Hearing Research, 157,* 1-42.

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

Elilhali, M. (2004). *Neural basis and computational strategies for auditory processing.* PhD thesis, University of Maryland, College Park, MD.

Elilhali, M., Fritz, J.B., Klein, D.J., Simon, S.Z., Shamma, S.A. (2004). Dynamics of precise spike timing in primary auditory cortex. *Journal of Neuroscience, 24,* 1159-1172.

Elman, J.L., McClelland, J.L. (1986). Exploiting lawful variability in the speech waveform. In: J.S. Perkell, D.H. Klatt (eds.), *Invariance and Variability in Speech Processing*. Hillsdale, NJ: Lawrence Erlbaum, 360-385.

Escudero, P. (2002). The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts. In: J. Leather, A. James (eds.), *Proceedings of the 4th Symposium on the Acquisition of Second Language Speech, New Sounds 2000,* University of Klagenfurt, 122-131.

Eulitz, C., Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience, 16,* 577-583.

Evers, V., Reetz, H., Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics, 26,* 345-370.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Flege, J.E. (1981). The phonological basis of foreign accents. *TESOL Quarterly, 15,* 443-455.

Flege, J.E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15,* 47-65.

Flege, J.E. (1991). Age of learning affects the authenticity of voice-onset (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 80,.* 395-441.

Flege, J.E. (1993). Production and perception of a novel, second-langage phonetic contrast. *Journal of the Acoustical Society of America, 93,* 1589-1608.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 233-273.

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

Flemming, E.S. (2002). *Auditory Representations in Phonology.* New York, London: Routledge.

Forrest, K., Weismer, G., Milenkovic, P., Dougall, R.N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America, 84,* 115-123.

Fowler, C.A. (1989). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fosker, T., Thierry, G. (2004). P300 investigation of phoneme change detection in dyslexic adults. *Neuroscience Letters, 357,* 171-174.

Fox, R.A., Nissen, S. (2002). Age-related changes in the acoustic characteristics of voiceless English fricatives. In: *Proceedings of the 2002 Conference of Linguistics and Phonetics,* Meikai University, Chiba City, Japan.

Franks, R.D., Pecevich, M., Waldo, M.C., Freedman, R., Adler, L.E., Pachtman, E. (1982). Neuropsychological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biological Psychiatry, 17,* 639-654.

Frese, I. (2004). Phrasenstruktur im Gehirn: Neuroanatomische Korrelate syntaktischer Konstituentenanalyse. *AIMS, 10,* University of Stuttgart.

Fry, D.B., Abramson, A.S., Eimas, P.D., Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171-189.

Fujisaki, H., Kawashima, T. (1968). Identification and discrimination of synthetic speech sounds. *Reports of the 6th International Congress on Acoustics, 2,* Tokio, 95-98.

Gage, N.M., Poeppel, D., Roberts, T.P.L., Hickok, G. (1998). Auditory evoked M100 reflects onset acoustics of speech sounds. *Brain Research, 814,* 236-239.

Gage, N.M., Roberts, T.P. (2000). Temporal integration: Reflections in the M100 of the auditory evoked field. *Neuroreport, 11,* 2723-2726.

Gazzaniga, M.S., Ivry, R.B., Mangun, G.R. (1998). *Cognitive Neuroscience.* New York, London: Norton and Company.

Gibson, E.J. (1991). *An Odyssey in Learning and Perception.* Cambridge, MA: Bradford Books.

Giard, M.H., Lavikainen, J., Perrin, F., Rinikainen, K., Bertrand, O., Thevenet, M., Pernier, J., Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory. *Journal of Cognitive Neuroscience, 27,* 133-143.

Goldinger, S.D., Pisoni, D.B., Luce, P.A. (1996). Speech perception and spoken word recognition: Research and theory. In: N.J. Lass (ed.), *Principles of Experimental Phonetics.* St. Louis: Mosby, 277-327.

Gomot, M., Giard, M.H., Adrien, J.L., Barthelemy, C., Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology, 39,* 577-584.

Gordon, M., Barthmaier, P., Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association, 32,* 141-174.

Gottfried, T.L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12,* 91-114.

Gow, D. W., Caplan, D. (1996). An examination of impaired acoustic - phonetic processing in aphasia. *Brain and Language, 52*, 386-407.

Gray, H.M., Ambady, N., Lowenthal, W.T., Deklin, P. (2004). P300 as an index of attention to self-relevant stimuli. *Journal of Experimental Social Psychology, 40,* 216-224.

Grieser, D., Kuhl, P.K. (1989). Categorization of speech by infants: Support for speech sound prototypes. *Developmental Psychology, 25,* 577-588.

Guenther, F.H., Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100,* 1111-1121.

Guion, S.G., Flege, J.E., Akahane-Yamada, R., Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107,* 2711-2724.

Hall, T.A. (1992). *Syllable Structure and Syllable-Related Processes in German.* Tübingen: Niemeyer.

Halle, M. (2002). Introduction. In: M. Halle (ed.), *From Memory to Speech and Back.* London: MacMillan, 1-17.

Hämäläinen, M.S., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65,* 413-497.

Hämäläinen, M.S., Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions on Bio-medical Engineering, 36,* 165-171.

Hari, R. (1990). The neuromagnetic method on the study of the human auditory cortex. In: F. Grandori, M. Hoke (eds.), *Auditory Evoked Magnetic Fields and Electric Potentials*. Basel: Karger, 222-282.

Hari, R. (1991). On the brain's magnetic responses to sensory stimuli. *Journal of Clinical Neurophysiology, 8,* 157-169.

Hari, R., Kaila, K., Kaila, T., Tuomitso, T., Varbpula T. (1982). Interstimulus dependence of the auditory cortex to interruptions in a steady rhythm. *Neuroscience Letters, 99,* 164-168.

Hari, R., Levnen, S., Raij, T. (2000). Timing of human cortical functions during cognition: Role of MEG. *Trends in Cognitive Sciences, 4,* 455-462.

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

Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley.

Hedrick, M., Ohde, R. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America, 94*, 2005-2023.

Heinz, J.M. Stevens, K.N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America, 33,* 589-596.

Hertrich, I., Mathiak, K., Lutzenberger, W., Ackermann, H. (2000). Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. *NeuroReport, 11,* 4017-4020.

Hertrich, I., Mathiak, K., Lutzenberger, W., Ackermann, H. (2002). Hemisperic lateralization of the processing of consonant-vowel syllables (formant transitions): Effects of stimulus characteristics and attentional demands on evoked magnetic fields. *Neuropsychologia, 40,* 1902-1917.

Hickok, G. (2001). Functional anatomy of speech perception and speech production: Psycholinguistic implications. *Journal of Psycholinguistic Research, 30,* 225-235.

Hickok, G., Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131-138.

Hickok, G., Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92,* 67-99.

Hillenbrand, J., Getty, L., Clark, M., Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099-3111.

Holt, L.L., Lotto, A.J., Diehl, R. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America, 116,* 1763-1773.

Horwitz, B., Poeppel, D. (2002). How can EEG/MEG and fMRI/PET data be combined? *Human Brain Mapping, 17,* 1-3.

Hugdahl, K. (1998). Dichotic listening: Probing temporal lobe functional integrity. In: R. J. Davidson, K. Hugdahl (eds.), *Brain Asymmetry.* Cambridge, MA: MIT Press, 123-156.

Hume, E., Johnson, K., Seo, M., Tserdanelis, G., Winters, S. (1999). A cross-linguistic study of stop place perception. *Proceedings of the 14th International Conference on Phonetic Sciences,* 2069-2072.

Hume, E., Johnson, K. (2001). A model on the interplay of speech perception and phonology. In: E. Hume, K. Johnson (eds.), *The Role of Speech Perception in Phonology*. Academic Press, 3-26.

Hume, E., Johnson, K. (2003). The impact of partial phonological contrast on speech perception. In M.J. Sole, D. Recasens, J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, 2385-2388.

Huotilainen, M., Kujala, A., Hotakainen, M., Shestakova, A., Kushnerenko, E., Parkkonen, L., Fellman, V., Näätänen, R. (2003). Auditory magnetic responses of healthy newborns. *Auditory and Vestibular Systems, 14,* 1871-1875.

Hutsler, J.J. (2003). The specialized structure of human language cortex: Pyramidal cell size asymmetries within auditory and language-associated regions of the temporal lobes. *Brain and Language, 86,* 226-242.

Hutsler, J.J., Galuske, R.A. (2003). Hemispheric asymmetries in cerebral cortical networks. *Trends in Neurosciences, 26*, 429-435.

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.

Ivry, R.B., Robertson, L.C. (1998). *The Two Sides of Perception.* MIT Press.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levnen, S., et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *PNAS, 101,* 6809-6814.

Jacobsen, R., Fant, G., Halle, M. (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Jacobson, T., Schröger, E., Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41,* 654-659.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience, 23,* 9541-9546.

Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. In: B. Lindblom, S. Ohman (eds.), *Frontiers of Speech Communication Research.* London: Academic Press, 77-91.

Johnsrude, I.S., Penhune, V.B., Zatorre, R.J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123,* 155-163.

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

Jusczyk, P. (1995). On characterizing the development of speech perception. In: J. Mehler, R. Fox (eds.), *Beyond the Blooming, Buzzing Confusion.* Hillsdale, NJ: Lawrence Erlbaum.

Kaas, J.H., Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *PNAS, 97,* 11793-11799.

Kager, R. (1999). *Optimality Theory.* Cambridge: Cambridge University Press.

Kanno, A., Nakasato, N., Nagamatsu, K, Iwasaki, M., Hatanaka, K, Murayama, N., Yoshimoto, T. (2001). Comparison of source localization for the P30m, P50m, and N100m peaks of the auditory evoked fields. In: J. Nenonen, R.J. Ilmoniemi, T. Katila (eds.), *Proceedings of the 12th International Conference on Biomagnetism,* Espoo, Finnland, Helsinki University of Technology, 25-28.

Kaukoranta, E., Hari, R., Lounasmaa, O.V. (1987). Responses of the human auditory cortex to vowel onset after fricative consonants. *Experimental Brain Research, 69,* 19-23.

Kayser, J., Tenke, C.E., Bruder, G.E. (1998). Dissociation of brain ERP topographies for tonal and phonetic oddball tasks. *Psychophysiology, 35,* 576-590.

Kenstowicz, M. (1994). *Phonology in Generative Grammar.* Cambridge, MA, Oxford, UK: Blackwell.

Kiparsky, P. (1982). Lexical phonology and morphology. In: I.S. Yang (ed.), *Linguistics in the Morning Calm.* Seoul: Hanshin, 3-91.

Kohler, K.J. (1977). *Einführung in die Phonetik des Deutschen.* Berlin: E. Schmidt.

Korzyukov, O.A., Winkler, I., Gumenyuk, V.I., Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *Neuroimage, 20,* 2245-2258.

Kraus, N., McGee, T. (1992). Electrophysiology of the human auditory system. R.R. Fay, A. N. Popper (eds.), *The Mammalian Auditory Pathway: Neurophysiology.* New York: Springer.

Kraus, N., McGee T., Micco A., Sharma A., Carrell T., Nicol T. (1993). Mismatch negativity in school-age children to speech stimuli that are just perceptibly different. *Electroencephalography and Clinical Neurophysiology, 88,* 123-130.

Kudela, K. (1968). Spectral analysis of Polish fricative consonants. In: W. Jassem (ed.), *Speech Analysis and Synthesis.* Warszawa: PAN.

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

Kuhl, P.K. (1992). Infants' perception and representation of speech: Development of a new theory. In: J.J. Ohala, T.M. Nearey, B.L. Berwing, M.M. Hodge, G.E. Wiebe (eds.), *Proceedings of the International Conference on Spoken Language Processing,* Edmonton, Alberta.

Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97,* 11850-11857.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831-843.

Kuhl, P.K., Iverson, P. (1995). Linguistic experience and the 'Perceptual Magnet Effect'. W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 121-154.

Kujala, T. Näätänen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience and Biobehavioral Reviews, 25,* 535-543.

Lacerda, F. (2001). Re-assessing the perceptual consequences of CV-transition speed. *Lund Linguistic Working Papers, 49,* 98-101.

Ladefoged, P., Maddieson, I. (2002). *The Sounds of the World's Languages.* Blackwell.

Ladefoged, P., Shalev, M. Bhaskararao, P. (1994). Phonetics of Toda. *PILC Journal of Dravidic studies, 4,* 19-55.

Lahiri, A., Marslen-Wilson, W.D. (1991). The mental representation of lexical form: A phonological approach. *Cognition, 38,* 245-294.

Lahiri, A., Marslen-Wilson, W.D. (1992). Lexical processing and phonological representation. In: G.J. Docherty, D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 229-254.

Lahiri, A., Reetz, H. (2002). Underspecified recognition. In: C. Gussenhoven, N. Werner (eds.), *Laboratory Phonology VIII*. Berlin: Mouton, 637-675.

Liberman, A.M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman, A.M., Harris, K. S. Hoffman, H.S., Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 89,* 358-368.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review, 74,* 431-461.

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

Lisker, L. (2001). Hearing the Polish sibilants /s, š, ś/: Phonetic and auditory judgements. In: N. Grønnum, J. Rischel (eds.), *To honour Eli Fischer-Jørgensen*. Copenhagen: C.A. Reitzel, 226-237.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In: K. Elenius, P. Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences, 4,* Stockholm, 92-99.

Lütkenhöner, B., Pantev, C., Hoke, M. (1990). Comparison between different methods to approximate an area of the human head by a sphere. In: F. Grandori, M. Hoke (eds.), *Auditory Evoked Magnetic Fields and Electric Potentials*. Basel: Karger, 103-118.

Macmillan, N.A., Kaplan, H.L., Creelman, C.D. (1977). The psychophysics of categorical perception. *Psychological Review, 84,* 452-471.

Mäkelä, J.P., Hari, R., Leinonen, L. (1988). Magnetic responses of the human auditory cortex to noise/square wave transitions. *Electroencephalography and Clinical Neurophysiology, 69,* 423-430.

Malmivuo, J., Plonsey, R. (1995). *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*. New York: Oxford University Press.

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

Markham, D. (1997). Phonetic imitation, accent, and the learner. *Traveaux de l'institut de linguistice de Lund, 33,* Lund University.

Marshall, P,J., ar-Haim, Y., Fox, N.A. (2004). The development of P50 suppression in the auditory event-related potential. *International Journal of Psychophysiology, 51,* 135-141.

Massaro, D.W. (1998). Categorical Perception: Important phenomenon or lasting myth? In: R.H. Mannell, J. Robert-Ribes (eds.), *Proceedings of the 5th International Congress of Spoken Language Processing, 6,* Sydney, 2275-2279.

Mathiak, K. (2002). *Neuroimaging of cortical processing of complex auditory stimuli: speech, time, and space.* Habilitationsschrift, University of Tübingen.

Mathiak, K., Hertrich, I., Lutzenberger, W., Ackermann, H. (1999). Preattentive processing of consonant vowel syllables at the level of the supratemporal plane: A whole-head magnetencephalography study. *Cognitive Brain Research, 8,* 251-257.

Mathiak, K., Hertrich, I., Lutzenberger, W., Ackermann, H. (2001). Neural correlates of duplex perception: A whole-head magnetoencephalography study. *NeuroReport, 12,* 501-506.

Mathiak, K., Hertrich, I., Lutzenberger, W., Ackermann, H. (2002). Functional cerebral asymmetries of pitch processing during dichotic stimulus application: A whole-head magnetoencephalography study. *Neuropsychologia, 40,* 585-593.

Matthies, M.L., Svirsky, M.A., Perkell, J.S., Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech and Hearing Research, 39,* 936-946.

May, J.G. (1981). Acoustic factors that may contribute to categorical perception. *Language and Speech, 24,* 273-284.

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., and Taylor, J.G., Näätänen, R. (1999). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience, 6,* 99-120.

McQueen, J. (2004). Speech perception. In: K. Lamberts, R. Goldstone (eds.), *Handbook of Cognition.* Sage Publications, 255-275.

Mehler, J., Segui, J., Frauenfelder, U. (1981). The role of syllable in language acquisition and perception. In: T.F. Myers, J. Laver, J. Anderson (eds.), *The Cognitive Representation of Speech.* Amsterdam: North-Holland, 295-333.

Micheyl, C., Carlyon, R.P., Shtyrov, Y., Hauk, O., Dodson, T., Pulvermüller, F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *Journal of Cognitive Neuroscience, 15,* 747-758.

Michie, P. (2001). What has MMN revealed about the auditory system in schizophrenia? *International Journal of Psychophysiology, 42,* 83-100.

Miller, J.L., Wier, C.C., Pastore, R., Kelly, W.J., Dooling, R.J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 410-417.

Milner, A.D., Goodale, M.A. (1995). *The Visual Brain in Action*. Oxford University Press.

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.

Mitterer, H., Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics, 65,* 956-969.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J., Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics, 18,* 331-340.

Möbius, B., van Santen, J. (1996). Modeling segmental duration in German text-to-speech synthesis. *Proceedings of the International Conference on Spoken Language Processing, 4,* 2395-2398.

Morais, J. (1985). Literacy and awareness of the units of speech: Implications for research on the units of perception. *Linguistics, 23,* 707-721.

Morais, J., Kolinsky, R. (1995). Levels of processing in the phonological segmentation of speech. *Language and Cognitive Processes, 12,* 871-876.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage, 13,* 684-701.

Morrongiello, B., Robson, R., Best, C.T. , Clifton, R.K. (1984). Trading relations in the perception of speech by 5-year-old children. *Journal of Experimental Child Psychology, 37,* 231-250.

Munro, M., Flege, J.E., MacKay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17,* 313-334.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functioning. *Behavioral Brain Science, 13,* 201-288.

Näätänen, R. (1992). *Attention and Brain Function.* Hillsdale, NJ: Lawrence Erlbaum.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing, 16,* 6-18.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negitivity MMN and its magnetic equivalent MMNm. *Psychophysiology, 38,* 1-21.

Näätänen, R. (2003). Mismatch negativity: Clinical research and possible applications. *International Journal of Psychophysiology, 48,* 179-188.

Näätänen, R. Schröger, E., Karakas, S., Tervaniemi, M., Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport, 4,* 503-506.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences, 24,* 283-288.

Näätänen, R., Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24,* 375-425.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M. (1989). Do event-related potential reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters, 98,* 217-221.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtinainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432-434.

Näätänen, R., Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin, 125,* 826-859.

Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology, 115,* 140-144.

Näätänen, R., Jacobsen, T., Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology, 42,* 25-32.

Nawrocki, G., Gonet, W. (2004). *Realizations of /x/ in intervocalic contexts in Southern Polish.* Unpublished manuscript.

Nenonen, S., Shestakova, A., Huotilainen, M., Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language, 92,* 26-32.

Neppert, J.M.H. (1999). *Elemente einer Akustischen Phonetik*. Hamburg: Buske.

Nguyen, N., Hoole, P., Marchal, A. (1994). Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters. *Journal of the Acoustical Society of America, 96,* 33-39.

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., Manning, C., Meyer, G. (1993). The perceptual weighting of acoustic cues changes with linguistic experience. *Journal of the Acoustical Society of America, 94,* S1865.

Nittrouer, S., Miller, M.E. (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America, 102,* 572-580.

Nittrouer, S., Studdert-Kennedy, M., McGowan, R.S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120-132.

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

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

Nowak, P.M. (2003). The role of vocalic context in the perception of Polish sibilants. In M.J. Sole, D. Recasens, J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (p. 107). Barcelona.

Obleser, J., Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience, 16,* 31-39.

Ohde, R.N., Haley, K.L. (1997). Stop-consonant and vowel perception in 3- and 4- year-old children. *Journal of the Acoustical Society of America, 102,* 3711-3722.

Paavilainen, P., Jaramillo, M., Näätänen, R., Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters, 265,* 179-182.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., Winkler, I. (2001a). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology, 38,* 359-365.

Paavilainen, P., Valppu, S., Näätänen, R. (2001b). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1+1~2 but 1+1+1<3. *Neuroscience Letters, 301,* 179-182.

Padgett, J., Zygis, M. (2003). The evolution of sibilants in Polish and Russian. In: T.A. Hall, S. Hamann (eds.), *ZAS Papers in Linguistics, 32*, Berlin, 155-174.

Parnell, M.M., Amerman, J.D. (1978). Maturational influences on perception of coarticulatory effects. Journal of Speech and Hearing Research, 21, 682-701.

Peltola, M.S., Kujala, T., Toumainen, J., Ek, M., Aaltonen, O., Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. Neuroscience Letters, 352, 25-28.

Penhune, V.B., Zatorre, R.J., MacDonald, J.D., Evans, A.C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. Cerebral Cortex, 6, 661-672.

Peperkamp, S., Pattinato M., Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In: B. Beachley, A. Brown, F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development, 2,* Sommerville, MA: Cascadilla Press, 650-661.

Phillips, C., Marantz, A., McGinnis, M. Pesetsky, D. Wexler, K. Yellin, E. (1995). Brain mechanisms of speech perception: A preliminary report. In: C.T. Schtze, J.B. Ganger, K. Brohier (eds*.), Papers on Language Processing and Acquisition, 26.* MIT Press, 153-191.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12,* 1038-1055.

Picton, T.W., Alain, C., Otten, L., Ritter, W., Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neurootology, 5,* 111-139.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In: J. Bybee, P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure.* Amsterdam: Benjamins.

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

Pisoni, D.B. (1997). Some thoughts on "normalization" in speech perception. In: K. Johnson, J.W. Mullennix (eds.), *Talker Variability in Speech Processing*. San Diego, CA: Academic Press, 9-32.

Pisoni, D.B., Luce, P.A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25,* 21-52.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25,* 679-693.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time". *Speech Communication, 41,* 245-255.

Poeppel, D., Phillips, C. Yellin, E., Rowley, H.A., Roberts, T.P., Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters, 221,* 145-148.

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

Polnau, D., Kossel, A. (2002). Zelluläre Grundlagen des Gedächtnisses. *DNP, 11,* 40-42.

Rauschecker, J.P. (1998). Parallel processing in the auditory cortex of primates. *Journal of Audiology and Neuro-Otololaryngology, 3,* 86-103.

Read, C., Zhang, Y.F., Nie, H.Y., Ding, B.Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24,* 31-44.

Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine.* New York: Elsevier Science Publishing.

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

Repp, B.H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81-110.

Repp, B.H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language, 10,* 243-335.

Repp, B.H., Mann, V.A. (1978). Influence of vocalic context on perception of the [s]-[ ] distinction. *Journal of the Acoustical Society of America, 64,* S1.

Rif, J., Hari, R., Hämäläinen, M.S., Sams, M. (1991). Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology, 79,* 464-472.

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., Näätänen, R. (1999). Analysis of speech sounds is left-hemisphere predominant at 100-150ms after sound onset. *Neuroreport, 10,* 1113-1117.

Ritter, W., Deacon, D., Gomes, M., Javitt, D.J., Vaughan, H.G. (1995). The mismatch negativity of event-related potential as a probe of transient auditory memory: A review. *Ear and Hearing, 16,* 51-67.

Rivera-Gaxiola, M., Csibra, G., Johnson, M.H., Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research, 111,* 13-23.

Roberts, T., Poeppel, D. (1996). Latency of auditory evoked {M}100 as a function of tone frequency. *Neuroreport, 7,* 1138-1140.

Roberts, T.P.L., Poeppel, D., Rowley, H.A. (1998). Magnetoencephalography and functional MRI: A quantitative study of speech perception. In: *20th Annual International Conference, 20, Hong Kong,* 2120-2123.

Romani, G.L., Williamson, S.J., Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science, 216,* 1339-1340.

Rosburg T., Haueisen J., Sauer H. (2002). Stimulus duration influences the dipole location shift within the auditory evoked field component N100m. *Brain Topography, 15,* 37-41.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology, 104,* 573-605.

Rosch, E., Lloyd, B.B. (1978). *Cognition and Categorization.* New Jersey: Lawrence Erlbaum Associates.

Rowan, A.J., Tolunsky, E. (2003). *Primer of EEG.* Butterworth Heinemann.

Rubach, J. (1984). *Cyclic and Lexical Phonology. The Structure of Polish.* Dordrecht: Foris.

Saffran, E.M., Marin, O.S.M., Yeni-Komishian, G.H. (1976). An analysis of speech perception in word deafness. *Brain and Language, 3,* 209-228.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology, 62,* 437-448.

Samson, S., Zatorre, R.J. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia, 32,* 231-240.

Sattler, J.B. (1998). Zur Testung der Händigkeit. *Left Hand Corner, 2,* 6-12.

Schandry, R. (1998). *Lehrbuch Psychophysiologie.* München: Psychologie Verlagsunion.

Scherg, M. (1990). Fundamentals of dipole source potential analysis. In: F. Grandori (ed.), Auditory Evoked Fields and Potentials, Advances in Audiology, 6. Karger, 40-69.

Scherg, M. (1991). *Akustisch evozierte Potentiale.* Stuttgart, Berlin, Köln: Kohlhammer.

Scherg, M., Vajsar, J., Picton, T.W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience, 1,* 336-355.

Schröger, E., Winkler, I. (1995). Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. *Neuroscience Letters, 193,* 185-188.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400-2406.

Scott, S.K., Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neuroscience, 26,* 100-107.

Scott, S.K., Wise, R.J.S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition, 92,* 13-45.

Shadle, C.H. (1985). The acoustics of fricative consonants. *R.L.E. Technical Report, 506,* MIT.

Shadle, C.H. (1990). Articulatory-acoustic relationships in fricative consonants. In: W.J. Hardcastle, A. Marchal (eds.), *Proceedings of Speech Production and Speech Modelling,* 187-209.

Shadle, C.H., Mair, S.J. (1996). Quantifying spectral characteristics of fricatives. In: C.H. Shadle, S.J. Mair (eds.), *Proceedings of the 4th International Conference on Language Processing,* 187-195.

Shafer, V.L., Schwartz, R.G., Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research, 18,* 242-254.

Sharma, A., Dorman, M.F. (1998). Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *Journal of the Acoustical Society of America, 104,* 511-517.

Sharma, A., Dorman, M.F. (2000). Neurophysiological correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107,* 2697-2703.

Shestakova, A., Brattico, E., Houtilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Journal of Cognitive Neuroscience and Neuropsychology, 13,* 1813-1816.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J., Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage, 12,* 657-663.

Sinex, D.G., McDonald, L.P. (1989). Synchronized discharge rate representation of voice onset time in the chinchilla auditory nerve. *Journal of the Acoustical Society of America, 85,* 1995-2004.

Steinschneider, M., Fishman, Y.I., Arezzo, J.C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *Journal of the Acoustical Society of America, 114,* 307-321.

Stevens, K.N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America, 68,* 836-842.

Stevens, K.N. (1985). Evidence for the role of acoustic boundaries in the perception of speech sounds. In: V. Fromkin (ed.), *Phonetic Linguistics: Essays in honor of Peter Ladefoged*. New York: Academic Press.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17,* 3-46.

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. (2005). Features in speech perception and lexical access. In: D.B. Pisoni, R.E. Remez (eds.), *The Handbook of Speech Perception.* Blackwell, 125-155.

Strange, W. (1995). Cross-linguistic studies of speech percpetion: A historical review. In: W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore, MD: York Press, 3-48.

Stufflebeam, S.M., Poeppel, D., Rowley, H., Roberts, T.P.L. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *Neuroreport, 9,* 91-94.

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

Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America, 111,* 2887-2896.

Tecchio, F., Salustri, C., Thaut, M.H., Pasqualetti, P. Rossini, P.M. (2000). Conscious and preconscious adaptation to rhythmic auditory stimuli: A magnetoencephalographic study of human brain responses. *Experimental Brain Research, 135,* 222-230.

Thyer, N., Hickson, L. (2000). The perceptual magnet effect in Australian English vowels. *Perception and Psychophysics, 62,* 1-20.

Tiitinen, H. Sivonen, P., Alku, P.,Virtanen, J., Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research, 8,* 355-363.

Titova, N., Näätänen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neuroscience Letters, 308,* 63-65.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., Best, C. (1994). English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. Paper presented at the International Conference on Spoken Language, Yokohama, Japan.

Ulanovsky, N., Las, L., Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience, 6,* 391-398.

Underbakke, M., Polka, L., Gottfried, T.L., Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84,* 90-100.

Valbret, H., Moulines, E., Tubach, J. (1991). Voice transformation using PSOLA technique. *Proceedings of Eurospeech, 91,* 345-348.

Van Hessen, A.J., Schouten, M.E.H. (1999). Categorical perception as a function of stimulus quality. *Phonetica, 56,* 56-72.

Van Heuven, V.J.J.P. (1983). Rise time and duration of friction as perceptual cues in the affricate-fricative contrast in English. In: M.P.R. van den Broecke, V.J.J.P. van den Heuven, W. Zonnefeld (eds.), *Sound Structures, Studies for Antonie Cohen.* Foris, 141-157.

Van Son, R.J.J.H., Pols, L.C.W. (1996). A comparison between the acoustics of vowel and consonant reduction. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 20,* 13-25.

Vasama, J., Mäkelä, J.P., Tissari, S.O., Hämäläinen, M.S. (1995). Effects of intensity variation on human auditory cortex: N100 topography and multiple dipole model analysis. *Acta Otolaryngology, 115,* 616-621.

Velenovsky, D.S., Cetas, J.S., Price, R.O., Sinex, D.G., McMullen, N.T. (2003). Functional subregions in primary auditory cortex defined by thalamocortical terminal arbors: An electrophysiological and anterograde labeling study. *Journal of Neuroscience, 23,* 308-316.

Von Kriegstein, K., Eger E., Kleinschmidt A., Giraud A.L. (2003). Modulation of neural responses to speech by directing attention to voices or semantic content. *Cognitive Brain Research, 17,* 48-55.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., Liddle, P.F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994-1005.

Vroeijenstijn, M. (1998). *The importance of correct geometry in neuronal source localization.* Master's Thesis, University of Utrecht.

Wagner, A., Ernestus, M. (2004). Language-specific relevance of formant transitions for fricative. *Journal of the Acoustical Society of America, 115,* 2392.

Webster, M.A., Werner, J.S., Field, D.J. (in press). Adaptation and the phenomenology of perception. In: C. Clifford, G. Rhodes (eds.), *Fitting the Mind to the World: Adaptation and Aftereffects in High Level Vision: Advances in Visual Cognition Series, 2.* Oxford University Press.

Werker, J.F., Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Wessinger, C.M., Buonocore, M.H., Kussmaul, C.L., Mangun, G.R. (1997). Tonotopy in human auditory cortex examined with functional neuroimaging. *Human Brain Mapping, 5,* 18-25.

Wheeler, M.W. (2005). *The Phonology of Catalan.* Oxford: Oxford University Press.

Wible, B., Nicol, T., Kraus N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain, 128,* 417-423.

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

Wilde, L. (1993). Inferring articulatory movements from acoustic properties at fricative-vowel boundaries. *Journal of the Acoustical Society of America, 94,* 1881.

Williamson, S.J., Kaufman, L. (1990). Theory of neuroelectric and neuromagnetic fields. In: F. Grandori, M. Hoke, G.L. Romani (eds.), *Auditory Evoked Magnetic Fields and Electric Potentials.* Basel: Karger, 1-39.

Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., Näätänen, R. (1990). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology, 27,* 228-235.

Winkler, I., Karmos, G., Näätänen, R. (1996). Adaptive modelling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742,* 239-252.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., et al. (1999a). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357-369.

Winkler, I., Kujala, T.,Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R.J., Näätänen, R. (1999b). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36,* 638-642.

Winkler, I., Czigler, I. (1998). Mismatch negativity: Deviance detection or the maintenance of the "standard". *Neuroreport, 9,* 3809-3813.

Winkler, I., Korzyukov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Ilmoniemi, R.J., Alho, K., Näätänen, R. (2002). Temporary and longer term retention of acoustic information. *Psychophysiology, 39,* 530-534.

Woods, D. (1992). Auditory selective attention in middle-aged and elderly subjects: An event-related brain potential study. *Electroencephalography and Clinical Neurophysiology, 84,* 456-468.

Woods, D.L. (1995). The component structure of the N1 wave of the human auditory evoked potential. In: G. Karmos, M. Molnar, V. Csepe, I. Czigler, J.E. Desmedt (eds.), *Perspectives of Event-related Potentials Research, EEG Supplement, 44.* Elsevier Science (BV), 102-108.

Yao, J., Dewald, J.P.A. (2005). Evaluation of different cortical source localization methods using simulated and experimental EEG data. *Neuroimage, 25,* 369-382.

You, H.-Y. (1979). *An acoustical and perceptual study of English fricatives.* Master's Thesis, University of Edmonton, Canada.

Zatorre, R.J. (2001). Neural specializations for tonal processing. *Annals of the New York Academy of Sciences, 930,* 193-210.

Zatorre, R.J., Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946-953.

Zatorre, R.J., Belin, P., Penhune, V.B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6,* 37-46.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., Stevens, E. B. (2001). MMFs to native and non-native syllables show an effect of linguistic experience. In: J. Nenonen, R.J. Ilmoniemi, T. Katila (eds.), Proceedings of the 12th International Conference on Biomagnetism, Espoo, Finnland, Helsinki University of Technology, 379-382.

Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *Neuroimage, 26,* 703-720.

Zygis, M. (2003). Phonetic and phonological aspects of Slavic sibilant fricatives. In T.A. Hall, S. Hamann (eds.), *ZAS Papers in Linguistics, 32,* Berlin, 175-213.

Zygis, M., Hamann, S. (2003). Perceptual and acoustic cues of Polish coronal fricatives. In: *Proceedings of the 15th International Conference on Phonetic Sciences,* 395-398.