

Distinguishing between prelexical levels in speech perception: an adaptation-fMRI study

Attila Andics^{1,2}

Supervisors: Dr. Miranda van Turenhout², Dr. James M. McQueen³

Tutor: Prof. Dr. Peter Hagoort²

²F. C. Donders Centre for Cognitive Neuroimaging, ³Max Planck Institute for Psycholinguistics

Behavioural studies of speech perception indicate a dissociation of acoustic-phonetic and a more abstract level of prelexical processing. Here we used functional magnetic resonance imaging, in combination with an adaptation paradigm to distinguish between the neural correlates of acoustic-phonetic and phonological processing of the fricative speech sounds [S] and [s] by Dutch listeners. To manipulate the perceived quality of the fricatives while keeping their acoustic properties constant, we varied the sounds' vowel-context: Dutch listeners are more likely to classify an ambiguous fricative from an [S] – [s] continuum as [S] when it is followed by [i], and as [s] when it is followed by [y] (Smits, 2001a). Syllable pairs were presented consisting of an [S], an [s], or an ambiguous fricative, followed by an [y] or an [i] vowel. The most ambiguous fricative was identified individually for each subject in a pretest. During the fMRI scanning session syllable pairs were presented rapidly and randomly intermixed. A sparse scanning paradigm was used, in which auditory stimuli were presented during silent gaps between image acquisitions. We hypothesized that syllable pairs in which the fricatives were perceived as the same would lead to larger adaptation in brain regions responsible for phonological processing as compared to syllable pairs in which the fricatives were perceived as different. Distinct fMRI adaptation patterns were observed corresponding to acoustic-phonetic and phonological processing respectively, supporting the functional distinction between these processing levels. No cortical regions showed both patterns: the cortical regions underlying the acoustic-phonetic and phonological stages in prelexical processing appear to be anatomically distinct.

¹ Correspondence to: Attila Andics, F. C. Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, 6500 HB Nijmegen

Introduction

Prelexical processing levels in speech perception

During speech perception, continuous and infinitely variable sound waves must be mapped onto stored lexical representations in the listener's brain. These lexical items are discrete. The task the listener has is to match every particular token to a word type. Analogue representations therefore have to turn into discrete representations, and how exactly this happens is one of the greatest mysteries of speech perception. One solution is that there are prelexical levels, where representations of speech sounds mediate between the speech signal and the lexicon. As McQueen (2005) has argued, prelexical levels could be an efficient way to partially solve the mapping problem.

A possibly important step on the prelexical stage could be the phenomenon referred to as categorical perception. The definition of categorical perception is based on identification and discrimination experiments with a continuum of speech sounds varied in an acoustic parameter and ranging across two (or more) unambiguous consonants. The defining properties are (1) an abrupt increase of steepness in the identification function for a speech sound at the phoneme boundary, and (2) an abrupt increase of the discrimination accuracy function for across-category stimulus pairs from the continuum as contrasted to the poor discrimination accuracy of within-category pairs. The phenomenon of categorical perception was first described by Liberman (1957), and was claimed to be a speech-specific phenomenon from the beginning. Later, general auditory mechanisms were shown to contribute significantly to the categorical perception of speech sounds, questioning that it is specific to speech and even that it is specific to humans. But it was repeatedly shown that language experience does play an important role in categorical perception. Today it is hardly debated that at least a certain kind of categorical perception is speech- and language-specific: although languages, on the one hand, exploit natural boundaries, but, on the other hand, they are also able to modify them (for an overview, see Diehl et al., 2004).

But the explanatory power of categorical perception is often exaggerated. Models with a prelexical level that makes discrete categorical phoneme decisions are clearly simplistic (McQueen, 2005). Indeed, categorical perception

seems to reveal categories that are organised around prototypes and have an internal structure (Phillips et al., 2000). But these categories cannot be discrete phonological categories, because the defining property of discrete categories is that all within-category contrasts are lost. Therefore, categorical perception in itself is not a sufficient mechanism to phonologically categorize speech sounds, and this does not appear to be the right phenomenon to demonstrate the presence of phoneme categories either. To make things even worse, in a general critic of categorical perception research Schouten et al. (2003) have pointed out that the discrimination task might be measuring a task-inherent bias, in fact, an artefact only. All in all, categorical perception does not solve the invariance problem.

We will argue for a more plausible view of prelexical processing which consists of minimally two processing stages. There were several attempts to make similar distinctions. Phillips (2001) has distinguished three levels of prelexical processing: acoustic, phonetic and phonological processing levels. According to his view, acoustic processing is a not speech-specific, not language-dependent auditory processing, which is sensitive to fine-grained differences of speech sounds, independently of their phoneme category membership; phonetic processing is a speech-specific, language-dependent processing, where the processing function of the input sounds is nonlinear, but within-category differences are still relevant; and finally, phonological processing is a speech-specific, language-dependent processing, where the processing function of the input sounds is discrete categorical, and within-category differences are irrelevant. We believe that this analysis rightfully shows that acoustic-phonetic processing involves various, both speech-specific and not speech-specific sub-processes; and it also rightfully points out that phonological processing is necessarily more than just speech-specific processing, but it fails to present distinctive behavioural evidence for each stage and it also fails to note that there is no clear proof for the presence of discrete phonological categories at the prelexical processing stage, even though Scott and Wise (2004) have warned that a phonemic level of representation is not logically necessary. Similarly, Indefrey and Cutler (2005) have noted that there is evidence compatible with models involving alternative intervening representations or no intervening representations.

Here we test a simpler framework for prelexical processing. On logical grounds, assuming that there are prelexical phonological representations to help solve the invariance problem, it is motivated clearly that there should also be at least one earlier stage of

acoustic-phonetic processing. The role of this early processing stage would be to generate more abstract phonological output from the speech signal for the later stage (McQueen, 2005). Behavioural evidence suggests there are early processes, which are sensitive to fine physical differences in the speech signal; and speech-specific processes, which depend on the categorical identity of speech sounds (e.g., Samuel and Kat, 1996). Smits (2001b) has proposed that the decoding of continuous, coarticulated speech is based on hierarchical categorization dependencies, suggesting there are at least two distinct, cascading processing stages and no discrete phonemic representations on the prelexical level. We hypothesize that there are (minimally) two processing stages at the prelexical level, one acoustic-phonetic and one more phonologically abstract, with these two stages acting in cascade and passing information on continuously to the lexicon. In the present study we will examine whether there are neural correlates of this two-way distinction (acoustic-phonetic processing versus phonological processing).

Neuroanatomical correlates of prelexical processing levels

This section reviews recent neuroimaging literature in light of the proposed two-way distinction to see what previous studies have said about neural correlates of separate prelexical processing levels. Most models of speech processing assume that processing is hierarchically organised and that this hierarchy might be mapped onto auditory anatomy. Primate and human studies indicate the interconnectedness of adjacent regions in the auditory cortex, with an information flow from core, to belt, to parabelt, and to more distal regions, such as more distant parts of the superior and middle temporal gyri and the left inferior frontal gyrus (Davis and Johnsrude, 2003; Scott and Johnsrude, 2003). Scott and Wise (2004) proposed that there are two distinct processing pathways involved in speech perception, both having a network of connections spreading from primary auditory areas: an anterior stream of processing, running lateral and anterior to primary auditory cortex is implicated in the mapping of sound onto meaning, while a divided posterior stream including the temporo-parietal junction and the posterior superior temporal sulcus is involved in a mapping of speech sounds onto motor representations of articulation.

A large number of studies have reported the involvement of superior temporal regions in speech-specific processing in adults (Zatorre et al., 1992; Binder et al., 2000, Scott and Wise, 2003), in young children (Ahmad et al., 2003) and even in infants (Dehaene-Lambertz et al., 2002). The most commonly found area is the left posterior superior temporal gyrus (Benson et al., 2001; Ruff et al., 2003; Zevin and McCandliss, 2005; Dehaene-Lambertz et al., 2005). There is evidence for its involvement in speech sound discrimination (Jacquemot et al., 2003; Callan et al., 2004) and in using phonetic experience (Liebenthal et al., 2003). The superior temporal gyrus was also claimed to have more general and more specific roles, like processing spectrally complex sounds (Belin et al., 1999; Joanisse et al., 2003), processing transient acoustic features of speech (Poldrack et al., 2001) or detecting rhymes (Seghier et al., 2004; Burton et al., 2005). Bilateral activation of the superior temporal gyri in speech processing tasks was shown too (Demonet et al., 1992; Hugdahl et al., 2003). The superior temporal sulcus also seems to have a role in speech-specific processing (Jäncke et al., 2002; Dehaene-Lambertz et al., 2005), and its posterior part is even claimed to be involved in the manipulation of phonological representations (Seki et al., 2004). Although these studies have provided clear evidence for the involvement of the superior temporal regions in prelexical stages of speech processing, their exact function is still largely unknown.

There is also considerable evidence for the involvement of inferior frontal regions, with left dominance, in phonetic or phonological processing (Zatorre et al., 1992; Demonet et al., 1992; Hsieh et al., 2001; Gandour et al., 2002; Sharp et al., 2001; Golestani and Zatorre, 2004; Seghier et al., 2004; Burton et al., 2005; Dehaene-Lambertz et al., 2005), but how and how much is this involvement related to speech-specific processes and how much is it a consequence of more general processes, is debated. It is often claimed that the inferior frontal gyrus has a more general role and its activation reveals a general mechanism for selecting among competing phonetic categories (Thompson-Schill et al., 1997; Zhang et al., 2004) or a phonologically based working memory mechanism (Nixon et al., 2004). Poldrack et al. (2001) have found that a subset of phonological processing left inferior frontal regions is also sensitive to acoustic-phonetic features. Jacquemot et al. (2003) suggested that activations in the inferior frontal regions are related to explicit extraction of abstract linguistic features, or to the segmentation of the auditory stimuli.

These results show that functional imaging studies aiming to distinguish prelexical levels successfully

showed that processing speech and non-speech has separate neural correlates (Zatorre et al., 1992; Demonet et al., 1992; Binder et al., 2000; Benson et al., 2001; Jäncke et al., 2002; Scott and Johnsrude, 2003; Zevin and McCandliss, 2005). But regions that are found to be speech-specific are often presented as findings of phonological processing areas (Poldrack et al., 2001; Burton et al., 2005; Dehaene-Lambertz et al., 2005), without making a further distinction between speech-specific processing levels, or without suggesting the possible relevance of such a separation. Other studies even consider any acoustic-change related activations as correlates of phonological representations, in case the stimuli are speech stimuli (e.g., Seki et al., 2004). Although it is mainly an issue of clarity in defining processing levels, the inconsequent use of terms often results in claims about phonological representations and phoneme categories in these studies whose design did not allow for an acoustic-phonetic versus phonological distinction. Phonological processing refers to something qualitatively more than just speech-specific processing. But the precise role of cortical areas that are claimed to be speech-specific, and whether it is possible to assign certain areas to certain prelexical, speech-specific levels, remains unclear in all these studies.

Only few neuroscientific studies, and especially few functional magnetic resonance imaging studies aimed for the distinction of phonological processing and earlier prelexical, acoustic-phonetic processing levels. Phillips et al. (2000) conducted an MEG mismatch study in which they tried to cancel out acoustic effects by grouping several acoustically different stimuli from the same phoneme category in one condition and claimed to have found that auditory cortex accesses phonological categories, although they did not claim that they had found neural correlates of phoneme categorization or phonological processing there. In a cross-linguistic design, Dehaene-Lambertz et al. (2000) exploited the mismatch negativity effect with EEG, and showed the effects of a phonological change in the brain. Using a very similar design to that of Dehaene-Lambertz et al., Jacquemot et al. (2003) carried out an fMRI experiment. In their study, French and Japanese volunteers were scanned while performing a discrimination task. Three pseudowords were presented in each trial; the first two were always identical, and the third one was either identical (acoustically the same) or different. The key of the manipulation was that while the change was phonological for one

population, it was only acoustic for the other. This way, they could subtract the activations involved in the phonological versus the acoustic discriminations. Jacquemot et al. have found superior temporal activation as a cortical response for a phonological change, and therefore they have claimed that the language-specific phonological grammar can shape the auditory cortex. Their design, however, could not entirely exclude the effects of the fact that the phonological change was a not well-controlled or well-balanced acoustic change at the same time.

To conclude, most previous neuroimaging studies to date are inadequate in that they do not allow for conclusions about the neural correlates of an acoustic-phonetic versus phonological distinction to be made. In fact, there seems to be an agreement about the mysterious nature of phonemes in the brain. As Scott and Wise (2004) have noted: “there is a lack of evidence for post-acoustic, prelexical, perceptual processing of phonemes in the functional imaging literature” (p. 21). Similarly, in a recent meta-analysis of hemodynamic studies on prelexical and lexical levels in listening, Indefrey and Cutler (2005) have argued that “at present, neuroimaging data do not allow for a distinction between phonetic and phonological processing” (p. 14).

In the present study, we aim to show that a phonological level of processing can be anatomically distinguished from earlier processing levels. More specifically, we attempt to find a functional and anatomical distinction of acoustic-phonetic processes which are sensitive to fine physical differences in the speech signal; and phonological processes which depend on the categorical identity of speech sounds. Consequentially, we aim to localize acoustic-phonetic and phonological processes and describe the differential activation pattern of the responsible cortical areas.

Vowel-context effect as a tool

Speech sound perception studies are sensitive to stimulus properties, to the precise nature of contrasts. For example, the difficulty in testing acoustic versus phonetic-phonological processing was that stimuli had to be manipulated in a way that the only difference between conditions is their perception as speech or non-speech. Testing acoustic-phonetic versus phonological processing seems to be even more difficult: one has to manipulate within-category / across-category property of stimulus-pairs without manipulating any other properties of the stimuli. For that, a strict definition of the phonological contrast is needed. This contrast should be balanced in all other, irrelevant aspects (such as the physical distance of the stimuli). At the same time, all stimuli in a phonological contrast should be contrasted to an

unambiguously identified speech sound. Thus, phonemic identity of the speech sound stimuli should be manipulated while their acoustic properties remain constant.

Our solution to the problem of finding a sensitive contrast to test phonological processing is based upon the fact that the phonemic categorization of a consonant can depend on the neighboring vowel. The phenomenon which is responsible for such vowel-context effects is referred to as compensation for coarticulation. Phonemes are coarticulated in continuous speech, and coarticulation results in context-sensitive acoustics of phonemes. Listeners compensate for the effects of coarticulation: perception of a speech sound depends both on its acoustic properties and its context, and the direction of this context-moderated perception is opposite to the effects of coarticulation (Diehl et al., 2004).

The underlying mechanisms of compensation for coarticulation, whether it is caused by general auditory processes or is phonologically mediated, were long debated. Mitterer (submitted) recently suggested that compensation mechanisms depend on the type of coarticulation. He showed evidence that certain context-effects are based in audition (see also Mitterer et al., 2003; Mitterer and Blomert, 2003), while other context-effects, mainly those in which acoustic cues are more dissimilar and thus little room is left for auditory interactions, act on the phonetic/phonemic level. One such phonetically or phonologically mediated compensation effect is the vowel-context effect with fricative-vowel syllables. In his sushi experiment, Whalen (1989) has shown with English listeners that phonemic categorization of an ambiguous fricative depends on the quality of the following vowel. Similarly, Smits (2001a) has found that Dutch listeners classify the same ambiguous fricative as [ʃ] (as in “sjaal”) in [i]-contexts but as [ʃ] (as in “saai”) in [y]-contexts. Smits has suggested that sounds in the syllable are categorized in a hierarchical manner, as a consequence of a statistical learning mechanism: the fricative identification is preceded by the vowel identification. Manipulating sine-wave speech and audio-visual speech stimuli Mitterer (submitted) has recently found supporting evidence for Smits’ conclusion that the fricative-vowel context effect depends on the phonemic identity of the vowel, and as such, it necessarily acts on a phonological processing level. We used the same fricative-vowel syllables as stimuli in our experiment.

The vowel-context effect causes a shift of the identification function, a shift of the fricative

category boundary for fricative-vowel syllables. For our purposes, the value of this late context-dependency in perception is that we can manipulate phonemic categorization of a fricative without actually changing the sound. As a consequence, we can create a “within-category versus across-category” type contrast while (1) keeping the same stimulus distance within a comparison, between the conditions; and at the same time (2) comparing each and every stimulus to an unambiguous phoneme. This way, we can define a strictly phonological contrast. Vowel-context effect also makes it possible to create a well-balanced “within-category versus identical” type contrast, so acoustic-phonetic processing can also be tested.

Adaptation-fMRI in a sparse scanning paradigm

Neural dishabituation as a sign of change detection in a neuronal population on different speech processing levels was found with EEG (and with MEG) earlier in mismatch negativity studies (e.g., Näätänen et al., 1997). A great advantage of MMN is that it can be elicited in the absence of parallel behavioural responses. Functional magnetic resonance imaging research was long lacking a method that combines the advantages of the MMN paradigm (being a similarly sensitive measure of neuronal change detection) and those of fMRI (providing a high spatial resolution). Grill-Spector and Malach (2001) have recently suggested a novel experimental paradigm, called adaptation-fMRI to study the invariant properties of specific neuronal populations to changes in the environment. Its application proceeds in two stages. As they describe: “First, the neuronal population is adapted by repeated presentation of a single stimulus. Second, some property of the stimulus is varied and the recovery from adaptation is assessed. If the signal remains adapted, it will indicate that the neurons are invariant to that attribute. However, if the fMRI signal will recover from the adapted state it would imply that the neurons are sensitive to the property that was varied.” (p. 293). The adaptation-fMRI paradigm was first used in object recognition studies, but after combining it with the technique of sparse scanning (Hall et al., 1999) in order to avoid interference with the scanner noise, auditory adaptation-fMRI paradigms (or auditory habituation paradigms) are also being developed and are used efficiently (e.g., Zevin and McCandliss, 2005). The value of adaptation-fMRI for our purposes is that it may efficiently be used as a tool to test fast neuronal responses to fine acoustic-phonetic or phonological changes on anatomically precisely defined loci in the human brain.

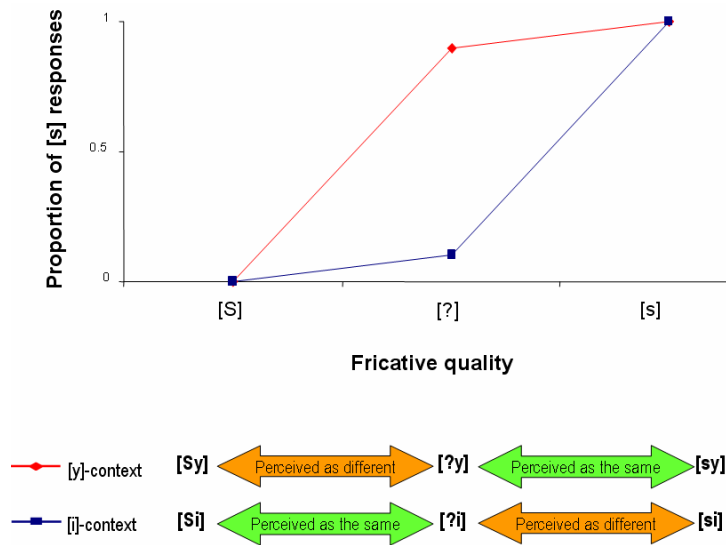


FIGURE 1 : Vowel-context effect, illustration. An ambiguous fricative [ʔ] from the [S] – [s] continuum is usually classified as [S] when heard in context of an [i] vowel, but classified as [s] when heard in context of an [y] vowel. Arrows show which syllable pairs are usually perceived as the same and which ones are usually perceived as different.

Design of the present study

In this study, we have combined the advantages of a phonological vowel-context effect and those of the auditory adaptation-fMRI paradigm. The units of our stimuli were fricative-vowel syllables. We used three fricatives: [S], [s] and an ambiguous fricative [ʔ] from the [S] – [s] continuum, each of them in [i] and [y] vowel contexts. The most ambiguous fricative [ʔ] was identified individually for each subject in a pretest in a phonemic categorisation task, using a 9-step [S] – [s] continuum.

In the fMRI experiment, pairs of fricative-vowel syllables were presented auditorily. In the test conditions we varied the relation of the members within a syllable pair: in certain trials, fricatives in the pair-members were previously identified as the same or as different phonemes. Brain responses to syllable pairs were recorded during pseudo-passive listening. Following the logic of adaptation-fMRI, we expected that if pair-members are treated as the same by a cortical area, the repeated firing of neurons will

lead to adaptation and thus to a reduction in overall activity in that area, as compared to pair-members that are treated as different by that area. We expected that already one presentation of a syllable establishes sufficient context to generate this adaptation effect. We could manipulate acoustic properties and distance of the fricative sounds while keeping the perception of members of a syllable pair unaffected. For example, fricatives of both the syllable pair Si-Si and the syllable pair Si-ʔi were perceived as the same, although the acoustic-phonetic relations were different in the two pairs. We used this manipulation for

testing the sensitivity of a cortical area to acoustic-phonetic changes.

But the key property of the individually identified ambiguous sound was that it was classified as [S] in [i]-contexts, but as [s] in [y]-contexts by the listener. And by exploiting this vowel-context effect, we could perform the inverse test as well: we could manipulate the perception of members of a syllable pair as the same or different while keeping the acoustic properties and distance of the sounds constant. For example, fricatives of the syllable pair Si-ʔi were perceived as the same, while fricatives of the syllable pair Sy-ʔy were perceived as different, although the acoustic-phonetic relations were not different in the two pairs (see Figure 1). This was our critical manipulation for testing the sensitivity of a cortical area to phonological changes.

The present study has focused on the tests of specific contrasts of four conditions, which were the following (see also Table 1):

SAME (identical pairs): pairs of syllables with perceptually unambiguous, acoustically identical fricatives, referred to as “SAME”.

CRITICAL-SAME (within-category pairs): pairs of

Conditions	Unambiguous sound compared to unambiguous	Ambiguous sound compared to unambiguous
Pair-members identified as the same phoneme	SAME: identical pairs (Si-Si, si-si, Sy-Sy, sy-sy)	CRSA: Critical-same: within-category pairs (Si-ʔi, ʔi-Si, sy-ʔy, ʔy-sy)
Pair-members identified as different phonemes	DIFF: different endpoint stimuli (Si-si, si-Si, Sy-sy, sy-Sy)	CRDI: Critical-different: across-category pairs (si-ʔi, ʔi-si, Sy-ʔy, ʔy-Sy)

TABLE 1 : Experimental conditions

syllables with a context-independently perceived (unambiguous) and a context-dependently perceived (ambiguous) fricative which are in a large proportion of trials identified identically in the pretest, referred to as “CRSA”:

CRITICAL-DIFFERENT (across-category pairs): pairs of syllables with a context-independently perceived (unambiguous) and a context-dependently perceived (ambiguous) fricative which are in a large proportion of trials identified differently in the pretest, referred to as “CRDI”.

DIFFERENT (different endpoint stimuli): pairs of syllables with perceptually unambiguous, acoustically and phonologically different fricatives, referred to as “DIFF”.

Importantly, the vowel did not change within a pair, and every pair consisted of a comparison of a given syllable to an unambiguous syllable. (We use the term “unambiguous syllable” for syllables with an endpoint fricative [S] or [s], whose classification is context-independent.)

Acoustic-phonetic processing areas were assumed to be sensitive to any physical differences of speech sounds. Phonological processing areas, or areas that code a phonologically abstract representation, were assumed to follow the behavioural pattern in making distinctions between syllables perceived as the same and syllables perceived as different. When defining the contrasts, we followed two principles: first, we used the minimal possible comparison; and second, we required that our contrasts contain the SAME condition as the only legitimate base for an interpretable comparison. For acoustic-phonetic processing, the minimal contrast was SAME (identical pairs) < CRSA (within-category pairs) – and we also expected that these areas would show less adaptation in all other conditions than in the SAME condition: indeed, it would be difficult to interpret areas that show a SAME < CRSA effect but not a SAME < CRDI or SAME < DIFF effect. For phonological processing, the minimal contrast was the conjunction of two simple contrasts: SAME (identical pairs) < CRDI (across-category pairs) and CRSA (within-category pairs) < CRDI (across-category pairs).

Following this logic, we aimed to distinguish a phonological processing level from an earlier, acoustic-phonetic processing level along two simple predictions: First, if a cortical area shows adaptation in the identical (SAME) condition as compared to all other conditions, then this area plays a role in acoustic-phonetic processing. Second, if a cortical area shows adaptation both in the identical (SAME) and within-category

(CRSA) conditions as compared to the across-category (CRDI) condition, then this area is involved in phonological processing. We also hypothesized that the areas characterized this way are both functionally distinct (thus follow different SAME-CRSA-CRDI patterns) and anatomically distinct.

Method

Participants

Thirty subjects participated in the study (19 female). None of them had participated in any of our pilot studies. All participants were right-handed, native Dutch speakers with no known hearing disorders. Mean age was 23 years (range 19-26). All participants gave informed written consent before participating in the experiment.

Materials and procedure

In a series of pilot experiments we optimized the stimulus set of fricative-vowel syllables and learned that (1) compensation for coarticulation was a strong effect when using synthetic stimuli, and that (2) there was a large intersubject variability both in phoneme categorization and in the size of the vowel-context effect. Thus we needed to know on an individual base how each subject perceives each stimulus type from a fine-grained continuum without actually asking them during the fMRI experiment. At the same time, we wanted to minimize the number of stimuli used in the experiment, in order to maximize the number of repetitions for each stimulus type. To solve this problem we ran a pretest before the fMRI experiment with each subject, in which both the unambiguity of the endpoints was checked and the most ambiguous fricative was identified individually in a phonemic categorization task. In the fMRI experiment, only these individually fine-tuned stimuli were presented.

Therefore, the experiment consisted of two parts: a behavioural pretest phase and an fMRI experiment phase. The pretest was immediately followed by the fMRI experiment.

Pretest

Materials

Syllables in the pretest consisted of a fricative from an [S] – [s] continuum and an [i] or an [y] vowel. Both the male voice vowels (mean F0 was 127 Hz) and a 9-step fricative continuum were synthesized by Holger Mitterer, using Praat 4.0 (Boersma and Weenink, 2004). The fricatives were varied in the frequency at which the most energy in the frication noise was concentrated. For an exact physical

SILENCE 370ms	SYLLABLE 1 380ms	SILENCE 370ms	SYLLABLE 2 380ms	SILENCE 370ms	SCANNING 1800ms
------------------	---------------------	------------------	---------------------	------------------	--------------------

Figure 2: Scanning with silent gaps

description of the stimulus synthesis, see Mitterer (submitted). Endpoints of the continuum were unambiguously identified as [S] and [s] in a preliminary identification test with the stimulus set by 6 native Dutch speaking subjects. Seven middle steps of the continuum were linearly distributed in a narrower interval which covered enough of the continuum to contain the most ambiguous sound as determined by 14 native Dutch speaking subjects' individually averaged responses in our pilot studies. 18 syllables were synthesized by concatenating each of the 9 fricatives with each of the 2 vowels. Length of the fricative and the vowel were 180 ms and 200 ms, respectively, adding up into 380 ms long syllables.

Procedure

Stimuli were displayed in a sound-proof room through headphones at a constant, comfortable listening level. A trial consisted of one syllable presentation and one button response. The pretest consisted of a short practice (6 trials) and the experiment with one self-terminated break after the first 8 chunks. Participants had a phoneme identification task: they were instructed to decide whether the first sound in the syllable heard was an [s] (described as “s” in the Dutch word “saai”) or an [S] (described as “sj” in the Dutch word “sjaal”). Participants had to give a button press response with their left or right index finger, according to the “s” or “sj” sign on the corresponding side. The coupling of sound and side was balanced across subjects. Participants were instructed to respond as accurately and as fast as possible. Each trial began with the stimulus presentation and lasted 2500 ms. The offset of the response-window was determined by the onset of the next trial. The pretest lasted 12 minutes.

Experimental design

Each of the 18 syllable types was presented 15 times. Trials were clustered into 15 chunks of 18, where each chunk contained each syllable once. These chunks were then randomized and ordered with the constraint that the last trial of each chunk had to be different from the first trial of the consecutive chunk. Thus, two consecutive syllables were never identical during the whole test, and the distance of two identical trials was never larger than 35, with an average distance of 18 for each syllable-type. This was to

minimize the biasing effects of task-internal probability learning. The experiment was run and data were collected by a Presentation software.

fMRI experiment

Materials

Stimuli in this test were syllable pairs. Three levels of the pretest's 9-step fricative continuum were used: the two unambiguous endpoints [S] and [s]; and one of the seven middle steps, the most ambiguous level [ʔ], which was determined individually for each participant on the base of the pretest. The fricative was followed by an [i] or an [y] vowel: either by the same vowels as the ones in the pretest, or by the gender-changed version of the same vowels (for that, PRAAT's change gender function was used, the resulting new mean F0 was 260 Hz).

Syllable pairs consisted of two 380 ms long fricative-vowel syllables, separated by a 370 ms long silent gap. Trial types and conditions were defined on the basis of the relation of the two syllables within a pair. Test trials and their types, distractor trials and their types, and silent trials were distinguished. In silent trials, both syllables were substituted by a 380 ms long silence. The vowel quality did not change within a pair in any of the trial types: in test trials, vowels from the pretest were used and they were identical in the two syllables; in distractor trials, the vowel of the first syllable was one from the pretest, while the vowel of the second syllable was the gender-changed version of the same vowel quality.

Test trial types varied in vowel quality across pairs and in the relation of the fricatives within a pair. Given that there were 3 fricatives, there were 9 possible combinations of fricative-fricative couplings for each vowel, so the 2 vowel contexts resulted in 18 different types of syllable pairs. This way every syllable-pair appeared equally often with both possible orderings (e.g., Si-ʔi and ʔi-Si). Each of these 18 types were repeated 35 times during the experiment.

Distractor trial types, similarly to the test trial types varied in vowel quality across pairs and in the relation of the fricatives within a pair. Only the two unambiguous fricatives were used in these trials, the 4 possible fricative-fricative couplings and the 2 vowel contexts resulted in 8 different types. Each of these 8 types was repeated 10 times during the experiment.

The silent trial was repeated 70 times during the experiment.

Procedure

Participants had a distractor task during scanning: they were instructed to listen to the syllable pairs and to press a button only when the second syllable is uttered by a different speaker, a female voice. Participants were instructed to respond with the left index finger. There was a short practice session after the pretest.

Button response was expected for 80 out of 780 trials (10%), with an average time-gap of 36 s between 2 distractor trials. There was a 22 second (6 pulses) long break after every 9,5 minutes (two chunks). During the break the scanner was on, but no stimuli were presented. Participants were informed about the beginning and the end of the break visually.

Experimental design.

We applied a rapid event-related paradigm. But in order to increase the power of the statistical analyses, most of the test trials and the silent trials were presented in clusters of 3, meaning that the same syllable pair or the same silent trial was presented three times in a row. In order to reduce predictability of the consecutive trials on the other hand, one out of 7 test trials and one out of 7 silent trials were presented non-clustered. Thus, 3-clusters of a given test or silent trial type appeared twice as often as their non-clustered variant, and 86% of a given type appeared in 3-clusters. Distractor trials always appeared non-clustered. The experiment consisted of 780 trials.

To minimize probability learning effects and the effects of attentional fluctuation, an equal distribution of trial types was ensured throughout the experiment. Furthermore, for ordering purposes, trials were grouped into five categories: same, critical, different, distractor and silent categories. Trial ordering was based on these categories. Units (3-clusters and non-clustered trials) were randomized with the restriction that no 2 consecutive units came from the same category.

The clustering, equal distribution and restricted randomization of trials were done in order to create a well-balanced, powerful, rapid event-related design.

fMRI scanning parameters

Whole-brain functional images were acquired on a 3-Tesla MRI system (Siemens TRIO). Using a gradient echo planar scanning sequence, 26 axial slices were obtained for each subject (voxel size $3,5 \times 3,5 \times 4 \text{ mm}^3$, field of view = 224 mm, TA

= 1800 ms, TR = 3670 ms, TE = 35 ms, flip angle = 70°). To avoid interference with the scanner noise, we used the technique of sparse scanning (Hall et al., 1999). All stimuli were presented between volume acquisitions, during silence (see Figure 2).

We expected that the BOLD signal would peak between 4 and 6 sec, and thus we assumed that the largest proportion of activation corresponding to a given trial would be acquired in the second volume acquisition following a syllable pair (the volume acquisition between 4040 ms and 5840 ms after the beginning of the second syllable within a pair).

All functional images were acquired in one run that lasted 50 minutes and contained 820 pulses, including 3 dummy scans at the beginning and 5 extra scans at the end of the run. After the acquisition of functional images, a high-resolution anatomical scan (T1-weighted MP-RAGE, 192 slices) was acquired.

Data analysis

Pretest

Boundary determination

The most ambiguous sound of the fricative continuum (i.e., the category boundary) was determined individually, on the basis of the fricative identification curves for the two vowels (see Figure 3 for an example for an individual identification curve). The selection algorithm had five steps, which were applied sequentially, but the algorithm terminated immediately when the boundary level was determined. The five consecutive steps were the following:

(1) Normally, that sound level was chosen for which the [S] responses were the closest to 50%, for vowel-contexts collapsed.

(2) In ambiguous cases, that level was chosen for which the neighbouring levels were more ambiguous, more precisely that one of the competing ambiguous levels for which the least ambiguous, non-shared neighboring level had responses closer to 50%, for vowel-contexts collapsed.

(3) In still ambiguous cases, that level was chosen for which the difference of the identification curves for the two vowel-contexts was greater.

(4) In still ambiguous cases, the level closer to the middle of the continuum was chosen.

(5) In still ambiguous cases, one of the competing levels was chosen randomly.

To have a sensitive tool for checking the category boundary, based on our pilot studies the differences between neighbours of the seven middle steps of the fricative continuum were set rather small. However, it was theoretically possible that a subject's fricative

category boundary falls between two stimulus levels, and the identification curve is so steep that both levels neighbouring the hypothesized position of this category boundary are too far, and the perceptual balance of the stimulus set is not ensured by choosing any of the levels. For this reason, an inclusion criterion was set to check whether our pretest continuum was fine-grained enough for the participant's perceptual system.

Furthermore, we expected great intersubject variability in the position of the category boundary and in the size of the vowel-context effect. The selection algorithm of the most ambiguous sound was optimized firstly for finding the most ambiguously identified fricative level, for vowel-contexts collapsed; and not, or only secondarily for finding the fricative level with the greatest identification difference between the two vowel-contexts. This choice was made to ensure a perceptually balanced stimulus set for the fMRI experiment.

But on the other hand the manipulation in the fMRI experiment was sensitively based on the vowel-context effect, and thus the presence of unambiguously, context-independently and ambiguously, context-dependently perceived fricatives was going to be critically important there. For this reason, two further inclusion criteria were set: a criterion to check whether endpoint sounds were indeed unambiguously identified in each vowel-context; and another criterion to check whether the vowel-context effect is strong enough for the selected ambiguous level.

Only those subjects entered the fMRI data analysis whose pretest responses to the selected stimuli had passed all three criteria.

Inclusion criteria

Criterion 1 - fine-grained continuum: the percentage of [s] decisions at the boundary level is between 33% and 66%, for the vowels collapsed.

Criterion 2 - unambiguity of phonemic decision at the extremes: the percentage of [s] decisions at level 0 ([S] endpoint) is less than 40%, at level 8 ([s] endpoint) is greater than 60%, for the vowels collapsed; the percentage difference of [s] decisions at level 0 and level 8 is greater than 33%; and both endpoints receive decisions that are less ambiguous than the boundary level.

Criterion 3 - vowel-context effect: either at the boundary level or at one of its non-endpoint neighbouring levels the percentage of [s] decisions for vowel [y] is at least 25% greater than for vowel [i].

fMRI experiment

fMRI image preprocessing

fMRI data were analyzed with BrainVoyager QX 1.2.6 and 1.3.8 (Brain Innovation). Functional images were corrected for motion and slice time acquisition. Functional images were coregistered with the anatomical scan and transformed into Talaraich coordinate space (Talaraich and Tournoux, 1988). Images were spatially smoothed with a full-width at half-maximum (FWHM) Gaussian kernel of 6 mm.

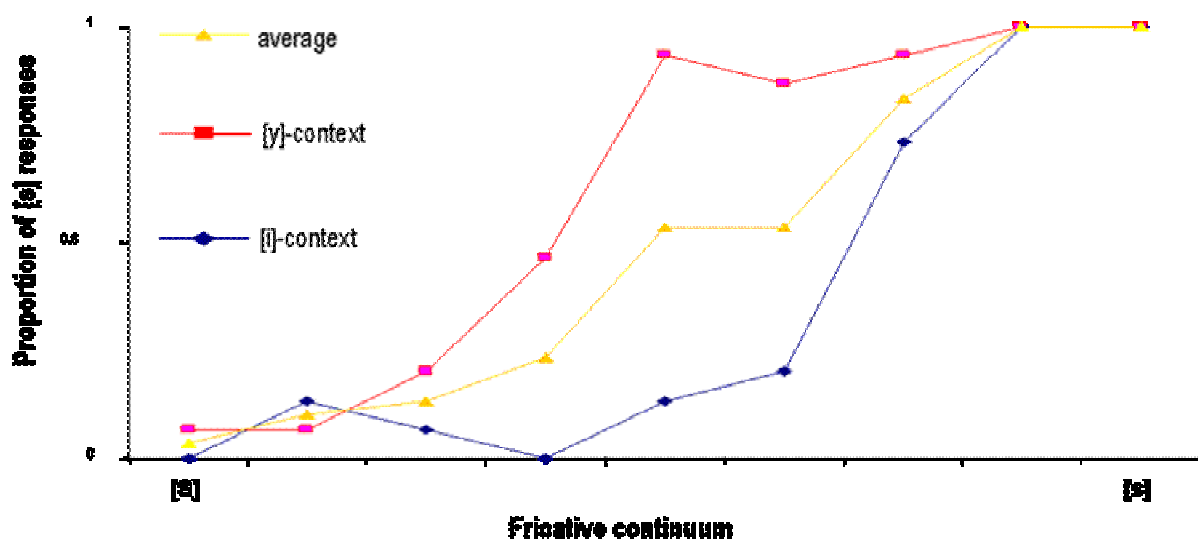


FIGURE 3: Vowel-context effect for an individual subject. Responses in the phoneme identification task (pretest) for fricatives from an [S] – [s] continuum in [i] and [y] contexts.

Statistical analyses

Statistical analyses were performed in the context of the general linear model. Regions of interest were defined based on fixed effects analyses. Maximum cluster spread range was 25 mm. Single contrast analysis was performed with a statistical threshold set at $P < 0.001$ at the voxel level, uncorrected for multiple comparisons. Conjunction analysis of two contrasts for fixed effects was performed with a standard “minimal t-statistic” approach (Nichols et al., in press) that BrainVoyager uses, which is equivalent to a logical AND of the contrasts at the voxel level. In order to get a statistical threshold for the conjunction analysis that is comparable to our single contrast analysis, we estimated the probability of finding a voxel that is independently significant in each and both contrasts (i.e., the joint probability), by multiplying the probabilities for each contrast (e.g., Allan et al., 2000; Cabeza et al., 2002). Thus, both contrasts in the conjunction were

assigned a threshold set at $P < 0.033$, the square-root of $P < 0.001$.

Beta weights (regression coefficients) as indices of effect size were then obtained for all voxels included in these regions of interest, per subject and per trial type. These beta weights, averaged first regionally and then across those trial types that were collapsed within one condition in a specific contrast, were normalized for all conditions (that is, the average of the beta weights for all conditions was set as baseline, it was assigned the weight zero, and the set of all conditions was linearly shifted accordingly) per ROI and per subject. Random effects analyses were performed applying t-tests to the averaged and normalized beta weights, with a threshold set at $P < 0.05$.

Only regions that were defined in a fixed effects analysis and then passed the t-test for a specific contrast are reported. To reduce the risk of reporting false positive activations, only regions including more than 10 contiguous voxels were considered.

Fricative level	[i]-context	[y]-context	average	difference between vowel-contexts
level 0	0.049	0.137	0.093	0.088
level 1	0.080	0.341	0.211	0.261
level 2	0.096	0.509	0.303	0.413
level 3	0.126	0.565	0.346	0.439
level 4	0.197	0.726	0.462	0.529
level 5	0.324	0.813	0.569	0.488
level 6	0.478	0.864	0.671	0.386
level 7	0.707	0.915	0.811	0.208
level 8	0.928	0.979	0.953	0.051

TABLE 2: Proportion of [s] decisions (25 subjects)

Fricative level	[i]-context	[y]-context	average	difference between vowel-contexts	number of included subjects
boundary -6	0.000	0.033	0.017	0.033	2
boundary -5	0.028	0.156	0.092	0.128	12
boundary -4	0.061	0.245	0.153	0.183	23
boundary -3	0.075	0.410	0.242	0.334	23
boundary -2	0.106	0.511	0.308	0.406	24
boundary -1	0.155	0.640	0.369	0.485	25
boundary	0.221	0.781	0.501	0.560	25
boundary +1	0.432	0.851	0.649	0.419	25
boundary +2	0.624	0.891	0.757	0.266	25
boundary +3	0.814	0.947	0.868	0.132	23
boundary +4	0.856	0.990	0.923	0.133	13
boundary +5	0.500	1.000	0.633	0.733	2
boundary +6	0.667	1.000	0.833	0.333	2
boundary +7	0.600	0.867	0.733	0.267	1

TABLE 3: Proportion of [s] decisions, levels centered around the selected boundary level for each subject (25 subjects)

Results

Pretest results

All subjects passed the “fine-grained continuum” test and the “unambiguous endpoints” test, but 5 subjects failed on the “vowel-context effect” test. Therefore, 25 out of the 30 subjects passed all 3 inclusion criteria. Our inclusion criteria were quite liberal. In fact, half of the subjects would have passed much stricter criteria as well: 15 out of the 30 participants had a boundary level with 50 +/- 10% [s] decisions (fine-grained continuum), had identified the endpoint in each vowel-context with an unambiguity greater than 80% (unambiguous endpoints) and finally either at the boundary level or at one of its non-endpoint neighbouring levels the percentage of [s] decisions for vowel [y] was at least 33% greater than for vowel [i] (vowel-context effect). Figure 3 shows the vowel-context effect for an individual subject.

The selection of the fricative category boundary level, which could have been between level 1 and level 7 from the fricative continuum (level 0 and level 8 were the endpoints) showed little variance (mean = 4.5; sd = 1.01) for the 25 subjects who passed the inclusion criteria. However, picking a middle step from the continuum as a standard boundary level for all

participants would have clearly weakened the phonetic balance of the stimulus set used in the fMRI experiment. Choosing level 4 as the boundary level for all participants would have resulted in 46% [s] decisions; choosing level 5 for all would have resulted in 57% [s] decisions at that level. By applying individual fine-tuning, the overall percentage of [s] decisions was 50% at the individually selected boundary level, thus it was indeed ambiguous.

As a secondary effect, individual fine-tuning also increased the context-effect at the selected ambiguous level: the percentage difference of [s] decisions for the two vowels was higher (56%) at the individually selected ambiguous level as compared to the hypothetical case of choosing level 4 (53%) or level 5 (49%) of the continuum as the ambiguous sound for all participants. (See the Tables 2 and 3 for an overview of the pretest results.)

The context-effect at the selected ambiguous level was strong, behavioural responses to [ʔi] and to [ʔy] were highly different (see Figure 4). This confirmed that the critical CRSA and CRDI conditions are indeed different.

fMRI results

Performance data

Two subjects had to be excluded because of movement in the fMRI scanner, a third subject was

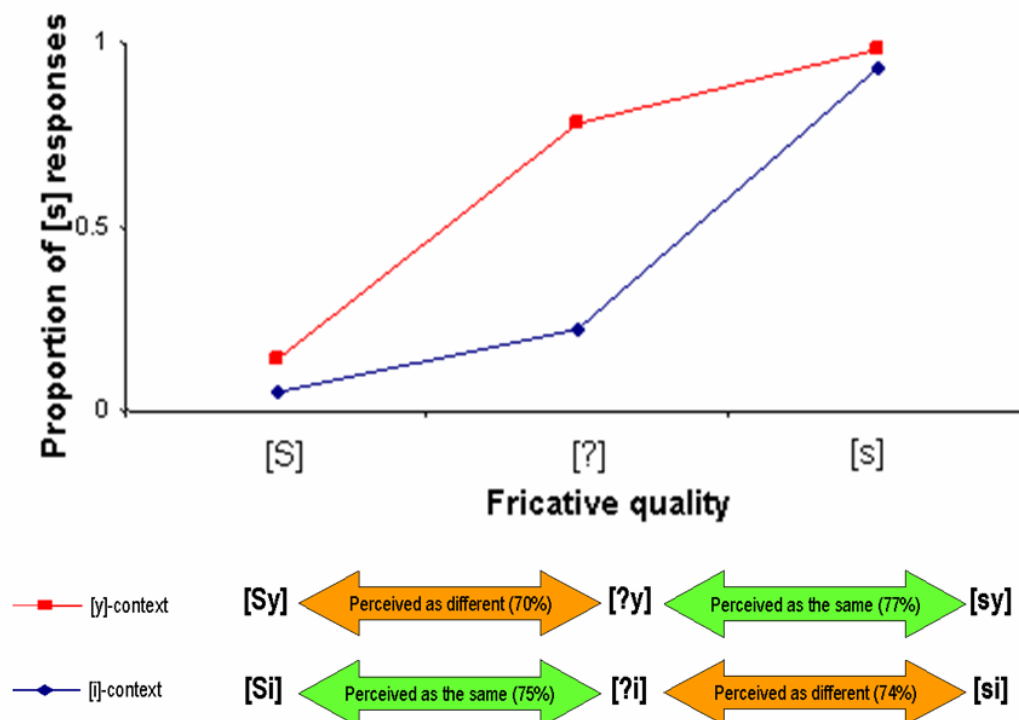
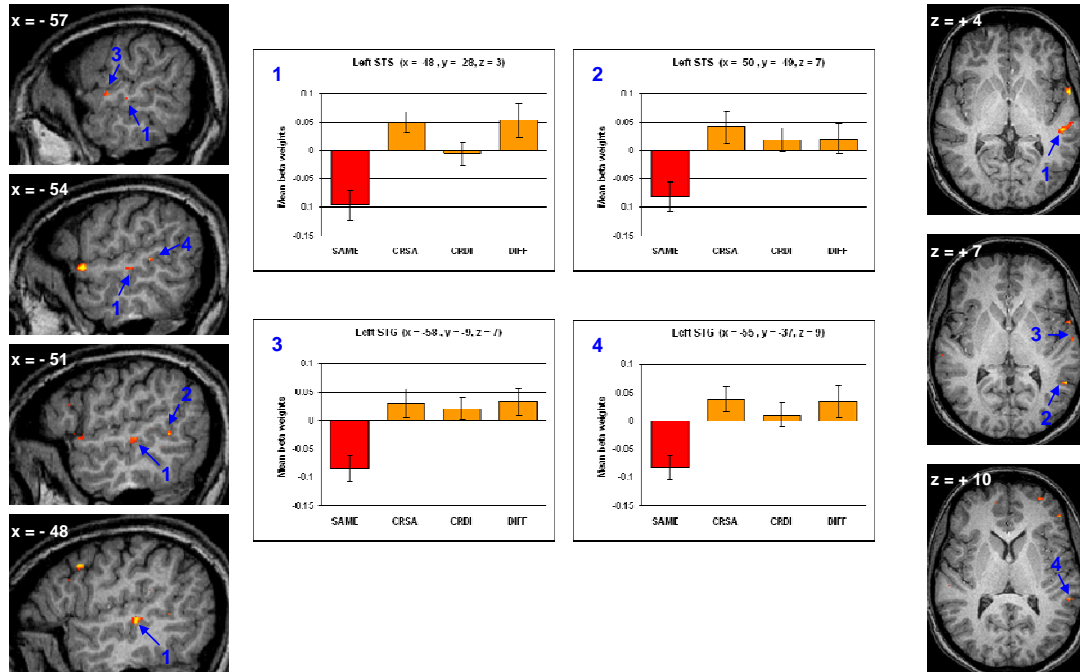


FIGURE 4: Vowel-context effect, group average (25 subjects)

FIGURE 5. Four left superior temporal regions showing an acoustic-phonetic pattern. Sagittal and axial maps showing significant clusters ($P < 0.001$) of the fixed effects analysis superimposed on a representative brain. Bar diagrams indicate mean beta weights and standard errors for the four conditions. Color coding of the bars signals significant comparisons between the conditions.

█ < █ ($p < 0.05$ for each pair of conditions)



excluded because his scanning was accidentally interrupted. So in the context of the group analyses below, the fMRI data of 22 subjects are presented.

Participants performed well on the voice-change detection task. Only one subject made more than one false alarms, 17 out of the 22 subjects made no false alarms during the whole fMRI experiment. The percentage of false alarms was below 0.7%.

This shows that the task was indeed very easy.

Acoustic-phonetic test

The first test aimed at identifying cortical areas that are involved in the acoustic-phonetic

processing of speech sounds. We assumed that these areas have undergone habituation in the SAME condition, caused by the repetition of the syllable, while remained dishabituated in the CRSA, CRDI, DIFF conditions. To focus on an acoustic-phonetic level of processing, a minimal (SAME < CRSA) contrast was used: regions being significantly less activated when the fricatives in a syllable pair were physically the same opposed to when they were physically different but perceived as the same fricatives, were determined in a single contrast fixed effects analysis. Paired-samples t-tests – a random effects analysis – were then applied on these regions (SAME < CRSA, SAME < CRDI, SAME < DIFF) to filter those with an interpretable acoustic-

Acoustic-phonetic processing	voxels	Talx	Taly	Talz	BA	SA<CS	SA<CD	SA<DI
L STS	129	-48	-28	3	22	.001	.029	.005
L STS	16	-50	-49	7	22	.012	.017	.017
L STG (anterior)	14	-58	-9	7	22	.016	.003	.004
L STG	15	-55	-37	9	22	.000	.011	.016
L middle frontal gyrus	87	-30	11	38	8	.000	.012	.001
L superior frontal gyrus	29	-8	10	49	6	.000	.012	.012

(Talx, Taly, Talz refers to the Talaraich coordinates, BA refers to the Brodmann areas. The last three columns show significance levels in the paired-samples t-tests for the comparison of the denoted conditions. SA: same, CS: critical-same, CD: critical-different; D: different. See the text for details.)

TABLE 4: List of brain areas, acoustic-phonetic test

phonetic pattern as characterized by our four conditions.

Areas following an acoustic-phonetic pattern (SAME < CRSA, CRDI, DIFF) were all left lateralized and included regions of the superior temporal gyrus and the superior temporal sulcus (BA 22), the middle frontal gyrus (BA 8) and the superior frontal gyrus (BA 6). See Table 4 for details.

Figure 5 shows the similar acoustic-phonetic patterns of four clusters from the left superior temporal sulcus and gyrus.

Phonological test

The second test aimed at identifying the regions involved in phonological processing. We assumed that these areas work with discrete, categorical representations of speech sounds, and are insensitive to within-category variations. Thus, a similar level of adaptation and therefore a similar amount of activation decrease was expected in the SAME and CRSA conditions as compared to the CRDI condition in all phonological processing regions. To make the determination of the regions of interest specific, a conjunction analysis of the minimal contrast (CRSA < CRDI) and the contrast to the no-change condition (SAME < CRDI) was performed for fixed effects. Then the behaviour of each identified area was further checked in t-tests (SAME < CRDI, CRSA < CRDI).

Areas showing a phonological pattern as characterized by these three conditions, were the following: the left superior temporal gyrus (BA 22), the right anterior superior temporal gyrus (BA 38), the right inferior temporal gyrus (BA

20), the right fusiform gyrus (BA 20), the left inferior frontal gyrus (BA 9), regions in the right inferior frontal gyrus, the right superior frontal gyrus, regions in the medial frontal cortex and in the basal ganglia. See Table 5 for details.

Interaction of acoustic-phonetic and phonological patterns

No regions passed both the acoustic-phonetic and the phonological tests.

The SAME-CRSA-CRDI pattern for each ROI from the acoustic-phonetic test was paired with the SAME-CRSA-CRDI pattern for each ROI from the phonological test, resulting in 270 ROI-pairs altogether. These ROI-pairs entered a repeated measures ANOVA in which the interaction of the two main effects (effect of test, effect of condition) was tested.

85% of all ROI-pairs showed a significant interaction ($p < 0.05$), and 96% of all ROI-pairs showed a trend for an interaction ($p < 0.1$). Thus, on a global level activation patterns of areas from the two tests are distinct.

Specific interactions of neighbouring areas determined in different tests were also tested. We identified superior temporal areas, close to the superior temporal sulcus with an acoustic-phonetic pattern; and areas with a phonological pattern in the left superior temporal sulcus and in the right anterior superior temporal gyrus. These subregions were distinct both anatomically and functionally. For instance, Figure 6 shows the two largest left superior temporal area in each contrast (both from BA 22). Both areas followed the curve of the superior temporal sulcus, the area in the phonological contrast was slightly more anterior. Their activation

<i>Phonological processing</i>	voxels	Talx	Taly	Talz	BA	SA<CD	CS<CD	SA<DI	CS<DI
L STS	75	-45	-23	-8	22	.017	.024	.365	.426
R STG (anterior)	57	44	12	-11	38	.009	.006	.205	.339
R inferior temporal gyrus	23	40	-5	-20	20	.021	.005	.946	.783
R fusiform gyrus	20	36	-21	-14	37	.003	.045	.740	.733
L inferior frontal gyrus	184	-39	9	30	9	.017	.011	.108	.272
R inferior frontal gyrus	24	39	27	-11	47	.006	.050	.139	.331
R inferior frontal gyrus	21	32	20	16	45/46	.002	.039	.015	.019
R inferior frontal gyrus	14	42	12	16	45/46	.011	.022	.075	.015
R superior frontal gyrus	44	16	34	46	8	.029	.002	.126	.028
Medial frontal cortex	72	3	23	-11	25	.009	.001	.064	.126
Medial frontal cortex	58	3	14	1	25	.032	.004	.047	.039
Medial frontal cortex	50	5	-11	27	23	.009	.013	.057	.021
Basal ganglia	38	28	-15	26		.026	.018	.779	.763
Basal ganglia	36	26	2	22		.001	.006	.126	.129

(Talx, Taly, Talz refers to the Talaraich coordinates, BA refers to the Brodmann areas. The last four columns show significance levels in the paired-samples t-tests for the comparison of the denoted conditions. SA: same, CS: critical-same, CD: critical-different; D: different. See the text for details.)

TABLE 5: List of brain areas, phonological test

patterns showed a significant interaction ($p < 0.01$).

We also found near sites of the left inferior/middle frontal gyri that were functioning differentially. Figure 7 shows two larger inferior/middle frontal areas: a cluster from the left middle frontal gyrus (BA 8) shows an acoustic-phonetic pattern, while a cluster on the superior part of the left inferior frontal gyrus (BA 9) shows a phonological pattern. Their activation patterns also showed a significant interaction ($p < 0.02$).

Discussion

Acoustic-phonetic and phonological contrasts

An overwhelming majority of neuroimaging studies of prelexical processing was aiming to make a distinction between not speech-specific (acoustic) and speech-specific (phonetic-phonological) processing levels, but were not designed to separate a phonological processing level. Our experiment was designed to make a distinction between acoustic-phonetic and phonological processing levels.

The acoustic-phonetic test (SAME < CRSA) had a controlling function: it was designed to check for the effects of all non-phonological changes in our speech stimuli. On the first level of group analyses we made a minimal comparison and contrasted syllable pairs with an acoustic-phonetic (but not phonological) change (CRSA) to unambiguous syllable pairs with no change (SAME). If a cortical area is involved in acoustic or phonetic processes, then it should be sensitive to fine changes of speech sounds and thus it should be differentially activated by the two conditions. We expected neuronal adaptation and thus a decreased level of activity in the identical condition.

For the phonological test we applied a strict and minimal critical contrast (CRSA < CRDI) and combined it with a comparison to the no-change condition (SAME < CRDI) in a conservative conjunction analysis (Nichols et al., in press) on the first level of group analyses. Below I discuss the validity of this test in more detail.

The critical contrast (CRSA < CRDI) was indeed a minimal comparison. In all syllable pairs of the contrast, the same ambiguous fricative sound was coupled with one of the two unambiguous fricative sounds. Thus, acoustic properties of the displayed stimuli and physical distances of the fricatives within a syllable pair were identical in the two conditions. If a cortical area is involved in acoustic processes, then it should not be differentially activated by the two conditions. Inversely, if an area is differentially activated by the two conditions, then there is good reason to assume that this area is involved in a process which makes a difference between certain groups of syllable pairs in synchrony with our grouping into the two conditions, along more abstract than acoustic dimensions.

Further characterizing the syllables in the critical contrast, it is important to point out that the ambiguous fricative sound was indeed ambiguous, it was identified as [s] or as [ʃ] in exactly 50% of all cases, also as a result of the individual fine-tuning of the critical stimulus at the category boundary. That is, phonetic properties of the displayed stimuli were balanced between the two conditions. If a cortical area is involved in phonetic (speech-specific, but not yet phonological) processes, then again, this area should not be differentially activated by the two conditions. That is, if an area is differentially activated, then we can assume its involvement in a process which differentiates syllable pairs along other than acoustic or phonetic dimensions.

However, there was a clear difference between CRSA and CRDI conditions. Fricatives in the CRSA condition were identified as the same phonemes, while fricatives in the CRDI condition were identified as different phonemes. It was the vowel-context effect on the fricatives that made this crucial difference. As we have seen, there is evidence that this specific compensation effect acts in a late phonetic phase or more probably on a phonological level, since phonemic categorization of the fricative seems to be preceded by the phonemic categorization of the vowel (Smits, 2001a; Mitterer, submitted). Since this is the effect that makes our two critical conditions contrastive, we can suppose that only dimensions higher than the dimension of the vowel-context effect can be causally involved in eliciting differential cortical responses to CRSA and CRDI conditions.

FIGURE 6. Regions in the left STS with different activation patterns. Sagittal, coronal and axial maps showing significant clusters ($P < 0.001$) of the fixed effects analysis superimposed on a representative brain. Yellow/red: acoustic-phonetic contrast. Light/dark green: phonological contrast. A: area with acoustic-phonetic pattern. P: area with phonological pattern. Bar diagrams indicate mean beta weights and standard errors for the four conditions. Color coding of the bars signals significant comparisons between the conditions.

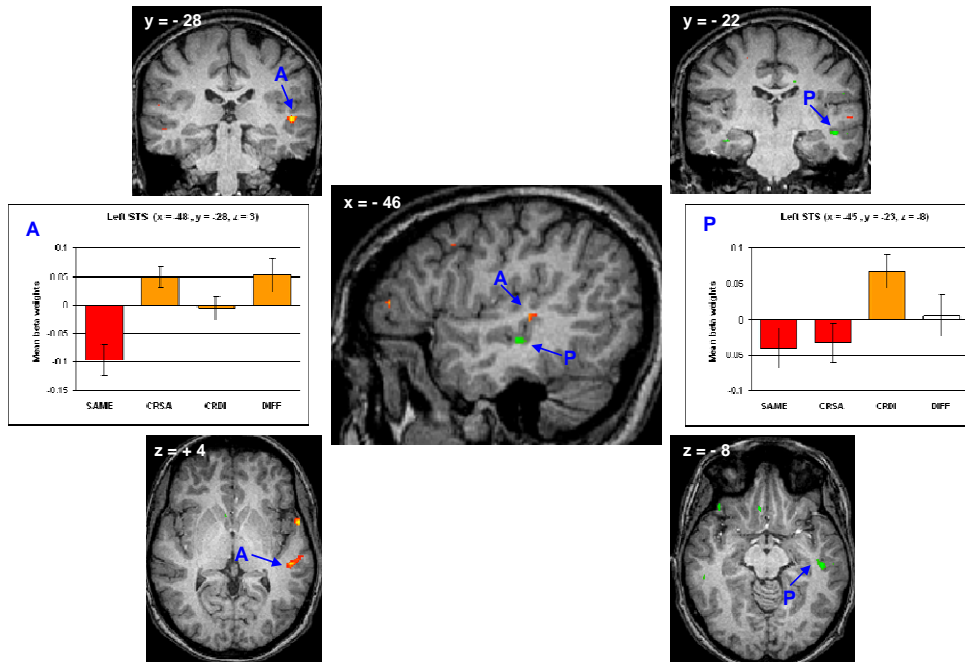
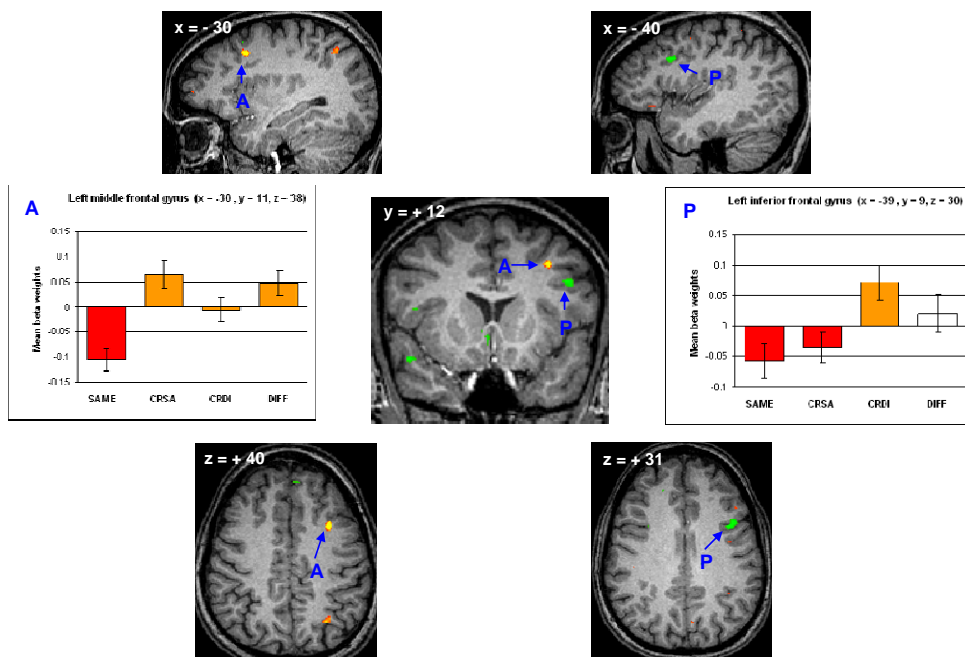


FIGURE 7. Regions in the left inferior and middle frontal gyrus with different activation patterns. Sagittal, coronal and axial maps showing significant clusters ($P < 0.001$) of the fixed effects analysis superimposed on a representative brain. Yellow/red: acoustic-phonetic contrast. Light/dark green: phonological contrast. A: area with acoustic-phonetic pattern. P: area with phonological pattern. Bar diagrams indicate mean beta weights and standard errors for the four conditions. Color coding of the bars signals significant comparisons between the conditions.



It has to be noted that unambiguous speech sounds played a central role in our study: every syllable pair presentation consisted of a comparison of a certain speech sound to a context-independently unambiguously identified speech sound. In case of the critical contrast, it increased the chance that although our comparison was strict, the resulting findings remained relevant for speech. To further reduce the possible danger of reporting artifacts resulting from contrasting activations in different conditions instead of contrasting them to any sort of baseline activation, we considered the critical contrast in conjunction with a second contrast (SAME < CRDI) in the phonological test. This contrast ensured that only such areas were characterized as areas following a phonological pattern which also survived a comparison to the pair of identical, unambiguous syllables, meaning that the activation difference observed for that area indeed reflects a response to a change versus no-change relevant in normal speech processing.

On the basis of the above listed reasons we suggest that activation increase for the critical-different (CRDI) condition as compared to both the critical-same (CRSA) and identical (SAME) conditions at a certain cortical area would be caused by the dishabituation of a group of neurons as a response to a phonologically (but not acoustically or phonetically) relevant change. All in all, we defined a strictly phonological contrast, where the acoustic properties were identical and the phonetic properties were balanced in the conditions; and all the stimuli were compared to an unambiguous phoneme (instead of simply being compared to each other). We claim that our phonological test was sufficiently strict and had the potential to reveal the neural correlates of phonological processing.

Cortical regions in prelexical processing

We showed that certain subregions of the superior temporal gyrus and sulcus seem to have a role in acoustic-phonetic processing of speech sounds; while other subregions here seem to be involved in phonological processing. This way, we found supporting evidence for the claims that there are phonetic processing areas (e.g., Demonet et al., 1992) and also for the claims that there are phonological processing areas in the superior temporal gyrus (e.g., Jacquemot et al., 2003), with the important note that these are highly specialized, differentially functioning subregions.

Similarly, we have found differentially functioning, adjacent regions in the left inferior and middle frontal gyri, suggesting that certain subregions here play a role in acoustic or phonetic processing, while others in phonological processing. This result is in line with Poldrack et al. (2001) who have found that a subset of phonological processing left inferior frontal regions is also sensitive to acoustic-phonetic features. Importantly, we also found a clear anatomical separation.

The supramarginal gyrus was often found in studies of phonetic or phonological processing (e.g., Benson et al., 2001; Golestani and Zatorre, 2004; Jacquemot et al., 2003; Dehaene-Lambertz et al., 2005). Here we found no differential supramarginal activity in any of the tests. One reason might be that none of those studies used fricatives as critical stimuli. It can be that the supramarginal gyrus is involved in the phonological processing of certain speech sound categories only.

Distinct processing levels

It is theoretically possible that areas showing an adaptation effect in the phonological test (decreased activity in certain conditions as compared to other conditions) also show an adaptation effect in the acoustic-phonetic test. The finding of cortical areas that show the expected adaptation effect in both tests; or finding that a considerable proportion of the areas that showed the effect in one of the tests follow a similar pattern, would question the relevance and plausibility of separating speech processing levels. We expected, however, that this would not be the case, for two independent, complementary reasons. First, our phonological test was supposed to be free of acoustic-phonetic confounds; and second, the phonological processing level was characterized by its insensitivity to within-category differences (that is, differences between the SAME and CRSA conditions).

Our expectations were met. There were several regions showing the expected adaptation effect in each contrast, but no regions passed both the acoustic-phonetic and the phonological tests. That is, the processing levels described by our tests are anatomically distinct. Also, we found significant interaction between the regions with acoustic-phonetic pattern and the regions with phonological pattern, and it clearly suggested that the acoustic-phonetic and phonological processing levels as characterized by the activation patterns are functionally distinct. The functional distinction of areas determined in our two tests also suggested that in the CRSA and SAME conditions phonological areas did not show a significantly different level of activation, that is, within-category contrasts were lost. It is, according to Phillips (2001), a defining

property of phonological categories. In short, these results mean that it is plausible to assume that separate speech-specific processing levels exist; they have neural correlates with separate localizations and can be characterized by different neural activation patterns.

On the basis of our results we can distinguish two levels of speech sound processing. The first level is clearly an earlier than phonological level: when being processed by these areas, speech sounds are not yet categorized phonologically. This is a prelexical, acoustic-phonetic processing level. The second level is a clearly phonological level: when being processed by these areas, speech sounds are either being categorized or are already categorized phonologically. This is the level of phonological processing. Although there is no logical reason to assume the involvement of lexical processes, it may be that some regions what we have identified as phonological processing areas may reflect lexical or postlexical stages. But it is reasonable to suppose that at least some of the phonological processing regions are involved in prelexical processing, given that compensation for coarticulation and hence the vowel-context effect are presumably learned so that they can assist in speech decoding during word recognition.

It is important to point out what we do claim and what we do not claim on the basis of this distinction concerning phonemic representations. We claim that our second level is involved in prelexical processing; and also that our second level is involved in phonological categorization. We do not claim, however, that there is a distinct prelexical level in speech processing that represents discrete phoneme categories.

What we suggest, in short, is that the separate localization of the neural correlates of acoustic-phonetic processing and phonological processing can be better described by our study than by those, which did not define a phonological processing level as strict as we did here. Nevertheless, to specify more precisely what various cortical areas underlying acoustic-phonetic and phonological processing are actually doing, and to clarify what further prelexical processing stages can be distinguished, will be subject of future research.

Methodological notes

Our study has demonstrated that the novel combination of a phonetic phenomenon (vowel-context effect) and a neuronal phenomenon (adaptation) resulted in an effective and

powerful paradigm. We have also shown that the adaptation-fMRI method works efficiently with item-trains which are only 2 items long.

Individual fine-tuning has proven to be very useful in our study: it has significantly increased the behavioural effect that was exploited in the fMRI scanner. Speech perception studies that did not take intersubject variability into consideration, might fail to draw proper conclusions because of the decreased signal-to-noise ratio originated in the stimulus set and because of the ineffective exclusion of confounding effects in the design.

Acknowledgements

I thank my supervisors for their careful, promoting and demanding guidance – I have gained much more during our discussions than what this paper can reflect. I thank my fellow students in the master's programme for sharing the pioneering experience. I also thank Holger Mitterer for his help with the speech sound stimuli. Finally, I thank my wife, Agnes for her patience and her continuous support, and my family in Hungary for bearing the lack of me.

References

- Ahmad, Z., Balsamo, L. M., Sachs, B. C., Xu, B., and Gaillard, W. D. (2003). Auditory comprehension of language in young children. Neural networks identified with fMRI. *Neurology*, 60, 1598-1605.
- Allan, K., Dolan, R. J., Fletcher, P. C., and Rugg, M. D. (2000). The role of the right anterior prefrontal cortex in episodic retrieval. *NeuroImage*, 11, 217-227.
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., and Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*, 10, 417-429.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., and Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78, 364-396.
- Binder, J., Frost, J. A., Hammeke, T. A., Bellowgan, P., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex*, 10, 512-528.
- Boersma, P., and Weenink, D. (2004). Praat: doing phonetics by computer (Version 4.0) [Computer program]. Retrieved September 9, 2004, from <http://www.praat.org/>
- Burton, M. W., LoCasto, P. C., Krebs-Noble, D., and Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *NeuroImage*.
- Cabeza, R., Dolcos, F., Graham, R., and Nyberg, L. (2002). Similarities and differences in the neural

- correlates of episodic memory retrieval and working memory. *NeuroImage*, 16, 317-330.
- Callan, D., Jones, J., Callan, A., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, 22, 1182-1194.
- Cheesman, M. F., and Greenwood, K. G. (1995). Selective adaptation by context-conditioned fricatives. *Journal of Acoustical Society of America*, 97(1), 531-538.
- Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23, 3423-3431.
- 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., Dehaene, S., and Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298, 2013-2015.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech mode. *NeuroImage*, 24(1), 21-33.
- Demonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115, 1753-1768.
- Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annual Review Psychol.*, 55, 149-179.
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Saththamnuwong, N., Tong, Y., and Li, X. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience*, 7, 1076-1087.
- Golestani, N., and Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage*, 21, 494-506.
- Grill-Spector, K., and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)*, 107, 293-321.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., and Bowtell, R. W. (1999). „Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, 7(3), 213-223.
- Hsieh, L., Gandour, J., Wong, D., and Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang*, 3, 227-252.
- Hugdahl, K., Thomsen, T., Erslund, L., Rimol, L. M., and Niemi, J. (2003). The effects of attention on speech perception: an fMRI study. *Brain and Language*, 85, 37-48.
- Indefrey, P., and Cutler, A. (2005). Prelexical and lexical processing in listening. In: Gazzaniga, M. (ed.) *The Cognitive Neurosciences - 3rd Edition*. Cambridge, MA: MIT Press.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, 23(29), 9541-9546.
- Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H.-J. (2002). Phonetic perception and the temporal cortex. *NeuroImage*, 15, 733-746.
- Joanisse, M. F., and Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *NeuroImage*, 19, 64-79.
- Lieberman, A. L. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.
- Liebenthal, E., Binder, J. R., Piorkowski, R. L., and Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience*, 15(4), 549-558.
- McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255--275). London: Sage Publications.
- Mitterer, H. (submitted). On the causes of compensation for coarticulation: convergent evidence for hierarchical categorization.
- Mitterer, H., and Blomert, L. (2003). Coping with phonological assimilation in speech perception: evidence for early compensation. *Perception and Psychophysics*, 65(6), 956-969.
- Mitterer, H., Csépe, V., and Blomert, L. (2003). Compensation for phonological assimilation in perception: evidence from Hungarian liquid assimilation. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2321-2324.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemä, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432-434.
- Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J.-B. (in press). Valid conjunction inference with the minimum statistic. *NeuroImage*.
- Nixon, P., Lazarova, J., Hodinott-Hill, I., Gough, P., and Passingham, R. (2004). The inferior frontal gyrus and

- phonological processing: an investigation using rTMS. *Journal of Cognitive Neuroscience*, 16(2), 289-300.
- Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science*, 25, 711-731.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience* 12(6), 1038-1055.
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., and Gabrieli, J. D. E. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *Journal of Cognitive Neuroscience*, 13(5), 687-697.
- Ruff, S., Marie, N., Celsis, P., Cardebat, D., and Démonet, J.-F. (2003). Neural substrates of impaired categorical perception of phonemes in adult dyslexics: an fMRI study. *Brain and Cognition*, 53, 331-334.
- Samuel, A. G., and Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 676-694.
- Schouten, B., Gerrits, E., and van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41, 71-80.
- Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *TRENDS in Neurosciences*, 26(2), 100-107.
- Scott, S. K., and Wise, R. J. S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, 41, 23-34.
- Scott, S. K., and Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92, 13-45.
- Seghier, M. L., Lazeyras, F., Pegna, A. J., Annoni, J. M., Zimine, I., Mayer, E., Michel, C. M., and Khateb, A. (2004). Variability of fMRI activation during a phonological and semantic language task in healthy subjects. *Human Brain Mapping*, 23(3), 140-155.
- Seki, A., Okada, T., Koeda, T., and Sadato, N. (2004). Phonemic manipulation in Japanese: an fMRI study. *Cognitive Brain Research*, 20, 261-272.
- Sharp, D., Scott, S., Cutler, A., and Wise, R. (2001). Distinct cortical processing of vowels and consonants during lexical access? *NeuroImage*, 13(6), S599.
- Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 1145-1162.
- Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics* 63, 1109-1139.
- Talaraich, P., and Tournoux, J. A stereotactic coplanar atlas of the human brain. Stuttgart: Thieme, 1988.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 14792-14797.
- Whalen, D. H. (1989). Vowel and consonant judgements are not independent when cued by the same information. *Perception and Psychophysics*, 46, 284-292.
- Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846-849.
- Zevin, J. D., and McCandliss, B. D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, 1(4).
- Zhang, J. X., Feng, C. M., Fox, P. T., Gao, J. H., and Tan, L. H. (2004). Is left inferior frontal gyrus a general mechanism for selection? *NeuroImage*, 23(2), 596-603.