

Department of Modern Languages
Faculty of Arts
University of Helsinki

**CATEGORICAL REPRESENTATIONS OF PHONEMIC
VOWELS INVESTIGATED WITH FMRI**

Kirsi Harinen

ACADEMIC DISSERTATION

To be publicly discussed,
by due permission of the Faculty of Arts
at the University of Helsinki in Auditorium 229 (Aurora), Siltavuorenpenger 10,
on the 27th of October, 2017, at 12 o'clock.

Helsinki 2017

ISBN 978-951-51-3694-7 (pbk.)
ISBN 978-951-51-3695-4 (PDF)
<http://ethesis.helsinki.fi>

Unigrafia
Helsinki 2017

CONTENTS

ABSTRACT	5
ACKNOWLEDGEMENTS	6
LIST OF ORIGINAL PUBLICATIONS.....	7
1 INTRODUCTION.....	9
1.1 Organization of the speech sound system	9
1.1.1 Vowels	9
1.1.2 Vowel phonemes.....	10
1.1.3 Categorical perception of phonemes: prototypes and nonprototypes	11
1.2 Neural substrates of speech processing	12
1.2.1 Speech processing in the human auditory cortex.....	12
1.2.2 Theories and models of speech processing	13
1.2.2.1 Phonetic theories of speech perception.....	13
1.2.2.2 The Rauschecker model	14
1.2.2.3 The Hickok and Poeppel model	14
1.3 Processing of phonemic vowels in AC.....	15
1.4 Categorical processing in IPL.....	16
1.5 The effects of attention and tasks on AC activations	17
2 AIMS OF THE STUDY	18
3 METHODS AND RESULTS	19
3.1 Functional magnetic resonance imaging (fMRI).....	19
3.2 Common procedures in Studies I–III	19
3.2.1 Subjects	19
3.2.2 Stimuli	20
3.2.3 Procedure.....	20
3.2.4 fMRI data acquisition and analysis.....	21
3.3 Study I. Task-dependent activations of human auditory cortex to prototypical and nonprototypical vowels	22
3.3.1 Vowels	22
3.3.2 Tasks and stimulus streams	23
3.3.3 Results and discussion	24
3.4 Study II. Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks	25
3.4.1 Vowels	25
3.4.2 Tasks and stimulus streams	26
3.4.3 Results and discussion	26
3.5 Study III. Acoustical and categorical tasks differently modulate activations of human auditory cortex to vowels	28
3.5.1 Vowels	28
3.5.2 Tasks and stimulus streams	29
3.5.3 Results and discussion	31

4 GENERAL DISCUSSION	34
4.1 Categorical phoneme representations in human AC	34
4.2 Categorical processing in IPL?	36
4.3 The importance of stimulus and task control.....	37
4.4 Conclusions	38
REFERENCES	39

ABSTRACT

The present thesis investigates the sensitivity of the human auditory cortex (AC) to the contrast between prototype and nonprototype vowels as well as between phonemic and nonphonemic vowels. Activations to vowels were measured with functional magnetic resonance imaging (fMRI), which was also used to analyze the effect of categorical processing on modulations in AC and adjacent inferior parietal lobule (IPL) observed during active listening tasks. A prominent theoretical view suggests that native phonemic vowels (i.e., phonemes) are represented in the human brain as categories organized around a best representative of the category (i.e., phoneme prototype). This view predicts systematic differences in the neural representations and processing of phoneme prototypes, nonprototypes and nonphonemic vowels.

In three separate studies, subjects were presented with vowel pairs and visual stimuli during demanding auditory and visual tasks. Study I compared activations to prototypical and nonprototypical vowels, whereas Study II focused on the contrast between phonemic and nonphonemic vowels. Study II also tested whether activations in IPL during a categorical vowel memory task depend on whether the task is performed on phonemic (easy to categorize) or nonphonemic (harder to categorize) vowels. Study III was designed to replicate the key findings of Studies I and II. Further, Study III compared activations to identical vowels presented during a number of different task conditions requiring analysis of the acoustical or categorical differences between the vowels.

The results of this thesis are in line with the general theoretical view that phonemic vowels are represented in a categorical manner in the human brain. Studies I–III showed that information about categorical vowel representations is present in human AC during active listening tasks. Areas of IPL, in turn, were implicated in general operations on categorical representations rather than in categorization of speech sounds as such. Further, the present results demonstrate that task-dependent activations in AC and adjacent IPL strongly depend on whether the task requires analysis of the acoustical or categorical features of the vowels. It is important to note that, in the present studies, surprisingly small differences in the characteristics of the vowel stimuli or the tasks performed on these vowels resulted in significant and widespread activation differences in AC and adjacent regions. As the key findings of Studies I and II were also quite successfully replicated in Study III, these results highlight the importance of carefully controlled experiments and replications in fMRI research.

ACKNOWLEDGEMENTS

This work was carried out in the @brain research group, Institute of Behavioural Sciences, University of Helsinki. All fMRI measurements were conducted in the Advanced Magnetic Imaging (AMI) Centre, Aalto University School of Science.

I would like to express my warmest gratitude to my supervisors Dr. Teemu Rinne and Professor Emeritus Olli Aaltonen for their guidance and support throughout this work. I am deeply grateful to Dr. Rinne for the opportunity to be a full-time doctoral student (2009–2012) in his group. I give my warmest thanks to Professor Stefan Werner for agreeing to be my opponent and Professor Martti Vainio for agreeing to be my custos. I extend my thanks to the official reviewers of this thesis, Dr. Iiro Jääskeläinen and Dr. Riikka Möttönen, for their constructive comments on this thesis.

I thank my coauthors Docent Oili Salonen and Mrs. Emma Salo. I also wish to thank other members of the @brain group, Mrs. Suvi Häkkinen and Mr. Patrik Wikman. Thank you for being friends and sharing your knowledge with me. Special thanks go to Mrs. Marita Kattelus (AMI) for her kind guidance in fMRI at the beginning of my thesis work.

This work was partly funded by the Emil Aaltonen Foundation (six months during 2013-2014).

Vantaa, 1.9.2017
Kirsi Harinen

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

- I Harinen, K. & Aaltonen, O. & Salo, E. & Salonen, O. & Rinne, T. 2013. Task-dependent activations of human auditory cortex to prototypical and nonprototypical vowels. *Human Brain Mapping*, 34(6), 1272-1281. doi:10.1002/hbm.21506

- II Harinen, K. & Rinne, T. 2013. Activations of human auditory cortex to phonemic and nonphonemic vowels during discrimination and memory tasks. *Neuroimage*, 77, 279-287. doi:10.1016/j.neuroimage.2013.03.064

- III Harinen, K. & Rinne, T. 2014. Acoustical and categorical tasks differently modulate activations of human auditory cortex to vowels. *Brain and Language*, 138, 71-79. doi:10.1016/j.bandl.2014.09.006

The publications are referred to in the text by their roman numerals.

1 INTRODUCTION

Subjectively, our perception of speech seems to be based on static, letter-like units. However, speech sounds vary considerably depending on the speaker and phonetic context (Peterson and Barney, 1952; Liberman et al., 1957). Despite this acoustic variation, listeners are able to identify and name isolated speech sounds of their native language with ease as perceptually distinct categories (Best, 1994; Flege, 1995; Goto, 1971; Werker and Tees, 1984). This phenomenon, categorical perception of phonemes, has evoked extensive interest in speech sciences as it provides a window into the fundamental organization of language. The present thesis investigates the neural basis of categorical phoneme perception by measuring activations in the human auditory cortex to Finnish phonemic and nonphonemic vowels.

1.1 ORGANIZATION OF THE SPEECH SOUND SYSTEM

Each human language is based on a distinct set of abstract contrastive categories of speech sounds (vowels and consonants). These categories are called phonemes. It has been suggested that phonemic categories are organized around prototypical sounds (best representative of a category; Kuhl, 1991, 1994; Samuel, 1982). The following paragraphs will introduce categorical perception of phonemic vowels in more detail.

1.1.1 VOWELS

Vowels are produced with a relatively open vocal tract (e.g., Fant, 1960). The position of the tongue and lips vary the shape and acoustical characteristics of the vocal tract resulting in the amplification of certain frequencies (vocal tract resonances) of sounds. These amplifications are called formants (dark horizontal stripes in Figure 1). Each vowel has a distinct combination of the two lowest formants, F1 (associated with the height of the tongue) and F2 (the place of articulation; Assman and Summerfield, 1989; Delattre et al., 1952; Hillenbrand et al., 1995; Klatt, 1982; Miller, 1989; Peterson and Barney, 1952; Rosner and Pickering, 1994). The frequencies of the formants F3, F4 and other upper formants vary relatively little between different vowels, thus vowels are identified mainly according to formants F1 and F2. The formant structure of the vowels is relatively stable over time.

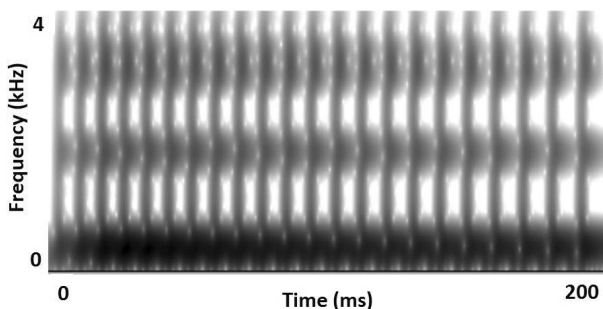


Figure 1. Spectrogram of a typical Finnish /y/ vowel.

1.1.2 VOWEL PHONEMES

Phonemes are abstract perceptual classes of speech sounds (phones). That is, a phoneme category consists of phones that differ in their acoustic-phonetic structure. Speech sounds are classified as belonging to different phoneme categories if replacing one phoneme with another changes the meaning of an otherwise identical word. According to this rule, /i/ and /a/ are different phonemes in Finnish (and in English) as, for example, ‘pila’ and ‘pala’ (‘bit’ and ‘bat’) differ in meaning. Although the term is difficult to define unambiguously, phonemes are usually seen as the smallest semantically significant units of language.

The number of phonemic vowels in human languages vary considerably. In Finnish, there are eight different phonemic vowels: /i/, /y/, /u/, /e/, /ø/, /o/, /æ/ and /a/. Figure 2 illustrates examples of Finnish /i/, /u/ and /a/ vowels (black circles) and three groups of other vowels that are not phonemes in Finnish (gray diamonds) in F1/F2 space. Note that, despite the clear acoustic differences, all sounds illustrated as black circles in the top left, top right and bottom right corners of Figure 2 are perceived categorically by a native Finnish speaker as /i/, /u/ and /a/ vowels, respectively. The nonphonemic groups (gray diamonds) in Figure 2 are not systematically associated with any Finnish phonemic vowels and are not perceived categorically (without extensive training).

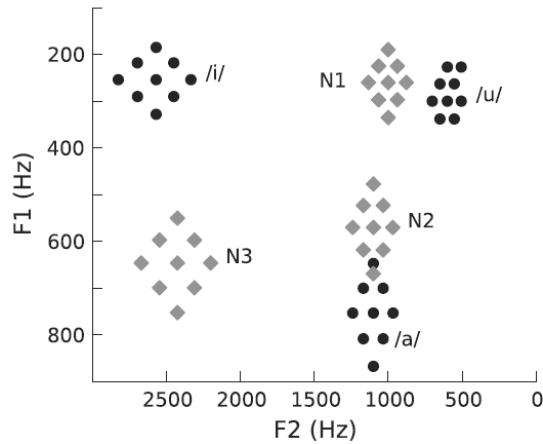


Figure 2. Examples of Finnish phonemic vowels (black circles) /i/, /u/ and /a/ and three groups (N1–N3) of nonphonemic vowels (gray diamonds) in F1/F2 space.

1.1.3 CATEGORICAL PERCEPTION OF PHONEMES: PROTOTYPES AND NONPROTOTYPES

Categorical perception of phonemes has been approached from two slightly different theoretical viewpoints. According to the first theoretical viewpoint (boundary-based categorical perception of phonemes), two speech sounds with equal acoustic distance (e.g., in F1/F2 space) are easier to differentiate from each other when they belong to different phoneme categories than when the sounds belong to the same phoneme category (Eimas, 1963; Harnad, 1987, 2003; Liberman et al., 1957; Pisoni, 1973; Repp, 1984). The most conservative and strongest version of this view even argued that it is not possible to tell two different speech sounds apart within the same phonemic category.

The second viewpoint emphasizes the inner structure of the vowel categories and the importance of the category prototypes in vowel perception and categorization. According to this latter view point, phonemic vowel categories are organized around an ideal or best representative phone (i.e., prototype) of that category (Iverson and Kuhl, 1995; Kuhl, 1991, 1994; Samuel, 1982). Kuhl and colleagues (Iverson and Kuhl, 1995; Kuhl, 1991; Kuhl et al., 1992) demonstrated that a phonemic vowel judged to be prototypical was more difficult to discriminate from its neighbors than a less prototypical one (i.e., a nonprototype phonemic vowel) within the same phonemic category. This is because the prototypes act as “perceptual magnets” pulling adjacent vowels together so that the perceptual space shrinks around prototypes even though the physical distance between the

sounds is kept constant. Figure 3 illustrates an individually defined vowel category border and a prototype vowel in F1/F2 space.

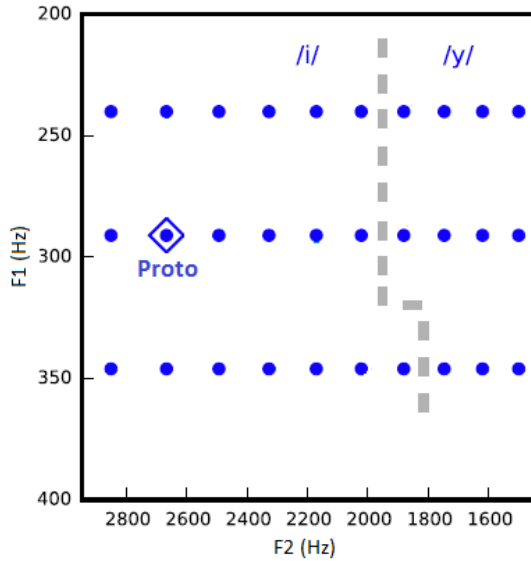


Figure 3. Examples of vowels (blue circles) in F1/F2 space. The gray dashed line shows an individually defined category boundary between /i/ and /y/ vowels. The best (for this subject) representative vowel of the phoneme category /i/, i.e., the prototype, is highlighted with a diamond.

1.2 NEURAL SUBSTRATES OF SPEECH PROCESSING

1.2.1 SPEECH PROCESSING IN THE HUMAN AUDITORY CORTEX

The anatomical and functional organization of the human auditory cortex (AC) is widely investigated but still many details have remained quite unknown (Hackett, 2011). Primary AC (A1) occupies regions in and near the medial tip of Heschl’s gyrus (HG; Moerel et al., 2014; Saenz and Langers, 2014) in the superior temporal lobe. Other auditory cortical areas extend from HG to planum temporale (PT) and to anterior and posterior parts of the superior temporal gyrus and sulcus (STG/STS; Woods et al., 2010; Hackett, 2011). Prior imaging literature implicates wide AC regions in the processing of different aspects of speech (Benson et al., 2001; Binder et al., 2000; Chang et al., 2010; Dehaene-Lamberg et al., 2005; Desai et al., 2008; Friederici, 2011; Hickok, 2009; Hickok and Poeppel, 2007; Jäncke et al., 2002; Liebenthal et al., 2005; Obleser et al., 2006; Raizada and Poldrack, 2007;

Rauschecker and Scott, 2009; Scott and Johnsrude, 2003; Turkeltaub and Coslett, 2010, review; Weinberger, 2011; Woods et al., 2011) and suggests that speech and nonspeech sounds are processed separately already at the level of primary AC (e.g., Bruder et al., 2011; Edmonds et al., 2010; Kilian-Hütten et al., 2011; Rinne et al., 1999; Uppenkamp et al., 2006; Staeren et al., 2009; Whalen et al., 2006; Woods et al., 2011).

Previous imaging studies have typically investigated language-sensitivity in AC by comparing activations during speech and nonspeech (e.g., tones, pseudospeech, rotated/manipulated speech analogs) sounds (Benson et al., 2001; Binder et al., 2000; Demonet et al., 1992, 2005; Hutchison et al., 2008; Narain et al., 2003; Obleser et al., 2006; Uppenkamp et al., 2006; Vouloumanos et al., 2001; Whalen et al., 2006; Zatorre et al., 1992). In addition to language-specific activations, however, such comparisons are typically affected by the (unavoidable) acoustic difference between speech and nonspeech stimuli (Desai et al., 2008; Liu and Holt, 2011). Some earlier studies have used acoustically distorted (e.g., sine-wave) speech to compare activations associated with speech and nonspeech conditions. The acoustically identical stimulus in uninformed (nonspeech) vs. informed (speech) condition is perceived either as unintelligible or intelligible speech, respectively (Giraud et al., 2004; Hakonen et al., 2016; Möttönen et al., 2006; Tiitinen et al., 2012). Further, speech and nonspeech stimuli may also differ in other dimensions (e.g., speech stimuli are likely to be more familiar than nonspeech stimuli; Desai et al., 2008). Thus, the differences between activations to speech and nonspeech stimuli should be interpreted with caution.

1.2.2 THEORIES AND MODELS OF SPEECH PROCESSING

The classic auditory and motor theories of speech perception are mainly based on phonetic research. More recent theories incorporate results from animal, lesion and neuroimaging studies, and they suggest the interaction of auditory and motor operations. The following paragraphs introduce the main ideas of classic phonetic theories of speech perception and two prevailing brain-level models of speech processing.

1.2.2.1 Phonetic theories of speech perception

Researchers have long searched for explanations for the human ability to link variant speech sounds effortlessly into abstract linguistic constructs such as phonemes. Auditory theories of speech perception (Diehl, 1987; Diehl and Kluender, 1989; Fant, 1960; Stevens 1981, 1989; Stevens and Blumstein, 1978) emphasize that speech and nonspeech sounds are represented and

processed in common systems. Auditory theories suggest that acoustic properties of both speech and nonspeech sounds are similarly analyzed in the early auditory cortical areas and that speech sounds are matched with their phonetic representations (e.g., acoustic target of vowel prototypes) in some higher-level auditory areas (e.g., Diehl and Kluender, 1989; Stevens, 1989). Motor theories of speech perception (Liberman et al., 1967; revised Liberman and Mattingly, 1985), in turn, maintain that speech sounds are processed in specialized cortical networks (Iverson et al., 2003; Whalen et al., 2006). Further, motor theories presuppose a close linkage between speech perception and production so that we perceive speech sounds based on the knowledge of how these sounds are produced (i.e., motor programs for articulation). Thus, the auditory and motor theories disagree on how and when the transformation from acoustic signals to some form of abstract categories (e.g., phonemes) is accomplished. The auditory theories predict that acoustic signals are transformed into categorical phoneme representations in some higher-level nonprimary auditory and auditory-related cortical regions (e.g., posterior STG, posterior STS and IPL), whereas, according to motor theories, speech is already processed differently than other sounds at the earliest cortical level.

1.2.2.2 The Rauschecker model

The dual stream model introduced by Rauschecker and colleagues (Leaver and Rauschecker, 2010; Rauschecker, 2011; Rauschecker and Scott, 2009) is mainly based on neurophysiological recordings in animals and noninvasive brain imaging results in humans. In this model, a ventral stream originates from AC and projects to the premotor cortex (PMC) via the anterior STG/STS and inferior frontal gyrus (IFG). This stream is involved in auditory object analysis and decoding speech. A dorsal stream, projecting posteriorly from AC to IPL and further to PCM and IFG, processes information related to auditory space, motion and auditory sequences (e.g., in speech and music). The streams from AC to PCM form feedforward and feedback loops for the integration of auditory-sensory and motor-articulatory information. Thus, Rauschecker's model combines ideas from auditory and motor theories and emphasizes the importance of motor cortices in speech (DeWitt and Rauschecker, 2012).

1.2.2.3 The Hickok and Poeppel model

The model proposed by Hickok and Poeppel (2004, 2007) also assumes that speech is processed along two parallel streams. In their model, a ventral

stream projects from superior STG to posterior STS, middle temporal gyrus (MTG) and IFG. The ventral stream supports speech comprehension. The dorsal stream, in turn, projects from STG to regions in the parieto-temporal junction (Sylvian-parietal-temporal area; Spt), IFG and premotor cortex. The dorsal stream, and particularly area Spt, supports auditory-motor integration for speech production. Further, Hickok and Poeppel suggest that the dorsal stream is involved in auditory sub-lexical speech perception tasks (such as vowel discrimination and n-back task), as access to phonemic segments is available within the dorsal stream.

1.3 PROCESSING OF PHONEMIC VOWELS IN AC

Prior EEG studies have reported that the mismatch negativity (MMN) component of the event-related potential (ERP) generated at least partly in AC is sensitive to differences between phonemic (native) and nonphonemic (nonnative) vowels (Bruder et al., 2011; Huotilainen et al., 2001; Näätänen et al., 1997; Sharma and Dorman, 1998). Aaltonen et al. (1997) showed that MMN is also sensitive to the contrast between prototypical and nonprototypical (/i/) vowels. Based on these results, it has been suggested that language-specific phoneme representations reside in AC.

At least one previous study has used functional magnetic resonance imaging (fMRI) to measure activation differences for prototypical and nonprototypical vowels in AC. Guenther and colleagues (2004) presented their subjects with stimulus blocks consisting of either one repetitive prototypical or nonprototypical /i/ vowel (judged as good or bad exemplar of /i/ by most listeners, respectively). Their subjects were required to attend to the vowels and note the differences from sound to sound (despite the fact that only one repetitive vowel was presented in a block). They reported stronger activations in posterior AC to nonprototype than prototype vowel blocks. The authors suggested that this difference was because the nonprototype vowel was associated with a larger neural representation than the prototype vowel. However, it may be that the activation difference between nonprototype and prototype blocks was affected by additional (uncontrolled) factors. For example, as the subjects heard only one repetitive vowel during prototype and one during nonprototype block, the vowel in the nonprototype block, taken from the category boundary, may have been perceptually more ambiguous than the vowel in the prototype block. Consequently, the subjects' task to detect differences from sound to sound may have been more attention-engaging during nonprototype blocks. Thus, it is possible that the enhanced activation during nonphonemic vowel blocks was related to task-level effects.

Until now the studies on categorical representations of the vowels are still quite sparse and the results leave many open questions. First, previous

studies have often compared the activation to one vowel with the activation to another (in few conditions) so that the results could be related to acoustic differences between the stimuli. Second, the characteristics of the speech stimuli have varied from one study to another but also within a study (natural vs. synthetic speech, native vs. nonnative vowels, isolated vowels vs. syllable context, sinewaves, noise, manipulated speech analogs, number and duration of the stimuli, rate of presentation etc.). It is not known how such variation affects the interpretation and comparison of the results obtained in different studies. Third, the effects of attention and different tasks on vowel processing have not been systematically studied although attention and tasks strongly modulate auditory processing in AC (Fritz et al., 2005; Hall et al., 2000; Lomber and Malhotra, 2008; Petkov et al., 2004; Rinne et al., 2009, 2012). Fourth, systematic replications of the key findings are, to date, very rare.

1.4 CATEGORICAL PROCESSING IN IPL

Previous studies have implicated areas of IPL in categorical processing of auditory stimuli (e.g., Raizada and Poldrack, 2007; Rinne et al., 2009; 2012; review see Turkeltaub and Coslett, 2010). Categorical processing is essential in mapping variant acoustic signals into invariant phoneme concepts. Raizada and Poldrack (2007) investigated categorical perception of phonemes /b/ and /d/ in a syllable (consonant+vowel) context with fMRI. They reported that activations in the supramarginal gyrus (SMG; part of IPL) were stronger during processing of between-category pairs than within-category pairs. Turkeltaub and Coslett (2010) reported in their meta-analysis on sublexical speech perception that IPL was systematically activated in fMRI studies (N=8) on categorical phoneme perception. Rinne et al. (2009, 2012) compared activations in STG and IPL during categorical pitch/spatial memory (n-back) and discrimination tasks. They reported enhanced activations during categorical memory tasks in IPL, while discrimination tasks performed on similar stimuli were associated with increased activations in STG but not in IPL. Taken together, these results suggest that areas of IPL are involved in the processing of categorical information. IPL, however, is not necessarily specialized in the categorical processing of speech. As areas in posterior STG, parieto-temporal junction and IPL are implicated in speech processing, a better understanding of the functional significance of activation in these regions during active listening tasks is needed.

1.5 THE EFFECTS OF ATTENTION AND TASKS ON AC ACTIVATIONS

Earlier studies have demonstrated that auditory attention strongly modulates activation in AC (Grady et al., 1997; Hall et al., 2000; Petkov et al. 2004; Rinne et al., 2007, 2009; Woods et al., 2009). Further, AC activations are also modulated by the characteristics of the auditory tasks performed on the speech and nonspeech sounds (Angenstein et al., 2012; Hickok and Poeppel, 2007; Hickok and Saberi, 2012; Leung and Alain, 2011; Petkov et al., 2004; Rinne et al., 2009, 2012; Scheich et al., 2007). For example, discrimination and n-back memory tasks performed on identical sounds are associated with distinct activation patterns in STG and adjacent IPL (Rinne et al., 2009). Therefore, in the present thesis, activations to vowels were measured during a visual condition (no directed auditory attention) and during different auditory tasks.

2 AIMS OF THE STUDY

The present thesis consists of three studies using fMRI to systematically investigate the sensitivity of human auditory cortical areas to phenomena associated with categorical perception of phonemic vowels. As vowels are the fundamental building blocks of speech and as AC plays an important role in speech processing, it was hypothesized that activations in AC would be sensitive to the language-specific contrast between prototype and nonprototype vowels as well as to the difference between phonemic and nonphonemic vowels. Activations in AC were investigated during task conditions that required the subjects to focus on acoustic or categorical aspects of the vowels or to focus on a visual task (no directed attention to vowels).

Studies I and II tested the contrasts between prototypical and nonprototypical as well as between phonemic and nonphonemic vowels, respectively. Further, Study II also investigated the idea that activations in posterior STG and IPL observed in prior studies during n-back pitch and spatial memory tasks (Rinne et al., 2009, 2012) are due to categorical processing required by these tasks. It was hypothesized that if activations in the posterior STG and IPL during a categorical 2-back task are associated with categorical processing, then these activations should be higher when this task is performed with nonphonemic (hard to categorize) than on phonemic (easy to categorize) vowels. Study III was designed to replicate the results of the Studies I and II. Further, this study investigated the role of posterior STG and IPL in categorical processing by comparing activations to identical vowels presented during a number of different task conditions requiring the analysis of acoustical or categorical differences between vowels. The key hypothesis was that activations during a vowel discrimination task would depend on whether the task was performed on the basis of acoustical or categorical features of the vowels.

3 METHODS AND RESULTS

3.1 FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI)

Functional magnetic resonance imaging (fMRI) was used in all studies of the present thesis as it allows one to investigate activations in all auditory cortical areas during one recording session (ca. 1 h) with a relatively high spatial resolution (ca. 2 mm x 2 mm x 2 mm). In fMRI, a strong static magnetic field (in the present studies 3 T) and radio frequency magnetic pulses are used to measure changes in blood oxygenation and blood flow that occur when a brain area is activated (Ogawa et al., 1990). It is assumed that neural activity and the blood oxygenation level dependent (BOLD) signal are coupled so that increased neural activity is associated with an enhancement of the BOLD signal in that area (Logothetis et al., 2001, 2008; Mukamel et al., 2005).

3.2 COMMON PROCEDURES IN STUDIES I–III

3.2.1 SUBJECTS

In all studies, subjects (Table 1) were healthy right-handed adults with normal hearing. An informed written consent was obtained from each subject before the experiment. The study protocol was approved by the Ethics Committee of the Helsinki Hospital District (Studies I and II) or by the Ethics Committee of the Institute of the Behavioural Sciences, University of Helsinki (Study III). All subjects were native Finnish speakers.

Table 1. Subjects in Studies I–III

Study	N	Females	Mean age
I	20	12	24
II	22	13	24
III	21	11	24

3.2.2 STIMULI

Vowels in all studies were synthesized using formant synthesis in a Praat software package (www.praat.org). Vowels were defined by their two lowest formants F1 and F2 (Table 2). The other formants (F3–F5 in Study I, F3–F7 in Study II and F3–F8 in Study III) were the same for all vowels in all studies. A linear falling contour from 150 to 100 Hz was used as F0 to give the synthetic vowels a more natural impression. All vowels were 200 ms in duration (including a linear 5 ms onset and offset ramp). Visual stimuli consisted of Gabor gratings (duration 100 ms) with varying orientation presented at fixation.

Table 2. Formant frequencies (Hz) in Studies I–III.

formants	Study I		Study II			Study III		
	/a/-/æ/	/y/-/i/	/i/	/a/	/u/	/y-i/	/u-o/	/a-æ/
F1	720	250	185–328	647–867	227–378	240–346	270–566	660–896
F2	1046–2119	1518–2882	2334–2825	967–1240	506–702	1500–2852	500–952	950–2264
F3	3010	3010	3010	3010	3010	3010	3010	3010
F4	3300	3300	3300	3300	3300	3300	3300	3300
F5	3850	3850	3850	3850	3850	3850	3850	3850
F6	-	-	4850	4850	4850	4850	4850	4850
F7	-	-	5850	5850	5850	5850	5850	5850
F8	-	-	-	-	-	6850	6850	6850

3.2.3 PROCEDURE

In all studies, subjects were presented with blocks of concurrent (but asynchronous) auditory and visual stimuli. During a task block, subjects responded to either auditory or visual targets by pressing one of two buttons with their right index and middle fingers (Study I) or by pressing one button with their right index finger (Studies II and III). Each task block was followed by a rest period during which subjects focused on a fixation mark (x) presented in the middle of a screen and waited for the next task block. During auditory task blocks, subjects were instructed to ignore the concurrent visual stimuli, while in the visual task blocks subjects focused on the visual task and ignored the vowels. The tasks used in Studies I–III are summarized in Table 3. Before fMRI, each subject was carefully trained (60–90 min) to perform the demanding tasks according to visual task-instruction symbols.

The auditory stimuli were delivered binaurally with an UNIDES ADU2a audio system (Unides Design, Helsinki, Finland) via plastic tubes through a

porous EAR-tip (ER3, Etymotic Research, Elk Grove Village, IL, USA) acting as an earphone. The scanner noise was attenuated by the earplugs, circumaural ear protectors (Bilsom Mach 1, SNR 23 dB) and viscous foam pads attached to the sides of the head coil. The visual stimuli were presented in the middle of a screen viewed through a mirror fixed to the head coil. During fMRI, behavioral responses were recorded in order to verify that subjects performed the demanding tasks in the scanner as expected (for details, see the original articles).

Table 3. Experimental design in Studies I–III.

Study	Tasks	Stimuli (type)	Conditions (N)	Blocks (N)	Block duration (s)	Vowel pairs/block (N)	Targets/block (N)
I	vowel rating vowel discrimination IRN pitch discrimination visual	vowel pairs IRN sounds Gabors	8	66	30	22	22
II	vowel discrimination vowel categorical n-back visual	vowel pairs Gabors	9	128	15	15	2-4
III	vowel discrimination categorical discrimination vowel categorical n-back visual	vowel pairs Gabors	13	144	13	13	3-5

3.2.4 FMRI DATA ACQUISITION AND ANALYSIS

Functional images were acquired using a gradient-echo echo-planar imaging sequence (GE-EPI; Table 4). The middle EPI slices were aligned along the Sylvian fissures based on a high-resolution anatomical image (MPRAGE, voxel matrix 256 x 256, FOV 25.6, resolution 1.0 mm x 1.0 mm x 1.0 mm). The imaging area covered the superior temporal lobe, insula, and most of the inferior parietal lobe in both hemispheres.

First level statistical analysis (within run) was performed using general linear modeling in FSL (www.fmrib.ox.ac.uk/fsl). The data were motion-corrected, high-pass filtered (cutoff 100 s), and spatially smoothed (Gaussian kernel of 5 mm full-width half-maximum). A second-level statistical analysis was used to combine the data from the two runs.

For analysis across subjects, the high-resolution anatomical images were normalized in spherical standard space using FreeSurfer (<http://freesurfer.net>). The anatomically normalized three-dimensional (3D) cortical surfaces were rotated and projected to a two-dimensional (2D) space separately for each hemisphere using equal-area Mollweide projection (Python libraries matplotlib and basemap, <http://matplotlib.sourceforge.net>). This procedure was then applied

separately for each subject to transform the results of the 3D second-level statistical analysis to 2D. Finally, the group analysis (FSL) was run on the flattened data. Z-statistic images were thresholded using clusters determined by $Z > 2.3$ and a (corrected) cluster significance threshold of $P < 0.05$ (using Gaussian random field theory). Activations are shown on a flattened mean 2D cortical surface.

Table 4. fMRI acquisition parameters in Studies I–III.

Study	Scanner 3.0 T	Year of data acq.	Head coil, (N of channels)	TR (ms)	TE (ms)	Slice thickness (mm)	In-plane resolution (mm)	Slices (N)	FOV (cm ²)	Volumes (N)
I	GE Signa	2010	8	2048	32	2.1	2.1 x 2.1	24	20 x 20	1 x 1322
II	GE Signa	2010– 2011	16	2048	32	2.1	2.1 x 2.1	24	20 x 20	1 x 1470
III	Siemens MAGNETOM Skyra	2012	20	2070	30	2.0	2.0 x 2.0	27	18.9 x 18.9	2 x 712

3.3 STUDY I. TASK-DEPENDENT ACTIVATIONS OF HUMAN AUDITORY CORTEX TO PROTOTYPICAL AND NONPROTOTYPICAL VOWELS

3.3.1 VOWELS

Two vowel continua (/a/–/æ/ and /y/–/i/) with 19 vowels in each were synthesized. The frequency of the first formant (F1) and upper formants (F3–F5) in both continua were fixed, but the second formant (F2) varied in steps of 30 mels (Figure 4). First, each vowel variant was presented 20 times in random order and subjects indicated by pressing two response buttons whether they heard /a/ or /æ/ and, in another run, /y/ or /i/. Next, subjects rated (scale 1–4; 1 = poor category exemplar, 4 = good category exemplar) the vowel variants that were consistently categorized as /æ/ or /i/ relative to a good Finnish pronunciation of /æ:/ and /i:/. For each subject, the /æ/ and /i/ vowels with the highest rating score were used as prototypes and the ones with lowest rating as nonprototypes.

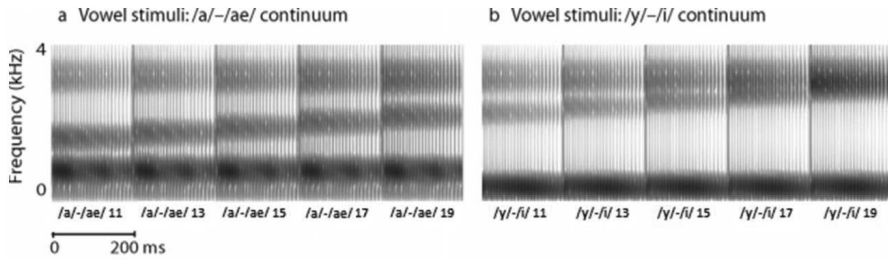


Figure 4. Spectrograms of five vowel stimuli from (a) /a/-/æ/ continuum and (b) /y/-/i/ continuum. Both continua consisted of 19 vowels separated by 30 mel steps in F2.

3.3.2 TASKS AND STIMULUS STREAMS

Study I compared activations to prototypical and nonprototypical vowel pairs presented during pitch discrimination (performed on noise bursts with pitch; Figure 5a), visual task (b), vowel discrimination, and vowel rating tasks. In the vowel discrimination task, subjects indicated whether the vowel pair consisted of same or different vowels. In the vowel rating task, subjects indicated whether the first or second vowel of the pair was a better exemplar of the phoneme category. The difference between the first and second part of the vowel was 30, 60, 90 or 120 mel.

Prototype and nonprototype vowel pairs were presented in separate task blocks. In prototype blocks, subjects were presented with vowel pairs in which one vowel was always prototypical /i/ or /æ/ and the other vowel was either the same (50 %) or a different vowel from the same vowel continuum (/y/-/i/ or /a/-/æ/; Figure 5c). In nonprototype blocks, the vowel pairs were constructed in the same way around a nonprototype (d). Vowel pairs were presented with a 1200–1600 ms onset-to-onset interval. Concurrently with the vowel pairs, IRN bursts and Gabor gratings were presented in each task block. IRN bursts (duration 100 ms) were presented with a 110–190 ms onset-to-onset interval so that one pitch was repeated 8–10 times after which the pitch slightly increased or decreased (target; Figure 5a). Analogously, Gabor gratings were presented with a 140–200 ms onset-to-onset interval so that one orientation was repeated seven to nine times after which the orientation slightly changed (target; Figure 5b).

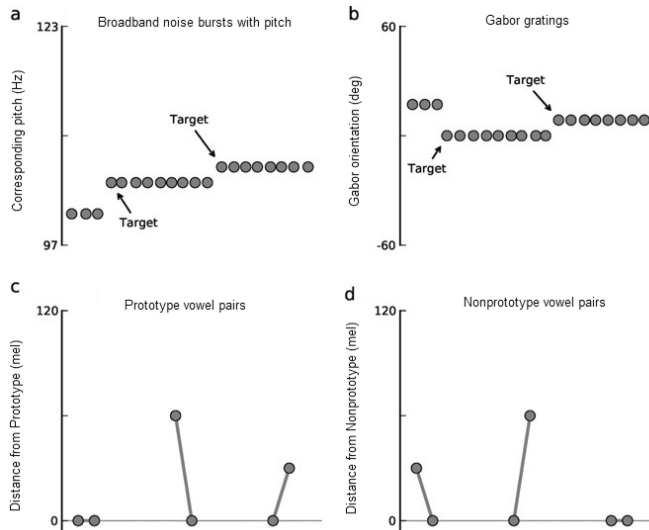


Figure 5. Task design in Study I. Structure of the stimulus streams (time scale on the horizontal axis is schematic). (a) In 30 s blocks (alternating with 10 s rest with no stimuli), subjects were presented with concurrent and asynchronous streams of iterated rippled noise (IRN) bursts with varying pitch, (b) Gabor gratings with varying orientation, and (c, d) vowel pairs. In pitch and Gabor discrimination tasks (a, b), subjects were required to detect pitch or orientation changes and indicate the direction of the change. During the vowel discrimination task, subjects indicated whether vowels in a pair were the same or different and during the vowel rating task, subjects indicated whether the first or second vowel in a pair was a better exemplar of the phoneme category (c, d).

3.3.3 RESULTS AND DISCUSSION

Activations in auditory cortical areas near HG were stronger during vowel task blocks with prototype than nonprototype vowel pairs (Figure 6a, red). This effect was mainly due to enhanced activations during vowel discrimination task with prototype vowels (b, red) as no significant differences were detected between prototype and nonprototype blocks in the vowel rating task. Acoustically, the stimuli in prototype and nonprototype blocks were highly similar as all vowels were taken from the same two vowel continua and some vowel pairs could even appear in both prototype and nonprototype blocks. Consistently, no significant stimulus-dependent activation differences were observed between prototype and nonprototype blocks presented during the visual task (no directed auditory attention). Together these results show that auditory cortical areas near HG are sensitive to the language-specific difference between a vowel prototype and nonprototype in a task-dependent manner. The enhanced activations observed during the discrimination task with prototype vowel pairs were probably because the vowels in the prototype pairs were perceptually more

similar to each other than to the vowels in the nonprototype pairs (the magnet effect; Kuhl, 1991). Yet, it is also possible that, due to the perceptual similarity, the discrimination task with prototype vowel pairs required more elaborate acoustic analysis of the vowels. Thus, although the enhanced activation observed during the vowel discrimination blocks with prototype vowel pairs is a consequence of the language-specific difference between the prototype and nonprototype pairs, this activation enhancement could be related to more elaborate acoustic processing of prototype pairs and not to language-specific processing as such.

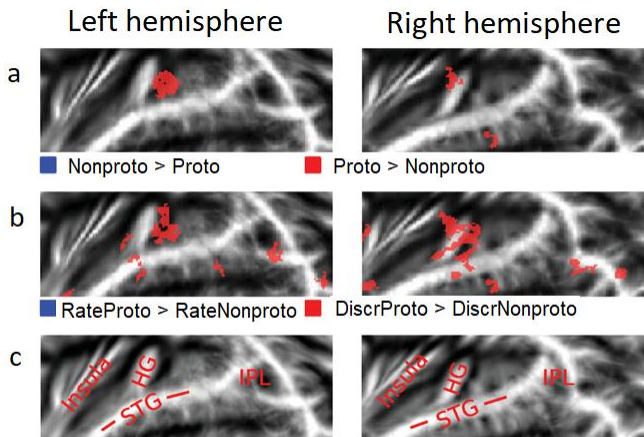


Figure 6. (a) Comparison of activations during nonprototype and prototype blocks collapsed across the vowel rating and discrimination tasks. (b) Areas where activations were stronger during prototype than nonprototype blocks during vowel rating (blue) and vowel discrimination (red). (c) Anatomical labels. STG superior temporal gyrus, HG Heschl's gyrus, IPL inferior parietal lobule.

3.4 STUDY II. ACTIVATIONS OF HUMAN AUDITORY CORTEX TO PHONEMIC AND NONPHONEMIC VOWELS DURING DISCRIMINATION AND MEMORY TASKS

3.4.1 VOWELS

Three groups of phonemic and nonphonemic vowel categories were synthesized (9 different vowels in each; see Figure 2). Each category contained 9 different vowels. The phonemic categories were defined based on typical Finnish /a/, /i/ and /u/ phonemes. The nonphonemic categories (N1, N2 and N3) were organized in regions of F1/F2 space where no prototypical

Finnish phonemes exist so that the categories were not systematically associated with only one Finnish phoneme. Within a category, the vowels were separated by at least 60 mel in F1/F2 space.

3.4.2 TASKS AND STIMULUS STREAMS

Study II compared activations to phonemic and nonphonemic vowels during vowel discrimination task, vowel categorical memory task (n-back), and visual task with identical but task-irrelevant vowel stimuli (Figure 7).

In the vowel discrimination task (Figure 7a), subjects were required to indicate when the first and second part of the vowel pair were the same (the vowels in a pair were either the same or separated by approximately 60 or 120 mel). In the n-back memory tasks (b), subjects indicated when the vowel pair (the vowels in a pair were always the same) belonged to the same category as the one presented 1, 2 or 3 trials (depending on the difficulty level) before. Phonemic and nonphonemic vowel pairs were presented in separate blocks. The auditory stream consisted of within-category vowel pairs from three phonemic or nonphonemic categories in all conditions. The vowel pairs were presented with 900–1100 ms onset-to-onset interval (with concurrent visual stimuli in every task block).

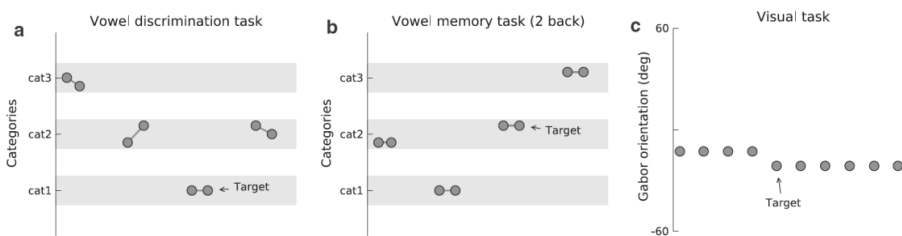


Figure 7. In 15 s blocks (alternating with 8 s rest with no stimuli), subjects were concurrently presented with vowel pairs (from three Finnish phonemic or three nonphonemic vowel categories) and Gabor gratings (onset-to-onset interval 300–500 ms). In the vowel discrimination task (a), they were required to indicate when the first and the second part of the vowel pair were the same. In the n-back vowel memory task (b), subjects indicated when the vowel pair belonged to the same vowel category as the one presented 1, 2 or 3 trials (depending on the difficulty level) before (2-back task is illustrated). In the visual task (c), subjects were required to detect Gabor orientation changes. Time scale on the horizontal axis is schematic.

3.4.3 RESULTS AND DISCUSSION

Significant activation differences were detected between task blocks with phonemic and nonphonemic vowels. Auditory tasks with phonemic vowels were associated with enhanced activations in areas of IPL (Figure 8a–c,

blue). Auditory tasks with nonphonemic vowels, in turn, enhanced activations in anterior and posterior STG (red). Comparisons of activations to blocks with phonemic and nonphonemic vowels presented during the visual task (i.e., no directed auditory attention) revealed significantly stronger activations during phonemic blocks in areas of posterior STG/STS and IPL, but no areas showed stronger activations to nonphonemic vowels (d). Together these results suggest that a more thorough acoustic analysis in AC was required to complete the vowel tasks on nonphonemic than on phonemic vowels and that phonemic vowels are processed in language-specific networks in posterior STG and IPL.

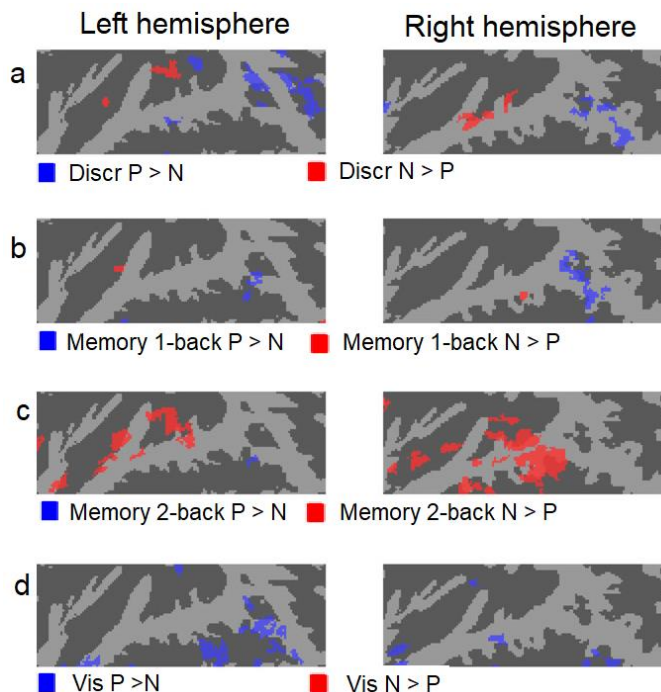


Figure 8. Comparisons of activations in task blocks with phonemic (P) and nonphonemic vowels (N) during (a) vowel discrimination, (b) 1-back vowel memory, (c) 2-back vowel memory task, and (d) visual task (with task-irrelevant vowels).

Similar to previous studies comparing activations in AC during discrimination and n-back memory tasks (Rinne et al., 2009, 2012), vowel discrimination tasks were associated with enhanced activations in anterior/posterior STG, whereas activations during vowel n-back tasks were enhanced in IPL (Figure 9). Activations during vowel n-back tasks with phonemic vowels also increased with memory load (i.e., 3-back > 1-back). In Study II, it was hypothesized that if IPL activations during n-back tasks are

due to categorical processing required in these tasks, then these activations should be higher during tasks performed on nonphonemic (hard to categorize) than on phonemic (easy to categorize) vowels. However, enhanced activations during 2-back task performed on nonphonemic vowels were detected in STG, but not in IPL (Figure 8c). This suggests that the processing requirements in the 2-back memory task performed on nonphonemic or phonemic vowels were different only during the earlier processing stages when auditory information is analyzed to achieve category labels for each vowel.

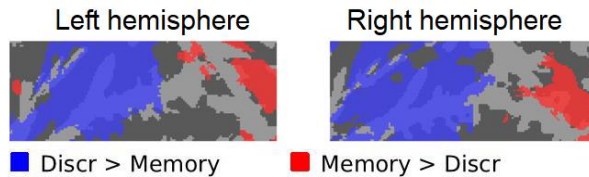


Figure 9. Areas where activations were stronger during vowel discrimination than vowel memory tasks (blue) and areas where activations were stronger during vowel memory than vowel discrimination tasks (red).

3.5 STUDY III. ACOUSTICAL AND CATEGORICAL TASKS DIFFERENTLY MODULATE ACTIVATIONS OF HUMAN AUDITORY CORTEX TO VOWELS

3.5.1 VOWELS

Three groups of phonemic (/i–y/, /u–o/ and /æ–a/; 30, 28 and 45 vowels in each group, respectively) and nonphonemic vowels (NPh1, NPh2 and NPh3; 12 vowels in each) were synthesized (Figure 10a). In each group, the vowels were separated by 60 mel in either F1, F2 or both. In order to define individual phoneme categories and category boundaries, subjects classified the vowels in each phonemic vowel group into two phonemes. Based on the results, three phoneme boundaries (/i–/y/, /u–/o/ and /æ–/a/) and six corresponding phoneme categories were defined. Next, subjects rated the goodness of the vowels in these categories. The vowel with the highest rating in each category was selected as the prototype and the vowel with the lowest rating as the nonprototype. These vowels were then paired with adjacent vowels (within the same category) to construct prototype or nonprototype vowel pairs (Figure 10b). Further, cross-category vowel pairs were constructed based on the individually defined phoneme boundaries. In the cross-category vowel pairs, one vowel was a nonprototype next to the vowel

boundary and the other vowel was an adjacent nonprototype vowel at the same or opposite side of the phoneme boundary (b).

3.5.2 TASKS AND STIMULUS STREAMS

Study III compared activations to prototypical, nonprototypical and nonphonemic vowels during vowel discrimination, vowel 2-back, category discrimination, and visual tasks (Figure 10c–f).

Prototype, nonprototype, nonphonemic and cross-category vowel pairs were presented in separate task blocks. Vowel discrimination, vowel 2-back memory and visual tasks with Gabor gratings were similar to the tasks used in Study II. In the category discrimination task subjects were required to indicate when both parts of the vowel pair belonged to the same phonemic category.

Each task block contained vowels from all the three vowel categories (/i/, /u/ or /æ/), vowel groups (/i–y/, /u–o/ or /æ–a/) or nonphonemic categories (NPh1–3). In all cases, the vowels in a pair were either the same or separated by approximately 60 mel (in F1, F2 or both).

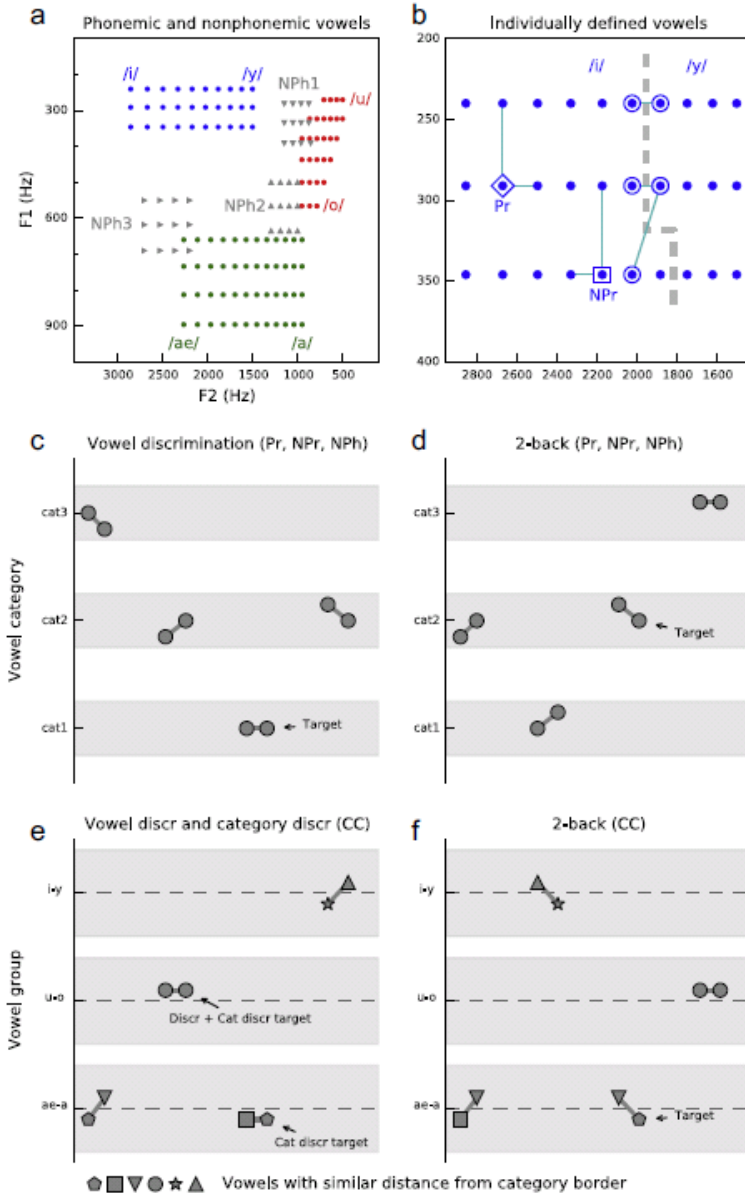


Figure 10. Experimental design. (a) In 13 s blocks (alternating with 8 s rest with no stimuli), subjects were presented with vowel pairs from three Finnish phonemic vowel continua (blue, red and green) and three nonphonemic (NPh) vowel categories (gray) defined in F1/F2 space. (b) Individually defined category boundary between */i/* and */y/* vowels (gray dashed line), prototype (P, diamond) */i/*, nonprototype (NPr, square) */i/*, and three nonprototype */i/* and two */y/* vowels (circle) near the category boundary. Examples of prototype (Pr + adjacent vowel), nonprototype (nonprototype + adjacent vowel) and cross-category vowel pairs are shown (a thin line connects vowels in a pair). (c) In the vowel discrimination and (d) in the vowel 2-back memory task, subjects were presented with prototype, nonprototype or nonphonemic vowel pairs. In the vowel discrimination task subjects were required to indicate when the vowels in a pair were the same. In the vowel 2-back memory task, subjects indicated when the vowel pair belonged to the same vowel category as the one presented 2 trials before. (e and f) Three different tasks were performed on cross-category and within-category vowel pairs (CC) which comprised of vowels

near the category border. In the discrimination task, the target was a pair in which the parts were identical (e). In the category discrimination task, the target was a pair in which both parts belonged to the same vowel category (but were not always identical vowels) (e). In the 2-back tasks, the target was a vowel pair that belonged to the same vowel group (*i–y/*, */u–o/* or */æ–a/*) as the pair presented two trials before (f). In e and f, the markers represent different vowels at a similar distance from the vowel category border.

3.5.3 RESULTS AND DISCUSSION

Consistent with Study I, stronger activations in AC were observed for prototype than for nonprototype vowel pairs (Figure 11a) during discrimination task. However, while in Study I the activation enhancements to prototype pairs were restricted to areas near HG, in Study III these activation enhancements were observed in wider areas in STG and IPL. In Study III, the difference between prototype and nonprototype blocks was also observed during the vowel 2-back task (b). As in Study I, only minor stimulus-dependent activation differences between prototype and nonprototype vowel blocks were observed when the prototype and nonprototype vowels were presented during visual tasks (i.e., no directed auditory attention; c). Thus, stimulus-dependent activations alone cannot explain the activation differences observed between prototype and nonprototype vowel blocks during active listening tasks.

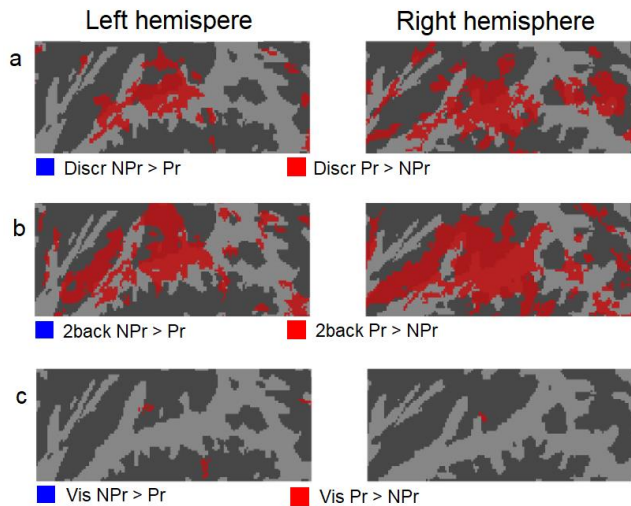


Figure 11. Activation differences during discrimination, 2-back and visual tasks performed on nonprototype (NPr) and prototype (Pr) vowel pairs.

Consistent with Study II, stronger activations in AC were observed for nonphonemic than for phonemic (nonprototype) vowel pairs during discrimination task (Figure 12a). This effect was also present during 2-back task (b). Furthermore, nonphonemic vowel blocks were associated with stronger stimulus-dependent activations (during visual task) than phonemic blocks (c). These stimulus-dependent effects were probably due to the fact that in Study III there were more different vowels in nonphonemic (vowels in nonphonemic pairs were randomly selected) than in phonemic (the pairs of phonemic vowels were organized around a nonprototype) blocks. These stimulus-dependent effects in medial STG, however, cannot explain the differences between phonemic and nonphonemic vowel blocks in more lateral STG areas during auditory tasks. Thus, Study III successfully replicated the main findings of Studies I and II showing that AC activations during active tasks are sensitive to the language-level differences between prototype and nonprototype as well as between phonemic and nonphonemic vowels.

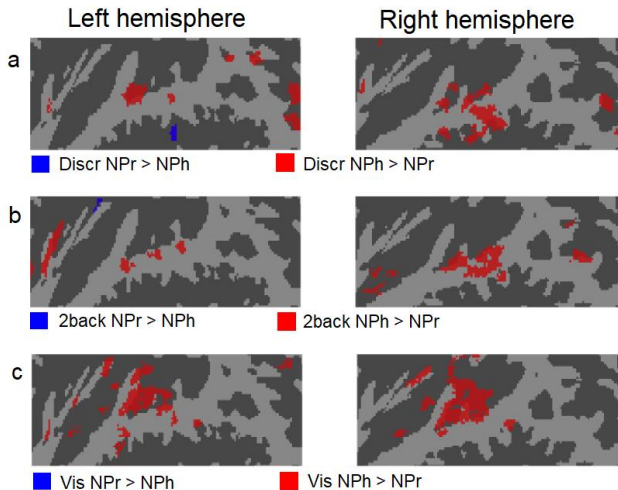


Figure 12. Activation differences during discrimination, 2-back and visual tasks performed on nonprototype (Npr) and nonphonemic (Nph) vowel pairs.

Consistent with Study II, vowel discrimination and vowel 2-back tasks enhanced activations in anterior-posterior STG and IPL, respectively (Figure 13a). These task-dependent activation patterns were quite similar irrespective of whether the tasks were performed on prototype, nonprototype, nonphonemic or cross-category vowel pairs. Study III also compared activations to identical cross-category (CC) vowel pairs presented during discrimination, 2-back memory and category discrimination tasks (b, c). In general, vowel discrimination and category discrimination were

associated with quite similar activation patterns in STG regions. The category discrimination task, however, was associated with enhanced activations in IPL (b). Further, although quite similar IPL areas were activated during the category discrimination and 2-back tasks, the category discrimination task was associated with stronger activations in areas of the insula and STG, whereas during the 2-back task activations were stronger in IPL (c). These activation patterns observed during three different tasks performed on identical stimuli show that activations in areas of AC and IPL strongly depend on whether the task requires analysis of the acoustical or categorical features of the sounds. More specifically, these results support the view of Rinne and colleagues (2009, 2012) that during listening tasks areas of STG are implicated in analysis of detailed acoustic information, whereas activations in IPL are associated with operations on categorical representations.

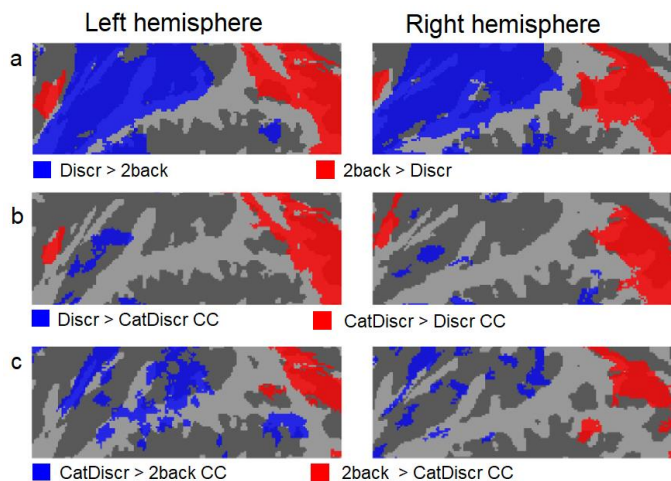


Figure 13. (a) Comparison of activations during all vowel discrimination (blue) vs. all vowel 2-back (red) tasks. (b) Activation differences during discrimination and category discrimination and (c) during category discrimination and 2-back tasks performed on identical cross-category (CC) vowel pairs.

4 GENERAL DISCUSSION

The results of the present thesis are consistent with the view that phonemic vowels are represented in a categorical manner and that categorical vowel information is available in human AC during active listening tasks. First, Studies I and II found that activation in AC is sensitive to the language-level difference between prototypical and nonprototypical phonemic vowels and between phonemic and nonphonemic vowels. Importantly, these results were successfully replicated in Study III. The results of the present thesis also shed new light on the role of IPL in categorical processing. The results implicate IPL in tasks requiring operations on categorical representations rather than categorization as such.

4.1 CATEGORICAL PHONEME REPRESENTATIONS IN HUMAN AC

The results of Studies I–III showed that activations in AC during active listening tasks were stronger during prototype compared to nonprototype vowel blocks and during nonphonemic compared to phonemic vowel blocks. These results cannot be explained by stimulus-dependent activations as all vowel groups used in Studies I–III were acoustically highly similar, even partly overlapping in F1/F2 space. Consistently, the stimulus-dependent activation differences during visual tasks with identical stimuli (prototype vs. nonprototype and phonemic vs. nonphonemic vowels) were absent or negligible.

Previous studies have shown that auditory attention strongly modulates activation in AC (Grady et al, 1997; Petkov et al., 2004; Rinne et al., 2007, 2009; Woods et al., 2009). It could be argued that native speech sounds are more attention-engaging than nonspeech sounds. This could result in stronger attention effects observed during the presentation of phonemic vowels versus nonphonemic vowels. Similar attention-related effects could modulate activations to prototype and nonprototype vowels. In all the studies of the present thesis, however, the vowels were introduced during demanding active tasks to engage attention. Thus, the activation differences between prototype vs. nonprototype and phonemic vs. nonphonemic vowel blocks cannot be easily explained by attention-related differences. Hence, the most parsimonious interpretation of the present results is that activations in human AC are sensitive to the phenomena associated with categorical perception of phonemic vowels.

This interpretation is in concordance with previous studies using noninvasive brain imaging methods, suggesting different activations to

acoustically similar but perceptually distinct speech sounds in AC (Chang et al., 2010; Guenther et al., 2004; Kilian-Hütten et al., 2011). Importantly, the present results are fully consistent with the predictions based on Kuhl's (1991) work. According to her view, phonemic vowel categories are organized around vowel prototypes. That is, speech sounds near an individual vowel prototype are perceived more similar to each other than speech sounds near a nonprototype, which makes the discrimination of the vowels more difficult around prototypes. This concerns only native phonemic categories, whereas the representations of nonphonemic (or non-native) vowel categories are not organized in a similar manner. Kuhl's ideas were based mainly on behavioral results. The present fMRI results support the idea that native phonemes are represented in a categorical manner in the brain and that information about such categories is available in human AC.

A large number of previous studies have aimed to map the brain regions involved in the processing of speech (Alho et al., 2014; Binder et al., 2000; Chang et al., 2010; Davis and Johnsrude, 2003; Desai et al., 2008; Friederici, 2011; Hickok and Poeppel, 2007; Jäncke et al., 2002; Lee et al., 2012; Liebenthal et al., 2005; Obleser et al., 2006, 2007, Okada et al., 2010; Rauschecker and Scott, 2009; Scott and Johnsrude, 2003; Turkeltaub and Coslett, 2010; Uppenkamp et al., 2006). Wide regions of anterior and posterior STG/STS and adjacent auditory related areas, such as IPL and inferior frontal areas, have been implicated in different aspects of speech processing. Specifically, areas of STG have been connected to acoustic analysis of the speech sounds. Accordingly, it could be argued that in the present study wide regions in STG, including areas in or near primary AC showing sensitivity to prototypic or nonphonemic status of the vowels, are involved in speech processing. However, the present results do not necessarily reveal speech specific processing regions as the sensitivity to prototype vs. nonprototype and phonemic vs. nonphonemic vowels could be related to differences in acoustical processing requirements for the individual task conditions. For example, during vowel discrimination task on prototypical vowel pairs, more activation was observed in primary AC regions probably because the vowels around a prototype are perceived more similar than vowels around a nonprototype (the perceptual magnet effect, Kuhl 1991). Due to this perceptual similarity, the discrimination task on prototype vowel pairs may have required more detailed acoustic analysis of the vowels. Similarly, nonphonemic vowels could require more processing, as the processing of phonemic vowels could be facilitated by existing phonemic representations. Thus, although the distinction between prototype and nonprototype vowels as well as between phonemic and nonphonemic vowels exists only at the language-specific level, the activation differences observed during discrimination tasks with prototypical vs. nonprototypical vowels are not necessarily due to language-specific processing as such.

4.2 CATEGORICAL PROCESSING IN IPL?

Prior fMRI studies have implicated areas in IPL in categorical processing of auditory stimuli. For example, Dehaene-Lambertz et al. (2005) and Raizada and Poldrack (2007) report that IPL was more strongly activated by a phoneme category change than by an acoustic change of the speech sounds. Turkeltaub and Coslett (2010) showed in their meta-analysis on sublexical speech perception (i.e., phonemes and syllables) that IPL was systematically activated in fMRI studies related to categorical phoneme perception. In contrast, Rinne and colleagues (2009, 2012) showed that categorical n-back tasks performed on pitch-varying or location-varying (non-vowel) sounds were associated with enhanced activations in IPL, suggesting that categorical processing in IPL is not specific to speech stimuli.

In Study II, IPL activations were stronger during categorical n-back tasks (in comparison with discrimination tasks) and these activations increased with increasing load in the n-back task. No significant differences in IPL activations were found, however, when the categorical 2-back task was performed on phonemic or nonphonemic vowels although it was expected that the load on categorical processing would be higher when the task was performed on nonphonemic vowels that should be harder to categorize than the phonemic vowels. Thus, if the requirements for categorical processing were higher during categorical 2-back task with nonphonemic vowels, these results suggest that IPL is not associated with categorization as such but rather with general operations on the categorical representations. The enhanced STG activations observed during categorical n-back tasks performed on nonphonemic vowels suggest that the categories are resolved in these regions.

Study III further tested the role of IPL in categorical processing from a slightly different perspective. Would IPL show enhanced activations also during a discrimination task if the task required categorical processing (in addition to acoustic analysis)? To investigate this question, activations to identical vowel pairs were compared during discrimination (subjects indicated whether vowels in a pair were acoustically the same or different) and category discrimination (subjects indicated whether vowels in a pair belonged to the same or different category) task. During a categorical vowel discrimination task, activations were significantly enhanced in IPL, whereas stronger activations were observed in STG during (acoustic) discrimination task. These results show that task-dependent activations of AC and IPL are strongly modulated depending on whether the task requires acoustic or categorical analysis of the speech sounds. To summarize, the present results indicate that while IPL areas are strongly activated during task requiring processing of categorical information, IPL does not seem to be involved in categorization as such. Rather, the present results suggest that categories are resolved and category labels are obtained earlier in the auditory system (in STG).

4.3 THE IMPORTANCE OF STIMULUS AND TASK CONTROL

Previous studies on speech processing in human AC have often compared activation to speech and nonspeech stimuli during passive listening or easy vigilance tasks. In such studies, the interpretation of the results may be affected by the effects associated with the (unavoidable) acoustic differences between the speech and nonspeech stimuli. If the speech and nonspeech stimuli are not well matched acoustically, it is difficult to resolve, whether the possible activation differences are associated with phonetic or acoustic processing of the stimuli (Desai et al., 2008; Liu and Holt, 2011). Further, activations during passive listening conditions could easily be affected by uncontrolled task and attention effects. AC activations are strongly modulated by auditory attention (Petkov et al., 2004; Rinne et al., 2007, 2008, 2009; Woodruff et al., 1996; Woods et al., 2009). Behaviorally relevant and familiar stimuli, such as speech, could be associated with involuntary attention effects.

In the present thesis, stimulus-related effects and differences were carefully controlled. All vowels were synthesized using a similar method and they were acoustically quite similar. The prototypical and nonprototypical vowels as well as phonemic categories and category boundaries (in Studies I and III) were individually defined for each subject. In Study III discrimination, category discrimination and categorical 2-back tasks were all performed with identical within-category and cross-category vowel pairs. Importantly, all critical comparisons were made across conditions in which the vowel stimuli were acoustically very similar or identical.

Task-related effects, in turn, were controlled as follows. In all studies, (1) activations to vowels were investigated during demanding visual (no directed auditory attention) and auditory tasks in order to separately investigate stimulus-dependent activations and effects associated with active listening. (2) In Study II, task difficulty in categorical n-back tasks was systematically modulated. Task-difficulty was associated with similar IPL activation increase irrespective of whether the task was performed on phonemic or nonphonemic vowel pairs. Such systematic task-difficulty effect supports the interpretation that similar resources in IPL were used in both cases. (3) Comparison of activations during different tasks makes the interpretation of the results easier. For example, any stimulus-dependent effects should be equally present in all auditory and visual task conditions. Further, this allows one to separate general auditory attention effects (present during all auditory tasks) from task-specific effects (e.g., difference between discrimination and n-back tasks). Furthermore, manipulation of task requirements allowed also analysis of the functional significance of the activations. (4) In the present studies, the critical comparisons were made across conditions requiring identical motor responding. This is important as previous studies have

shown that AC activations are modulated by motor operations (Ehrsson et al., 2000; Grezès et al., 2003; Schneider et al., 2014; Wikman et al., 2015). Thus, activation differences in the present studies cannot be due to different motor operations. It is noteworthy that, in the present thesis, Study III successfully replicated the main findings of Studies I and II. This is highly important especially in the light of recent debate on inflated false-positive rates in fMRI studies (Monti, 2011; Eklund et al., 2016).

4.4 CONCLUSIONS

The present thesis used fMRI to investigate the sensitivity of AC to phenomena related to categorical vowel representations. The main results of Studies I–III are as follows: First, the results support the theoretical view that phonemic vowels are represented in a categorical manner in the brain. Further, the studies show that information about categorical vowel representations is present in human AC. Second, AC activations for vowels strongly depend on the characteristics of the listening tasks. The effects associated with the contrasts between prototype vs. nonprototype and phonemic vs. nonphonemic vowels are most prominent during active tasks. Third, the present results strongly emphasize the importance of carefully controlled experiments as surprisingly small differences in the characteristics of vowel stimuli or tasks performed on these vowels can result in strong activation differences in AC and adjacent areas. Fourth, the results support the view that IPL is involved in categorical processing. IPL activations, however, are not associated with categorization as such but rather with operations on categorical representations. Finally, the results of the present studies suggest that replications, albeit often neglected, are a fundamental part of systematic research.

REFERENCES

- Aaltonen, O. & Eerola, O. & Hellstrom, A. & Uusipaikka, E. & Lang, A. H. 1997. Perceptual magnet effect in the light of behavioral and psychophysiological data. *The Journal of the Acoustical Society of America*, 101(2), 1090-1105.
- Alho, K. & Rinne, T. & Herron, T. J. & Woods, D. L. 2014. Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. *Hearing Research*, 307, 29-41.
- Angenstein, N. & Scheich, H. & Brechmann, A. 2012. Interaction between bottom-up and top-down effects during the processing of pitch intervals in sequences of spoken and sung syllables. *Neuroimage*, 61(3), 715-722.
- Assmann, P. F. & Summerfield, Q. 1989. Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency. *The Journal of the Acoustical Society of America*, 85(1), 327-338.
- Benson, R. R. & Richardson, M. & Whalen, D. H. & Lai, S. 2006. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage*, 31, 342-353.
- Benson, R. R. & Whalen, D. H. & Richardson, M. & Swainson, B. & Clark, V. P. & Lai, S. & Liberman, A. M. 2001. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78(3), 364-396.
- Best, C. T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, 167(224), 233-277.
- Binder, J. R. & Frost, J. A. & Hammeke, T. A. & Bellgowan, P. S. & Springer, J. A. & Kaufman, J. N. & Possing, E. T. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10(5), 512-528.
- Bruder, J. & Leppanen, P. H. T. & Bartling, J. & Csepe, V. & Demonet, J. F. & Schulte-Korne, G. 2011. An investigation of prototypical and atypical within-category vowels and non-speech analogues on cortical auditory

- evoked related potentials (AERPs) in 9year old children. *International Journal of Psychophysiology*, 79(2), 106-117.
- Chang, E. F. & Rieger, J. W. & Johnson, K. & Berger, M. S. & Barbaro, N. M. & Knight, R. T. 2010. Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428-1432.
- Cooper, F. S. & Delattre, P. C. & Liberman, A. M. & Borst, J. M. & Gerstman, L. J. 1952. Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, 24(6), 597-606.
- 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(1), 21-33.
- Delattre, P. & Liberman, A. M. & Cooper, F. S. & Gerstman, L. J. 1952. An experimental study of the acoustic determinants of vowel color; observations on one-and two-formant vowels synthesized from spectrographic patterns. *Word*, 8(3), 195-210.
- Demonet, J. F. & Chollet, F. & Ramsay, S. & Cardebat, D. & Nespoulous, J. L. & Wise, R. & Rascol, A. & Frackowiak, R. 1992. The anatomy of phonological and semantic processing in normal subjects. *Brain*, 115(6), 1753-1768.
- Demonet, J. F. & Thierry, G. & Cardebat, D. 2005. Renewal of the neurophysiology of language: Functional neuroimaging. *Physiological Reviews*, 85(1), 49-95.
- Desai, R. & Liebenthal, E. & Waldron, E. & Binder, J. R. 2008. Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, 20(7), 1174-1188.
- DeWitt, I. & Rauschecker, J. P. 2012. Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, 109(8), E505-E514.
- Diehl, R. L. 1987. Auditory constraints on speech perception. *The psychophysics of speech perception* (pp. 210-219). Springer Netherlands.

- Diehl, R. L. & Kluender, K. R. 1989. On the objects of speech perception. *Ecological Psychology*, 1(2), 121-144.
- Edmonds, B. A. & James, R. E. & Utev, A. & Vestergaard, M. D. & Patterson, R. D. & Krumbholz, K. 2010. Evidence for early specialized processing of speech formant information in anterior and posterior human auditory cortex. *The European Journal of Neuroscience*, 32(4), 684-692.
- Ehrsson, H. H. & Fagergren, A. & Jonsson, T. & Westling, G. & Johansson, R. S. & Forssberg, H. 2000. Cortical activity in precision- versus power-grip tasks: An fMRI study. *Journal of Neurophysiology*, 83(1), 528-536.
- Eimas, P. D. 1963. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 6(4), 206-217.
- Eklund, A. & Nichols, T. E. & Knutsson, H. 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(28), 7900-7905.
- Fant, G. 1960 *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton.
- Fant, G. 1971. *Acoustic Theory of Speech Production: With calculations based on X-ray studies of russian articulations* (Vol 2). Walter de Gruyter.
- Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277.
- Friederici, A. D. 2011. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4), 1357-1392.
- Fritz, J. B. & Elhilali, M. & Shamma, S. A. 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *The Journal of Neuroscience*, 25(33), 7623-7635.
- Giraud, A. L. & Kell, C. & Thierfelder, C. & Sterzer, P. & Russ, M. O. & Preibisch, C. & Kleinschmidt, A. 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral cortex*, 14(3), 247-255.

- Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, 9(3), 317-323.
- Grady, C. L. & Van Meter, J. W. & Maisog, J. M. & Pietrini, P. & Krasuski, J. & Rauschecker, J. P. 1997. Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport*, 8(11), 2511-2516.
- Grezes, J. & Tucker, M. & Armony, J. & Ellis, R. & Passingham, R. E. 2003. Objects automatically potentiate action: An fMRI study of implicit processing. *The European Journal of Neuroscience*, 17(12), 2735-2740.
- Guenther, F. H. & Nieto-Castanon, A. & Ghosh, S. S. & Tourville, J. A. 2004. Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research*, 47(1), 46-57.
- Hackett, T. A. 2011. Information flow in the auditory cortical network. *Hearing Research*, 271(1), 133-146.
- Hakonen, M. & May, P. J. & Alho, J. & Alku, P. & Jokinen, E. & Jääskeläinen, I. P. & Tiitinen, H. 2016. Previous exposure to intact speech increases intelligibility of its digitally degraded counterpart as a function of stimulus complexity. *NeuroImage*, 125, 131-143.
- Hall, D. A. & Haggard, M. P. & Akeroyd, M. A. & Summerfield, A. Q. & Palmer, A. R. & Elliott, M. R. & Bowtell, R. W. 2000. Modulation and task effects in auditory processing measured using fMRI. *Human Brain Mapping*, 10(3), 107-119.
- Harnad, S. 1987. Psychophysical and cognitive aspects of categorical perception: A critical overview. *Categorical Perception: The Groundwork of Cognition* (pp. 1-52).
- Harnad, S. 2003. Categorical perception. *Encyclopedia of Cognitive Science*, 67(4).
- Hickok, G. 2009. The functional neuroanatomy of language. *Physics of Life Reviews*, 6(3), 121-143.
- Hickok, G. & Okada, K. & Serences, J. T. 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, 101(5), 2725-2732.

- Hickok, G. & Poeppel, D. 2004. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
- Hickok, G. & Poeppel, D. 2007. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- Hickok, G. & Saberi, K. 2012. Redefining the functional organization of the planum temporale region: space, objects, and sensory-motor integration. In *The human auditory cortex* (pp. 333-350). Springer New York.
- Hillenbrand, J. & Getty, L. A. & Clark, M. J. & Wheeler, K. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5), 3099-3111.
- Huotilainen, M. & Kujala, A. & Alku, P. 2001. Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neuroscience Letters*, 310(2), 133-136.
- Husain, F. T. & Fromm, S. J. & Pursley, R. H. & Hosey, L. A. & Braun, A. R. & Horwitz, B. 2006. Neural bases of categorization of simple speech and nonspeech sounds. *Human Brain Mapping*, 27(8), 636-651.
- Hutchison, E. R. & Blumstein, S. E. & Myers, E. B. 2008. An event-related fMRI investigation of voice-onset time discrimination. *Neuroimage*, 40(1), 342-352.
- Iverson, P. & Kuhl, P. K. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*, 97(1), 553-562.
- Iverson, P. & Kuhl, P. K. 1996. Influences of phonetic identification and category goodness on american listeners' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*, 99(2), 1130-1140.
- Iverson, P. & Kuhl, P. K. 2000. Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, 62(4), 874-886.
- 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(1), B47-57.

- Jäncke, L. & Wustenberg, T. & Scheich, H. & Heinze, H. J. 2002. Phonetic perception and the temporal cortex. *Neuroimage*, 15(4), 733-746.
- Kilian-Hütten, N. & Valente, G. & Vroomen, J. & Formisano, E. 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. *The Journal of Neuroscience*, 31(5), 1715-1720.
- Klatt, D. 1982. Prediction of perceived phonetic distance from critical-band spectra: A first step. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82*, 7, 1278-1281.
- Kuhl, P. K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2), 93-107.
- Kuhl, P. K. 1994. Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812-822.
- Kuhl, P. K. & Williams, K. A. & Lacerda, F. & Stevens, K. N. & Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.
- Leaver, A. M. & Rauschecker, J. P. 2010. Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience*, 30(22), 7604-7612.
- Lee, Y. S. & Turkeltaub, P. & Granger, R. & Raizada, R. D. 2012. Categorical speech processing in broca's area: An fMRI study using multivariate pattern-based analysis. *The Journal of Neuroscience*, 32(11), 3942-3948.
- Leung, A. W. & Alain, C. 2011. Working memory load modulates the auditory "What" and "Where" neural networks. *Neuroimage*, 55(3), 1260-1269.
- Lieberman, 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* 54(5), 358-368.
- Lieberman, A. M. & Cooper, F. S. & Shankweiler, D. P. & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review*, 74(6), 431.

- Lieberman, A. M. & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.
- Liebenthal, E. & Binder, J. R. & Spitzer, S. M. & Possing, E. T. & Medler, D. A. 2005. Neural substrates of phonemic perception. *Cerebral Cortex*, 15(10), 1621-1631.
- Liu, R. & Holt, L. L. 2011. Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. *Journal of Cognitive Neuroscience*, 23(3), 683-698.
- Logothetis, N. K. & Pauls, J. & Augath, M. & Trinath, T. & Oeltermann, A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150-157.
- Logothetis, N. K. 2008. What we can do and what we cannot do with fMRI. *Nature*, 453(7197), 869-878.
- Lomber, S. G. & Malhotra, S. 2008. Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature Neuroscience*, 11(5), 609-616.
- Miller, J. D. 1989. Auditory-perceptual interpretation of the vowel. *The Journal of the Acoustical Society of America*, 85(5), 2114-2134.
- Moerel, M. & De Martino, F. & Formisano, E. 2014. An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, 8, 225.
- Monti, M. M. 2011. Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, 5, 28.
- Mukamel, R. & Gelbard, H. & Arieli, A. & Hasson, U. & Fried, I. & Malach, R. 2005. Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science*, 309(5736), 951-954.
- Möttönen, R. & Calvert, G. A. & Jääskeläinen, I. P. & Matthews, P. M. & Thesen, T. & Tuomainen, J. & Sams, M. 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage*, 30(2), 563-569.
- Näätänen, R. & Lehtokoski, A. & Lennes, M. & Cheour, M. & Huotilainen, M. & Iivonen, A. & Vainio, M. & Alku, P. & Ilmoniemi, R. J. & Luuk, A. 1997.

- Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432-434.
- Narain, C. & Scott, S. K. & Wise, R. J. & Rosen, S. & Leff, A. & Iversen, S. D. & Matthews, P. M. 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13(12), 1362-1368.
- Obleser, J. & Boecker, H. & Drzezga, A. & Haslinger, B. & Hennenlotter, A. & Roettinger, M. & Eulitz, C. & Rauschecker, J. P. 2006. Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*, 27(7), 562-571.
- Obleser, J. & Wise, R. J. & Alex Dresner, M. & Scott, S. K. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, 27(9), 2283-2289.
- Ogawa, S. & Lee, T. M. & Kay, A. R. & Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24), 9868-9872.
- Okada, K. & Rong, F. & Venezia, J. & Matchin, W. & Hsieh, I. H. & Saberi, K. & Serences, J. T. & Hickok, G. 2010. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, 20(10), 2486-2495.
- Peterson, G. E. & Barney, H. L. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175-184.
- Petkov, C. I. & Kang, X. & Alho, K. & Bertrand, O. & Yund, E. W. & Woods, D. L. 2004. Attentional modulation of human auditory cortex. *Nature Neuroscience*, 7(6), 658-663.
- Pisoni, D. B. 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.
- Poeppl, D. & Hickok, G. 2004. Towards a new functional anatomy of language. *Cognition*, 92(1), 1-12.

- Raizada, R. D. & Poldrack, R. A. 2007. Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, 56(4), 726-740.
- Rauschecker, J. P. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, 271(1-2), 16-25.
- Rauschecker, J. P. & Scott, S. K. 2009. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-724.
- Repp, B. H. 1984. Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice*, 10, 243-335.
- 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(5), 1113-1117.
- Rinne, T. & Balk, M. H. & Koistinen, S. & Autti, T. & Alho, K. & Sams, M. 2008. Auditory selective attention modulates activation of human inferior colliculus. *Journal of Neurophysiology*, 100(6), 3323-3327.
- Rinne, T. & Kirjavainen, S. & Salonen, O. & Degerman, A. & Kang, X. & Woods, D. L. & Alho, K. 2007. Distributed cortical networks for focused auditory attention and distraction. *Neuroscience Letters*, 416(3), 247-251.
- Rinne, T. & Koistinen, S. & Salonen, O. & Alho, K. 2009. Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *The Journal of Neuroscience*, 29(42), 13338-13343.
- Rinne, T. & Koistinen, S. & Talja, S. & Wikman, P. & Salonen, O. 2012. Task-dependent activations of human auditory cortex during spatial discrimination and spatial memory tasks. *Neuroimage*, 59(4), 4126-4131.
- Rosner, B. S. & Pickering, J. B. 1994. *Vowel perception and production*. Oxford University Press.
- Saenz, M. & Langers, D. R. 2014. Tonotopic mapping of human auditory cortex. *Hearing Research*, 307, 42-52.

- Samuel, A. G. 1982. Phonetic prototypes. *Perception & Psychophysics*, 31, 307-314.
- Scheich, H. & Brechmann, A. & Brosch, M. & Budinger, E. & Ohl, F. W. 2007. The cognitive auditory cortex: task-specificity of stimulus representations. *Hearing research*, 229(1), 213-224.
- Schneider, D. M. & Nelson, A. & Mooney, R. 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517), 189-194.
- Scott, S. K. & Johnsrude, I. S. 2003. The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100-107.
- Sharma, A. & Dorman, M. F. 1998. Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *The Journal of the Acoustic Society of America*, 104(1), 511-517.
- Staeren, N. & Renvall, H. & De Martino, F. & Goebel, R. & Formisano, E. 2009. Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology*, 19(6), 498-502.
- Stevens, K. N. 1981. Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics, and psychoacoustics. *Advances in Psychology*, 7, 61-74.
- Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.
- Tiitinen, H. & Miettinen, I. & Alku, P. & May, P. J. 2012. Transient and sustained cortical activity elicited by connected speech of varying intelligibility. *BMC neuroscience*, 13(1), 157.
- Turkeltaub, P. E. & Coslett, H. B. 2010. Localization of sublexical speech perception components. *Brain and Language*, 114(1), 1-15.
- Uppenkamp, S. & Johnsrude, I. S. & Norris, D. & Marslen-Wilson, W. & Patterson, R. D. 2006. Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage*, 31(3), 1284-1296.

- 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(7), 994-1005.
- Weinberger, N. M. 2011. Reconceptualizing the primary auditory cortex: Learning, memory and specific plasticity. *The auditory cortex* (pp. 465-491). Springer US.
- Werker, J. F. & Tees, R. C. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866-1878.
- Whalen, D. H. & Benson, R. R. & Richardson, M. & Swainson, B. & Clark, V. P. & Lai, S. & Mencl, W. E. & Fulbright, R. K. & Constable, R. T. & Liberman, A. M. 2006. Differentiation of speech and nonspeech processing within primary auditory cortex. *The Journal of the Acoustical Society of America*, 119(1), 575-581.
- Wikman, P. A. & Vainio, L. & Rinne, T. 2015. The effect of precision and power grips on activations in human auditory cortex. *Frontiers in Neuroscience*, 9, 378.
- Winkler, I. & Lehtokoski, A. & Alku, P. & Vainio, M. & Czigler, I. & Csepe, V. & Aaltonen, O. & Raimo, I. & Alho, K. & Lang, H. & Iivonen, A. & Näätänen, R. 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Brain Research. Cognitive Brain Research*, 7(3), 357-369.
- Woodruff, P. W. & Benson, R. R. & Bandettini, P. A. & Kwong, K. K. & Howard, R. J. & Talavage, T. & Belliveau, J. & Rosen, B. R. 1996. Modulation of auditory and visual cortex by selective attention is modality-dependent. *Neuroreport*, 7(12), 1909-1913.
- Woods, D. L. & Herron, T. J. & Cate, A. D. & Kang, X. & Yund, E. W. 2011. Phonological processing in human auditory cortical fields. *Frontiers in Human Neuroscience*, 5, 42.
- Woods, D. L. & Herron, T. J. & Cate, A. D. & Yund, E. W. & Stecker, G. C. & Rinne, T. & Kang, X. 2010. Functional properties of human auditory cortical fields. *Frontiers in Systems Neuroscience*, 4, 155.

Zatorre, R. J. & Evans, A. C. & Meyer, E. & Gjedde, A. 1992. Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846-849.