HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2020 July 24.

Published in final edited form as:

Neuroimage. 2016 April 01; 129: 214–223. doi:10.1016/j.neuroimage.2016.01.016.

Early-latency categorical speech sound representations in the left inferior frontal gyrus

Jussi Alho[a,*], **Brannon M. Green**[b], **Patrick J. C. May**[c], **Mikko Sams**[a], **Hannu Tiitinen**[a], **Josef P. Rauschecker**[a,b,d], **Iiro P. Jääskeläinen**[a,e,f,*]

[a]Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, 00076 AALTO, Espoo, Finland [b]Laboratory of Integrated Neuroscience and Cognition, Interdisciplinary Program in Neuroscience, Georgetown University Medical Center, Washington, DC 20057 [c]Special Laboratory Non-Invasive Brain Imaging, Leibniz Institute for Neurobiology, Brenneckestraße 6, D-39118 Magdeburg, Germany [d]Institute for Advanced Study, TUM, Munich-Garching, 80333 Munich, Germany [e]MEG Core, Aalto NeuroImaging, Aalto University, 00076 AALTO, Espoo, Finland [f]AMI Centre, Aalto NeuroImaging, Aalto University, 00076 AALTO, Espoo, Finland

Abstract

Efficient speech perception requires the mapping of highly variable acoustic signals to distinct phonetic categories. How the brain overcomes this many-to-one mapping problem has remained unresolved. To infer the cortical location, latency, and dependency on attention of categorical speech sound representations in the human brain, we measured stimulus-specific adaptation of neuromagnetic responses to sounds from a phonetic continuum. The participants attended to the sounds while performing a non-phonetic listening task and, in a separate recording condition, ignored the sounds while watching a silent film. Neural adaptation indicative of phoneme category selectivity was found only during the attentive condition in the pars opercularis (POp) of the left inferior frontal gyrus, where the degree of selectivity correlated with the ability of the participants to categorize the phonetic stimuli. Importantly, these category-specific representations were activated at an early latency of 115–140 ms, which is compatible with the speed of perceptual phonetic categorization. Further, concurrent functional connectivity was observed between POp and posterior auditory cortical areas. These novel findings suggest that when humans attend to speech, the left POp mediates phonetic categorization through integration of auditory and motor information *via* the dorsal auditory stream.

Keywords

speech perception; categorical perception; magnetoencephalography; inferior frontal gyrus; stimulus-specific adaptation

[*]**Correspondence**: Jussi Alho, Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, PO Box 12200, FI-00076 Aalto, Finland. jussi.alho@aalto.fi, Iiro Jääskeläinen, Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, PO Box 12200, FI-00076 Aalto, Finland. iiro.jaaskelainen@aalto.fi.

1 Introduction

Phonemes, the elementary units of speech, greatly vary in their acoustic structure when produced by different speakers in different contexts. The brain therefore faces a fundamental challenge of mapping highly variable acoustic signals to distinct phonetic categories. Although categorical perception of speech sounds is well documented, it remains unresolved how the brain accomplishes this many-to-one mapping. Specifically, it is unclear which cortical areas exhibit categorical speech processing and at what latencies from sound onset this occurs.

Current theories postulate that speech is cortically processed by parallel ventral and dorsal auditory streams (Rauschecker, 1998a, b; Wise, 2003; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). The ventral stream, involving superior-to-middle temporal areas and terminating in pars triangularis (PTr; roughly corresponding to Brodmann area [BA] 45) of the inferior frontal gyrus (IFG), has been suggested to process speech signals for comprehension, whereas the dorsal stream, projecting from auditory cortex via the temporoparietal junction to premotor cortex (PMC) and pars opercularis (POp; BA 44) of IFG, has been proposed to mediate a mapping between auditory and articulatory-motor representations (Rauschecker, 2011). Given that each human has a repertoire of potential speech gestures which is less variable than the mass of acoustic speech signals one has to categorize (Liberman et al., 1967; Liberman and Mattingly, 1985), it can be hypothesized that categorical speech representations (CSR) are found in the speech-motor areas (e.g. POp/PMC) and that they guide speech categorization *via* the dorsal stream.

The sensorimotor nature of speech processing is supported by empirical findings whereby disrupting speech-motor areas with transcranial magnetic stimulation (TMS) impairs speech sound discrimination or categorization (Meister et al., 2007; Möttönen and Watkins, 2009; Sato et al., 2009; D'Ausilio et al., 2012; Grabski et al., 2013). Furthermore, stimulus-specific adaptation (SSA) of functional magnetic resonance imaging (fMRI) signals revealed CSR in the left PMC (Chevillet et al., 2013), POp (Myers et al., 2009; Lee et al., 2012) and anterior insula (aINS) (Myers et al., 2009). Importantly, the categorical processing in these fMRI studies was task-independent, as subjects engaged in a listening task wherein phoneme category information was irrelevant.

Lower-level phonological processing areas in temporal and parietal lobes have also been implicated in categorical speech processing. FMRI-adaptation revealed CSR in the left supramarginal gyrus (SMG) when the auditory input was attended to (Raizada and Poldrack, 2007). However, re-analysis of these same data with multivariate rather than univariate techniques revealed CSR in POp, rather than in SMG, with the discrepancy supposedly due to different spatial scales of cortical representations in different dorsal-stream areas (Lee et al., 2012). A further fMRI-adaptation experiment where subjects watched a film without the soundtrack and were under instruction to ignore the sounds found CSR in the left superior temporal sulcus (STS) (Joanisse et al., 2007). Electrocorticographic (ECoG) recordings from the posterior superior temporal gyrus (STG) revealed CSR during passive listening to speech sounds with small acoustic differences (Chang et al., 2010). The left mSTS was associated with phonemic perception being more strongly activated by familiar speech sounds than

acoustically (i.e. spectro-temporally) matched non-phonemic sounds (Liebenthal et al., 2005). Further, a study using combined fMRI and electroencephalography (EEG) suggested that categorization of highly familiar (e.g. native) and newly acquired speech sounds rely on long-term representations in mSTS and short-term representations in pSTS, respectively (Liebenthal et al., 2010). Another fMRI-study identified category-selective responses to speech sounds in anterior superior temporal regions (Leaver and Rauschecker, 2010). In line with these findings, a recent meta-analysis localized invariant phoneme representations consistently in anterior-to-mid STG (DeWitt and Rauschecker, 2012).

As evidence for categorical perception of phonemes has been found both within the ventral and dorsal streams, it seems plausible that invariant representations are formed independently based on both spectro-temporal and articulatory-motor information (for review, see Rauschecker, 2012). An intriguing question is what determines the engagement of the two respective streams in speech categorization. Previous research has proposed a modulatory role for the dorsal stream in speech perception, particularly in the learning of new sound categories (Liebenthal et al., 2010), under adverse listening conditions (Osnes et al., 2011; Du et al., 2014), or during sublexical tasks, such as syllable discrimination (Hickok and Poeppel, 2007). However, none of these conditions were present in the above-mentioned studies reporting CSR in the dorsal stream areas, which raises the possibility that the discrepant results between the ventral and dorsal stream involvement in speech categorization could be explained by differences in allocation of auditory attention. In support of this interpretation, a recent study using TMS and magnetoencephalography (MEG) demonstrated that the involvement of articulatory-motor areas in the early (<100 ms) processing of acoustic-phonetic features of speech depended on attention, while the longer-latency auditory-motor interaction (>170 ms) occurred even when the subjects were under instruction to ignore the sounds and to focus on watching a silent film (Möttönen et al., 2014).

Here, we used SSA and cortically-constrained MEG source estimates to infer the location, latency, and attention-dependence of CSR. Sounds from a phonetic continuum were presented to participants while they were performing a non-phonetic listening task and, in a separate passive recording condition, ignoring the sounds while watching a film without the soundtrack. The following questions were addressed: Are CSR observed in speech-motor areas regardless of auditory attention? Is the latency of CSR compatible with the proposal that phonological categories are accessed ~150 ms after sound onset (Salmelin, 2007)? Does the neural selectivity underlying CSR correlate with behavioral categorization?

2 Materials and Methods

2.1 Participants

All 22 subjects were right-handed and reported neither a history of hearing problems nor neurological illnesses. MEG data from four subjects were excluded from analyses due to poor a signal-to-noise ratio (SNR). This resulted in a final sample of 18 subjects (6 females; age mean ± SD = 25.3 ± 4.0, range 21–38 years). The experiment was approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa, Finland.

2.2 Stimuli

The present study utilized the same stimulus material as that in Chevillet et al. (2013), that is, a place-of-articulation continuum between the natural utterances /da/ and /ga/ (Fig. 1). Place-of-articulation refers to the point of maximum obstruction in the vocal tract in the articulation of a consonant. The stimuli were produced with the STRAIGHT toolbox (Kawahara and Matsui, 2003) for MATLAB (MathWorks), which allows for the parametric manipulating of the acoustic and acoustic-phonetic structure of natural voice recordings. The natural utterances were taken from recordings provided by Shannon et al. (1999). Two phonetic continua (or "morphlines") were generated at 0.5% intervals between the /da/ and /ga/ prototypes: one for a male voice and the other for a female voice. Morphed stimuli were generated up to 25% beyond each natural utterance (i.e. from −25% /ga/ to +125% /ga/), for a total of 301 stimuli per morphline. The stimuli created beyond the natural utterances were qualitatively assessed to ensure their intelligibility and behaviorally verified in a categorization test (described below). All stimuli were resampled to 48 kHz, trimmed to 300-ms duration, and root-mean-square normalized in amplitude. A linear amplitude ramp of 10-ms duration was applied to sound offsets to avoid auditory artefacts. Amplitude ramps were not applied to sound onsets so as to avoid interfering with the natural features of the consonant sound.

2.3 Discrimination behavior

Prior to brain imaging, the subjects completed a discrimination test to identify individual category boundaries. The discrimination thresholds of the subjects were measured at 10% intervals along both male and female voice continua. The adaptive staircase algorithm QUEST (Watson and Pelli, 1983), implemented in MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), was used to adjust the difference between paired stimuli based on subject performance. This allowed the measurement of the just-noticeable difference (JND) at each location (for both morph directions), which is known to have its minimum value at category boundaries. To diminish the risk that the subjects would categorize the sounds during MEG, the task on each trial was to report as quickly and accurately as they could whether the two sounds were exactly the same or in any way different without assigning them to a specific phonetic category. A maximum period of 3 seconds was allowed for a response before the next trial started. In half of the trials, the paired stimuli were identical and in the other half they were different from each other. Of the pairs where the stimuli differed from each other, half represented a displacement in one direction along the continuum, and the other half a displacement in the opposing direction. In total, 560 trials were presented, with 20 conditions (10% intervals from 0 to 90% with displacements toward 100%, and 10% intervals from 10 to 100% with displacement toward 0%) and 28 trials per condition.

2.4 Categorization behavior

After brain imaging, the subjects were asked to categorize the auditory stimuli along both morphlines to confirm the location of their individual category boundary as well as to measure its sharpness. Categorization was tested at 10% intervals from −25% (i.e. 25% past /da/ away from /ga/) to +125% (25% past /ga/ away from /da/). On each trial, the

subject was presented with a single sound and given up to 3 seconds to indicate as quickly and accurately as possible whether s/he had heard /da/ or /ga/. Each subject completed 15 runs of 20 trials per condition, for a total of 300 trials per morphline. The resulting data were fitted with subject-specific sigmoidal functions to estimate boundary locations as well as boundary sharpness. The sigmoid was given by the generalized logistic curve:

$$f(x) = \frac{1}{1 + e^{-(x-\alpha)/\beta}},$$

where x is the location along the morph line, α is the location of the boundary along the morph line, and 1/β is the steepness of the boundary (with lower values of β resulting in sharper boundaries).

2.5 MEG paradigm

To infer neuronal stimulus selectivity, the SSA paradigm was used. In SSA, two stimuli – an adaptor and a probe – are presented in succession in each trial, and the similarity between the two stimuli is varied between trials to investigate neuronal tuning along the dimensions of interest (Butler, 1972). In this setup, the attenuation of the response to the probe reflects the overlap between the neural populations responding to the adaptor and the probe, respectively. In the present study, the silent interval between the adaptor and the probe was 500 ms and the interval between successive adaptor-probe pairs varied randomly and uniformly in the 5–7 second range. For each subject, four sounds along the morphline were selected on an individual basis according to the pre-imaging behavioral discrimination test. These sounds were combined into the following four adaptor-probe pairs defined by the acoustic-phonetic change and the phoneme category change between the adaptor and the probe: (1) identical sounds (ID), (2) 33% acoustic-phonetic difference, same category (33S), (3) 33% acoustic-phonetic difference, different category (33D), and (4) 67% acoustic-phonetic difference, different category (67D) (Fig. 1). Thus, any difference in adaptation between 33S and 33D, which were equalized with regard to acoustic-phonetic dissimilarity, can be attributed to an explicit representation of the phoneme categories.

Brain regions containing category-selective neurons should show larger adaptation (i.e. reductions of the probe response in relation to the adaptor response) in the 33S trials than in the 33D trials, as the stimuli in each pair in the latter condition would activate different neuronal populations, whereas the same neuronal populations would be activated in the former condition. In this way, SSA enables the dissociation of phoneme category selectivity from mere tuning to acoustic-phonetic differences. On the one hand, category-selective neurons respond similarly to dissimilar stimuli from the same category but differently to similar stimuli belonging to different categories (Freedman et al., 2003; Jiang et al., 2007). On the other hand, in the case of tuning to acoustic-phonetic differences, neuronal responses gradually drop off with acoustic dissimilarity, without the sharp transition at the category boundary that is the hallmark of perceptual categorization. Morphlines were extended beyond the prototypes (25% in each direction) so that the actual stimuli used to create the stimulus pairs for each subject would span 100% of the difference between /da/ and /ga/ but could be shifted so that they were centered at the category boundary for each subject.

To observe responses independent of overt phoneme categorization, the subjects were scanned while they performed an attention-demanding distractor task for which phoneme category information was irrelevant (from here on referred to as the ATTEND condition). For this reason, each presented sound in a pair persisted 30 ms longer in one ear than in the other. In ATTEND, the subject was asked to listen out for these offsets and to report whether the two sounds in the pair persisted longer in either the same ear or in different ears. The reporting was done by pressing a button either with the left or right hand, indicating either "same" or "different". To disentangle activation resulting from categorical decisions from that associated with categorical motor activity (i.e. to average out the motor responses), the label of the left- and right-hand side responses alternated on each run (i.e. whether left-hand side response indicated "same" or "different"). For controlling the effect of attention, the subjects were additionally scanned while they watched a film without soundtrack (a wildlife documentary) and were under instruction to ignore the sounds (from here on referred to as the IGNORE condition). To diminish the possibility that data obtained in the IGNORE condition could be partly contaminated by prior exposure of the stimuli, the interval between the discrimination test and brain imaging was at least two days and on average 21 days.

Both ATTEND and IGNORE conditions comprised 512 trials (128 for each adaptor-probe pair). The trial order was randomized and the number of presentations between all stimuli was equalized. With an average 6-s trial duration, the measurement time per condition totaled ~51 min, which was divided into eight ~6.4-min blocks to prevent fatigue. The measurements were divided over two days (one condition per day), with the order counterbalanced across subjects. The auditory stimuli were delivered through insert earphones (Etymotic Research Inc., IL, USA), comprising plastic tubing and earplugs, with the sound level set at 65 dB.

2.6 MEG data acquisition and preprocessing

The MEG data were acquired with a whole-head 306-channel neuromagnetometer (VectorView, Elekta-Neuromag, Helsinki, Finland) of the MEG Core of Aalto NeuroImaging infrastructure at Aalto University. The device was situated in a magnetically shielded room, with a three-layer μ-metal and aluminum cover to attenuate effects of outside magnetic fields. An additional active noise-cancelation system was used. Before each MEG recording session, the locations of five head-position indicator (HPI) coils attached to the scalp were recorded with respect to three anatomical landmark points (the nasion and two preauricular points) using a 3-D digitizer (Isotrak, Polhemus, Colchester, VT). Additional scalp surface points (~30) were digitized to facilitate the coregistration with anatomical magnetic resonance (MR) images. Vertical and horizontal electro-oculograms (EOG) were used to detect eye blinks and movements. The MEG signals were band-pass filtered at 0.03–330 Hz and digitized at 1 kHz sampling frequency. To compensate for the MEG signal changes due to movements, the head position of the subject was continuously tracked during the data acquisition by exciting the HPI coils at high frequencies (290–330 Hz).

During preprocessing, external noise was suppressed and head movements (estimated continuously at 200-ms intervals) were compensated using the signal-space separation method (Taulu and Simola, 2006) (Maxfilter, Elekta-Neuromag, Helsinki, Finland). MEG

signals time-locked to the onset of the stimuli were averaged across trials for each condition. Amplitudes were measured with respect to a 200-ms pre-adaptor baseline. Trials where the MEG gradiometer, MEG magnetometer, or EOG channel peak-to-peak amplitude exceeded 3000 fT/cm, 4000 fT, or 150 μV, respectively, were rejected from the average. For each subject, more than 100 trials per adaptor-probe pair and per condition were included in the analyses. The averaged signals were band-pass filtered between 1 and 40 Hz.

2.7 Structural magnetic resonance imaging (MRI)

The individual MR images were acquired with a 3T scanner (Magnetom Skyra, Siemens) of the AMI Centre of Aalto NeuroImaging infrastructure at Aalto University. Coregistration between MEG data and MRIs was done by identifying the fiducial point locations in the MRIs. FreeSurfer software was used to reconstruct the cortical mantle from the MRI data (Dale et al., 1999; Fischl et al., 1999a).

2.8 MEG source estimation

The source currents were estimated at each cortical location by computing a depth-weighted minimum-norm estimate (MNE) (Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006b). The MNE solution does not represent true brain activation but rather recovers a source distribution with minimum overall power that is consistent with the measured MEG signals. The forward solutions for all source locations were computed using a single-compartment boundary-element model (BEM) based on the information from individual structural MRIs and MEG sensor locations (Hämäläinen and Sarvas, 1989). The cortical surface of each subject was decimated to ca. 7000 source locations per hemisphere with an average 5-mm spacing between adjacent locations. A noise covariance matrix was estimated from 200-ms pre-adaptor stimulus baselines of the raw MEG data. Activity at each source location was estimated for each time point of the evoked response using an inverse operator computed from the forward solution and the noise covariance matrix. The source orientations were controlled with a loose orientation constraint (Lin et al., 2006a). In addition to MNEs, dynamic statistical parametric map (dSPM) estimates were generated (Dale et al., 2000). As a measure of signal-to-noise (derived through normalizing the MNE by the noise sensitivity at each cortical location), dSPM indicates the locations where MNE amplitudes are above noise level. Since individual MRI could not be obtained for one subject, a FreeSurfer average brain was used as a surrogate in this subject (by aligning the individual fiducial points to the fiducial points of the average head).

2.9 Spatiotemporal cluster analysis of evoked responses

Source analysis of evoked responses was conducted with a nonparametric randomization test based on spatiotemporal clustering (Maris and Oostenveld, 2007). The data were downsampled to 200 Hz and the individual cortical surfaces were morphed to a FreeSurfer average brain with 10242 dipoles per hemisphere (Fischl et al., 1999b). The medial wall of the cerebral cortex, as defined by an automatic parcellation (Desikan et al., 2006), was excluded from the analysis due to low SNR. To quantify the SSA effect, the adaptor-probe reduction rate R (i.e. strength of adaptation) was defined for a 500-ms time window as R = [adaptor response - probe response] / adaptor response.

A *t*-value was calculated for each data point (i.e. dipole/time point) for the given contrast between pairing conditions with a two-sided paired-samples *t*-test. All data points with *p*-value < 0.05 (uncorrected for multiple comparisons) were clustered on the basis of spatial and temporal adjacency. Cluster-level statistics were calculated by summing the *t*-values within every cluster and the maximum of the cluster-level statistics was used as the actual test statistic. A reference distribution of test statistics was produced by taking a thousand random partitions of the combined data across the conditions and calculating a test statistic for each partition. A multiple-comparisons-corrected cluster *p*-value was obtained by comparing the test statistic of the contrast of interest against the reference distribution. The null hypothesis of no difference between the conditions was rejected if *p*-value < 0.05. The analysis was performed in five consecutive 100-ms time windows starting from the probe onset and implemented with the MNE-Python toolbox (Gramfort et al., 2013; Gramfort et al., 2014).

For any observed cluster, a paired-samples *t*-test was conducted to reveal the differences in the reduction rates between the adaptor-probe pairs. The reduction rates were obtained by first morphing the observed cluster onto the individual cortical surface (Fischl et al., 1999c), extracting time courses by averaging over the source locations within the region of the cluster, averaging the time courses over the temporal extent of the cluster with respect to both adaptor and probe onsets, and finally calculating the adaptor-probe reduction rate as described above.

For any observed cluster exhibiting phoneme category selectivity, a correlation test was conducted to examine whether subjects who exhibited better behavioral categorization performance also exhibited stronger neural selectivity to phoneme categories. The behavioral categorization performance was measured after the scanning in a categorization test and quantified as the sharpness of category boundary (see Materials and Methods; Fig. 2). The selectivity to phoneme categories was quantified as the percent change between the reduction rates in 33S and 33D (i.e. $[R_{33S} - R_{33D}] / R_{33S}$). The reduction rates were obtained as in the paired-samples *t*-test (described in section 4.9). The Spearman rank correlation test was applied to examine the level of correlation.

2.10 Connectivity analysis

For estimating inter-areal connectivity, the phase slope index (PSI; Nolte et al., 2008) was computed from single-trial MNEs. PSI is based on the notions that the imaginary part of the coherency/cross-spectrum is insensitive to false connectivity caused by volume conduction (Nolte et al., 2004) and that the direction of information flow can be derived from the slope of the phase of the cross-spectrum. Accounting for the inevitable delay when distinct brain areas interact through a physical medium, PSI provides a robust measure of effective connectivity insensitive to common challenges in MEG connectivity analyses, such as low SNR and signal mixing due to volume conduction. MNE inverse solutions were first computed for all (max 512) artefact-free epochs (−200–500 ms with respect to adaptor onset) in the ATTEND condition. For estimating the cross-spectra, the epochs were then filtered with the continuous Morlet wavelet transform between 8–40 Hz with 4 Hz steps

(with the wavelet width linearly increasing from 1.5 to 7.5 cycles for the lowest to the highest frequency).

Since PSI is a signed quantity, indicating both the connectivity strength as well as the direction of information flow, the null hypothesis is simply that the values are drawn from a zero-mean distribution (Haufe et al., 2013). The statistical significance was determined with a nonparametric permutation test based on spatiotemporal clustering (Maris and Oostenveld, 2007). A *t*-value was first calculated for each PSI (i.e. to test whether the mean differs from 0) and all PSIs with *p*-value < 0.05 (uncorrected for multiple comparisons) were then clustered on the basis of spatial and temporal adjacency. Cluster-level statistics were calculated by summing the *t*-values within every obtained cluster and the maximum of the cluster-level statistics was used as the actual test statistic. A reference distribution of test statistics was produced by calculating a test statistic for one thousand random partitions of the data. A multiple-comparisons-corrected cluster *p*-value was obtained by comparing the test statistic against the reference distribution, with the null hypothesis that the mean equals zero rejected if *p*-value < 0.05.

PSIs were calculated between any observed speech-selective cluster and all other source locations (excluding the medial wall). Individual seed regions-of-interest (ROIs) were defined by first morphing the clusters onto the individual cortical surface and then applying functional constraints by selecting the sources within the clusters where the individual dSPM values exceeded a threshold of 6 (F-statistic) at any time within 500 ms after adaptor onset. Seed time courses were obtained by averaging across source locations within the seed ROI. Only the radial components of the seed time courses were kept and, depending on source orientation, sign-flips were applied to reduce signal cancellations. The analysis was performed with the MNE-Python toolbox (Gramfort et al., 2013; Gramfort et al., 2014).

3 Results

3.1 Behavior

The syllable discrimination test showed clear minima in the JND for morph differences in each direction for all subjects. The category boundary was inferred to be halfway between the two smallest JND measurements (one in each direction). The explicit phoneme category boundary measured in the syllable categorization test conformed with the JND minima (Fig. 2). During the MEG scanning, the average performance in the non-phonetic listening task of the *ATTEND* condition across subjects was 84.8% (ID: 84.7 %, 33S: 85.2%, 33D: 85.4%, 67D: 84.0%; with no statistically significant differences between the pairing conditions), indicating that the task was demanding and therefore minimized the chance that the subjects covertly categorized the stimuli in addition to performing the task. The long response times in this task (mean ± SD, all: 1267 ± 293 ms, ID: 1222 ± 280 ms, 33S: 1271 ± 312 ms, 33D: 1286 ± 313 ms, 67D: 1305 ± 330 ms) in this task further validate its difficulty. The response times for ID were shorter than for 33S, 33D, or 67D (ID *vs.* 33S: $t_{(17)} = 3.20$, $p < 0.05$; ID *vs.* 33D: $t_{(17)} = 3.99$, $p < 0.001$; ID *vs.* 67D: $t_{(17)} = 4.28$, $p < 0.001$). No statistically significant differences were observed between any other pairing conditions.

3.2 Spatiotemporal cluster analysis of evoked responses

3.2.1 Selectivity for phoneme category—In the *ATTEND* condition, the contrast 33S > 33D, indicative of phoneme category selectivity, yielded one cluster involving parts of the left aINS and POp (115–140 ms, $p < 0.05$; Fig. 3). Paired-samples *t*-tests revealed stronger release from adaptation in 33D compared to 33S ($t_{(17)} = 3.41$, $p < 0.01$) as well as in ID compared to 33D ($t_{(17)} = 2.59$, $p < 0.05$) and in 33S compared to 67D ($t_{(17)} = 2.17$, $p < 0.05$). The contrast yielded no significant clusters in the *IGNORE* condition.

3.2.2 Selectivity for acoustic-phonetic features regardless of phoneme category—Sharp acoustic-phonetic selectivity (i.e. 3*ID > 33S + 33D + 67D) was found in the *ATTEND* condition in an area involving parts of the left aSTG and pINS (235–275 ms, $p < 0.05$; Fig. 3). Paired-samples *t*-test showed that the release from adaptation in 33S, 33D, and 67D were each significantly stronger than that in ID (33S *vs.* ID: $t_{(17)} = 2.77$, $p < 0.05$; 33D *vs.* ID: $t_{(17)} = 3.13$, $p < 0.01$; 67D *vs.* ID: $t_{(17)} = 3.87$, $p < 0.01$). Two clusters exhibited broad acoustic-phonetic selectivity (i.e. ID + 33S + 33D > 3*67D) in the *ATTEND* condition: one in the left anterior temporal cortex (300–400 ms, $p < 0.01$; *t*-test, 67D *vs.* ID: $t_{(17)} = 2.16$, $p < 0.05$; 67D *vs.* 33S: $t_{(17)} = 2.58$, $p < 0.05$; 67D *vs.* 33D: $t_{(17)} = 2.23$, $p < 0.05$) and the other in the left posterior temporal cortex (300–345 ms, $p < 0.05$; *t*-test, 67D *vs.* 33S: $t_{(17)} = 2.23$, $p < 0.05$; Fig. 3).

Broad acoustic-phonetic selectivity was found also in the IGNORE condition, with one cluster located in the left middle temporal cortex (315–400 ms, $p < 0.05$; *t*-test, 67D *vs.* ID: $t_{(17)} = 2.37$, $p < 0.05$; 67D *vs.* 33S: $t_{(17)} = 4.11$, $p < 0.001$; 67D *vs.* 33D: $t_{(17)} = 3.41$, $p < 0.01$) and another in the left posterior temporal cortex (300–400 ms, $p < 0.01$; *t*-test, 67D *vs.* ID: $t_{(17)} = 2.25$, $p < 0.05$; 67D *vs.* 33S: $t_{(17)} = 2.15$, $p < 0.05$; 67D *vs.* 33D: $t_{(17)} = 2.64$, $p < 0.05$; Fig. 3). Sharp acoustic-phonetic selectivity was not found in the IGNORE condition.

3.3 Correlation between neural phoneme category selectivity and behavioral phoneme categorization

A significant positive correlation was found between the individual degree of phoneme category selectivity in the left aINS/POp and the ability to categorize the phonetic stimuli (Spearman $r = 0.61$, $p < 0.05$; Fig. 4). The categorization ability was measured behaviorally after brain imaging in a syllable categorization test (see section 2.4). Two outliers were removed from the analysis.

3.4 The effect of attending vs. ignoring the auditory input to speech processing

As differential results were obtained between the ATTEND and IGNORE conditions, an additional analysis was performed to examine the differences in the processing of speech sounds when the auditory input was attended vs. ignored. The stimulus categories were first combined into one category and responses between ATTEND and IGNORE conditions were contrasted in five consecutive 100-ms time windows starting from the adaptor (not probe) onset (and therefore being unaffected by the repetition suppression). The contrast yielded five clusters, all indicating stronger responses in the ATTEND condition (Fig. 5): (1) ~ left IFG (120–170 ms, $p < 0.05$), (2) ~ left IFG / PMC (200–300 ms, $p < 0.01$), (3) ~ right PMC / middle frontal gyrus (200–300 ms, $p < 0.05$), (4) ~ right inferior temporoparietal cortex

(200–300 ms, $p < 0.05$), and (5) ~ left dorsal PMC / superior parietal cortex (300–400 ms, $p < 0.05$).

3.5 Effective connectivity during speech processing

Connectivity tests were conducted to determine the cortical areas to which the speech-selective areas are effectively connected. The analysis was restricted to the range 100–400 ms after adaptor sound onset and performed in 100-ms time windows. In the 120–170 ms range, the speech-selective region in the posterior temporal cortex showed significant connectivity with an area involving the left POp and PMC ($p < 0.05$; Fig. 6). The direction of information flow was estimated as going from the posterior temporal cortex to POp/PMC. No significant connectivity was found with respect to any other speech-selective cluster.

4 Discussion

The present study investigated which cortical areas support categorical speech processing, at what latencies from sound onset such processing occurs, and whether it depends on auditory attention. MEG was recorded while participants, presented with paired sounds from a phonetic continuum, (1) engaged in an attention-demanding listening task wherein phoneme category information was irrelevant and (2) ignored the same sounds while watching a film without the soundtrack. Recent findings imply that frontal premotor structures, and more generally the dorsal auditory stream, support speech categorization (e.g. Alho et al., 2012; Chevillet et al., 2013; Alho et al., 2014). The results presented here corroborate these findings, revealing CSR in left inferior frontal areas. For the first time, our findings show that these category-specific representations are activated at early latencies (115–140 ms), compatible with the known speed of perceptual phonetic categorization (Salmelin, 2007; Bidelman et al., 2013). As further novel findings, we observed that these representations depend on auditory attention, correlate with the participants' ability to categorize the phonetic stimuli, and show concurrent functional connectivity with left posterior auditory cortical areas.

4.1 Neural selectivity for phoneme category in POp

Phoneme category selectivity was found in a left-hemisphere area involving POp and aINS (Fig. 3). The effect was present only when attention was directed to the auditory input, not when the subjects ignored the sounds while watching a silent film. Importantly, as the task in the attention condition diverted attention from the phonetic features, the observed category selectivity can still be considered task-independent. The relatively early latency of the effect (<140 ms) is in agreement with the proposal that access to phonological categories occurs at ~150 ms after stimulus onset (Salmelin, 2007) and suggests that these representations might drive the categorization.

Our findings are consistent with previous fMRI-adaptation studies showing CSR in speech-motor areas when the paradigm required attention to the auditory input (Chevillet et al., 2013; Lee et al., 2012; Myers et al., 2009), but not when the auditory input was ignored (Joanisse et al., 2007). A recent MEG study reported adaptation effects indicative of CSR in the left pSTS/pSTG between 430–500 ms after sound onset (Altmann et al., 2014). However,

the participants were actively discriminating the sounds, which may explain the discrepancy with our results, considering that the left STG is more strongly activated in sublexical speech perception experiments where participants engage in an active decision task compared to passive listening or non-phonetic listening tasks (Turkeltaub and Coslett, 2010). Also, due to the long latency of CSR in Altmann et al. (2014), the study does not decisively answer the question which cortical region(s) drive the categorization rather than reflect CSR as a result of projections from other regions. The latter is a concern especially in view of the fact that late evoked responses (>200 ms) seem to depend on cortical feedback connections (Garrido et al., 2007).

In interpreting the results, it has to be considered that in case any covert phonetic categorization occurred in the ATTEND condition, the discrimination of within-category sounds (33S) would presumably be more difficult than that of between-category sounds (33D), therefore making it possible that the differential activity in POp and aINS observed in the 33S *vs.* 33D contrast represents a difference in the engagement of executive functions rather than processing of phoneme categories *per se*. However, this alternative interpretation is refuted by the apparent difficulty of the non-phonetic duration discrimination task in the ATTEND condition (as reflected in the hit rate and response times, see section 3.1) together with the fact that no differences were observed in the performance (i.e. hit rate or response times) between the 33S and 33D sound pairs in this task. Further, the early latency (115–140 ms) makes it unlikely that the difference between 33S and 33D could reflect any postperceptual processes, such as response selection or decision making.

Supporting the attention-dependence of the early premotor contribution to speech processing, we observed stronger responses to speech sounds in the left IFG between 120–170 ms in the ATTEND compared to the IGNORE condition (Fig. 5). Previous studies have reported similar findings; for example, a recent TMS-MEG study demonstrated that involvement of articulatory motor areas in the early processing of acoustic-phonetic features of speech depended on auditory attention, whereas longer-latency auditory-motor interaction (>170 ms) occurred even when the subjects were told to ignore the sounds and to focus on watching a silent film (Möttönen et al., 2014). Another study showed that the IFG was engaged in the processing of degraded speech only when the subjects were attending to the stimulation (Wild et al., 2012). A potential confounding factor that needs to be acknowledged in the present ATTEND vs. IGNORE comparison is the acquisition of the data for these conditions on different days. The acquisition on separate days was deemed necessary to avoid subject fatigue (potentially compromising the data quality) resulting from overly long recording sessions. Any day-to-day fluctuation in responses within subjects should have increased noise and thus reduced possibility to observe significant effects. However, the high replicability of MEG responses (Hari and Parkkonen, 2008; Nenonen et al., 2010) together with the interference elimination and motion correction used in the present study mitigates this concern.

4.2 Neural selectivity for acoustic-phonetic features in temporal cortex

Selectivity to acoustic-phonetic features (without category-boundary effects) was found in the left temporal cortex (Fig. 3). An area just anterior to the auditory core (Heschl's gyrus)

exhibited sharp acoustic-phonetic selectivity (i.e. release from adaptation for all except identical sounds). This finding is compatible with a previous fMRI-adaptation study showing acoustic-phonetic selectivity in an anterior region of auditory cortex (Chevillet et al., 2013) and with findings showing sharper tuning to sound frequency in anterior than in posterior auditory cortical areas (Jääskeläinen et al., 2004). However, similarly to the observed phoneme category selectivity, the effect was present only when the auditory input was attended to, therefore further highlighting the importance of auditory attention in the early processing of speech stimuli. Broad acoustic-phonetic selectivity (i.e. release from adaptation only in 67D) was observed, regardless of attention to the auditory input, in extensive lateral temporal lobe areas, involving parts of STG, STS, middle temporal gyrus, and inferior temporal sulcus. When interpreting the source localization of MEG signals, one has to take into account the selective sensitivity of MEG to sulcal sources (although see Hillebrand and Barnes, 2002).

Together, these results are consistent with studies indicating that phonetic recognition occurs in left anterolateral superior temporal cortex (i.e. the ventral auditory stream) (Binder et al., 2000; Scott et al., 2000; Leaver and Rauschecker, 2010; DeWitt and Rauschecker, 2012). Our findings also support a functional distinction between temporal and frontal areas, in that temporal areas exhibit sensitivity to acoustic variation both within and between phonetic categories, whereas frontal areas show selectivity to phoneme category with insensitivity to within-category acoustic variation (e.g. Myers, 2007; Myers et al., 2009).

4.3 Correlation between neural phoneme category selectivity and behavioral categorization

The phoneme category selectivity found in our MEG measurements was positively correlated with performance in a categorization test conducted after the MEG session (Fig. 4), thus implying that the underlying neural tuning properties predict behavioral phoneme categorization performance. Similarly, Chevillet et al. (2013) reported a correlation between behavioral categorization and neural phoneme category selectivity in the left PMC (located adjacent to POp). These results are in agreement with TMS studies demonstrating that disturbing the speech-motor system impairs speech sound discrimination (Meister et al., 2007; Möttönen and Watkins, 2009; Sato et al., 2009; D'Ausilio et al., 2012; Grabski et al., 2013) as well as with our recent MEG studies showing that speech categorization during scanning correlates positively with left PMC response amplitudes (Alho et al., 2012) and with functional connectivity strength between left PMC and auditory cortex (Alho et al., 2014). Even though the present effect was located in the aINS and POp instead of the adjacent PMC (the site of stimulation in the abovementioned TMS-studies), these frontal regions are considered to be parts of the same articulatory network (e.g. Hickok and Poeppel, 2007). Further, modeling of TMS-pulse generated cortical currents has shown that the effects of TMS are more widespread on the cortical surface than has traditionally been believed (Ahveninen et al., 2013).

4.4 Effective connection from posterior temporal lobe to POp

Directed functional connectivity was observed from the left posterior temporal lobe to POp/PMC between 120–170 ms after sound onset (Fig. 6). We interpret this to reflect the

engagement of the dorsal auditory stream during speech processing and, consequently, the observed CSR in POp to reflect the result of sensorimotor integration whereby speech sounds are mapped onto the motor articulations likely to have produced them. This interpretation is supported by combined fMRI and diffusion tensor imaging (DTI) studies demonstrating a dorsal connection from posterior superior temporal cortex to POp (BA 44) and PMC (Frey et al., 2008; Frey et al., 2014), which supports sensorimotor mapping of sound-to-articulation, while a ventral pathway connects the anterior superior temporal cortex with PTr (BA 45) (Frey et al., 2008) and may be responsible for sound-to-meaning mapping. Also, a recent study using combined electrical microstimulation and fMRI to investigate primate frontotemporal effective connectivity found that while stimulating the frontal operculum (corresponding to human PTr and pars orbitalis of IFG) activated the anterior temporal lobe, stimulation of the more dorsal area F5 (corresponding to human POp and ventral PMC) activated more posterior temporal areas (Petkov et al., 2015).

Thus, a possible scenario to account for the present findings is that the CSR in POp constrain the acoustic-phonetic interpretation of the phonetic continuum into discrete categories *via* feedforward and feedback connections between auditory and premotor areas (Callan et al., 2004; Davis and Johnsrude, 2007; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Rauschecker, 2011; Schwartz et al., 2012; Bornkessel-Schlesewsky et al., 2015).

In addition to POp, we found CSR also in aINS, which has long been implicated in articulatory planning (Dronkers, 1996), auditory processing of vocalizations (Sander and Scheich, 2005; Remedios et al., 2009), as well as integration of auditory and motor information (Mutschler et al., 2009). Thus, the present data help to reinvigorate the view that the insular cortex represents an important, yet often overlooked, brain region involved in speech processing (Ardila et al., 2014; Oh et al., 2014).

4.5 Conclusions

Our results are very clear in identifying the left POp as the most distinct brain region for phoneme categorization. Indeed, the left POp exhibited early-latency (115–140 ms) phoneme category selectivity that was positively correlated with behavioral phonetic categorization. Taking advantage of our recording techniques, this finding demonstrates - for the first time - phoneme-category specific responses in the left POp that occurs early enough to coincide with phonetic perception. Furthermore, concurrent functional connectivity was observed between the left POp and posterior auditory cortical areas, which implies that phoneme category invariance arises from dorsal-stream-mediated integration of auditory and motor information. At the same time, due to its proximity with PTr (BA 45), which constitutes the endpoint of the auditory ventral stream, POp (BA 44) is in an excellent position to enable the transformation between an articulatory and a phonological code in the inferior frontal gyrus, as previously proposed by Rauschecker and Scott (2009). Finally, as an additional novel result, we found that the category-specific representations in the left POp depend on auditory attention.

Acknowledgements

This research was supported by grants from the Emil Aaltonen Foundation to J.A. and from the Academy of Finland (grant No. 138145 to I.P.J. and FiDiPro to J.P.R.). The authors wish to thank Dr. Mark Chevillet and Dr. Maximilian Riesenhuber for their contribution to the experimental design, Marita Kattelus for helping with MRI acquisition, and Matt Huszagh for helping with MEG data acquisition and processing. The authors declare no competing financial interests.

References

Ahveninen J, Huang S, Nummenmaa A, Belliveau JW, Hung A-Y, Jääskeläinen IP, Rauschecker JP, Rossi S, Tiitinen H, Raij T, 2013 Evidence for distinct human auditory cortex regions for sound location versus identity processing. Nature Communications 4: 2585.

Alho J, Sato M, Sams M, Schwartz JL, Tiitinen H, Jääskeläinen IP (2012) Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. NeuroImage 60:1937–1946. [PubMed: 22361165]

Alho J, Lin FH, Sato M, Sams M, Tiitinen H, Jääskeläinen IP (2014) Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. Front Psychol 5:394. [PubMed: 24834062]

Altmann CF, Uesaki M, Ono K, Matsuhashi M, Mima T, Fukuyama H (2014) Categorical speech perception during active discrimination of consonants and vowels. Neuropsychologia 64C:13–23.

Ardila A, Bernal B, Rosselli M (2014) Participation of the insula in language revisited: A meta-analytic connectivity study. Journal of Neurolinguistics 29:31–41.

Bidelman GM, Moreno S, Alain C (2013) Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage 79:201–212. [PubMed: 23648960]

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528. [PubMed: 10847601]

Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP (2015) Neurobiological roots of language in primate audition: common computational properties. Trends Cogn Sci. 19(3):142–150. [PubMed: 25600585]

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. [PubMed: 9176952]

Butler RA (1972) The influence of spatial separation of sound sources on the auditory evoked response. Neuropsychologia 10:219–225. [PubMed: 5055228]

Callan DE, Jones JA, Callan AM, 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. [PubMed: 15219590]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. Nat Neurosci 13:1428–1432. [PubMed: 20890293]

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. J Neurosci 33:5208–5215. [PubMed: 23516286]

D'Ausilio A, Bufalari I, Salmas P, Fadiga L (2012) The role of the motor system in discriminating normal and degraded speech sounds. Cortex 48:882–887. [PubMed: 21676385]

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9:179–194. [PubMed: 9931268]

Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E (2000) Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26:55–67. [PubMed: 10798392]

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hearing research 229:132–147. [PubMed: 17317056]

Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for

subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31:968–980. [PubMed: 16530430]

DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci U S A 109:E505–514. [PubMed: 22308358]

Dronkers NF (1996) A new brain region for coordinating speech articulation. Nature 384:159–161. [PubMed: 8906789]

Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci U S A 111:7126–7131. [PubMed: 24778251]

Fischl B, Sereno MI, Dale AM (1999a) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9:195–207. [PubMed: 9931269]

Fischl B, Sereno MI, Tootell RBH, Dale AM (1999b) High-resolution intersubject averaging and a coordinate system for the cortical surface. Human brain mapping 8:272–284. [PubMed: 10619420]

Fischl B, Sereno MI, Tootell RB, Dale AM (1999c) High-resolution intersubject averaging and a coordinate system for the cortical surface. Human brain mapping 8:272–284. [PubMed: 10619420]

Frey S, Mackey S, Petrides M (2014) Cortico-cortical connections of areas 44 and 45B in the macaque monkey. Brain and language 131:36–55. [PubMed: 24182840]

Frey S, Campbell JS, Pike GB, Petrides M (2008) Dissociating the human language pathways with high angular resolution diffusion fiber tractography. J Neurosci 28:11435–11444. [PubMed: 18987180]

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ (2007) Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci U S A 104:20961–20966. [PubMed: 18087046]

Grabski K, Tremblay P, Gracco VL, Girin L, Sato M (2013) A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. Brain Res 1515:55–65. [PubMed: 23542585]

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. NeuroImage 86:446–460. [PubMed: 24161808]

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7:267. [PubMed: 24431986]

Hämäläinen MS, Sarvas J (1989) Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. IEEE transactions on bio-medical engineering 36:165–171. [PubMed: 2917762]

Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32:35–42. [PubMed: 8182960]

Hari R, Parkkonen L, 2008 MEG in the study of higher cortical functions. Progress in epileptic disorders 5, 103–112.

Haufe S, Nikulin VV, Muller KR, Nolte G (2013) A critical assessment of connectivity measures for EEG data: a simulation study. NeuroImage 64:120–133. [PubMed: 23006806]

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature reviews Neuroscience 8:393–402. [PubMed: 17431404]

Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. NeuroImage 16:638–650. [PubMed: 12169249]

Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci U S A 101:6809–6814. [PubMed: 15096618]

Joanisse MF, Zevin JD, McCandliss BD (2007) Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. Cereb Cortex 17:2084–2093. [PubMed: 17138597]

Kawahara H, Matsui H (2003) Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on, pp I-256–I-259 vol.251.

Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci 30:7604–7612. [PubMed: 20519535]

Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci 32:3942–3948. [PubMed: 22423114]

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1–36. [PubMed: 4075760]

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–461. [PubMed: 4170865]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631. [PubMed: 15703256]

Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR (2010) Specialization along the left superior temporal sulcus for auditory categorization. Cereb Cortex 20:2958–2970. [PubMed: 20382643]

Lin FH, Belliveau JW, Dale AM, Hämäläinen MS (2006a) Distributed current estimates using cortical orientation constraints. Human brain mapping 27:1–13. [PubMed: 16082624]

Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hämäläinen MS (2006b) Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. NeuroImage 31:160–171. [PubMed: 16520063]

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190. [PubMed: 17517438]

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. Current biology : CB 17:1692–1696. [PubMed: 17900904]

Möttönen R, Watkins KE (2009) Motor representations of articulators contribute to categorical perception of speech sounds. J Neurosci 29:9819–9825. [PubMed: 19657034]

Möttönen R, van de Ven GM, Watkins KE (2014) Attention fine-tunes auditory-motor processing of speech sounds. J Neurosci 34:4064–4069. [PubMed: 24623783]

Mutschler I, Wieckhorst B, Kowalevski S, Derix J, Wentlandt J, Schulze-Bonhage A, Ball T (2009) Functional organization of the human anterior insular cortex. Neurosci Lett 457:66–70. [PubMed: 19429164]

Myers EB (2007) Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: an fMRI investigation. Neuropsychologia 45:1463–1473. [PubMed: 17178420]

Myers EB, Blumstein SE, Walsh E, Eliassen J (2009) Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci 20:895–903. [PubMed: 19515116]

Nenonen J, Parkkonen L, Helle L, Taulu S, Ahonen A, 2010 Repeatability of AEF and SEF from static and moving head positions. 17th International Conference on Biomagnetism Advances in Biomagnetism–Biomag2010 Springer, pp. 306–309.

Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S, Hallett M (2004) Identifying true brain interaction from EEG data using the imaginary part of coherency. Clinical Neurophysiology 115:2292–2307. [PubMed: 15351371]

Nolte G, Ziehe A, Nikulin VV, Schlogl A, Kramer N, Brismar T, Muller KR (2008) Robustly estimating the flow direction of information in complex physical systems. Phys Rev Lett 100:234101. [PubMed: 18643502]

Oh A, Duerden EG, Pang EW (2014) The role of the insula in speech and language processing. Brain and language 135:96–103. [PubMed: 25016092]

Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. NeuroImage 54:2437–2445. [PubMed: 20932914]

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442. [PubMed: 9176953]

Petkov CI, Kikuchi Y, Milne AE, Mishkin M, Rauschecker JP, Logothetis NK (2015) Different forms of effective connectivity in primate frontotemporal pathways. Nat Commun 6:6000. doi: 10.1038/ncomms7000. [PubMed: 25613079]

Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. Neuron 56:726–740. [PubMed: 18031688]

Rauschecker JP (1998a) Parallel processing in the auditory cortex of primates. Audiol Neurootol 3:86–103. [PubMed: 9575379]

Rauschecker JP (1998b) Cortical processing of complex sounds. Curr Opin Neurobiol 8:516–521. [PubMed: 9751652]

Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hearing research 271:16–25. [PubMed: 20850511]

Rauschecker JP (2012) Ventral and dorsal streams in the evolution of speech and language. Front Evol Neurosci 4:7. [PubMed: 22615693]

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. [PubMed: 19471271]

Remedios R, Logothetis NK, Kayser C (2009) An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J Neurosci 29:1034–1045. [PubMed: 19176812]

Salmelin R (2007) Clinical neurophysiology of language: the MEG approach. Clin Neurophysiol 118:237–254. [PubMed: 17008126]

Sander K, Scheich H (2005) Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. J Cogn Neurosci 17:1519–1531. [PubMed: 16269094]

Sato M, Tremblay P, Gracco VL (2009) A mediating role of the premotor cortex in phoneme segmentation. Brain and language 111:1–7. [PubMed: 19362734]

Schwartz J-L, Basirat A, Ménard L, Sato M (2012) The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. Journal of Neurolinguistics 25:336–354.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 Pt 12:2400–2406. [PubMed: 11099443]

Shannon RV, Jensvold A, Padilla M, Robert ME, Wang X (1999) Consonant recordings for speech testing. The Journal of the Acoustical Society of America 106:L71–74. [PubMed: 10615713]

Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51:1759–1768. [PubMed: 16552102]

Turkeltaub PE, Coslett HB (2010) Localization of sublexical speech perception components. Brain and language 114:1–15. [PubMed: 20413149]

Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. Percept Psychophys 33:113–120. [PubMed: 6844102]

Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS (2012) Effortful listening: the processing of degraded speech depends critically on attention. J Neurosci 32:14010–14021. [PubMed: 23035108]

Wise RJ (2003) Language systems in normal and aphasic human subjects: functional imaging studies and inferences from animal studies. British Medical Bulletin 65:95–119. [PubMed: 12697619]

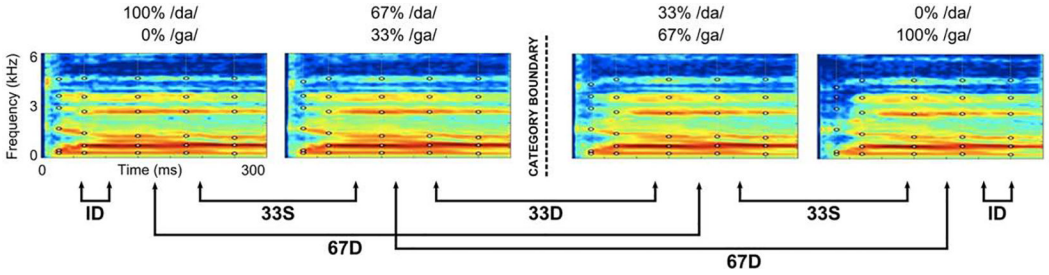

Figure 1:

Examples of the auditory stimuli (displayed as spectrograms) and adaptor-probe pairs used in the MEG experiment. The anchor points for mapping from one stimulus to the other are denoted with circles (o). The arrows show how the stimuli were paired to probe selectivity to acoustic-phonetic features and to phoneme categories: ID, identical sounds; 33S, 33% acoustic-phonetic difference and same category; 33D, 33% acoustic-phonetic difference and different category; 67D, 67% acoustic-phonetic difference and different category. Modified from Chevillet et al. (2013).

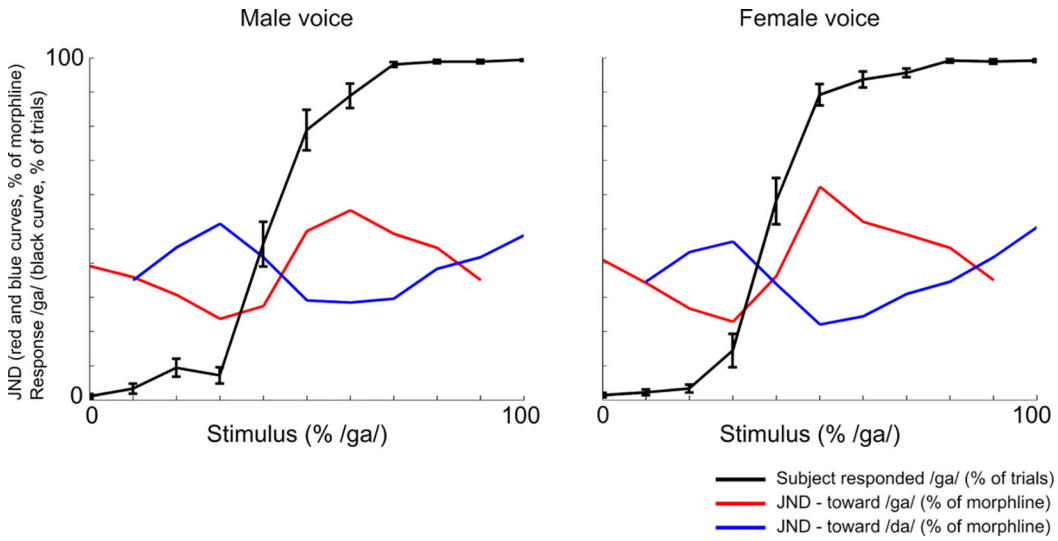

Figure 2:

Group average results from the syllable discrimination and categorization tasks. The JND was measured at 10 percentage intervals along each morphline (male and female voice) both toward /ga/ (red curve) and toward /da/ (blue curve). The halfway point between the minima of the two curves predicts the category boundary, measured in a separate categorization test (black curve). Error bars indicate SEM.

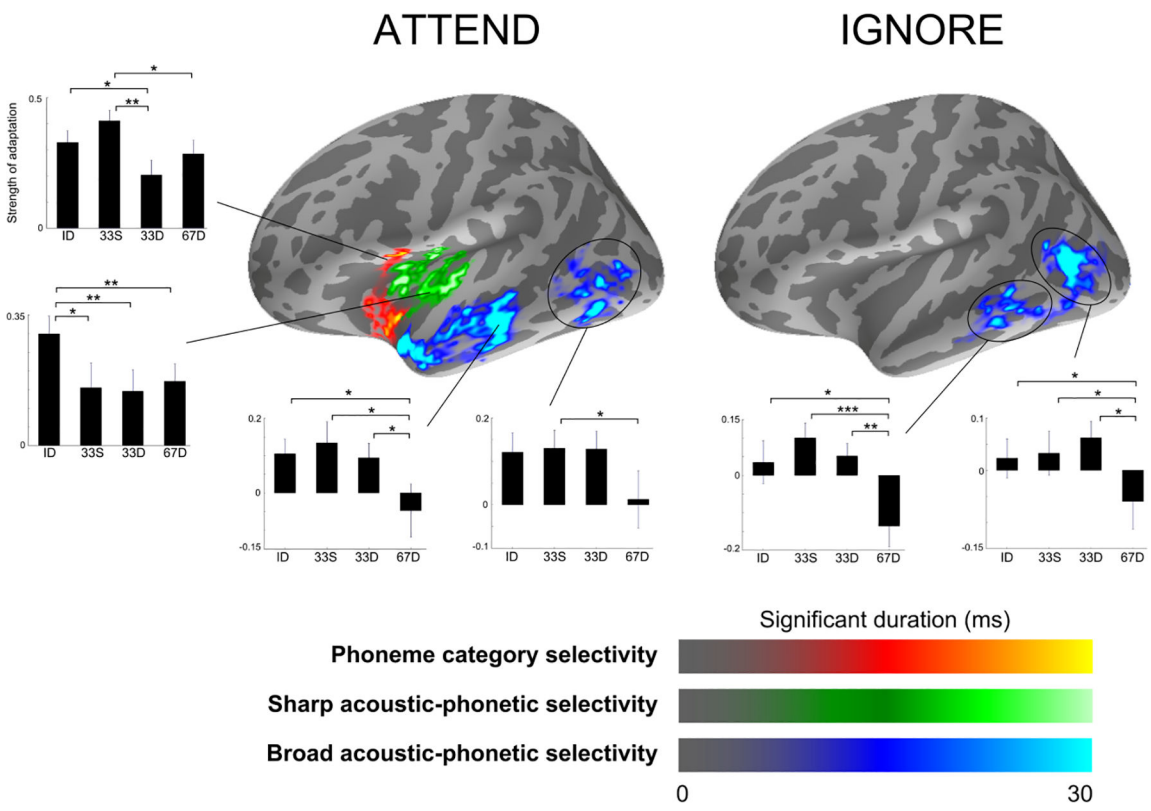

Figure 3:

Neural selectivity for phoneme category and acoustic-phonetic features determined with spatiotemporal cluster analysis and visualized on the left-hemisphere inflated surface. The bar charts show the reduction rates for the four adaptor-probe pairs as well as the t-tests for the differences between these rates. The color coding indicates the temporal extent of the clusters, error bars indicate SEM, and asterisks indicate significant differences (*p<0.05, **p<0.01, ***p<0.001). ID, identical sounds; 33S, same category with 33% acoustic-phonetic difference; 33D, different category with 33% acoustic-phonetic difference; 67D, different category with 67% acoustic-phonetic difference.

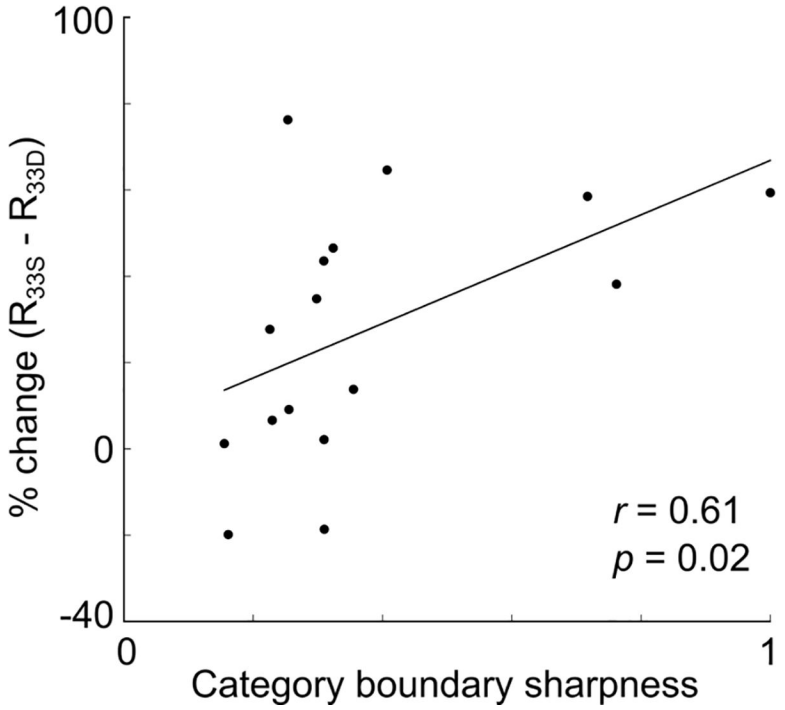

Figure 4:

Positive correlation between behavioral phoneme categorization and neural phoneme category selectivity in the left aINS/POp. Behavioral phoneme categorization was determined as category boundary sharpness, which was measured after scanning during a syllable categorization test (see section 2.4 and Fig. 2). The measurements were averaged over the male and female voice continua, and normalized. The level of neural phoneme category selectivity was quantified by the percent change in reduction rates between 33S and 33D. The *r* and *p* denote the Spearman rank correlation coefficient and the corresponding *p*-value, respectively. Black line represents the regression line.

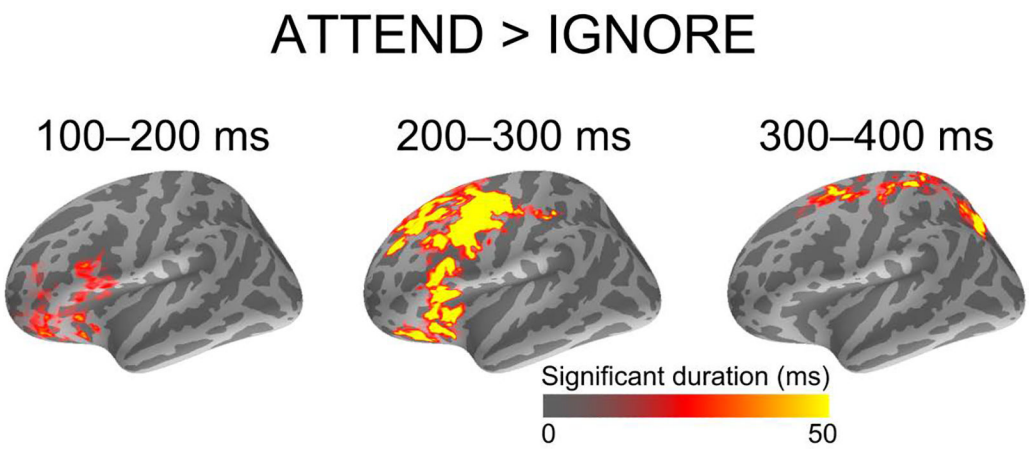

Figure 5:

Stronger responses to speech sounds when the auditory input was attended to vs. ignored. The analysis was performed with spatiotemporal clustering and visualized on the left-hemisphere inflated surface. The color coding indicates the temporal extent of the clusters. The time ranges indicate analysis time windows after the onset of the adaptor sound.

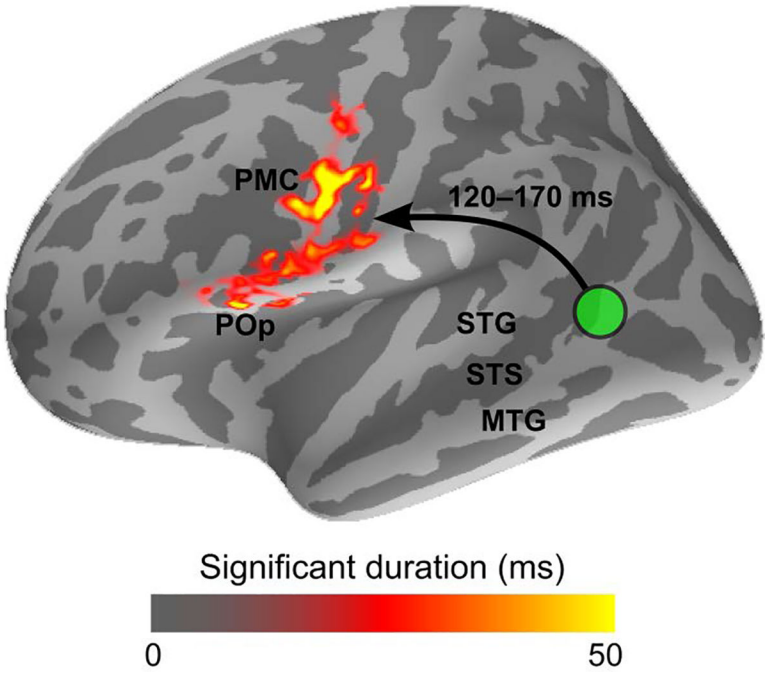

Figure 6:

Connectivity between the speech-selective cluster in the posterior temporal cortex (green circle) and the rest of the cortex during processing of speech sounds in the ATTEND condition. The observed positive PSIs indicate that information is flowing from the seed ROI to the cluster (arrow). The color coding indicates the temporal extent of the cluster, and the time ranges indicate latencies after sound onset. MTG, middle temporal gyrus; PMC, premotor cortex; POp, pars opercularis (of the inferior frontal gyrus); STG, superior temporal gyrus; STS, superior temporal sulcus.