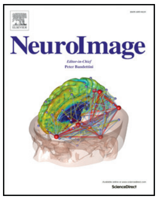
NeuroImage 129 (2016) 214–223

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



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, USA*
[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*

ARTICLE INFO

Article history:
Received 3 July 2015
Accepted 6 January 2016
Available online 14 January 2016

Keywords:
Speech perception
Categorical perception
Magnetoencephalography
Inferior frontal gyrus
Stimulus-specific adaptation

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.

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, 1998b; 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

---

[*] Corresponding authors at: Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, PO Box 12200, FI-00076, Aalto, Finland.

E-mail addresses: jussi.alho@aalto.fi (J. Alho), iiro.jaaskelainen@aalto.fi (I.P. Jääskeläinen).

http://dx.doi.org/10.1016/j.neuroimage.2016.01.016
1053-8119/

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?

Materials and methods

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.

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.

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 s 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.

Categorization behavior

After brain imaging, the subjects were asked to categorize the auditory stimuli along both morphlines to confirm the location of their