Brain & Language 163 (2016) 42–49

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Mental representations of vowel features asymmetrically modulate activity in superior temporal sulcus

Mathias Scharinger[a,b,*], Ulrike Domahs[c], Elise Klein[d], Frank Domahs[e]

[a] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany
[b] Institute of Psychology, Department of Cognitive and Biological Psychology, University of Leipzig, Germany
[c] Faculty of Education, Free University of Bozen-Bolzano, Italy
[d] Leibniz-Institut für Wissensmedien, Tübingen, Germany
[e] Institute for Germanic Linguistics, University of Marburg, Germany

ARTICLE INFO

Article history:
Received 29 February 2016
Revised 12 July 2016
Accepted 6 September 2016
Available online 24 September 2016

Keywords:
Phonological features
fMRI
Underspecification
Rounding
Place of articulation
Tongue height
Superior temporal sulcus
Superior temporal gyrus
Precuneus

ABSTRACT

Research in auditory neuroscience illustrated the importance of superior temporal sulcus (STS) for speech sound processing. However, evidence for abstract processing beyond the level of phonetics in STS has remained elusive. In this study, we follow an underspecification approach according to which the phonological representation of vowels is based on the presence vs. absence of abstract features. We hypothesized that phonological mismatch in a same/different task is governed by underspecification: A less specified vowel in second position of same/different minimal pairs (e.g. [e]) compared to its more specified counterpart in first position (e.g. [o]) should result in stronger activation in STS than in the reverse presentation. Whole-brain analyses confirmed this hypothesis in a bilateral cluster in STS. However, this effect interacted with the feature-distance between first and second vowel and was most pronounced for a minimal, one-feature distance, evidencing the benefit of phonological information for processing acoustically minimal sound differences.

1. Introduction

Human speech processing is a remarkable cognitive capacity that has inspired a plethora of research from different fields and different angles. Only recently, a better understanding of the neural bases of speech perception emerged (Davis & Johnsrude, 2003; Friederici, 2012; Hickok & Poeppel, 2007; Peelle, Johnsrude, & Davis, 2010; Scott & Johnsrude, 2003). There is growing agreement in assuming hierarchical processing steps from basic acoustic properties of the speech signal up to its rather abstract phonological representation (Davis & Johnsrude, 2003; Humphries, Sabri, Lewis, & Liebenthal, 2014; Obleser & Eisner, 2009; Okada et al., 2010; Peelle et al., 2010; Zhang et al., 2011). Linguistic theory has provided several means of describing and characterizing phonological representations, spanning from assumptions of near isomorphism between acoustic signal and mental code in *exemplar models* (Bybee, 2001; Pierrehumbert, 2002) to assumptions of high abstraction in *underspecification theory* (Featurally Underspecified Lexicon (FUL) by Lahiri and Reetz (2002), Lahiri and Reetz (2010) and Scharinger (2009)), where the mental code only consists of a minimum of acoustic-articulatory descriptions. This is exemplified in the distinction between voiced [d] and voiceless [t] by means of the absence or presence of a voicing feature (Hestvik & Durvasula, 2015).

Feature underspecification recently received increased attention from neurobiological approaches to phonology (e.g., Cornell, Lahiri, & Eulitz, 2011; Eulitz & Lahiri, 2004; Phillips et al., 2000), probably due to two reasons: First, the phonological feature as an abstract and categorical perceptual unit can adequately describe aspects of the cortical processing hierarchy, where detailed acoustic (within-category) information is processed in primary auditory areas (Heschl's gyrus), whereas more abstract and invariant (across-category) information is processed in surrounding areas, most prominently the superior temporal sulcus (STS, Davis & Johnsrude, 2003; Hickok & Poeppel, 2007; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Liebenthal et al., 2010; Okada et al., 2010; Poeppel, Idsardi, & van Wassenhove, 2008; Scott & Johnsrude, 2003). Second, underspecification provides an elegant way of describing phonological asymmetries and their perceptual

* Corresponding author at: Department of Language and Literature, Max-Planck-Institute for Empirical Aesthetics, Grüneburgweg 14, D-60322 Frankfurt, Germany.
E-mail address: mathias.scharinger@aesthetics.mpg.de (M. Scharinger).

http://dx.doi.org/10.1016/j.bandl.2016.09.002
0093-934X/

consequences (Friedrich, Lahiri, & Eulitz, 2008; Wheeldon & Waksler, 2004). For instance, the coronal nasal [n], which FUL considers to be underspecified for place of articulation (PoA), can assimilate to a following sound with a labial PoA, such as [b], resulting in *cra[m]berry* instead of *cra[n]berry*. On the other hand, the nasal [m] with a labial PoA specification, cannot assimilate to a following sound, illustrated by the hardly attested form *ru[n] toffee* for *ru[m] toffee*.

Previous neural evidence for phonological underspecification was predominantly acquired by means of electrophysiological methods, such as electro- and magnetoencephalography (Cornell, Lahiri, & Eulitz, 2013; Cornell et al., 2011; Eulitz & Lahiri, 2004; Friedrich et al., 2008; Hestvik & Durvasula, 2015; Scharinger, Bendixen, Trujillo-Barreto, & Obleser, 2012; Scharinger, Monahan, & Idsardi, 2012). Most findings were based on event-related potentials or fields from within the first 300 ms of speech processing, including the so-called Mismatch Negativity (MMN, Näätänen, Paavilainen, Rinne, & Alho, 2007; Schröger, 2005), an automatic change detection response of the brain that can index language-specific phoneme representations (Dehaene-Lambertz, Dupoux, & Gout, 2000).

Spatially more accurate imaging methods, such as functional magnetic resonance imaging (fMRI), have been used to confine regions that support the processing of phonological features such as voicing (distinguishing [d] from [t]) and place of articulation (distinguishing coronal [ø] from dorsal [o]). Feature-sensitive regions have been primarily found in STS (e.g., Arsenault & Buchsbaum, 2015; Lawyer & Corina, 2014; Obleser et al., 2006), but it is not clear whether activity in STS may also index feature underspecification. A possible activation pattern for underspecification can be deduced from Eulitz and Lahiri (2004). In this study, the authors presented the fully specified vowel [o] as standard, and the underspecified vowel [ø] as deviant, while in the reverse condition, underspecified [ø] was presented as standard and specified [o] as deviant. Both conditions elicited a discernible MMN that was minimally based on the acoustic difference between the two vowels. Crucially, however, the MMN response to a lesser specified, coronal vowel [ø] was enhanced when preceded by a more specified, dorsal vowel [o] (decreasing specificity from [o] to [ø]), compared to the reverse condition (increasing specificity, from [ø] to [o]). One interpretation of this asymmetric pattern is that the coronal surface feature of the deviant [ø] mismatched with the phonological dorsal feature of the standard [o], yielding an increased MMN response. By contrast, the dorsal surface feature of the deviant [o] did not mismatch with the underspecified phonological coronal feature of the standard [ø], thus, only resulting in an acoustic-based MMN. From this and similar findings (e.g. Cornell et al., 2011, 2013), we hypothesize that comparable asymmetries are to be found for the Blood Oxygenation Level Dependent (BOLD) response in phonological processing areas as measured by fMRI. This assumption is backed-up by studies showing that passive oddball designs carried out in an fMRI setting resulted in mismatch-related activations in superior temporal gyrus, extending into superior temporal sulcus (Shtyrov, Osswald, & Pulvermüller, 2008). A combined EEG/fMRI study furthermore revealed that the magnitude of the electrophysiological MMN response scaled with the BOLD response in superior temporal cortex (Liebenthal et al., 2003).

In the present fMRI study, we therefore compare vowel oppositions with differing degrees of phonological specificity, embedded in German pseudowords, intending to show that STS may be sensitive to phonological feature specifications. Notably, the activation patterns in our study revealed distinct asymmetries paralleling previous electrophysiological findings. Importantly, the current study allows for a precise localization of the asymmetry within STS, thereby demonstrating the specificity of this areas for phonological processing.

2. Materials and methods

2.1. Participants

A total of twenty-four native German-speaking volunteers participated in the experiment (12 female, mean age 28.2 ± 7 years). They all proved to be strongly right-handed on the Edinburgh handedness inventory (Oldfield, 1971) and did not report any hearing- or neurological problems. Participants were financially reimbursed for participation and provided their written informed consent in accordance with the Declarations of Helsinki and following the protocol of the local Ethics Committee of the RWTH Aachen Medical Faculty.

2.2. Material

Stimulus material stemmed from a previously published study (Klein, Domahs, Grande, & Domahs, 2011) and consisted of disyllabic German pseudowords, starting with a consonant-vowel syllable (CV) followed by a closed consonant-vowel-consonant syllable (CVC). Stimuli were created in pairs such that they differed in stress (first or second syllable stressed) or in the quality of the vowel in the first syllable. In the present study, the focus is only on first-syllable vowel differences, while stress differences are ignored. This means that we analyzed a subset of the stimuli presented by Klein et al. (2011), namely those blocks of pseudoword pairs for which the task focused on differences in phonological vowel quality.

In each pair, one pseudoword was spoken by a female voice, while the respective other one was spoken by a male voice, the order being counter-balanced across conditions. As a result, it was only on an abstract, phonological level that two pseudowords within a stimulus pair could be judged as identical, while they always differed substantially on a phonetic level. Mean stimulus duration was 1134 ± 201 ms.

Importantly, pseudoword stimuli contained the tense vowels [o], [ø], [e], and [u] in first syllable position. The vowels differed in specificity both in terms of phonological features as well as from the point of view of phonological underspecification (Lahiri & Reetz, 2002, 2010), with [e] being the least and [u] being the most specified vowel (see Table 1). Altogether, our stimulus and task design intended to invoke phonological processing. Apart from necessarily abstracting away from speaker differences, the use of natural and different renditions of each vowel according to context, and the pooling across different vowels in our comparisons in this study make a pure phonetic processing very unlikely and a phonetic interpretation not feasible.

In terms of specificity, [e] is underspecified for all three feature types rounding, place of articulation, and tongue height; thus, [e] is a non-round, coronal mid-vowel. On the other hand, [u] is fully specified, i.e., a round, dorsal high vowel. First-syllable vowels in pseudoword pairs differed either in 1 feature (contrast between [u]/[o]; and [e]/[ø]), 2 features (contrast between [ø]/[u] and [e]/[o]), or 3 features (contrast between [e]/[u]). For each level of *feature distance*, there were 48 pseudoword pairs. In 24 pairs, the first-syllable vowel of the second member had a *more specific* representation (more features specified, increasing specificity) than the first-syllable vowel of the first member. In contrast, in the respective other 24 pairs, the first-syllable vowel of the second member had a *less specific* representation (less features specified, decreasing specificity) than the first-syllable vowel of the first member (see Table 2). Thus, each level of specificity (increasing, decreasing) contained the same number of pairs, enabling a 3×2 design (feature distance $\times$ specificity) with a total of 144 pseudoword pairs. These 144 pairs of stimuli differing in vowel quality