Published in final edited form as:

J Neurolinguistics. 2014 January 1; 27(1): . doi:10.1016/j.jneuroling.2013.07.001.

An Investigation of Place and Voice Features Using fMRI-Adaptation

Laurel Lawyer^{1,1,*} and David Corina¹

¹Shields Ave, Department of Linguistics, University of California, Davis, CA, 95616

Abstract

A widely accepted view of speech perception holds that in order to comprehend language, the variable acoustic signal must be parsed into a set of abstract linguistic representations. However, the neural basis of early phonological processing, including the nature of featural encoding of speech, is still poorly understood. In part, progress in this domain has been constrained by the difficulty inherent in extricating the influence of acoustic modulations from those which can be ascribed to the abstract, featural content of the stimuli. A further concern is that group averaging techniques may obscure subtle individual differences in cortical regions involved in early language processing. In this paper we present the results of an fMRI-adaptation experiment which finds evidence of areas in the superior and medial temporal lobes which respond selectively to changes in the major feature categories of voicing and place of articulation. We present both single-subject and group-averaged analyses.

Keywords

fMRI-adaptation; phonological features; phonological processing; speech perception

1. Introduction

Speech perception is often characterized as the mapping from a variable acoustic input into a set of abstract linguistic representations which facilitate lexical access (Obleser & Eisner, 2008; Cutler, 2008; Poeppel et al., 2007). While the mechanism by which this mapping is achieved is still poorly understood, it is suggested that the most basic encoding unit is likely smaller than the recognizable speech segment or phoneme, at a level analogous to the acoustic or phonological feature. This concept has broad support within both theoretical linguistics and psycholinguistics, as feature-level units facilitate the description manifold linguistic processes including assimilation and vowel harmony (Jakobson et al., 1963; Clements & Hume, 1996) and serve as a frequent anchor point for models of speech perception (cf. McClelland & Elman (1986); Stevens (2006)). Though great progress has been made in identifying the neuroanatomical correlates of phonological processing, fewer studies have addressed abstract elements below the level of the single segment.

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author: lalawyer@ucdavis.edu (Laurel Lawyer), dpcorina@ucdavis.edu (David Corina).

¹Phone: +15307529933

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A primary issue in this research area is separating the processing effects which are purely acoustic from those that are abstract and therefore uniquely linguistic, a task which has proven to be quite difficult (see Kilian-Hütten et al. (2011) for discussion). Indeed, some researchers have suggested that acoustic processing is all that is necessary to facilitate the recognition of spoken syllables (Sussman, 2002). While such stances do provide compelling evidence for segment recognition as an emergent property of the general auditory system, there remains a great deal of work to do in connecting this into the larger body of literature. Again, given the rich history of research outside the purview of cognitive neuroscience which finds utility in encoding abstract elements below the level of the segment, detailed investigation into this topic using neurolinguistic methods is warranted. To this end, some progress has been made in understanding the processing of vowels, primarily owing to the fact that vowel stimuli are easy to control for acoustic/spectral content and thus provide more robust comparisons of abstract compared to acoustic encoding profiles. As consonants often have no one-to-one mapping from acoustic to phonological representations, these have proven much more difficult to manipulate, and have for the most part remained underutilized in neuroimaging studies (though see for example Rimol et al. (2005)).

A second issue which has contributed to the relative difficulty of establishing robust activations for low-level phonological elements is the considerable individual anatomical variation present in language areas including substantial variation in superior temporal and inferior frontal lobes (Ojemann, 1991). Studies which have succeeded in localizing activation specific to certain segments or featurally-defined groups of segments have for the most part converged on the superior temporal gyrus (STG) and sulcus (STS) bilaterally. Obleser *et al.* found discrete mapping of front and back vowels (distinguished by either high or low F_2 values) in the anterior superior temporal gyrus (aSTG) and inferior frontal gyrus (IFG) (Obleser et al., 2006). Utilizing the increased statistical power of multi-voxel pattern classification algorithms, Formisano *et al.* found the greatest degree of successful classification for specific vowels in the bilateral superior temporal cortex, particularly the STS/STG (Formisano et al., 2008). Formisano and colleagues have also traced consonant-specific effects to the posterior bank of Heschl's gyrus and sulcus, as well as the adjacent areas in the planum temporale (PT), bilaterally, with a left-hemisphere bias (Kilian-Hütten et al., 2011). These findings suggest that activations sensitive to the acoustic properties signaling sub-segmental content may be found quite early in the processing stream, and are consistent with contemporary models of speech perception that implicate the bilateral STG/STS as a locus for phonological encoding (cf. Poeppel & Hickok (2004); Poeppel et al. (2007); Hickok & Poeppel (2000, 2004)).

This study addresses both concerns over individual anatomic variation and the elicitation of abstract encoding of speech sounds. We utilize an fMRI-adaptation paradigm wherein subjects are presented with a single repeated stimulus item followed by a test item which differs from the repeated stimuli by some featural category. Adaptation works on the notion that a pool of neurons sensitive to a particular stimulus property become fatigued with repeated exposure, and are excited following the presentation of a stimulus item in which that property has changed. fMRI-adaptation has been used with success in other cognitive domains, particularly in vision, where adaptation effects have been found in early visual processing for stimulus features including orientation and color (Grill-Spector & Malach, 2001) as well as in higher-order processes such as facial recognition (Loffler et al., 2005). Adaptation has also been used in a handful of spoken language studies, in areas as diverse as phoneme discrimination and categorical perception (Wolmetz et al., 2010; Joanisse et al., 2007), speaker recognition (Belin & Zatorre, 2003), and emotional affect (Bestelmeyer et al., 2010). In this study, we investigate whether the featural content of consonants in CV syllables can be used as a basis to explain effects seen in an adaptation paradigm. The dissociation between acoustic similarity and featural similarity, verified by behavioral

testing of our stimuli, makes it possible to use this paradigm to investigate the abstract composition consonant stimuli without the confounding influence of acoustic processing.

Further, given the expectation of significant individual anatomical variation in the temporal lobes and thus potentially spatially distinct patterns of activation, we targeted a small number of subjects to facilitate the assessment of any group-level effects which may emerge, while also permitting analyses of single-subject activation patterns. While some individual variation is thus predicted, based on previous literature the variation is presumed to lie within fairly circumscribed areas. To wit, suppression of the hemodynamic response after repetition of syllables should reflect an adaptation of some portion of the linguistic system. As non-word stimuli are utilized, we expect this effect in the phonological system particularly and thus predict that adaptation effects will be seen in the bilateral temporal regions, specially in the STG/STS and potentially centered adjacent to Heschl's gyrus. Furthermore, if consonant stimuli do engage some level of abstract encoding, then the processing of differing featural categories are anticipated to engage different low-level representations. As a result, we expect discriminable patterns of activation unique to each featural change.

2. Material and methods

2.1. Participants

Eight right-handed native English speakers participated in this fMRI experiment. The data from two subjects were discarded after image acquisition (one for excessive movement during acquisition; the second for poor performance in a statistical assessment of speech-active regions, discussed below). The age range of subjects included in the following analysis was between 23 and 35 (average of 28.5), and gender was balanced (3 male, 3 female). All subjects were screened prior to participating and reported no history of neurological, auditory, or visual impairment. All subjects were self-reported as right-handed and monolingual English speakers. Subjects received no course credit or compensation for participation. Pursuant to the regulations of the Institutional Review Board of the University of California, Davis, informed consent was acquired from all participants.

2.2. Stimuli

Sixteen consonant-vowel (CV) nonwords were used, all of which conformed to English phonotactic rules. Four examples of each syllable were recorded by a male native English speaker. Each item was examined by the researcher both qualitatively and through spectral analysis with Praat software (Boersma & Weenink, 2011). Examples which were either spectrally atypical or which were perceived as unnatural sounding were discarded and a single optimal item from each set was selected for use in the experiment. Syllables were normalized with respect to length (50ms) and amplitude, and were enveloped using Audacity audio software (Audacity Team, 2010). The list of syllables used are in Table 2. A training set was developed concurrently which consisted of four additional syllables recorded and manipulated as above, but which were used only to familiarize the subjects with the experiment. In addition to syllables, white noise was generated in Audacity for use in the experimental paradigm discussed below. The raw white noise was filtered after generation to create high-frequency and low-frequency variants, which were then matched in apparent amplitude to the syllables above.

A set of off-line stimulus measures were obtained to quantify the perceptual confusability and discernibility of the items used in the fMRI experiment. Stimuli were identical to those used during scanning, with the exception of a parallel set of stimuli which had an added 14dB of white noise masking the original syllables. This level of noise was chosen based on the work by Phatak *et al.* which suggests stimuli with an added noise level between 12–

15dB results in frequent but not complete miscategorization of syllable identities (Phatak et al., 2008). Ten additional subjects participated in the stimuli assessment, which consisted of three tasks: two forced-choice categorization tasks, one with the original stimuli, and one with noise-masked stimuli, and a syllable identification task.

Results from these stimulus grading metrics indicated that subjects were able to correctly hear the intended syllables with an overall successful identification rate of 88.2%. In the forced-choice discrimination tasks, syllables presented in the clear resulted in few discrimination errors, with an average success rate across items of 96.22% (range = 91% – 100%, SD = 2.9%). Masked stimuli were more difficult to discriminate, with the success rate dropping to an average 74.1% with much larger variance across items (range = 57% – 84%, SD = 7.5%). An analysis of the categorization errors made with the masked stimuli suggests that the most frequently confused sounds were not always the featurally most similar. For instance, subjects made the same number of errors miscategorizing [ka] as [da] or as [ga], though [ka] shares fewer features with [da]. Taking the relative amount of confusability as an indicator of acoustic similarity, this data suggests that the degree of acoustic and featural correspondence between any two stimulus items in our test sets are not consistently equivalent. Table 1 lists the complete set of responses for both forced-choice categorization tasks.

2.3. Experimental design and image acquisition

The study utilized an fMRI-adaptation paradigm modeled after a visual orientation adaptation experiment by Fang *et al.* (Fang et al., 2005). This paradigm focuses on long-term adaptation effects created by the frequent repetition of single identical items which has been shown to result in a reduction in BOLD signal in neuronal populations that are reactive to the stimulus dimension of interest. Following the presentation of the adaptation items, a new stimulus item is presented which varies in some dimension from the previous items. The relative increase in BOLD signal at the presentation of the final stimulus allows one to infer whether a population of neurons in a given voxel exhibits sensitivity to the dimension along which the final stimulus item varies. This method increases the visibility of modulations in the neural signal due to small changes in stimuli characteristics.

In this experiment, single syllables ('Adapters') were repeated a number of times before the presentation of a final syllable ('Test') which varied in which phonological features were shared with the preceding stimuli. All Test syllables deviated from their respective Adapters by 0–3 major features, as measured by nodes following a contemporary view of phonological feature geometry (see Clements & Hume (1996) for discussion). Test items could vary on one or more specific value of Place, Voice, and/or Manner of articulation. The list distribution of the featural characteristics of the stimuli are in Table 2.

Stimuli were presented in four blocks of four trials each per run. Each trial began with two seconds of silence which allowed the program to sync with the scanner. This was followed by the adaptation series, test syllable, a fixed period of silence, four seconds of noise, and a final randomized length of silence. The adaptation series consisted of a single Adapter syllable randomly repeated 10, 12, 14, or 16 times. Test items were presented at the same rate and rhythm as the previous Adapter syllables. Test items were immediately followed by ten seconds of silence which allowed for a full decay of the hemodynamic response from the adaptation and test phases. This silence was followed by four seconds of either high or low frequency noise, the discrimination of which served as an explicit task for the subjects. All subjects performed above 85% in high-low noise categorization (average of 94%, SD = .05%). The final period of silence varied between 5 and 11 seconds, dependent on the length of the adaptation period, ensuring all trials were the same total length (38 seconds). This paradigm is schematized in Figure 1. One trial from each adaptation set was presented

randomly (without replacement) per block. Four blocks were presented in each run, which entailed the presentation of every Test from all four adaptation sets. All subjects participated in four runs, thus subjects heard each individual test item four times, resulting in a 48 total trials. Total run time for each subject was approximately 1 hour and 40 minutes.

Imaging data was acquired using a Siemens 3T scanner at the University of California, Davis Imaging Research Center in Sacramento, California using a standard Siemens 8-channel head coil. Four functional runs and one structural image were acquired from each subject. Functional runs consisted of 304 volumes and were collected using a gradient echo EPI sequence (37 slices, thickness = 2.8mm, TR = 2000ms, TE = 29ms, flip angle = 90°, FOV = 220mm × 220mm, voxel size = 2.8mm³) using GRAPPA. Functional volumes were aligned parallel to the anterior-posterior commissure (AC-PC line) and provided full coverage of the temporal and occipital lobes, as well as the inferior portions of the frontal and parietal lobes. All volumes acquired at the beginning and end of each trial were used only to sync the scanner with the protocol and were discarded from the analysis.

During scanning, stimuli were presented binaurally using noise-attenuating MR-compatible pneumatic headphones. Volume was adjusted to ensure stimuli were clearly audible above scanner noise and each subject verbally verified stimuli were audible and presented at a comfortable level. Subjects were asked to fixate on a white cross presented throughout the experiment. The experiment was run using Presentation software (v. 14.4, Neurobehavioral Systems). Following the functional runs, a high-resolution structural image covering the entire brain was acquired using an MPRAGE sequence (208 slices, thickness = 1mm, TR = 1900ms, TE = 2.8ms, flip angle = 7°, FOV = 256mm × 256mm, matrix = 256 × 256, voxel size = 1mm³).

2.4. Data Analysis

Data was preprocessed in an individual, whole-head analysis before being submitted to statistical analysis using SPM8 (Wellcome Department of Imaging Neuroscience, 2009). Functional volumes were slice time corrected, realigned to the mean image, and coregistered to each subject's high-resolution structural image. To enable group comparisons, a second set of images were processed as above, and additionally normalized to the MNI template and smoothed with an 8mm³ FWHM Gaussian kernel.

Within subjects, statistics were calculated using a fixed effects model. Seven conditions were included in the estimation: 1st Adapter, Last Adapter, 1-Feature Place Test, 1-Feature Voice Test, 2-Feature Test, 3-Feature Test, and Noise. The Last Adapter condition was composed only of those adapters which were followed by a 0-feature (identical) test item to avoid collinearity violations with respect to the remaining test items. For this reason, 0-Feature test items were not modeled. Each subject's realignment parameters were included as regressors. An additional wedge-shaped regressor was calculated which approximated the response for the remaining (unmodeled) Adapters and fit into the adaptation string in each trial to remove this variance from baseline calculations.

To examine effects restricted only to the linguistic system, the 1st Adapter > Noise contrast ($p < .001$, 10 vox cluster) was calculated for each subject. Due to the nature of the adaptation paradigm and the degrading response to repeated stimuli, this comparison was chosen as the best estimate of speech areas available to the model. The voxels which maintained statistical significance in this analysis were used as the basis of the Speech ROI used to mask later contrasts. To locate the regions exhibiting adaptation, voxels showing greater activation at the beginning of the adaptation series than to the end of the adaptation series were calculated (1st Adapter > Last Adapter; $p < .001$, 10 voxel cluster).

An individual ROI analysis was carried out within the speech regions selected above to investigate the degree of recovery exhibited by the various Test items. Test items were compared against their immediately preceding adaptation items (Last Adapters) to highlight the location and degree of release associated with each Test item modulation (1-Feature-Place > Last Adapter, 1-Feature-Voice > Last Adapter, 2-Feature > Last Adapter, 3-Feature > Last Adapter; all $p < .05$ with no cluster). These comparisons were also carried out using each subject's normalized data to estimate responses common to the group. As only six subjects were included in the group analysis, a fixed effects model was utilized which allows for the comparison of responses within this group but does not extend predictions to the populous as a whole.

3. Results

3.1. Group analysis

3.1.1. Speech-sensitive regions—Using the 1st Adapter > Noise comparison to elicit speech-sensitive regions, we find extensive activation extending throughout the temporal lobes bilaterally. This pattern of activation is in line with previous literature (Binder et al., 2000; Buchsbaum et al., 2001; Friederici & Alter, 2004; Rimol et al., 2005) and supports our initial prediction. Activation peaks were located in the left superior temporal gyrus extending into the sulcus (ISTG/ISTS) and middle temporal gyrus. This contrast also shows activation in the left precuneus, right superior occipital gyrus (rSOG), and well as the left medial frontal gyrus (lMedFG). Activation is shown in yellow in Figure 2, and activation foci for all contrasts can be found in Table 3.

3.1.2. Adaptation—Adaptation effects were found in the right STG/STS bilaterally, largely overlapping with the previously identified speech-sensitive regions. This pattern of activation points to the success of the paradigm utilized here to adapt a significant portion of the early speech perception system. Several other regions outside the Speech ROI also exhibited greater activation at the beginning of the adaptation string, including the cingulate, anterior cingulate, MedFG and several areas in the cerebellum. As these are well outside the Speech ROI and likely belie non-linguistic patterns of default network activation, they are not further discussed. Activation for this comparison is shown in green in Figure 2, and cluster foci are listed in Table 3

3.1.3. Place and Voice features—Several regions of activation which are unique to 1-Feature-Place and 1-Feature-Voice contrasts respectively are found in the ROI analysis. Activation for both 1-Feature-Place > Last Adapter and 1-Feature-Voice > Last Adapter are found in adjacent but largely non-overlapping regions of the right MTG and left STG. Additional activation for the 1-Feature-Place contrast is found in the left right STG. The 1-Feature-Voice contrast further includes a large region of activation in the left STG. These contrasts are illustrated in Figure 3, highlighting the contiguous but sparsely overlapping nature of this data. Locations of activation peaks are in Table 3.

3.2. Single-subject analysis

3.2.1. Speech-sensitive regions—All subjects showed activation in both left and right temporal regions for this contrast. Activation clusters were centered in the STG/STS, as would be expected given the group results. Individual subjects showed some variation with regards to how uniformly bilateral the response to speech stimuli were. Issues of response laterality are returned to in the discussion.

3.2.2. Adaptation—All six subjects showed a strong bilateral temporal response centered around Heschl's gyrus and extending into the adjacent temporal areas. In one subject, S5,

several additional regions were more active for 1st Adapters than Last Adapters, including areas well outside the Speech ROI. Increased activity was seen in frontal and parietal regions, as well as a more distributed pattern of activation throughout the temporal lobes.

3.2.3. Place and Voice features—Viewing the response patterns of the individual subjects for feature-level contrasts concurs with the group results showing discrete regions sensitive to changes in voice and place of articulation, though with great variability in specific responses. When comparing the activation on an individual basis for the 1-Feature-Place > Last Adapter and 1-Feature-Voice > Last Adapter contrasts, activation is found along the mid- and posterior STG in three subjects (S1, S4, and S5). Activation in the mid- and posterior STS was also common, found in four subjects (S1, S2, S3, and S5). (One subject, S6, showed no statistically significant activation to either Place > Last Adapter or Voice > Last Adapter contrasts, though at a lower threshold ($p < .1$) the data trends in the same direction.) There is no obvious pattern which distinguishes the activation in the STG versus that in the STS among the subjects who exhibit both. Nor is there a clear delineation between the locations of voice-sensitive compared to place-sensitive voxels in any subject. For instance, subject S3 shows a clearly interdigitated pattern of activation for both contrasts along the right STS, while S2 exhibits pattern of bilateral but highly circumscribed activation. Representative images from each subject are shown in Figure 4.

The individual subject data showed an overwhelming propensity for discrete responses to 1-Feature-Voice and 1-Feature-Place contrasts; however, there were some small areas of overlap in the responses of three subjects. In S1, the area of overlap was the posterior-most cluster in the right STS, adjacent to a somewhat larger area selective for 1-Feature-Voice contrasts. There were three areas of overlap in the data from S4: two small regions in left Heschl's gyrus, with an additional region in the posterior left STG adjacent to another larger voice-sensitive cluster. Finally, two areas of overlap were found in S5 which were somewhat more extensive than those of the other subjects. Here, overlap is found in the posterior STS bilaterally, extending over a small number of adjacent voxels. Neither S2 nor S3 had any areas of overlap between the two feature-sensitive contrasts.

4. Discussion

The present study sought to explore two central issues. First, is fMRI-adaptation a useful method for exploring the neural decomposition of consonant sounds? And second, what is the profile of feature-modulated responses in the listening brain? The results from this experiment indicate not only that we can successfully adapt some portion of the linguistic processing system, but that this paradigm has utility in discriminating small differences in low-level linguistic stimuli. The composition of speech sounds must include some level of abstraction below that of the segment, and the results presented here suggest that a featural analysis based on the major categories of place of articulation and of voicing (laryngeal state) may not be untenable. Both these positions illuminate the need for further research to delineate the particulars of the adapting response, and to clarify the contributions of various types of featural decomposition (i.e., by abstract category, or by something more acoustically- or gesturally-based). The results presented here are an encouraging preliminary step towards using an adaptation paradigm to investigate the neural implementation of feature-level phonological encoding.

4.1. Adapting the linguistic system

Several studies of linguistic perception have utilized adaptation studies to investigate a variety of phenomena. Belin & Zatorre found adaptation of the right anterior temporal lobes only for repetition of a single speaker's voice and reported no areas adapting to repeated syllables (Belin & Zatorre, 2003). The authors use this data to support the notion of a

bifurcation of linguistic processing across hemispheres. The current study cannot separate adaptation effects ascribed to the identity of the speaker compared to the content of the linguistic stimuli owing to the fact that only a single speaker was employed throughout. Despite this confound, the results from the current study do not replicate the previous findings from either the single-speaker or syllable conditions. The locus of speaker adaptation reported by Belin & Zatorre was in the right anterior temporal lobes, an area which is not engaged in any of the contrasts reported here. Adaptation effects were found not only bilaterally, but rather more posterior to those previously reported. Whether this adapting ROI reflects the adaptation of linguistic processing, or whether it is a combination of acoustic, linguistic, and meta-linguistic properties such as speaker identity is impossible to distinguish using the current paradigm and warrants further investigation.

More recent studies have had success adapting repeated syllables. Both Joanisse *et al.* and Wolmetz *et al.* utilized chains of syllables which varied in voice onset time. These studies utilized test items which maintained equal acoustic distance from adapting syllables but either remained within a phonemic category or crossed a phonemic category boundary. In both cases, areas more active in between-category discrimination were found in the left MTG, and left temporal pole/STS (Joanisse *et al.*, 2007; Wolmetz *et al.*, 2010). While neither study attempts to investigate properties of specific consonants, the activations elicited by phoneme discrimination tasks may be part of the same abstract phonological processing system. In this case, the data in the present study present an interesting point of comparison. Both previous studies converge on the left posterior STG/MTG as a nexus for phonemic discrimination, an area also implicated in voice- and place-sensitive feature contrasts, though this activation is much more medial than that reported by Joanisse *et al.* Further, some divergence in results is found in the contribution of the right temporal lobe. There is no appreciable contribution from the right hemisphere in Joanisse *et al.* (2007), but the right temporal lobe was suggested by Wolmetz *et al.* (2010) to contribute to both phonemic categorization as well as gradient acoustic responses. The results in the present study concord with this, finding voice- and place-sensitive regions bilaterally, supporting the suggestion that early phonological processing of spoken syllables engages both the right and left superior temporal lobes. This point is returned to below.

4.2. Lateralization of speech response

Though the response to speech stimuli across subjects implicated bilateral temporal regions, individual subjects present a mixed set of responses showing no generalizable pattern. Though canonically speech stimuli are presumed to elicit a left-dominant response for language processing, a large number of studies and meta-analyses report bilateral activation, particularly for low-level tasks such as passive listening to syllables (see for example (Binder *et al.*, 2000; Burton *et al.*, 2000; Blumstein *et al.*, 2005; Obleser *et al.*, 2010; Wolmetz *et al.*, 2010)). Contemporary models frequently appeal to bilateral phonological processing before higher-order linguistic information is processed or accessed (cf. Hickok & Poeppel (2004); Obleser & Eisner (2008)).

One prominent model put forward by Hickok and Poeppel draws on patient and imaging data to argue that processing of auditory stimuli takes place in the STG/STS bilaterally including all phonetic and phonological processing necessary for lexical access (Hickok & Poeppel, 2004). The current study can be taken to provide further evidence for the bilateral nature of low-level phonological processing. It has been argued by Poeppel that each hemisphere is specialized for the processing of different stimulus characteristics, with preference for shorter temporal elements in the left hemisphere, and longer temporal elements in the right hemisphere. Perception of place of articulation and voicing are both are cued by rapid acoustic elements such as formant transitions and voice onset time. These features would thus be included in the set of items processed within short time windows, and

as such, left-hemisphere processing would be expected. Among the place and voice contrasts, the group data shows a proclivity towards left-dominance, and though there is a large amount of variability across subjects, our data is not inconsistent with the broad predictions of this model.

The specific prediction that feature extraction should be an exclusively left-hemisphere task, however, is not entirely upheld in our data. Several subjects showed increased sensitivity to place and voice changes in the right hemisphere only, a result not predicted by the aforementioned models. As the feature responses do not reliably follow these expected patterns, it bears reflecting on what degree of generalizability is expected across subjects with respect to the specifics of low-level speech processing, particularly given the single-subject variability which has already been found in the superior temporal lobes (Ojemann, 1991).

4.3. Feature-level effects

Featural extraction has been tied to Heschl's gyrus (Kilian-Hütten et al., 2011), the planum temporale, and portions of the STG and STS surrounding Heschl's gyrus (Obleser et al., 2006). These findings have been supported by the work presented here. The loci of activation reported in this study lie within or adjacent to these areas, extending also into posterior portions of the STG/STS. Due to the ROI analysis which specifically restricted the search space to areas more active under speech conditions than noise, contributions from Heschl's gyrus and the PT may have been omitted or downplayed.

While areas selective for Voice or Place contrasts are always found in the region of the mid-to-posterior STG/STS, it is of particular interest to note that they rarely overlap. Both on a group level and in the individual subject data, Place and Voice selective regions provide robustly discrete areas of activation. In the group analysis the region of overlap is restricted to a single voxel in each of the right and left superior temporal gyri. Examining the individual subject data underscores this finding: only three subjects had any areas significantly active for both conditions, and in these three subjects overlapping regions were in the minority compared to regions sensitive to the single Voice or Place conditions. This suggests that there may be some amount of localized neural specialization for processing distinct portions of the phonological code.

The question as to whether this selective activation is modulated by the spectral/acoustic profile of the stimuli, or whether it is indicative of some low-level abstract linguistic processing is as yet unanswered. The acoustic reflection of the featural changes represented in the stimuli are quite distinct but complex. Voicing contrasts may be cued by a variety of elements such as voice onset time and degree of aspiration, whereas place of articulation information is frequently conveyed by rapid formant transitions which vary characteristically in each consonant-vowel pairing, but may also be modulated by voice onset time and the frequency of release burst (Massaro & Oden, 1980; Fant, 1967). Most theories of speech perception would suggest acoustic features such as these are mapped onto abstract linguistic features through some extraction mechanism. The acoustic and linguistic features are thus tightly coupled yet distinct, rendering it difficult if not impossible to separate the relative contributions from acoustic and linguistic processing. What is clear from the data presented here is that stimuli which specifically encode changes in the processing of place and voicing recruit separate processing circuits. Different manipulations to the speech stream engage different response types, both found in the bilateral STG/STS, but each exhibiting a unique pattern of activation.

5. Conclusions

In the present study, we employed an fMRI-adaptation paradigm to elicit discriminable responses to featurally disperse stimuli. The unique structure of an adaptation paradigm proved useful as a means to investigate speech processing below the level of the segment. Of greatest interest was the profile of responses to test items which changed either voicing or place of articulation features compared to the preceding string of repeated adaptation syllables. Our results show that areas of the brain with the greatest response to changes in place and voice features are distinct. In the group analysis, as well as in individual subjects, activation is found in the mid- and posterior STG/STS for both changes in voice and changes in place. Given the separability of these regions in the results, this research suggests that differing sets of linguistic or acoustic features may be represented in discrete though adjacent brain regions.

Acknowledgments

The authors would like to thank Lee Miller, Chris Bishop, Kevin Hill, Laura Chao and Alex Newhouse for their assistance with this project. This work was supported in part by grant NIH-NIDCD 2ROI-DC03099-11 awarded to David Corina.

References

- Audacity Team. Audacity, version 1.3.4. 2010
- Belin P, Zatorre RJ. Adaptation to speaker's voice in right temporal lobe. *NeuroReport*. 2003; 14:2105–2109. [PubMed: 14600506]
- Bestelmeyer PE, Rouger J, DeBruine LM, Belin P. Auditory adaptation in vocal affect perception. *Cognition*. 2010; 117:217–223. [PubMed: 20804977]
- Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing E. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*. 2000; 10:512528.
- Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*. 2005; 17:1353–1366. [PubMed: 16197689]
- Boersma P, Weenink D. Praat: doing phonetics by computer, version 5.2.35. 2011
- Buchsbaum BR, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*. 2001; 25:663–678.
- Burton MW, Small SL, Blumstein SE. The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*. 2000; 12:679–690. [PubMed: 10936919]
- Clements, GN.; Hume, EV. The internal organization of speech sounds. In: Goldsmith, JA., editor. *Handbook of Phonological Theory*. Blackwell; 1996. p. 245-306.
- Cutler A. The abstract representations in speech processing. *The Quarterly Journal of Experimental Psychology*. 2008; 61:1601–1619. [PubMed: 18671170]
- Fang F, Murray SO, Kersten D, He S. Orientation-tuned fMRI adaptation in human visual cortex. *Journal of Neurophysiology*. 2005; 94:4188–4195. [PubMed: 16120668]
- Fant, CGM. Descriptive analysis of the acoustic aspects of speech. In: Lehiste, I., editor. *Readings in Acoustic Phonetics*. Vol. 7. MIT; 1967. p. 93-107.
- Formisano E, Martino FD, Bonte M, Goebel R. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*. 2008; 5903:970–973. [PubMed: 18988858]
- Friederici AD, Alter K. Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*. 2004; 89:267–276. [PubMed: 15068909]
- Grill-Spector K, Malach R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*. 2001; 107:293–321.
- Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*. 2000; 4:131–138. [PubMed: 10740277]

- Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*. 2004; 92:67–99. [PubMed: 15037127]
- Jakobson, R.; Fant, CGM.; Halle, M. Preliminaries to Speech Analysis: The Distinctive Features and their Correlates. Cambridge, MA: MIT Press; 1963.
- Joanisse MF, Zevin JD, McCandliss BD. Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fmri and a short-interval habituation trial paradigm. *Cerebral Cortex*. 2007; 17:2084–2093. [PubMed: 17138597]
- Kilian-Hütten N, Valente G, Vrooman J, Formisano E. Cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*. 2011; 35:1715–1720. [PubMed: 21289180]
- Loffler G, Yourganov G, Wilkinson F, Wilson HR. fMRI evidence for the neural representation of faces. *Nature Neuroscience*. 2005; 8:1386–1390.
- Massaro DW, Oden GC. Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*. 1980; 67:996–1013. [PubMed: 7358922]
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cognitive Psychology*. 1986; 18:4–86.
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*. 2006; 27:562–571. [PubMed: 16281283]
- Obleser J, Eisner F. Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*. 2008; 13:14–19. [PubMed: 19070534]
- Obleser J, Leaver AM, VanMeter J, Rauschecker JP. Segregation of vowels and consonants in human auditory cortex: evidence for distributed hierarchical organization. *Frontiers in Psychology*. 2010; 1:1–14. [PubMed: 21833184]
- Ojemann GA. Cortical organization of language. *Journal of Neurosci*. 1991; 11:2281–2287.
- Phatak SA, Lovitt A, Allen JB. Consonant confusions in white noise. *Journal of the Acoustical Society of America*. 2008; 124:1220–1233. [PubMed: 18681609]
- Poeppel D, Hickok G. Towards a new functional anatomy of language. *Cognition*. 2004; 92:1–12. [PubMed: 15037124]
- Poeppel D, Idsardi WJ, van Wassenhove V. Speech perception at the interface of neurobiology and linguistics: prospects and problems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2007; 363:1071–1086.
- Rimol LM, Specht K, Weis S, Savoy R, Hugdahl K. Processing of sub-syllabic speech units in the posterior temporal lobe: An fmri study. *NeuroImage*. 2005; 26:1059–1067. [PubMed: 15894493]
- Stevens, KN. Features in speech perception and lexical access. In: Pisoni, DB.; Ramirez, RE., editors. *The Handbook of Speech Perception*. Blackwell Handbooks in Linguistics. Blackwell; 2006. p. 125-155.
- Sussman H. Representation of phonological categories: A functional role for auditory columns. *Brain and Language*. 2002; 80:1–13. [PubMed: 11817886]
- Wellcome Department of Imaging Neuroscience. Statistical parametric mapping. 2009; 8
- Wolmetz M, Poeppel D, Rapp B. What does the right hemisphere know about phoneme categories? *Journal of Cognitive Neuroscience*. 2010; 23:552569.

Highlights

1. fMRI-adaptation used to investigate featural properties of speech encoding
2. Feature-level encoding found in mid- and posterior-STG/STS bilaterally
3. Changes in place and voicing elicit spatially discrete representations
4. Findings supported in single-subject and group-level analysis

Figure 1.
Schematic of the design for the fMRI experiment.

Figure 2. Group-level activation for the 1st Adapter > Noise contrast (yellow) compared to the 1st Adapter > Last Adapter contrast (green). Both contrasts levels are set at $p < .001$, cluster size = 10.

Figure 3.

Activation for 1-Feature-Voice and 1-Feature-Place contrasts, showing both areas of overlap and areas distinct to each feature contrast. 1-Feature-Voice > Last Adapter contrast (red) and 1-Feature-Place > Last Adapter (green), are thresholded at $p < .05$ with no cluster. Both contrasts are masked by the 1st Adapter > Noise contrast to restrict the analysis to speech-sensitive regions.

Figure 4.

Activation for 1-Feature-Voice and 1-Feature-Place contrasts for each subject. 1-Feature-Place > Last Adapter contrast (green) and 1-Feature-Place > 1-Feature-Voice (red), are thresholded at $p < .05$ with no cluster. Both contrasts are masked by each subject's Speech ROI. Areas of overlap appear in yellow.

Table 1

Results for the Syllable Discrimination experiment. Each column corresponds to the stimuli presented, rows are subject responses in a forced-choice experiment.

ABX clear					ABX masked				
stimulus	ba	da	ta	a	stimulus	ba	da	ta	a
ba	63	2	1	0	ba	53	3	3	1
da	3	62	0	1	da	6	52	6	9
ta	1	1	64	0	ta	3	1	63	4
Sa	2	0	0	60	Sa	2	9	3	50
total	69	65	65	61	total	64	65	75	64
stimulus	ka	ga	da	za	stimulus	ka	ga	da	za
ka	76	0	0	0	ka	56	2	3	2
ga	1	45	0	0	ga	3	36	10	5
da	0	3	63	1	da	3	3	46	5
za	0	2	0	74	za	7	8	9	53
total	77	50	63	75	total	69	49	68	65
stimulus	pæ	kæ	dæ	ðæ	stimulus	pæ	kæ	dæ	ðæ
pæ	55	1	1	0	pæ	41	10	3	3
kæ	0	55	3	0	kæ	13	42	5	5
dæ	0	0	59	2	dæ	3	3	48	7
ðæ	0	2	0	62	ðæ	3	3	7	44
total	55	58	63	64	total	60	58	63	59
stimulus	gI	kI	dI	sI	stimulus	gI	kI	dI	sI
gI	56	1	0	2	gI	41	3	9	11
kI	2	62	1	0	kI	5	45	3	4
dI	0	1	54	1	dI	8	4	35	13
sI	1	0	1	58	sI	8	7	6	37

ABX clear					ABX masked				
stimulus	ba	da	ta	a	stimulus	ba	da	ta	a
total	59	64	56	61	total	62	59	53	65

Table 2

The featural relation between Adapters and Test items. Items on the left are the syllables used for adaptation. Test items are listed to the right. Number of features changed between Adapters and Test items are listed, as well as whether these were changes in Place (P), Voice (V), or Manner of articulation (M). Not included are the 0 Feature test items, which are identical to their respective Adapter. Stimuli are listed in the International Phonetic Alphabet (IPA).

Adapter	1 feat. P	1 feat. V	2 feat. P,V	3 feat. P,V,M
ba	da	—	ta	Sa
ka	—	ga	da	za
pæ	—	kæ	dæ	ðæ
gI	dI	kI	—	sI

Table 3

Table of supra-threshold activation values.

Contrast	k	Talairach coords	z	p	Approx. locations
1st Adapter > Noise	1206	62, -22, 0	Inf	0.000	R STG
	1043	-61, -14, 0	Inf	0.000	L STG
	95	-5, -59, 31	4.36	0.000	L Precuneus
	26	40, -76, 28	3.83	0.000	R SOG
	28	0, 59, 6	3.83	0.000	L MedFG
	45	-39, -67, 31	3.82	0.000	L Precuneus
	67	-2, 50, -14	3.78	0.000	L MedFG
1st Adapter > Last Adapter	1495	59, -22, 3	Inf	0.000	R STG
	1442	-56, -17, 0	Inf	0.000	L STG
	875	-5, -64, 0	4.74	0.000	L Cerebellum
	53	-16, -22, -8	4.58	0.000	Brainstem
	54	0, 17, 23	4.27	0.000	L AC
	31	3, -56, -39	4.21	0.000	R Cerebellum
	12	-33, -78, -39	4.04	0.000	L Cerebellum
	69	3, -20, 6	3.99	0.000	R Thalamus
	31	40, -73, -30	3.67	0.000	R Cerebellum
	32	0, -70, 42	3.66	0.000	L Precuneus
	18	3, 64, 6	3.64	0.000	MedFG
	37	23, -53, -5	3.54	0.000	R Parahippocampal Gyrus
	18	3, -20, 26	3.41	0.000	Cingulate
	12	48, -25, -5	2.77	0.003	R STG
	49	-54, -31, 14	2.71	0.003	L STG
1-Feature-Voice > Last Adapter	13	-36, -28, 9	2.65	0.004	L Transverse Temporal Gyrus
	2	-64, -11, 0	2.55	0.005	L STG
	1	54, -39, -2	2.36	0.009	R MTG
	5	48, -31, -5	2.68	0.004	R MTG
1-Feature-Place > Last Adapter	8	-36, -34, 14	2.64	0.004	L STG

Contrast	<i>k</i>	Talairach coords	<i>z</i>	<i>p</i>	Approx. locations
	4	56, -39, -5	2.54	0.006	R MTG
	5	54, -3, 0	2.43	0.007	R STG
	1	45, -39, 0	2.33	0.010	R STG