Cell PRESS

Neuron
Article

Abstract Coding of Audiovisual Speech: Beyond Sensory Representation

Uri Hasson,[1,2,*] Jeremy I. Skipper,[1,2] Howard C. Nusbaum,[2,3] and Steven L. Small[1,2,3]
[1]Department of Neurology
[2]Department of Psychology
[3]Center for Cognitive and Social Neuroscience
The University of Chicago, Chicago, IL 60637, USA
*Correspondence: uhasson@uchicago.edu
DOI 10.1016/j.neuron.2007.09.037

SUMMARY

Is there a neural representation of speech that transcends its sensory properties? Using fMRI, we investigated whether there are brain areas where neural activity during observation of sublexical audiovisual input corresponds to a listener's speech percept (what is "heard") independent of the sensory properties of the input. A target audiovisual stimulus was preceded by stimuli that (1) shared the target's auditory features (auditory overlap), (2) shared the target's visual features (visual overlap), or (3) shared neither the target's auditory or visual features but were perceived as the target (perceptual overlap). In two left-hemisphere regions (pars opercularis, planum polare), the target invoked less activity when it was preceded by the perceptually overlapping stimulus than when preceded by stimuli that shared one of its sensory components. This pattern of neural facilitation indicates that these regions code sublexical speech at an abstract level corresponding to that of the speech percept.

INTRODUCTION

Understanding the relationship between neural activity and the phenomenological perception of speech is one of the main challenges in the cognitive neuroscience of speech perception. A central question in this domain is whether there exists a level of representation in which speech is coded as abstract perceptual units that are distinct from the sensory cues from which they are derived. Decades of experimental research have argued for this possibility by showing that (1) different acoustic cues can be experienced as the same percept and (2) the same acoustic cue can be perceived differently in different contexts. However, whether there exists a neural layer that codes for speech at such an abstract level (sometimes referred to as a "half-mythical" level; Nelken and Ahissar, 2006), is an empirical question that is the objective of the research we report here.

We examined whether there are cortical regions in which neural activity tracks the perceived speech rather than its sensory properties. Specifically, such regions would code similarly two speech stimuli that differ in their sensory cues but that are experienced as the same percept. Consider the following two stimuli that exemplify this phenomenon: (1) an audiovisual video of a person saying "ta" and (2) a silent video of a person articulating /ka/ dubbed with an acoustic track of a person uttering /pa/ (/PK/). Both of these stimuli are perceived as "ta" (McGurk and MacDonald, 1976). Yet, this does not imply that the two stimuli are similarly coded at any neural level. For example, the experienced percept of /PK/ as "ta" may *just be* a result of simultaneous sensory coding for auditory /pa/ and visual /ka/, resulting in a "ta" percept.

Testing whether two different audiovisual stimuli are similarly coded at a certain neural level cannot be accomplished by directly contrasting the neural activity patterns evoked by these stimuli, because cortical regions in which the stimuli are coded similarly are predicted to show no reliable differences in activity, and such "null effects" are uninterpretable. Furthermore, even regions in which neural activity differentiates between the stimuli could be involved in abstract coding; they could mediate *active interpretation* of the stimulus (akin to a hypothesis testing process; Nusbaum and Schwab, 1986) in which the same percept is arrived at but via different computations. To circumvent these difficulties, we employed a method based on the logic that repeated processing of a stimulus is associated with decreased neural activity in regions involved in the processing of the stimulus (*repetition suppression*; Grill-Spector et al., 2006; Krekelberg et al., 2006). We made use of the audiovisual stimulus described above (/PK/; $PA_{Aud}KA_{Vid}$; often perceived as "ta"). During an fMRI study, this target stimulus (/PK/) was intermittently preceded by one of the following:

(1) an audiovisual /PA/ ($PA_{Aud}PA_{Vid}$; auditory overlap with /PK/),
(2) an audiovisual /KA/ ($KA_{Aud}KA_{Vid}$; visual overlap with /PK/),

Table 1. Overlap of Target Stimulus (/PK/) with Four Sorts of Preceding Stimulus

	Dimension of Overlap with Target /PK/		
Preceding Stimulus	Auditory	Visual	Perceptual
/PA/ (PA$_{\text{Aud}}$PA$_{\text{Vid}}$)	✓	×	×
/KA/ (KA$_{\text{Aud}}$KA$_{\text{Vid}}$)	×	✓	×
/TA/ (TA$_{\text{Aud}}$TA$_{\text{Vid}}$)	×	×	✓
/PK/ (PA$_{\text{Aud}}$KA$_{\text{Vid}}$)	✓	✓	✓

(3) an audiovisual /TA/ (TA$_{\text{Aud}}$TA$_{\text{Vid}}$; the same speech percept as /PK/, but nonoverlapping auditory and visual tracks).

This design enabled us to assess the relative facilitation for the target /PK/ as a function of the preceding stimulus (see Table 1 for summary of design).

A straightforward prediction was that regions known to be involved in processing relatively low-level auditory properties of the input would show less activity for the target /PK/ when it is preceded by /PA/ than when preceded by /TA/ or /KA/, since only in the first case does the preceding stimulus overlap with the target's auditory component (a low-level *auditory-repetition* pattern). Crucially, if there exist brain regions that represent the target stimulus in terms of the perceived speech category (i.e., what is "heard"), then these regions should show a very different pattern of activity. They should show less activity for the target /PK/ when it is preceded by /TA/ than when preceded by either /PA/ or /KA/. This finding would indicate that these regions code the utterance at an abstract level corresponding to perception, because the target stimulus /PK/ is equivalent to /TA/ in terms of the speech percept, but does not overlap with it either auditorily or visually (an *abstract-repetition* pattern).

Our test for identifying regions involved in abstract coding is stringent because it entails a reversal of the intuitive hypothesis on which processing an audiovisual stimulus should be associated with less neural activity when it shares an auditory or visual aspect with a preceding stimulus than when it shares neither. To validate the method, we also included a fourth (control) condition where the target stimulus was preceded by itself; this condition was expected to be associated with the least amount of neural activity as it constitutes a full repetition of the target's sensory and perceptual dimensions (see Table 1).

The design we employed borrows its logic from a number of studies of the visual system that examined coding of abstract and low-level properties of visually presented objects. In such studies, a cortical region is said to code for abstract properties of a stimulus (i.e., its *type*) if it shows reduced activity for a perceptually novel item that is drawn from the same category as a previously presented item. In contrast, a region is said to code for low-level sensory properties of a stimulus (i.e., a specific *token*) if it shows reduced activity for a given item only when it is identical to a previously presented item. Studies employing this logic have been pivotal in identifying cortical regions coding for view-invariant versus view-specific representations of faces (Grill-Spector et al., 1999; Pourtois et al., 2005) and objects (Vuilleumier et al., 2002) as well as coding for abstract versus low-level features of man-made objects and natural kinds (Simons et al., 2003).

We considered several cortical regions as candidates for abstract coding of speech. These included the auditory association cortex, inferior parietal regions, and ventrolateral prefrontal regions. Neuroimaging studies have shown that sublexical speech sounds like the ones used here evoke greater neural activity than nonspeech sounds (matched for spectral or temporal properties) in temporal and inferior parietal regions (e.g., Belin et al., 2000; Benson et al., 2006; Jäncke et al., 2002; Liebenthal et al., 2005; Vouloumanos et al., 2001). The interpretation of these results is a matter of debate. As noted by Belin et al. (2004), finding regions that show differential responses to speech versus nonspeech inputs does not address what sort of processing is indexed by this activity or the nature of the constructed representations. For instance, some have suggested that these cortical regions perform general auditory functions that support speech, such as representing fine spectral and temporal features of the stimulus, and that the increased activity for speech in these regions owes to the fact that it relies particularly strongly on these functions (e.g., Binder et al., 2000; Jäncke et al., 2002). Others have linked this activity to a more active process, involving "discriminations and categorization between highly similar exemplars of a sound category" (Belin et al., 2004). Yet, despite the difference in interpretation, such explanations have in common the notion that activity in temporal and inferior parietal regions during speech has to do with constructing an accurate representation of the sensory properties of speech. Thus, these explanations differ in a fundamental way from the putative model we are examining, on which these regions may code for properties that are independent of the sensory cues in the input.

Speaking more directly to the idea of abstract coding are findings showing that acoustic transitions are associated with increased activity in the left superior temporal and surpramarginal gyri (STG, SMG) when perceived as a categorical phonetic change than when perceived as an acoustic change (Jacquemot et al., 2003). Activity in left SMG has also been associated with acquiring a new phonetic category (Golestani and Zatorre, 2004). More generally, these regions have been associated with a "speech mode" of auditory processing, in that neural activity in these regions varies when artificial nonspeech stimuli are perceived as speech (Benson et al., 2006; Dehaene-Lambertz et al., 2005; Meyer et al., 2005).

Studies of multisensory perception suggest that temporal regions, but also ventrolateral prefrontal regions (VLPFC), are well positioned to take advantage of multisensory speech cues in constructing an abstract representation. Primate studies examining spiking activity,

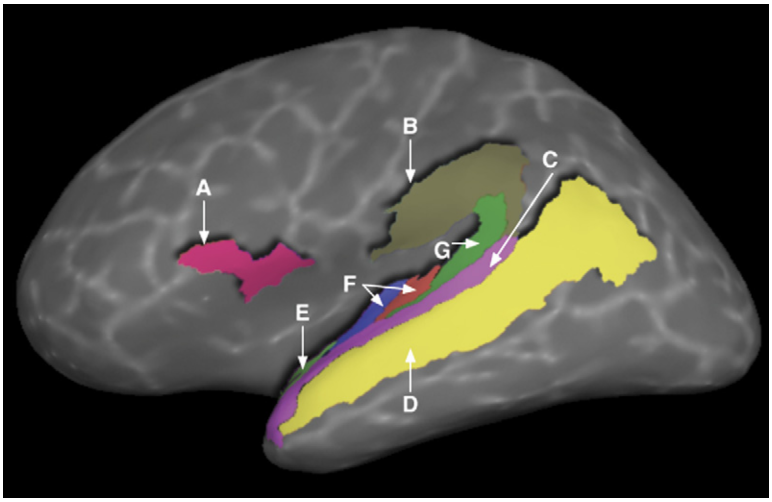

Figure 1. Anatomical Regions of Interest in the Current Study

(A) Pars opercularis of the inferior frontal G. (IFGOp); (B) supramarginal G.; (C) superior temporal G.; (D) superior temporal S.; (E) planum polare; (F) transverse temporal transverse temporal gyrus and sulcus; (G) planum temporale.

lateral field potentials, and the BOLD response have shown that temporal and prefrontal regions are sensitive to both auditory and visual information (Kayser et al., 2007), as well as to the match between them (Barraclough et al., 2005—superior temporal sulcus [STS]; Ghazanfar et al., 2005—primary auditory cortex and lateral belt with more multisensory sites on the lateral belt; Sugihara et al., 2006—VLPFC). Some of these studies have also reported unique responses in these regions for multimodal stimuli that include faces (Ghazanfar et al., 2005; Sugihara et al., 2006). In an imaging study with humans, Miller and D'Esposito (2005) found that temporal regions (left Heschl's gyrus, bilateral STS) and the left inferior frontal gyrus (IFG) show differential activity to audiovisual stimuli when these are perceived as "fused"; i.e., as having temporally synchronized audiovisual cues, independent of the actual synchrony of these cues. Finally, the notion that certain regions code speech stimuli at an abstract level is suggested by the finding that the population codes in prefrontal regions show greater similarity in BOLD response patterns for two stimuli when those are perceived similarly (/PK/ and /TA/; Pearson's *r* ~0.3) than when they are not perceived similarly (/PK/ versus /PA/ or /KA/; Skipper et al., 2007).

Our experiment, capitalizing on the well-established neural repetition suppression effect enabled us to investigate in a controlled manner whether activity in these regions tracks the perceived speech percept. If so, this would indicate that the neural coding of speech involves a departure from coding its sensory properties per se, in favor of a more abstract representation.

RESULTS

We conducted four analyses. First, as a validity check we determined if neural activity for the target /PK/ (collapsing over the preceding stimulus) was similar to previously reported findings (e.g., Ojanen et al., 2005; Pekkola et al., 2005a). We found that activation patterns closely matched that literature (see Figure S1 available online). The second analysis was a prerequisite for the main region of interest (ROI) analysis, and established whether the stimulus that preceded the target /PK/ affected the relative delay (i.e., phase shift) of the target's hemodynamic response function (HRF; cf., Formisano and Goebel, 2003; Taylor and Worsley, 2006; Thierry et al., 2003). Such delays could result in systematic over- or under-estimation of the hemodynamic response (see Experimental Procedures), and would therefore need to be quantified so that they could be accounted for in the ROI analysis proper.

The third analysis was the central analysis of the study, in which we identified regions coding for auditory- and abstract-level properties. This was a ROI analysis in which we tested for these repetition patterns in cerebral cortical regions associated with (1) early auditory processing: primary auditory cortex, located at the transverse temporal gyrus and sulcus (TTG bilaterally, cf., Hackett et al., 2001; Morosan et al., 2001); (2) secondary auditory cortices on the supratemporal plane associated with higher-level auditory processing: planum polare and planum temporale (PP and PT bilaterally, Griffiths and Warren, 2002); and (3) regions associated with sublexical speech processing: the pars opercularis of the IFG (IFGOp), STS, STG, and SMG. Anatomical regions were delineated on the surface representation of each participant's cortex using automatic parcellation tools whose accuracy has been shown to be similar to that of manual parcellation (Fischl et al., 2004; Figure 1 presents the ROIs delineated on the cortical surface of a single participant).

Within each anatomically defined ROI, we examined the repetition effect in brain regions identified in two independent functional localizer scans that identified regions sensitive to auditory and visual stimuli (cf. Miller and D'Esposito, 2005). These functional subdivisions within each ROI served as the unit of analysis, and we only probed for abstract-level facilitation effects in those subdivisions showing sensitivity to prior context as determined by an analysis of variance with participants modeled as a random

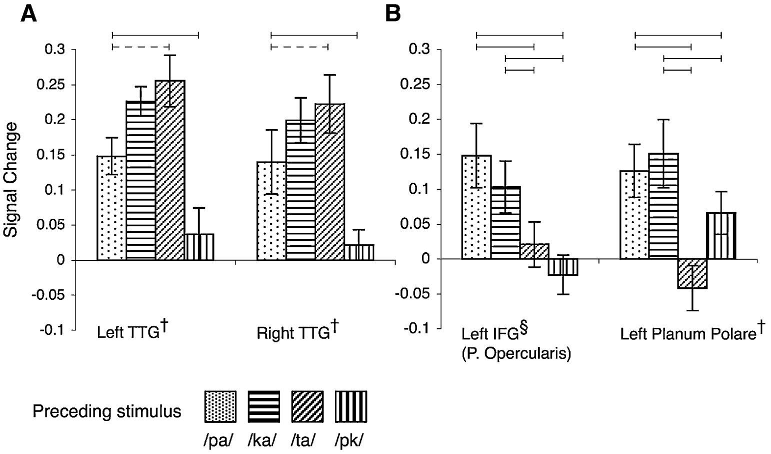

Figure 2. Neural Activity for the Target Audiovisual Stimulus /PK/ as a Function of Preceding Stimulus

Neural activity corresponded to an auditory-level repetition pattern in the transverse temporal gyrus bilaterally (A), but to an abstract-level repetition pattern in posterior left IFG and left planum polare (B). Conditions that differ reliably are connected via horizontal bars; continuous lines $p < 0.05$, dashed lines $p < 0.01$ (two-tailed t tests). † indicates a region independently identified as sensitive to auditory input in a localizer scan, § indicates a region independently identified as sensitive to visual input in a localizer scan. Error bars indicate SEM across participants.

factor (see Experimental Procedures). We expected that TTG, which is involved in early auditory processing, would demonstrate an auditory-repetition pattern. Our main question was whether an abstract-repetition pattern would be found in the other ROIs.

Determining the Effect of Preceding Stimuli on the Lag of the Target's HRF

As a prerequisite for the ROI analysis, we employed a procedure that established the relative lag of the HRF of the target /PK/ in each of the four experimental conditions, for each reliably active vertex (unit of surface area). This whole-brain analysis, while not being the main focus of the study, revealed that the phase of the HRF in voxels active for /PK/ was affected by preceding context (see Figure S2). In particular, it is important to note three patterns in these data. First, observation of /PK/ was generally associated with less neural activity when preceded by /TA/ or /PK/. Second, when /PK/ was preceded by either /PA/, /KA/, or /TA/, then activity in primary auditory regions was associated with a HRF peaking within ~4.5 s, but activity in secondary auditory cortices tended to be delayed by an additional 1.5 s. Finally, when /PK/ was preceded by /PK/, most HRFs were delayed by 1.5 s. These results are consistent with a small body of imaging studies showing that the delay of HRF may be modulated by factors such as working memory demands (Thierry et al., 2003) or the degree of habituation to a stimulus (Taylor and Worsley, 2006). However, currently there is not a clear model of the neurophysiological processes associated with such variations (Formisano and Goebel, 2003). The procedure highlights the importance of identifying differences in HRF lags when quantifying signal-change in experimental conditions.

ROI Analysis

All ROIs demonstrated lower neural activity for the target stimulus /PK/ when preceded by itself than when preceded by /PA/ or /KA/ (i.e., no ROI demonstrated "repetition enhancement," cf. Henson, 2003). As expected, TTG (bilaterally) showed an auditory-repetition pattern, indicating processing of the target's acoustic component (Figure 2A). Most importantly, two brain regions in the left hemisphere demonstrated an abstract-repetition pattern: both regions demonstrated reliably lower neural activity for the target syllable when preceded by /TA/ than when preceded by either /PA/ or /KA/. These regions were the pars opercularis in the posterior portion of left IFG (IFGOp, ~BA 44), and the left PP, an auditory association cortex (Figure 2B). To determine if this pattern was prevalent in cortical regions for which we had no a priori hypotheses, we examined the remaining 75 brain regions in the anatomical parcellation scheme we utilized (Fischl et al., 2004). If we were to find an abstract-repetition pattern in brain regions not typically implicated in speech comprehension, this would suggest that the two regions identified here are a subset of a larger network. However, no other brain region demonstrated an abstract repetition pattern, suggesting that such effects may be relatively limited to brain regions previously implicated in speech processing. For IFGOp and PP, we also conducted post-hoc analyses to determine whether preceding /PK/ by /PK/ or /TA/ resulted in different magnitudes of the BOLD response. In IFGOp, preceding /PK/ by /TA/ or /PK/ did not result in reliably different patterns of activity ($p > 0.3$), however, in PP, preceding /PK/ by /TA/ resulted in less neural activity for the target than when it was preceded by /PK/ $t(19) = 2.34$, $p < 0.05$.

In IFGOp, the abstract repetition pattern was found in cortical areas identified by the visual-only localizer, and in PP it was found in areas identified by the auditory-only localizer. To determine to what extent the areas identified in IFGOp by the visual localizer were sensitive to auditory information, we calculated the proportion of visually sensitive areas in IFGOp that were also sensitive to auditory input. We found that in this region, only a relatively small fraction of cortical areas active in the visual localizer were active during the auditory localizer (mean = 19%, median = 3%). This was the case even though the auditory localizer resulted in greater activity in IFGOp (mean proportion of active vertices = 10% [SEM = 2] versus 6% [SEM = 2] in the auditory and visual localizers, respectively;

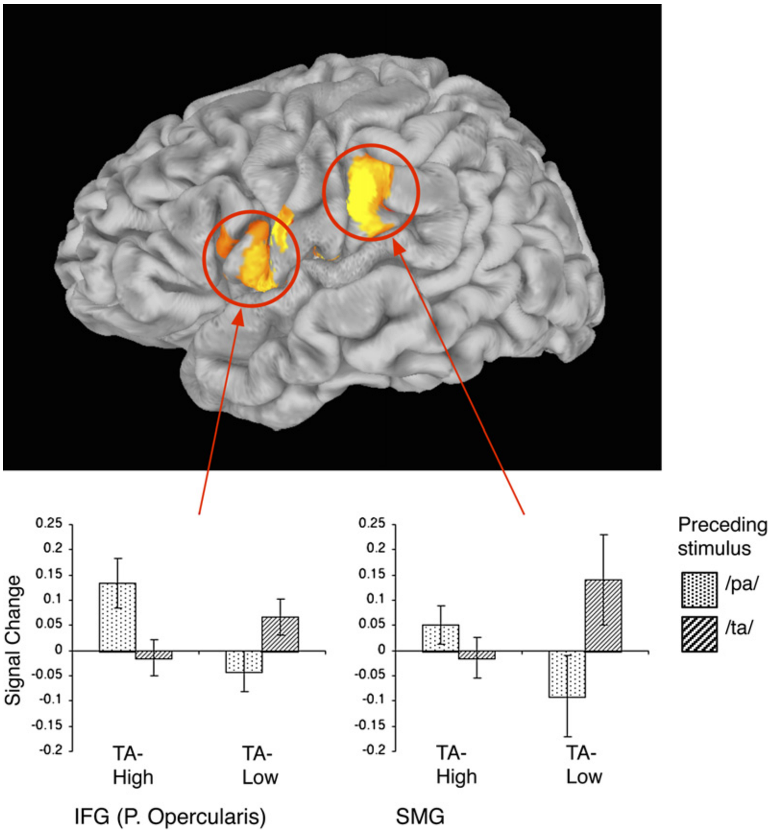

Figure 3. Region where TA-High and TA-Low Perceivers Showed Differential Sensitivity to the Preceding Stimulus when Processing the Target /PK/

During processing of /PK/, participants who perceived the target as "TA" (TA-High group, n = 11) showed different sensitivity to prior stimulus (/PA/, /TA/) than participants who perceived the target as "KA" (TA-Low group, n = 8). Such differential sensitivity was identified in a continuous cluster in the left hemisphere extending from anterior SMG to the pars opercularis in posterior IFG (individual vertex threshold for the interaction term $p < 0.05$, family-wise error corrected, $p < 0.05$). The bar graphs describe the nature of this interaction in two sample aspects of this cluster (SMG, posterior IFG): they show the mean activity for /PK/ as function of preceding stimulus (PA versus TA) and group-classification (TA-High versus TA-Low). Error bars indicate SEM across participants. TA-High participants were more facilitated by /TA/ than by /PA/ whereas TA-Low participants demonstrated the opposite pattern.

$p > 0.17$). In general, PP was not sensitive to visual information, with only 5 of the 22 participants showing responses to the visual-only localizer in this region.

Individual Differences

If the abstract-level repetition patterns found for IFGOp and PP on the group level are a result of the target stimulus being neurally coded as TA, then these effects should be stronger for individuals for which the /PK/ utterance is perceived as "ta" more often. That is, though the target /PK/ is usually perceived as "ta," some individuals perceive /PK/ as "ta" more often than others. To examine this prediction, we employed an independent behavioral procedure, carried out under identical scanner noise conditions, to partition the participants into "TA-High" or "TA-Low" perceivers (TA-Low perceivers were ones for which the modal percept was "ka"). We then conducted a reanalysis of fMRI data from the passive observation task to identify cortical areas where the effect of the preceding stimulus (/PA/, /TA/) on the processing of /PK/ differed for these two groups of participants. This whole-brain analysis revealed one cluster in the left hemisphere, extending from the anterior SMG via the subcentral gyrus to posterior IFG, where the effect of preceding stimulus varied as a function of participant group (see Figure 3). In this area, TA-High perceivers showed less neural activity for the target /PK/ when preceded by /TA/ than by /PA/, whereas TA-Low participants demonstrated the opposite pattern. We then examined the group effects specifically in those two ROIs where we found abstract-repetition patterns. For IFGOp, we found that participants classified as TA-High perceivers demonstrated a pronounced abstract-repetition pattern, whereas participants classified as TA-Low perceivers did not show this pattern and showed less differentiation between the four conditions. For PP, TA-High perceivers again showed an abstract-repetition pattern. However, only four individuals classified as TA-Low perceivers showed reliable activity in this region and so no definitive conclusion can be drawn for this group (see Figure S3). Thus, both the whole-brain and ROI reanalyses supported the hypothesis that abstract-repetition patterns in cortical regions are particularly strong for those participants that perceive the target stimulus as TA more often.

DISCUSSION

Using a well-established experimental paradigm, we examined whether there are neural substrates that code audiovisual speech utterances on an abstract level that transcends their sensory components and that corresponds to a linguistic speech category. Our results provide the first demonstration of such a coding for audiovisual speech and show that it takes place in the pars opercularis of the left IFG and the left PP. Importantly, such results were found in a passive task that was devoid of any explicitly defined decision judgment task. Further, consistent with the position that these effects index repetition at the abstract/speech-category level, such effects were more robust for participants classified as TA-High perceivers in left hemisphere cortical areas extending from the supramarginal gyrus to IFGOp.

Our findings are in accord with the primate and child development literature. In the macaque, the anterior-lateral portions of the auditory belt (roughly corresponding to PP in the human) project to the ventral prefrontal cortex (Deacon, 1992; Romanski et al., 1999a, 1999b), a region proposed to be homologous to Broca's area in humans (e.g., Rizzolatti and Arbib, 1998). Both these regions differentiate between different sorts of monkey calls, indicating relatively high-level auditory processing (Romanski and Goldman-Rakic, 2002; Tian et al., 2001). There is additional evidence that neurons in the ventral prefrontal cortex are tuned to higher-order auditory features of monkey vocalizations rather than to low-level spectrotemporal acoustic features (Cohen et al., 2007). Recent developmental data in humans acquired using MEG show that starting at around the age of 6 months, IFG begins responding to speech syllables, but is not responsive to tones; furthermore, this differentiation is accompanied by coupling of neural activity between IFG and superior temporal regions during syllable processing (Imada et al., 2006). Thus, both IFG and the anterior temporal regions are likely candidates for abstract sublexical speech processing in humans.

Our results strongly implicate left IFGOp in abstract representation of audiovisual speech. These results accord with prior findings showing that this region is sensitive to the temporal synchronization of auditory and visual inputs: specifically, it has been shown to demonstrate less activity for synchronized than nonsynchronized inputs (Miller and D'Esposito, 2005; Ojanen et al., 2005; Pekkola et al., 2005a). Our demonstration of the involvement of left IFGOp in abstract coding of audiovisual speech is also in line with several other studies attesting to its role in language processing. It is specifically this posterior region of IFG that has been causally associated with phonological processing using TMS (Gough et al., 2005), and it has been implicated numerous times in lower-level processing of written words (e.g., Paulesu et al., 1997; Poldrack et al., 1999; Wagner et al., 2000), consistent with its purported role in phonological encoding. Notwithstanding, it is important to recognize that reports of activity in IFGOp have been conspicuously absent from several studies that are strongly related to the current domain of inquiry, concerning the processing of sublexical auditory speech: first, two studies contrasting phonetic versus acoustic perception (Golestani and Zatorre, 2004; Jacquemot et al., 2003) failed to show involvement of this region in phonetic-level processing. Second, contrasts of sublexical speech versus nonspeech sounds have often *not* revealed differential activity in this region (e.g., Binder et al., 2000; Jäncke et al., 2002; Liebenthal et al., 2005; Vouloumanos et al., 2001). Below, we discuss possible constraints that could account for the current findings and the fact that this region has sometimes not been identified in prior studies.

One explanation for the absence of reliable results in this region in prior studies is that many imaging studies have employed strategic tasks when studying speech processing. Such tasks have been shown to modulate neural activity in left IFG (Binder et al., 2004; Blumstein et al., 2005; Burton et al., 2005; Hasson et al., 2006), which could reduce the magnitude of experimental effects in this region. Another possibility is that the neural substrates identified here are particularly involved in processing audiovisual input. Note that the abstract-repetition pattern in IFGOp was found in a functional area established by an independent visual-only localizer, but not in a selection established by independent auditory-only localizer. And importantly, in IFGOp we found little overlap between these functional areas. Finding the abstract repetition pattern in areas responsive to unimodal visual stimuli but less responsive to unimodal auditory stimuli suggests that these areas are more sensitive to auditory information in audiovisual contexts than when auditory information is presented alone. The sensitivity of these visually responsive areas to auditory information *in audiovisual contexts* is demonstrated by lower activity for the target /PK/ when it was preceded by /PK/ (a straightforward repetition) than when preceded by /KA/ (overlapping visually but not auditorily). Our findings accord with recent data from Sugihara et al. (2006), who reported such multisensory interactions in cells of the primate VLPFC. These cells might respond to visual-only input but not to auditory-only input and still respond much more strongly to audiovisual input than visual-only input. These findings suggest that visually sensitive areas in IFGOp may show potentiated responses to auditory stimulus in audiovisual contexts.

Another possibility is that the cortical areas demonstrating abstract-repetition effects in IFGOp may mediate categorization or identification of audiovisual or visual inputs. As mentioned, the localizer task that identified these areas consisted of passive observation of silent articulations of /PA/, /KA/, and /TA/, in absence of any explicit task. Though being a passive task, it could still be that participants were endogenously driven to lip-read and identify the silent articulations. Activity in this region has been previously reported during passive observation of silent articulations that are difficult to categorize (generated by reversing videos of word articulations; Paulesu et al., 2003) but also during active tasks where participants classified silent articulations (e.g., Calvert and Campbell, 2003). Similarly, the main experimental task for which we assessed activity in this area consisted of passive observation of audiovisual stimuli but could still have additionally prompted endogenously driven identification of these stimuli. Hickok and Poeppel (2004) suggest that frontal systems are utilized for speech processing when experimental tasks call for explicit analysis of the stimulus. Indeed, prior studies linking posterior IFG with audiovisual integration had used explicit tasks, e.g., reporting whether an audiovisual stimulus is perceived as fused (Miller and D'Esposito, 2005) or reporting whether the two modalities are matching or conflicting (Ojanen et al., 2005; Pekkola et al., 2005a). Nonetheless, this region is also active during passive listening to auditory speech in absence of task demands (e.g., Crinion and Price, 2005; Hasson et al., 2007). This region seems implicated in low-level speech processing, because in contrast to more anterior IFG regions, its activity is not modulated as function of the information communicated by sentences, nor does its activity predict memory for such contents (Hasson et al., 2007). Thus, while it is impossible to rule out the possibility that in the current study this region was driven by internally driven identification processes, there are also reasons to think it is regularly involved in speech comprehension. In other work, we have addressed the role of this region in the context of a larger functional network mediating audiovisual comprehension (Skipper et al., 2006, 2007). In particular, we have presented data suggesting that during audiovisual speech perception, inferior frontal and premotor regions play a role that is akin to generating an initial hypothesis about the communicated speech category and that these frontal regions are functionally connected to auditory and somatosensory areas. The information flow in this model is akin to that in a closed-control circuit, and the goal of processing is to minimize the discrepancy between the hypothesis generated in frontal regions and the sensory input registered in auditory and somatosensory regions. On this view then, activity in IFGOp is driven by an endogenously controlled ''active'' process.

Relatively few studies have specifically delineated the PP as a region of interest when examining human processing of nonsemantic sublexical speech; it is often conjoined for purposes of analysis with more lateral and anterior aspects of STG that are known to have different cytoarchitectonic structure (Morosan et al., 2005). Still, our results are strongly consistent with prior studies that have linked activity in this region to stimulus identification. This region, but not more posterior auditory association cortex, shows greater activity when individuals identify a specific environmental sound from within a stream of such sounds than when they attend to the location of acoustic stimuli (Viceic et al., 2006). This region also responds more strongly to changes in the acoustic properties of auditorily presented stimuli than to changes in their perceived location, and this sensitivity to acoustic change is increased when individuals are cued to pay attention to such changes (Ahveninen et al., 2006). Furthermore, this is one of few cortical regions that are involved in listening and producing both speech and song (Callan et al., 2006). It is important to note that the abstract-repetition result in PP was found in neural substrates independently identified by a localizer scan presenting auditory stimuli and that PP did not demonstrate reliable activity across individuals during the presentation of visual stimuli. This suggests that PP may be particularly sensitive to visual input when such input is paired with an ecologically matched and structured auditory stimulus. Such multisensory interactions have been reported in the primate auditory cortex. For example, Ghazanfar et al. (2005) recorded local field potentials in the lateral auditory belt and found that the vast majority of cortical sites in this region showed multisensory interactions (~90%). Importantly, such sites could be unresponsive to visual-only stimuli, strongly responsive to auditory-only stimuli, and still show reliably stronger (or reliably lower) activity to audiovisual stimuli than auditory-only stimuli. Interactions between the two modalities have also been found in the human auditory association cortex, though more posteriorly (Möttönen et al., 2002; Pekkola et al., 2006).

As expected, the region that showed the strongest effect of auditory-level priming was the TTG (Heschl's gyrus), thought to be the location of primary auditory cortex in the human (Hackett et al., 2001; Morosan et al., 2001). It showed less activity for /PK/ when the preceding stimulus was identical on the acoustic track (i.e., following /PK/ and /PA/). Interestingly, TTG showed less activity when the target /PK/ appeared after /PK/ than when it appeared after /PA/ (both of which share the exact same acoustic component), indicating that it is sensitive to whether or not the visual stimulus remained constant. This finding corroborates a body of prior results showing that TTG is activated by visual presentations of silent articulations (e.g., Pekkola et al., 2005a, 2005b) and is sensitive to changes in visual stimuli (Colin et al., 2002; Möttönen et al., 2002, 2004; Sams et al., 1991).

The other ROIs we examined demonstrated activation patterns that did not fall into the auditory- or abstract-level facilitation patterns. In particular, the planum temporale (PT) bilaterally demonstrated similar activity for the target when preceded by either /PA/, /KA/, or /TA/, and in all three cases, this activity was greater than in the condition when the target was preceded by itself. Note that when the target /PK/ was preceded by /PA/, /KA/, or /TA/, then in each case the transition from the preceding stimulus to the target constituted either a visual or auditory change, or both. Given that PT is sensitive to auditory but also visual input (Di Virgilio and Clarke, 1997), it is possible that any sensory change resulted in neural activity that overshadowed more subtle abstract-level facilitation effects in the region.

Do our findings suggest that the posterior left IFG and the planum polare are in some way ''specialized'' (sometimes interchangeably referred to as unique, vital, selective, dedicated, or fundamental) for speech? The answer to this often-raised question hinges on how ''specialization for speech'' is to be understood. On the one hand, our findings demonstrate that these left-hemisphere regions code audiovisual speech at an abstract level of representation that transcends its sensory properties and that corresponds to the perceived speech percept. Thus, these regions play a very important role in encoding audiovisual speech, as the coding of such information is likely to aid further language-processing stages involving lexical access. On the other hand, these regions could play a similar role in the perception of nonspeech audiovisual input. Speaking to this possibility, a few recent behavioral studies have shown that observing a musical performance can affect perception of the accompanying auditory track. For example, a performed note may be perceived as starting more abruptly when accompanied by a video of a person plucking the note than when accompanied by a video where the note is played with a bow (Saldaña and Rosenblum, 1993). Similarly, a video of a performer singing two tones affects the assessment of the interval between the tones; e.g., a 9-semitone interval is judged as a smaller interval when it is accompanied by a video of a singer performing a 2-semitone interval rather than when accompanied by a video of a singer performing a 9-semitone interval (Thompson et al., 2005). Thus, the neural mechanisms that integrate auditory and visual input toward a more abstract percept may or may not be unique to speech, and more research is necessary to answer this question (see Price et al., 2005, for a detailed discussion of this question).

Summary

We have shown that there are neurophysiological substrates that code properties of an audiovisual utterance at a level of abstraction that corresponds to the speech category that is ''heard,'' which can be independent of its sensory properties. We set out from the observation that there is no need to posit the existence of abstract coding to explain emergent features of audiovisual speech, because these features may just be the result of joint activity in lower-level unisensory regions. Yet, our results indicate that neural activity in left-hemisphere regions does indeed track the experienced speech percept, independent of its sensory properties.

EXPERIMENTAL PROCEDURES

Acquisition and Design

The analyses we report were conducted on a data set collected as part of prior research in our laboratory on the relation between audiovisual comprehension and production (Skipper et al., 2007). Here, we analyzed a rapid event-related fMRI experiment (3T), where participants (n = 22) passively observed a speaker pronouncing audiovisual syllables (TR = 1.5, TE = 24ms, flip angle = 71°, effective resolution = 3.75 × 3.75 × 3.8, 29 slices, no gap). The target syllable /PK/ ($PA_{Aud}KA_{Vid}$) was randomly preceded by one of four AV syllables: /PA/ ($PA_{Aud}PA_{Vid}$), /KA/ ($KA_{Aud}KA_{Vid}$), /TA/ ($TA_{Aud}TA_{Vid}$), or the target itself /PK/. The functional run was 7 min long and consisted of 280 volume acquisitions. The four types of condition (/PK/, /KA/, /TA/, and /PA/) were equally frequent, and the target /PK/ was preceded eight times by each stimulus. As such, the presentation method did not single out the target /PK/ as a stimulus of interest because it was embedded within equally frequent stimuli. The stimuli were presented in an event-related manner with a variable interstimulus interval (mean ISI = 3 s). Participants passively listened to and observed these stimuli because explicit judgments of linguistic materials have been shown to affect neural activity in brain regions involved in language comprehension (Binder et al., 2004; Blumstein et al., 2005; Hasson et al., 2006).

To determine which participants tended to perceive /PK/ as ''ta,'' following this passive task participants were again presented with the experimental protocol in the scanner, during functional acquisitions (i.e., under noise conditions identical to those in the passive listening task), but in the context of an active task that asked them to indicate for each stimulus whether they perceived it as ''pa,'' ''ka,'' or ''ta'' (34 trials of each stimulus, mean ISI = 3 s, 480 whole-brain acquisitions, 12 min length). On the basis of behavioral responses during this task, we partitioned participants according to whether their modal percept of /PK/ was ''ta'' (TA-High perceivers; n = 11) or ''ka'' (TA-Low perceivers; n = 8) (''pa'' was never the modal percept, and three participants were not included in these groups as they expressed uncertainty about whether they correctly remembered the response key assignments). Participants' responses for /PA/, /KA/, and /TA/ in the active task further showed that these stimuli were perceived accurately under the scanning noise conditions: mean identification accuracy was 95% (SEM = 0.02), indicating very good discrimination of the stimuli. Movement parameters in the passive and active scans did not differ reliably (a within-participant t test on the amount of movement in the two scans; $t(21) = 1.43$, $p > 0.16$).

Statistical Analysis: General Deconvolution and Surface-Mapping Procedures

On the individual level, statistical analysis was based on a regression model that partitioned the presentations of the target syllable as a function of preceding context. BOLD signals were acquired every 1.5 s (TR = 1.5 s) and the hemodynamic response function (HRF) for each condition was established via regression for the 12 s following the presentation of the target (eight acquisitions) without making a priori assumptions about its shape (Saad et al., 2006). Parcellation of cortical anatomy into ROIs was performed using the FreeSurfer software suite (Dale et al., 1999; Fischl et al., 1999, 2004). These tools delineate anatomical divisions via automatic parcellation methods (Fischl et al., 2004) in which the statistical knowledge base derives from a training set incorporating the anatomical landmarks and conventions described by Duvernoy (1991). The accuracy of these methods is similar to that of manual parcellation (Fischl et al., 2002, 2004).

Statistical Analysis: Whole-Brain Analysis

To establish which brain regions were sensitive to the target /PK/, collapsing over the preceding stimulus, we conducted the following vertex-wise analysis on the cortical surface. For each vertex, for each of the four context conditions, a mean HRF for /PK/ was created by averaging across participants. This resulted in four HRFs per vertex (each HRF modeled with eight data points). A 4 (context) × 8 (time) ANOVA was then conducted for each vertex, and a vertex was said to show activity for /PK/ if it demonstrated a main effect of time (that is, if the activity in that vertex demonstrated a systematic departure from baseline). Threshold was determined at $p < 0.001$ on an individual vertex level, uncorrected, to evaluate the data against the prior literature; prior studies—Ojanen et al. (2005) and Pekkola et al. (2005a)—set individual voxel thresholds at $Z > 3$ ($p < 0.0013$) and $Z > 1.8$ ($p < 0.036$), respectively.

Statistical Analysis: Regions of Interest

In addition to the main experiment, the participants were also scanned in two independent ''localizer'' scans to determine whether the locus of the repetition patterns was in neural substrates responsive to auditory information or visual information. In one localizer we presented just the auditory track of the audiovisual syllables used in the main experiment, and in the other localizer we presented just the visual track of these syllables. These localizer scans utilized the same fMRI acquisition parameters as those used in the passive task but included only the /PA/, /KA/, and /TA/ stimuli (7 min in length, 45 tokens of each stimulus). After parcellating the individual participants' brain surfaces into anatomical regions as described above, for each participant we used the localizer data to delineate functionally defined subdivisions within each region. That is, we established which parts of each anatomical region were sensitive to visual or auditory information. For each participant, we considered a functional (i.e., localizer-established) selection in an anatomical region as candidate for group-level analysis if, for that participant, the functional selection consisted of more than 20 reliably active vertices (reliability defined as $p < 0.05$, FDR corrected, Genovese et al. [2002]).

Auditory or visual functional selections (in SMG, STS, STG, IFGOp, PP, and PT) were analyzed at the group level only if at least 18 of the 22 participants each demonstrated reliable activity (i.e., at least 20 active vertices, $p < 0.05$ FDR corrected) for the respective localizer scan. There were therefore a maximum of 24 potential search regions (6 ROIs × 2 hemispheres, × 2 functional selections). Applying this initial filter, based solely on the localizer data, resulted in 21 candidate search regions, as IFGOp on the right was not sensitive to auditory information, and neither left nor right PP were sensitive to visual information. Of these 21 candidate search regions we identified those where activity for the target stimulus varied reliably as a function of preceding stimulus, using a repeated-measures ANOVA with participants as a random factor. We only probed for an abstract-repetition pattern in functional subdivisions where the ANOVA was reliable ($p < .01$). The ANOVA was reliable in 17 search regions as follows: SMG, auditory and visual selections bilaterally; STS, auditory and visual selections bilaterally; STG, visual selection bilaterally and auditory on the right; IFGOp, visual selection bilaterally and auditory on the left; PP, auditory selection bilaterally; PT, visual selection bilaterally. It was within these 17 functional subdivisions that were constrained by anatomical and functional criteria that we probed for an abstract-repetition pattern. The t tests conducted within each region did not assume equal variances and used Welch's approximation of the degrees of freedom. Because the four contrasts conducted within each region were not independent, we used Monte-Carlo simulations to estimate their joint probability in each region. Specifically, these simulations computed the chance probability of finding an abstract-repetition pattern for each region, while taking into account the covariance of the data in the region. In no region did the probability of finding an abstract-repetition pattern exceed $p = 0.00054$ (in IFGOp and PP the probabilities were $p = 0.00014$, and $p < 0.00001$, respectively). Assuming a probability of $p = 0.00054$ for the chance occurrence of the pattern in any single region, such a pattern is unlikely to occur by chance one or more times of 17 tests (see Establishing Probability for Finding an Abstract Repetition Pattern in at Least One Region of Interest in the Supplemental Data for details of the simulation procedure).

For group level analysis, the mean signal change in each functional subdivision was averaged across subjects in each of the four conditions. The *mean signal change* in each condition was defined as the mean signal estimate in the six imaging data collection time points starting at the ascent of the bold response; these six time points covered the ascent, peak, and descent of the HRF. The point of ascent of the HRF was established separately for each condition using a crosscorrelation procedure that we describe below. The importance of identifying the point of ascent separately for each experimental condition derives from the fact that certain conditions could be associated with a delayed, i.e., phase-shifted HRF, in which case averaging the signal values across the same (arbitrarily chosen) time points for all conditions will result in a systematic underestimation of activity in those conditions that peak at a relatively later point. The signal estimates in these six time points were averaged across surface vertices within each region for each participant, and then across participants. This averaging generated a neural response profile reflecting activity for the target /PK/ stimulus in each of the four experimental conditions for each functional subdivision in the anatomical regions of interest.

There are a number of advantages in this multistep procedure over an analysis procedure where data are first projected onto a common space and then averaged. Specifically, our procedure enabled us to conduct a group level analysis while (1) controlling for the inherent individual variance in anatomical structure across individuals and (2) controlling for the substantial variance that exists in the location of activation peaks during processing of auditory stimuli, particularly in belt regions outside the primary auditory cortex (Burton et al., 2000; Patterson et al., 2002; Wessinger et al., 2001).

To determine the point at which the HRF began to ascend in each vertex (for each condition) we used a crosscorrelation function (CCF). This function determined the correlation between the estimated HRF in each condition and a gamma-shaped response profile associated with a typical HRF (Cohen, 1997). This function also determined the phase-shift between the HRF and the canonical gamma function for which this correlation was maximal (i.e., the relative lag of the HRF). We could thus establish the time point where the HRF began to ascend.

Statistical Analysis: Individual Differences

In this analysis, participants' data from the passive task were registered onto a common surface template using FreeSurfer, and statistical analyses were performed in the surface domain using SUMA and AFNI (Cox and Hyde, 1997). We identified cortical regions where TA-High and TA-Low perceivers showed differential sensitivity to preceding stimulus during processing of /PK/. For each surface vertex we tested whether the difference between activity for /PK/ when presented after /TA/ (pk/ta) and activity for /PK/ when presented after /PA/ (pk/pa) differed between the two groups of participants (an interaction test: $\text{TA-High}_{pk/pa} - \text{TA-High}_{pk/ta} \neq \text{TA-Low}_{pk/pa} - \text{TA-Low}_{pk/ta}$, $p < 0.05$). We used a nonparametric Mann-Whitney test in each vertex because it was likely that data would not be normally distributed and because this test is robust against a small number of outlier values. We then searched for clusters where all vertices demonstrated a stronger abstract-repetition pattern for the TA-High group, $\text{TA-High}_{pk/pa} - \text{TA-High}_{pk/ta} > \text{TA-Low}_{pk/pa} - \text{TA-Low}_{pk/ta}$, and for clusters where all vertices demonstrated a weaker abstract-repetition pattern for the TA-High group, $\text{TA-High}_{pk/pa} - \text{TA-High}_{pk/ta} < \text{TA-Low}_{pk/pa} - \text{TA-Low}_{pk/ta}$. Cluster extent threshold was determined via permutation tests that indicated that reliable clusters would need to exceed 2100 surface vertices (~1% of the total number of vertices in a hemisphere's surface area).

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org.cgi.content/full/56/6/1116/DC1/.

ACKNOWLEDGMENTS

The fMRI study was supported by the National Institute of Deafness and Other Communication Disorders NIH RO1-DC03378. We would like to thank E. Chen for her advice regarding sampling from multivariate distributions, Anthony Dick for comments on drafts of this manuscript, and two anonymous reviewers for their insightful comments and suggestions.

Received: April 3, 2007
Revised: June 18, 2007
Accepted: September 17, 2007
Published: December 19, 2007

REFERENCES

Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., et al. (2006). Task-modulated" what" and" where" pathways in human auditory cortex. Proc. Natl. Acad. Sci. USA *103*, 14608–14613.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., and Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J. Cogn. Neurosci. *17*, 377–391.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309–312.

Belin, P., Fecteau, S., and Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. *8*, 129–135.

Benson, R.R., Richardson, M., Whalen, D.H., and Lai, S. (2006). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. Neuroimage *31*, 342–353.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., and Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex *10*, 512–528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., and Ward, B.D. (2004). Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. *7*, 295–301.

Blumstein, S.E., Myers, E.B., and Rissman, J. (2005). The perception of voice onset time: an fMRI investigation of phonetic category structure. J. Cogn. Neurosci. *17*, 1353–1366.

Burton, M.W., Small, S.L., and Blumstein, S.E. (2000). The role of segmentation in phonological processing: an fMRI investigation. J. Cogn. Neurosci. *12*, 679–690.

Burton, M.W., Locasto, P.C., Krebs-Noble, D., and Gullapalli, R.P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. Neuroimage *26*, 647–661.

Callan, D.E., Tsytsarev, V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H., and Turner, R. (2006). Song and speech: brain regions involved with perception and covert production. Neuroimage *31*, 1327–1342.

Calvert, G.A., and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. J. Cogn. Neurosci. *15*, 57–70.

Cohen, M.S. (1997). Parametric analysis of fMRI data using linear systems methods. Neuroimage *6*, 93–103.

Cohen, Y.E., Theunissen, F., Russ, B.E., and Gill, P. (2007). Acoustic features of rhesus vocalizations and their tepresentation in the ventrolateral prefrontal cortex. J. Neurophysiol. *97*, 1470–1484.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., and Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. Clin. Neurophysiol. *113*, 495–506.

Cox, R.W., and Hyde, J.S. (1997). Software tools for analysis and visualization of fMRI data. NMR Biomed. *10*, 171–178.

Crinion, J., and Price, C.J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. Brain *128*, 2858–2871.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis I: segmentation and surface reconstruction. Neuroimage *9*, 179–194.

Deacon, T.W. (1992). Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. Brain Res. *573*, 8–26.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. Neuroimage *24*, 21–33.

Di Virgilio, G., and Clarke, S. (1997). Direct interhemispheric visual input to human speech areas. Hum. Brain Mapp. *5*, 347–354.

Duvernoy, H.M. (1991). The Human Brain: Structure, Three-Dimensional Sectional Anatomy and MRI (New York: Springer-Verlag).

Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis II: inflation, flattening, and a surface-based coordinate system. Neuroimage *9*, 195–207.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al. (2002). Whole-brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron *33*, 341–355.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., and Kennedy, D. (2004). Automatically parcellating the human cerebral cortex. Cereb. Cortex *14*, 11–22.

Formisano, E., and Goebel, R. (2003). Tracking cognitive processes with functional MRI mental chronometry. Curr. Opin. Neurobiol. *13*, 174–181.

Genovese, C.R., Lazar, N.A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage *15*, 870–878.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., and Logothetis, N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J. Neurosci. *25*, 5004–5012.

Golestani, N., and Zatorre, R.J. (2004). Learning new sounds of speech: reallocation of neural substrates. Neuroimage *21*, 494–506.

Gough, P.M., Nobre, A.C., and Devlin, J.T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. J. Neurosci. *25*, 8010–8016.

Griffiths, T.D., and Warren, J.D. (2002). The planum temporale as a computational hub. Trends Neurosci. *25*, 348–353.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. *10*, 14–23.

Hackett, T.A., Preuss, T.M., and Kaas, J. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. *441*, 197–222.

Hasson, U., Nusbaum, H.C., and Small, S.L. (2006). Repetition suppression for spoken sentences and the effect of task demands. J. Cogn. Neurosci. *18*, 2013–2029.

Hasson, U., Nusbaum, H.C., and Small, S.L. (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. Cereb. Cortex, in press. Published online March 19, 2007. 10.1093/cercor/bhm016.

Henson, R.N.A. (2003). Neuroimaging studies of priming. Prog. Neurobiol. *70*, 53–81.

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition *92*, 67–99.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., and Kuhl, P.K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. Neuroreport *17*, 957–962.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J. Neurosci. *23*, 9541–9546.

Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H.J. (2002). Phonetic perception and the temporal cortex. Neuroimage *15*, 733–746.

Kayser, C., Petkov, C.I., Augath, M., and Logothetis, N.K. (2007). Functional imaging reveals visual modulation of specific fields in auditory cortex. J. Neurosci. *27*, 1824–1835.

Krekelberg, B., Boynton, G.M., and van Wezel, R.J. (2006). Adaptation: from single cells to BOLD signals. Trends Neurosci. *29*, 250–256.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., and Medler, D.A. (2005). Neural substrates of phonemic perception. Cereb. Cortex *15*, 1621–1631.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature *264*, 746–748.

Meyer, M., Zaehle, T., Gountouna, V.E., Barron, A., Jancke, L., and Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport *16*, 1985–1989.

Miller, L.M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J. Neurosci. *25*, 5884–5893.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., and Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage *13*, 684–701.

Morosan, P., Schleicher, A., Amunts, K., and Zilles, K. (2005). Multimodal architectonic mapping of human superior temporal gyrus. Anat. Embryol. (Berl.) *210*, 401–406.

Möttönen, R., Krause, C.M., Tiippana, K., and Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. Cog. Brain Res. *13*, 417–425.

Möttönen, R., Schurmann, M., and Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study. Neurosci. Lett. *363*, 112–115.

Nelken, I., and Ahissar, M. (2006). High-level and low-level processing in the auditory system: the role of primary auditory cortex. In Dynamics of Speech Production and Perception, P.L. Divenyi, S. Greenberg, and G. Meyer, eds. (Amsterdam: IOS Press).

Nusbaum, H.C., and Schwab, E.C. (1986). The role of attention and active processing in speech perception. In Pattern Recognition by Humans and Machines, Volume 1, Speech Perception, H.C. Nusbaum and E.C. Schwab, eds. (Orlando, FL: Academic Press), pp. 113–158.

Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I.P., Joensuu, R., Autti, T., and Sams, M. (2005). Processing of audiovisual speech in Broca's area. Neuroimage *25*, 333–338.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., and Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. Neuron *36*, 767–776.

Paulesu, E., Goldacre, B., Scifo, P., Cappa, S.F., Gilardi, M.C., Castiglioni, I., Perani, D., and Fazio, F. (1997). Functional heterogeneity of left inferior frontal cortex as revealed by fMRI. Neuroreport *8*, 2011–2017.

Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N.A., De Giovanni, U., Sensolo, S., and Fazio, F. (2003). A functional-anatomical model for lipreading. J. Neurophysiol. *90*, 2005–2013.

Pekkola, J., Laasonen, M., Ojanen, V., Autti, T., Jaaskelainen, I.P., Kujala, T., and Sams, M. (2005a). Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: an fMRI study at 3 T. Neuroimage *29*, 797–807.

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I.P., Mottonen, R., Tarkiainen, A., and Sams, M. (2005b). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport *16*, 125–128.

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I.P., Möttönen, R., and Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. Hum. Brain Mapp. *27*, 471–477.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., and Gabrieli, J.D.E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage *10*, 15–35.

Pourtois, G., Schwartz, S., Seghier, M.L., Lazeyras, F., and Vuilleumier, P. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. J. Cogn. Neurosci. *17*, 1043–1057.

Price, C., Thierry, G., and Griffiths, T. (2005). Speech-specific auditory processing: where is it? Trends Cogn. Sci. *9*, 271–276.

Rizzolatti, G., and Arbib, M.A. (1998). Language within our grasp. Trends Neurosci. *21*, 188–194.

Romanski, L.M., and Goldman-Rakic, P.S. (2002). An auditory domain in primate prefrontal cortex. Nat. Neurosci. *5*, 15–16.

Romanski, L.M., Bates, J.F., and Goldman-Rakic, P.S. (1999a). Auditory belt and parabelt projections to the prefrontal cortex in the Rhesus monkey. J. Comp. Neurol. *403*, 141–157.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., and Rauschecker, J.P. (1999b). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. *2*, 1131–1136.

Saad, Z.S., Chen, G., Reynolds, R.C., Christidis, P.P., Hammett, K.R., Bellgowan, P.S., and Cox, R.W. (2006). Functional imaging analysis contest (FIAC) analysis according to AFNI and SUMA. Hum. Brain Mapp. *27*, 417–424.

Saldaña, H.M., and Rosenblum, L.D. (1993). Visual influences on auditory pluck and bow judgments. Percept. Psychophys. *54*, 406–416.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Louassmaa, O.V., Lu, S.T., and Simola, J. (1991). Seeing speech: visual information from lip movement modifies activity in the auditory cortex. Neurosci. Lett. *127*, 141–145.

Simons, J.S., Koutstaal, W., Prince, S., Wagner, A.D., and Schacter, D.L. (2003). Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. Neuroimage *19*, 613–626.

Skipper, J.I., Nusbaum, H.C., and Small, S.L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In Action to Language via the Mirror Neuron System, M.A. Arbib, ed. (Cambridge: Cambridge University Press), pp. 250–285.

Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., and Small, S.L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cereb. Cortex, in press. Published online January 19, 2007. 10.1093/cercor/bhl147.

Sugihara, T., Diltz, M.D., Averbeck, B.B., and Romanski, L.M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J. Neurosci. *26*, 11138–11147.

Taylor, J.E., and Worsley, K.J. (2006). Inference for magnitudes and delays of responses in the FIAC data using BRAINSTAT/FMRISTAT. Hum. Brain Mapp. *27*, 434–441.

Thierry, G., Ibarrola, D., Demonet, J.F., and Cardebat, D. (2003). Demand on verbal working memory delays haemodynamic response in the inferior prefrontal cortex. Hum. Brain Mapp. *19*, 37–46.

Thompson, W.F., Graham, P., and Russo, F.A. (2005). Seeing music performance: visual influences on perception and experience. Semiotica *156*, 203–227.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J.P. (2001). Functional specialization in rhesus monkey auditory cortex. Science *292*, 290–293.

Viceic, D., Fornari, E., Thiran, J.P., Maeder, P.P., Meuli, R., Adriani, M., and Clarke, S. (2006). Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. Neuroreport *17*, 1659–1662.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., and Liddle, P.F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J. Cogn. Neurosci. *13*, 994–1005.

Vuilleumier, P., Henson, R.N., Driver, J., and Dolan, R.J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat. Neurosci. *5*, 491–499.

Wagner, A.D., Koutstaal, W., Maril, A., Schacter, D.L., and Buckner, R.L. (2000). Task-specific repetition priming in left inferior prefrontal cortex. Cereb. Cortex *10*, 1176–1184.

Wessinger, C.M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J.P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J. Cogn. Neurosci. *13*, 1–7.