Published in final edited form as:
J Cogn Neurosci. 2014 August ; 26(8): 1748–1763. doi:10.1162/jocn_a_00583.

Exploring the Roles of Spectral Detail and Intonation Contour in Speech Intelligibility: An fMRI Study

Jeong S. Kyong[1], **Sophie K. Scott**[1], **Stuart Rosen**[1], **Timothy B. Howe**[1], **Zarinah K. Agnew**[1], and **Carolyn McGettigan**[1,2]

[1]University College London

[2]Royal Holloway University of London

Abstract

The melodic contour of speech forms an important perceptual aspect of tonal and nontonal languages and an important limiting factor on the intelligibility of speech heard through a cochlear implant. Previous work exploring the neural correlates of speech comprehension identified a left-dominant pathway in the temporal lobes supporting the extraction of an intelligible linguistic message, whereas the right anterior temporal lobe showed an overall preference for signals clearly conveying dynamic pitch information. The current study combined modulations of overall intelligibility (through vocoding and spectral inversion) with a manipulation of pitch contour (normal vs. falling) to investigate the processing of spoken sentences in functional MRI. Our overall findings replicate and extend those of Scott et al., whereas greater sentence intelligibility was predominately associated with increased activity in the left STS, the greatest response to normal sentence melody was found right superior temporal gyrus. These data suggest a spatial distinction between brain areas associated with intelligibility and those involved in the processing of dynamic pitch information in speech. By including a set of complexity-matched unintelligible conditions created by spectral inversion, this is additionally the first study reporting a fully factorial exploration of spectrotemporal complexity and spectral inversion as they relate to the neural processing of speech intelligibility. Perhaps surprisingly, there was no evidence for an interaction between the two factors—we discuss the implications for the processing of sound and speech in the dorsolateral temporal lobes.

INTRODUCTION

Speech intonation can be defined as “the ensemble of pitch variations in speech caused by the varying periodicity in the vibration of the vocal cords” (Hart, Collier, & Cohen, 1990). Together with other prosodic cues such as stress and timing, intonation profiles are employed voluntarily by spoken language users for a variety of linguistic and metalinguistic purposes. For example, the rise or fall in pitch at the end of sentence in British English can signal declarative or interrogative intent (compare “You travelled by train.” with “You travelled by train?”) to resolve syntactic ambiguities (e.g., “That paper was available really

Reprint requests should be sent to Carolyn McGettigan, Department of Psychology, Royal Holloway, University of London, Egham Hill, Egham TW20 0EX, UK, or via Carolyn.McGettigan@rhul.ac.uk.

Europe PMC Funders Author Manuscripts

matters."; Cutler, Dahan, & van Donselaar, 1997) or to emphasize certain information within a statement (Ladd, 1996). Intonation assists the reception of speech more generally—speech is a fast-fading auditory signal, and the use of pitch to demarcate linguistic clauses or items in a list can assist in the encoding of spoken messages into working memory (Nooteboom, 1997). Beyond the transfer of the linguistic message, intonation profiles can also signal the talker's emotional state, as well as indexical aspects of the talker such as accent variations (e.g., declaratives in Northern Irish accents typically end with a rising tone).

Intonation and Speech Intelligibility

In the laboratory, the role of intonation in speech intelligibility has been investigated via manipulations of the fundamental frequency (F0) profiles of spoken sentences. When listening to English sentences against noise or a competing talker, participants recognize fewer words in the target sentences if the F0 contour is disrupted (e.g., via flattening, inversion, or exaggeration) compared with when the natural profile is maintained (although this trend is more marked when the manipulated contour creates misleading cues rather than reducing natural ones; Miller, Schlauch, & Watson, 2010; Watson & Schlauch, 2009; Binns & Culling, 2007). In a comprehensive study including word monitoring, lexical decision, and semantic categorization tasks, Braun, Dainora, and Ernestus (2011) found that performance was slowed when the test sentences possessed an unfamiliar intonation contour (a sinewave modulation).

In certain clinical groups, the influence of intonation (or lack thereof) in sentence comprehension is more marked. A cochlear implant restores sensorineural hearing in profoundly deaf individuals by direct electrical stimulation of the auditory nerve via an array of electrodes introduced into the cochlea (Rubinstein, 2004). Typically, the reduced number of effective channels of auditory nerve stimulation means that cochlear implants convey impoverished spectrotemporal detail from sound, and so recipients of the implants exhibit difficulties with the perception of pitch (Meister, Landwehr, Pyschny, Grugel, & Walger, 2011; Gfeller et al., 2007) and speech prosody (Nakata, Trehub, & Kanda, 2012; Meister, Landwehr, Pyschny, Wagner, & Walger, 2011; Meister, Landwehr, Pyschny, Walger, & Wedel, 2009). Carroll and colleagues showed, in both normal hearing participants listening to a simulation and in patients, that F0 contour is crucial to intelligibility of speech heard through an implant (Carroll, Tiaden, & Zeng, 2011).

Sentence Intonation in the Brain

Lesion data and neuroimaging studies have indicated a difference in the way the brain processes linguistic and melodic information (McGettigan & Scott, 2012; Zatorre & Baum, 2012; Zatorre & Belin, 2001; Johnsrude, Penhune, & Zatorre, 2000; Scott, Blank, Rosen, & Wise, 2000). The evidence suggests that speech intelligibility is predominately processed in an anterior-going pathway along the left dorsolateral temporal lobe (Evans et al., 2013; McGettigan, Evans, et al., 2012; McGettigan, Faulkner, et al., 2012; Peelle, Gross, & Davis, 2012; Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Scott, Rosen, Lang, & Wise, 2006; Narain et al., 2003; Scott et al., 2000), although some authors argue for a more bilateral system (Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010; Okada et al., 2010;

Hickok & Poeppel, 2007; Poeppel, 2003). These studies have contrasted the neural responses to intelligible forms of speech, including clear and degraded versions (e.g., noise-vocoded, noise-masked), with activation to unintelligible baselines that are matched for spectrotemporal complexity (e.g., spectrally rotated/inverted speech).

The literature on neural correlates of prosodic processing is rather less clear. The issue is, in part, complicated by whether the prosody in question is affective (e.g., “happy” vs. “sad” intonation) or linguistic (e.g., emphatic stress, question vs. statement). Although the patient literature historically pointed to an important role for the right hemisphere in affective prosody perception, the results were more ambiguous for sentence melody processing; however, there was evidence for a specific preference of the right hemisphere for certain acoustic cues (e.g., F0 contours) whereas the left hemisphere was more concerned with the processing of discernible linguistic information (Baum & Pell, 1999). In later neuroimaging work, Meyer, Alter, Friederici, Lohmann, and von Cramon (2002) presented normal sentences, “syntactic” sentences (where content words were replaced with pseudowords), and filtered speech preserving only prosodic cues. They observed that, whereas left superior temporal regions responded more strongly to the sentences in fMRI, the right temporal lobe preferentially responded to sentence melody alone. Studies using tonal languages (e.g., Mandarin) developed this finding, observing an overall preference for prosodic cues in the right hemisphere for native and nonnative speakers, but a left hemisphere sensitivity to linguistically relevant tonal information in the native group only (Tong et al., 2005; Gandour et al., 2004).

Scott et al. (2000) carried out an study of speech intelligibility, aimed at dissociating between neural responses to intelligible speech from those associated with acoustic complexity. They compared the neural responses to two types of intelligible speech using PET—untransformed (“clear”) speech and noise-vocoded speech (which simulates a cochlear implant)—with activation generated by two acoustically matched control conditions. This contrast revealed activation responding to intelligible speech in the left anterior STS. However, a comparison of the “clear” speech condition and its spectrally rotated control counterpart with the noise-excited conditions (noise-vocoded and rotated noise-vocoded speech) showed activation on the right anterior superior temporal gyrus (STG). Both of these findings were recently replicated using fMRI (Evans et al., 2013). This was interpreted by Scott et al. (2000) as evidence for a distinction between a preference for linguistic processing in the left and pitch processing in the right hemisphere when dealing with spoken sentences, as both the clear and rotated speech convey a sense of pitch and pitch variation. To date, however, no neuroimaging study has directly addressed at the role of sentence intonation in interaction with the intelligibility of connected speech in English.

The purpose of the current study was to carry out a fully factorial analysis of the neural correlates of spectrotemporal complexity and melody in spoken sentences to allow us to explore interactions between these processes previously associated with left- and right-dominant responses in superior temporal cortex, respectively. The current design investigated the effect of manipulating natural pitch contours in spoken English sentences, where the intelligibility of those sentences was additionally impoverished using vocoding—this manipulation creates speech of impoverished spectrotemporal complexity by

transforming the speech spectrum into small number of amplitude-modulated carrier bands. In pilot testing, we determined that modifying sentence intonation contours such that they become uninformative (i.e., giving no indication of lexical onsets) led to a significant reduction in the intelligibility of vocoded sentences. A marginally significant interaction with a vocoding factor, which parametrically modulated the spectrotemporal complexity of sentences, revealed that the gains in intelligibility with increased complexity were smaller for the falling tone sentences than for those with normal melody. Thus, a further aim of our functional imaging study was to, for the first time, explicitly interrogate the neural systems sensitive to speech intelligibility modulations because of both complexity and melody manipulations.

We predicted that preferential responses to the presence of normal sentence melody should be found in the right STG (reflecting the right hemisphere's preference for melodic and voice information; McGettigan & Scott, 2012) and in the left STS (reflecting the greater intelligibility of the sentences with normal melody; Scott et al., 2000). We hypothesized that any interaction of melody with spectrotemporal complexity (which is also positively related to intelligibility) would be expressed bilaterally, where left hemisphere responses would indicate the linguistic and acoustic changes concomitant with these manipulations, whereas the right hemisphere response would be more acoustic in nature. Finally, we predicted that responses tracking sentence intelligibility, regardless of acoustic complexity, would be seen in the left STS—an additional set of unintelligible, complexity-matched control conditions allowed us to test this hypothesis as well as to additionally carry out the first fully factorial exploration of the interaction of spectrotemporal complexity, intonation, and phonetic intelligibility for spoken sentences.

METHODS

Materials

Stimuli—The stimuli were 288 English sentences chosen from the BKB list (e.g., "They're buying some bread"; Bench et al., 1979). The sentences were recorded by a female speaker of British English, with mean duration of 1.55 sec (*SD* = 0.196 sec).

Signal Processing—The processing steps used to create the stimuli are illustrated schematically in Figure 1. The original speech signals were passed through a bank of bandpass filters (sixth-order Butterworth), whose outputs were full-wave rectified and low-pass filtered at a cutoff frequency of 30 Hz with fourth-order Butterworth filters to extract their amplitude envelopes. The number of channels was 2, 4, or 6, and the cutoff frequencies for each channel were calculated by evenly spacing them from 100 to 5500 Hz on a scale estimating the mapping of acoustic frequency into position on the basilar membrane (Greenwood, 1990).

For each channel in the filter bank, the extracted amplitude envelope was used to modulate a synthetic source function, the result of which was then filtered by a band-pass filter matching the initial analysis filter. The RMS (root mean square) level from each output filter was then set to be equal to the RMS level of the original analysis outputs before being summed together. The first step in the vocoding process involved creating a source wave, in

which a set of pulses was generated with a fundamental frequency (F0) contour matched that of the original speech. White noise at the same RMS level as the pulses was inserted in periods of voiceless speech or silence.

Standard vocoding techniques were used to manipulate both the degree of spectral detail (by varying the numbers of channels in the analysis/synthesis procedure) and the form of the voice pitch contour (by direct manipulation of fundamental frequency tracks). For each sentence, the F0 track was determined using STRAIGHT (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999). For tonal variation, two types of acoustic excitation were generated. Nx (where N represents “normal” or “natural” and x represents the number of channels) formed conditions with tone provided, whereas Fx were noninformative tone conditions. In the Nx conditions, the fundamental frequency was shifted up by a semitone. This preserved the shape of the F0 contour but accounted for any processing artifacts introduced by manipulation of the original F0 track. A brief observation confirmed that there were no significant difference in intelligibility between condition N and the speech with natural pitch variation. The Fx conditions preserved the patterning of voiced and voiceless excitation, but here the contour was replaced by a noninformative one, in which the F0 fell by a semitone over the course of the utterance (hence the label “F” for “falling”). The starting frequency was chosen randomly for each utterance, over a range determined by the distribution of voice fundamental frequencies for the speaker (192–212 Hz). A slightly falling contour was chosen in place of a monotone both because it seemed less artificial, but also because tone languages often have a flat tone, with which a monotone could be confused. It is observed in experimental phonetics that slight down-drift is clear in most languages and that the pitch of the voice is most commonly lower at the end of a sentence than it is at the beginning—having a falling contour in sentence-final position is one of the tonal universals (Cruttenden, 1986).

For nonspeech control stimuli, spectrally inverted vocoded speech was generated using the same process as above but flipping the relationship between input and output filter banks to render the stimuli unintelligible. Stimuli with two and six channels were prepared either with natural voice pitch or with a falling contour. Note that, although these control items are both vocoded and inverted, for the purposes of the manuscript they are referred to as “Inverted” for ease of comparison with the “Vocoded” (and noninverted) items from the partially intelligible Nx and Fx conditions. Examples of the stimuli from all 10 experimental conditions can be heard here: www.carolynmcgettigan.com/#!stimuli/c7zu.

Behavioral Pilot Experiment

The conditions used for the current experiment were determined in a pilot test. Eight right-handed native speakers of British English (aged 19–50 years old) took part in a behavioral experiment, which was conducted with approval from the University College London Research Ethics Committee. All listeners had audiometric thresholds within 20 dB HL at frequencies of 0.25, 0.5, 1, 2, and 4 kHz in both ears. Stimuli were delivered binaurally through headphones (Sennheiser, HD202) using special purpose-written Matlab scripts. Sound levels were set to a comfortable fixed level, as chosen by each listener. Listeners completed a test session measuring sentence repetition accuracy for novel items presented in

22 conditions: 16 vocoded [8 vocoding levels (1, 2, 3, 4, 5, 6, 8, 16 channels) × 2 excitation patterns (N,F)] and 6 inverted [3 vocoding levels (4, 8, 16 channels) × 2 excitation patterns (N,F)]. Please note that we refer here to excitation patterns in terms of the properties of the acoustic signals, rather than to the physiological excitation of the basilar membrane.

Stimuli were presented in a random order, and the participants gave self-paced repetition of the sentence heard. The number of key words correctly repeated for each sentence was used for scoring. Figure 2 shows a plot of mean sentence repetition accuracy, expressed as the percentage of correctly identified words, across all tested conditions (with Number of Channels plotted on a logarithmic scale). The mean scores for the inverted conditions were N4 2.78% (*SD* = 2.10), N8 0.00% (*SD* = 0.00), N16 1.39% (*SD* = 0.91), F4 2.08% (*SD* = 0.91), F8 1.39% (*SD* = 0.91), F16 0.69% (*SD* = 0.69), thus satisfying our prediction that these conditions would be essentially unintelligible (supported by one-sample *t* tests comparing condition averages with zero: all $p > .17$).

Using repetition data for the vocoded conditions only, a repeated-measures ANOVA was run to explore the main effects of Excitation (N,F) and Channels (1, 2, 3, 4, 5, 6, 8, 16) and their interaction. This gave a significant effect of Excitation, $F(1, 7) = 80.37$, $p < .0001$, and Channels, $F(7, 49) = 134.07$, $p < .0001$, and a significant interaction of the two, $F(7, 49) = 3.66$, $p < .005$. This indicated, as illustrated by Figure 2, that the relationship between intelligibility and the number of Channels is different for the two Excitation types, where the increase in intelligibility with added channels is more gradual for the sentences with a noninformative pitch contour (F).

A subset of the vocoded conditions were selected for use in the functional imaging study. These were N2, N4, N6, F2, F4 and F6. A repeated-measures ANOVA carried out on the pilot data from these conditions gave significant main effects of Excitation, $F(1, 7) = 22.02$, $p < .005$, and Channels, $F(2, 14) = 146.95$, $p < .0001$, and a marginally significant interaction of the two factors, $F(2, 14) = 3.64$, $p = .053$. To these six conditions, we added inverted conditions to match the range of spectrotemporal complexities in the vocoded sets: N2_Inv, N6_Inv, F2_Inv, and F6_Inv. These inverted conditions had not been pilot tested, but as they contained fewer channels than the inverted stimuli already included in the pilot, we were satisfied that these would not be intelligible to our participants.

fMRI

Participants—Participants in the study were 19 adults (18–40) who spoke English as their first language. All were right-handed, with healthy hearing and no history of neurological incidents, nor any problems with speech or language (self-reported). The study was approved by the University College London Research Ethics Committee.

Procedure—Before entering the scanner, each participant was given a brief period of familiarization with the conditions of the experiment. Stimuli from each of the 10 experimental conditions (N2, N4, N6, F2, F4, F6, N2_Inv, N6_Inv, F2_Inv, F6_Inv) were presented to the participant using MATLAB (Mathworks, Inc., Natick, MA) with the Psychophysics Toolbox extension (Brainard, 1997). The participant wore headphones (Sennheiser HD-201) and listened to a total of 30 BKB sentences (three examples from each

of conditions). Each stimulus was played first in its vocoded (or inverted) form and then repeated in an untransformed version, which provided clear feedback of the sentence content. The participant was not asked to perform a task, and no data were recorded from this session. None of the BKB sentences used in the training was repeated in the main fMRI run.

Functional imaging data were acquired on a Siemens Avanto 1.5-T MRI scanner (Siemens AG, Erlangen, Germany) with a 12-channel birdcage head coil. Auditory presentation of sentences took place in two runs of 146 echo-planar whole-brain volumes (repetition time = 8 sec, acquisition time = 3 sec, echo time = 50 msec, flip angle = 90°, 35 axial slices, 3 mm × 3 mm × 3 mm in-plane resolution, matrix size = 64 × 64). A sparse-sampling routine (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999) was employed, in which the auditory stimuli were presented in the quiet period between scans. Auditory onsets occurred 4.3 sec (±0.5 sec jitter) before the beginning of the next whole-brain volume acquisition. The condition order was pseudorandomized into fully randomized miniblocks of 48 trials, within which each of the vocoded conditions was presented six times, and each of the inverted conditions was presented three times—across the two runs, this came to a total of 288 trials, with 36 from each of the vocoded conditions and 18 from each of the inverted conditions. There was no active task, but participants were asked to listen carefully to the stimuli and to try and understand what was being said. Stimuli were presented using MATLAB with the Cogent toolbox extension (www.vislab.ucl.ac.uk) and routed through a Denon amplifier (Denon, Belfast, United Kingdom) to electrodynamic headphones worn by the participant (MR Confon GmbH, Magdeburg, Germany). After the functional run, a high-resolution T1-weighted anatomical image was acquired (HIRes MP-RAGE, 160 sagittal slices, voxel size = 1 mm^3). The total time in the scanner was around 50 min.

Analysis of fMRI Data—Data were preprocessed and analyzed in SPM8 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom). Functional images were realigned and unwarped, co-registered with the anatomical image, normalized using parameters obtained from unified segmentation of the anatomical image, rewritten with voxel dimensions 2 × 2 × 2 mm, and smoothed using a Gaussian kernel of 8 mm FWHM.

At the single-subject level, event onsets from all 10 conditions (N2, N4, N6, F2, F4, F6, N2_Inv, N6_Inv, F2_Inv, F6_Inv) were modeled as instantaneous (duration = 0) and coincident with the onset of the sound files and further convolved with the canonical hemodynamic response function in SPM8 along with six movement parameters of no interest. We note that as a very small number of images contributed to the model's implicit baseline (4 volumes in total), caution should be adopted when interpreting the scale on plots of parameter estimates. Contrast images were calculated in the single subject and used to carry out a set of planned second-level, random effects analyses:

A: 3 × 2 Factorial ANOVA—Excitation (N,F) × Channels (2,4,6): Contrast images for each condition (compared with the implicit baseline) were calculated at the first level and taken up to a second-level 3 × 2 flexible factorial, random effects ANOVA model including the factors Subject, Excitation, and Channels. Contrasts were estimated to describe the Main Effect of Channels ([kron ([1 1], orth(diff(eye(3))′)′]), Main Effect of Excitation (F contrast

of [−1 −1 1 1]), Interaction of Channels and Excitation ([kron([1 −1], orth(diff(eye(3))′)′]), Positive Effect of Channels ([−1 0.1 0.9 −1 0.1 0.9]), Negative Effect of Channels ([1 −0.1 −0.9 1 −0.1 −0.9]), Normal > Flattened melody ([−1 −1 1 1]), and Flattened > Normal melody ([1 1 −1 −1]).

B: 2 × 2 × 2 Factorial ANOVA— Excitation (N,F) × Channels (2,6) × Inversion (Vocoded, Inverted): This model used contrast images of Vocoded > Inverted (N2 > N2_Inv; N6 > N6_Inv; F2 > F2_Inv; F6 > F6_Inv) at the first level in a second-level full factorial ANOVA to explore the main effect of Inversion and its interactions with Excitation and Channels, in two-way and three-way interactions. Here, the effects of subject were not modeled explicitly as these were already accounted for in the difference contrasts between vocoded and inverted conditions at the single-subject level. We estimated the Main Effect of Inversion (F contrast of [1 1 1 1]), interaction of Inversion and Channels (F contrast of [−1 1 −1 1]), interaction of Inversion and Excitation (F contrast of [−1 −1 1 1]), the interaction of all three factors (F contrast of [−1 1 1 −1]), as well as the direct comparisons of Vocoded > Inverted ([1 1 1 1]) and Inverted > Vocoded ([−1 −1 −1 −1]).

C: 2 × 2 Factorial ANOVA—Excitation (N,F) × Channels (2,6) (Vocoded Only): Contrast images for the vocoded conditions N2, N6, F2, and F6 (compared with the implicit baseline) were calculated at the first level and taken up to a second-level 2 × 2 flexible factorial model including the factors Subject, Excitation, and Channels. Contrasts were constructed to estimate the Main Effect of Channels (F test: [−1 1 −1 1]), Main Effect of Excitation (F test: [−1 −1 1 1]), Interaction of Channels and Excitation F test: [−1 −1 1 1]), Positive Effect of Channels ([−1 1 −1 1]), Negative Effect of Channels ([1 −1 1 −1]), Normal > Flattened melody ([−1 −1 1 1]), Flattened > Normal melody ([1 1 −1 −1]).

D: 2 × 2 Factorial ANOVA—Excitation (N,F) × Channels (2,6) (Inverted Only): Contrast images for the inverted conditions N2_Inv, N6_Inv, F2_Inv and F6_Inv (compared with the implicit baseline) were calculated at the first level and taken up to a second-level 2 × 2 flexible factorial model including the factors Subject, Excitation, and Channels. Contrasts were constructed to estimate the Main Effect of Channels (F test: [−1 1 −1 1]), Main Effect of Excitation (F test: [−1 −1 1 1]), Interaction of Channels and Excitation F test: [−1 −1 1 1]), Positive Effect of Channels ([−1 1 −1 1]), Negative Effect of Channels ([1 −1 1 −1]), Normal > Flattened melody ([−1 −1 1 1]), Flattened > Normal melody ([1 1 −1 −1]).

E: One-way ANOVA—Effect of Intelligibility: Contrast images for all of the vocoded and inverted conditions were calculated at the first level and entered into a within-subject one-way ANOVA modeling the effects of Subject and Intelligibility. Using the mean intelligibility of each condition as measured in the behavioral pilot, zero-centered weighted contrasts described the main effect (F contrast: [−0.2 0.5 0.6 −0.3 0.2 0.5 −0.33 −0.33 −0.33 −0.31]) as well as the positive and negative correlates of sentence intelligibility (±*t* contrast: [−0.2 0.5 0.6 −0.3 0.2 0.5 −0.33 −0.33 −0.33 −0.31]).

All second-level models are reported at a voxelwise threshold of $p < .001$ (uncorrected). A cluster extent correction of 68 resampled voxels (2 × 2 × 2 mm) was applied for a whole-brain alpha of $p < .001$ using a Monte Carlo simulation (with 10,000 iterations) implemented

in MATLAB (Slotnick, Moo, Segal, & Hart, 2003). We took a conservative approach: as the estimated smoothness of the second-level model is often nonisometric in the three dimensions, and slightly variable across models, we implemented the single largest smoothness estimate in any dimension and across all models and used this to estimate the cluster extent threshold for correction. In the case of the current data set, the maximum FWHM value, which we entered into the simulation, was 13.0 (mm).

Second-level peak locations were used to extract condition-specific parameter estimates from 4-mm spherical ROIs built around the peak voxel (ROIs; using MarsBaR; Brett et al., 2002). The anatomical locations of peak and subpeak voxels (at least 8 mm apart) were labeled using the SPM Anatomy Toolbox (version 18) (Eickhoff et al., 2005).

Calculating Laterality Indices—To test the lateralization of the effects of Channels (2 < 4 < 6), Excitation (N > F), and increasing/decreasing Intelligibility, we used the LI toolbox in SPM8 (Wilke & Schmithorst, 2006). For each contrast of interest, the toolbox calculates laterality indices (LI) using the equation: $LI = (\Sigma activation_{left} - \Sigma activation_{right} / \Sigma activation_{left} - \Sigma activation_{right})$, where Σ refers to the sum of activation either in terms of the total voxel count or the sum of the voxel values within the statistical map of the contrast. Thus, values of LI can vary from +1 (*completely left lateralized*) to −1 (*completely left lateralized*). According to convention, an absolute LI value greater than 0.2 is taken to indicate a hemispheric dominance (Seghier, 2008). In this paper, “activation” in the LI formula was defined as the total voxel values within each hemisphere in the second-level *t* maps for our contrasts of interest, restricted in our case to the left and right temporal lobes (defined using an inclusive anatomical mask of the temporal lobes, which comes as part of the toolbox). To take account of thresholding effects, the toolbox calculates LIs at 20 thresholding intervals from 0 to the maximum value in the *t* map. At each level, the toolbox selects 100 bootstrap samples (5–1000 voxels) from each masked hemisphere, which are paired in all possible combinations (10,000) and used to calculate an equivalent number of LIs. From the final distribution of LIs, the toolbox reports trimmed means (where the top and bottom 25% of values have been discarded), as well as a single weighted mean based on these that is proportionally more affected by LI values from higher statistical thresholds.

RESULTS

Modulations of Spectrotemporal Complexity and Sentence Melody Engage Different Profiles of Superior Temporal Activation

The 3 × 2 ANOVA revealed significant main effects of Channels and Excitation, which reflected increased signal for greater numbers of channels and for the N excitation (compared with the falling-prosody F sentences). Activation showing a positive effect of Channels was found in two large clusters in the STG and STS of both hemispheres. In contrast, the positive effect N > F gave a marked asymmetry, with a large cluster in right STG/STS only. There were no voxels showing a significant interaction of the two factors nor any showing preferential responses to decreasing spectrotemporal complexity (6 < 4 < 2) or the falling/flattened sentences (F > N). The results of significant *t* contrasts, with plots of parameter estimates from ROIs built around the peak voxels, are shown in Figure 3. Table

1 lists the location and anatomical labels for the significant clusters, as well as F/T and *z* statistics for the main and local peak activations.

Superior Temporal Regions Are Sensitive to Spectral Inversion

This model gave a main effect of Inversion, which contained both positive and negative effects of this factor. Regions showing a greater signal for inverted conditions were confined to bilateral Heschl's gyrus and the planum temporale, whereas a cluster tracking posterior to anterior STS in the left hemisphere showed greater responses to the vocoded (noninverted) conditions (see Figure 4 and Table 2). There were no significant interactions of Inversion with either of the factors Excitation or Channels.

Modulations of Spectrotemporal Complexity and Sentence Melody Generated Differing Profiles of Activation for Vocoded and Inverted Sentences

A 2 × 2 model with factors Channels and Excitation for the vocoded conditions only revealed increased signal for sentences with greater numbers of channels, in bilateral STG/STS, and a preference for the normal prosody of the N-type conditions in right STG/STS (see Figure 5 and Table 3). A 2 × 2 model with factors Channels and Excitation including only the spectrally inverted conditions revealed effects of Excitation only. Bilateral regions of the STG gave greater responses for the N-type sentences bearing typical prosody, compared with the F-type sentences with uninformative pitch contours (see Figure 5 and Table 4).

Sentence Intelligibility Is Associated with Activation in Left STS

This model revealed a significant effect of Intelligibility, which broke down into positive effects (i.e., where signal positively correlated with the mean intelligibility of the conditions) in left STS and right STG/STS and negative effects in bilateral planum temporale (including Heschl's gyrus an extending into inferior parietal cortex [rolandic operculum] on the right; see Figure 6 and Table 5).

Laterality Indices Indicate Left-dominant Effects of Channels and Intelligibility and a Right-dominant Effect of Excitation

The results of the bootstrap lateralization analyses are summarized in Table 6. Using a conventional LI threshold of |0.2|, these indicated strong left-dominant effects of increasing Intelligibility (weighted mean: 0.71) and of the contrast of Vocoded > Inverted sentences (weighted mean: 0.84). The positive effects of Channels gave a weaker left lateralization (3 × 2 model weighted mean: 0.23, 2 × 2 model on vocoded conditions: 0.16). The contrast of N > F sentences gave strongly right-lateralized effects for the vocoded conditions (3 × 2 model weighted mean: −0.57, 2 × 2 model weighted mean: −0.57), but not for the spectrally inverted sentences (weighted mean: 0.0052).

DISCUSSION

Hemispheric Asymmetries in the Processing of Intonation and Speech Intelligibility

Sentences with a natural and informative intonation contour gave greater signal in right STG/STS compared with the flattened/falling sentences—a bootstrap analysis of laterality indices indicated that this was a right-dominant effect for the vocoded conditions. Figure 2 shows the main cluster overlaps with the regions responding to increased spectrotemporal complexity (i.e., Channels), where the overall effect was left-dominant in the temporal lobes. Interestingly, when only inverted sentences were included, there was a bilateral response to natural intonation in planum temporale (for Normal > Flattened, weighted mean LI: 0.0052). This may reflect that fact that the spectral inversion yields a slightly weakened pitch percept.

There were no significant clusters showing an interaction of the two factors. However, an analysis of the correlates of sentence intelligibility by condition gave some greater insight into the interacting effects of Channels and Excitation. This showed (as in the direct comparions of Vocoded > Inverted conditions) that responses to increasing intelligibility occurred in strongly left-dominant regions of the STS. Although plots of parameter estimates by condition showed a similar profile within the vocoded conditions in the left and right superior temporal cortex, the right hemisphere (and some left hemisphere) regions were also sensitive to differences in spectrotemporal complexity and intonation profile of the unintelligible, spectrally inverted conditions (see Figure 5A). This stands in contrast to the left STS peak at [−54 −42 4], where the plot of parameter estimates shows no such trend among the inverted conditions.

Our findings support those of Scott et al. (2000) and later work (Evans et al., 2013; McGettigan, Evans, et al., 2012; McGettigan, Faulkner, et al., 2012; Eisner et al., 2010; Davis & Johnsrude, 2003; Narain et al., 2003) in showing a strongly left-dominant response to speech intelligibility, whereas sensitivity to acoustic manipulations of spectrotemporal complexity is more bilaterally expressed (weight LIs of 0.71 and 0.23, respectively). Furthermore, we find that the neural response to the presence of an informative, natural intonation contour, regardless of other factors, is strongly right-dominant. The right temporal lobe has long been associated with the processing of prosodic information, as well as in the processing of vocal identity (McGettigan et al., 2013; von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005; Kriegstein & Giraud, 2004; Belin & Zatorre, 2003) —the current data may reflect some processing of the “voicelikeness” of the heard stimuli, where sentences bearing natural prosody and greater spectrotemporal complexity (regardless of intelligibility) more strongly resemble natural conspecific vocalizations (Rosen, Wise, Chadha, Conway, & Scott, 2011). This is in line with research describing right-dominant responses in superior temporal cortex to the processing of voices compared with other nonvocal sounds and the recognition of vocal identity (Bestelmeyer, Belin, & Grosbras, 2011; Kriegstein & Giraud, 2004; Belin & Zatorre, 2003; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; Belin, Zatorre, & Ahad, 2002).

The plots of condition-wise parameter estimates from intelligibility-sensitive left STS peaks in the current experiment indicate some trend level response to the manipulation of pitch

contour. In line with the behavioral pilot, which showed a significant deleterious effect of pitch flattening on sentence intelligibility, the activation profile in left STS was weaker for the pitch-flattened sentences compared with those with natural melody. A study of the interaction of prosodic and syntactic information in speech comprehension found that this process was impaired in patients with posterior, and not anterior, lesions to the corpus callosum (Sammler et al., 2010). This suggested that the posterior corpus callosum is essential for informational crosstalk between right hemisphere temporal regions predominately processing prosodic information and left temporal regions processing syntactic cues. A challenge for future work will be to interrogate the potential mechanisms for cross-talk in intact brains during the processing of cues to sentence intelligibility used in the current experiment.

Spectrotemporal Complexity versus Intelligibility: A Fully Factorial Investigation

The current study offered the first fully factorial statistical exploration of spectral inversion with intelligibility-related acoustic manipulations. This gave a clear indication of a preferential left-hemisphere response tracing the anterior-to-posterior extent of the left STS to speech that is at least partially intelligible, while fully controlling, condition by condition, for basic acoustic properties (spectrotemporal complexity and intonation contour). Importantly, this result cannot be ascribed to some metalinguistic effect such as greater attentional capture by the vocoded stimuli, as we found very strong effects showing the converse profile of a greater response to inverted sentences in the planum temporale bilaterally. Warren, Wise, and Warren (2005) suggested a role for the posteromedial planum temporale in the matching of incoming sounds to stored auditory templates for use in auditory–motor computations, for example, for the repetition of heard speech. The current result may indicate the engagement of planum temporal as part of the dorsal “how” pathway, passing through inferior parietal cortex and forward to frontal speech production regions (Scott & Johnsrude, 2003). Here, this pathway may be engaged in attempts to detect and transform “do-able” sensory information (Warren et al., 2005) from the unintelligible, inverted sounds. In contrast, computations in the anterior “what” pathway have been unable to map between the inverted acoustic input and learned linguistic representations. Note that this view of the “how” pathway is distinct from the discussion of posterior parts of the STG by Belin and Zatorre (2000).

Interestingly, there was no statistically significant interaction of the spectral inversion factor with the spectrotemporal complexity or intonation factors. We expected that there would be enhanced effects of the acoustic factors affecting intelligibility within the vocoded conditions, compared with the inverted conditions in which these factors should have only elicited responses related to acoustic processing. This fully factorial design raises new questions about the nature of hierarchical processing in the dorsolateral temporal lobes, particular in the left hemisphere, as it suggests that the difference between unintelligible and intelligible speech sounds is expressed as a step-level enhancement in response rather than a complex interaction of acoustic and phonetic/linguistic properties of the stimulus. This could indicate that, although acoustic properties of the stimulus modulate neural activity quite far along the hierarchical pathway (including STS, as indicated in the plots in Figure 2), there is an additional level of response that is engaged when the stimulus allows the extraction of

Europe PMC Funders Author Manuscripts

meaningful percepts (i.e., when it is not inverted; Figures 3 and 5). Whether this "stepwise" change occurs at the level of phonetic, phonemic, syllabic, or higher-order linguistic representations should form the subject of future investigations. In a recent MEG study, Sohoglu, Peelle, Carlyon, and Davis (2012) found that when a target degraded word matched that of a previous written prime, enhanced responses to that word in left inferior frontal cortex preceded signal changes in left superior temporal cortex. This indicates a top-down role for the inferior frontal cortex in tuning auditory representations to degraded speech based on higher-order expectations. In the current experiment, we see an overall greater signal along the STS for stimuli that provide some level of intelligible percept for listeners, compared with spectrally inverted, unintelligible stimuli, but no difference in how the vocoded and inverted conditions interacted with acoustic factors modulating intelligibility. Future work must address the network basis for this result—does the presence of some level of linguistic information engage "top–down" processes that lead to additional processing of potentially intelligible stimuli, above and beyond the computations of basic acoustic structure?

Future Directions: Tonal Languages

As described earlier, intonation plays a number of important roles in the comprehension of spoken English, which become even more pronounced when production or perception mechanisms are compromised by noise or distortion. This is even more marked for tonal languages such as Mandarin Chinese and Yoruba, where pitch has lexical or morphological relevance, and so the removal or disruption of pitch cues introduces linguistic ambiguities for the listener. We therefore propose to carry out a further investigation using modulations of spectrotemporal detail and intonation contour in a group of participants who are native speakers of a tonal language, such as Mandarin Chinese. We expect that the behavioral effects of disrupting the natural intonation contour would be more marked for these participants than the effects observed in the current pilot. Furthermore, as fundamental frequency plays an intrinsic role in the representation of linguistic units in tonal languages, we also predict that the observed effects of intonation may be less right-dominant for speakers of a tonal language, and indeed there may be stronger evidence for an interaction of the spectrotemporal and intonation factors in these participants.

Conclusion

We present the first neuroimaging study of the interactions of spectrotemporal complexity and melody in the processing of spoken English sentences. The results add further support to the now well-established view of a hierarchical pathway for the processing of intelligible speech in the left superior temporal cortex, with higher-order cortical fields in STS exhibiting some sensitivity to the role played by pitch information in sentence comprehension. In contrast, the right superior temporal cortex is driven by the basic acoustic properties of speech and shows a strong preference for sounds with natural dynamic pitch. Natural pitch variation occurs over a number of different levels, from sublexical to phrasal —future work will explore the neural correlates of intelligibility and sentence melody processing in tonal language speakers, where lexical pitch is central to the extracting the linguistic message.

Europe PMC Funders Author Manuscripts

REFERENCES

Baum SR, Pell MD. The neural bases of prosody: Insights from lesion studies and neuroimaging. Aphasiology. 1999; 13:581–608.

Belin P, Zatorre RJ. “What,” “where” and “how” in auditory cortex. Nature Neuroscience. 2000; 3:965–966.

Belin P, Zatorre RJ. Adaptation to speaker’s voice in right anterior temporal lobe. NeuroReport. 2003; 14:2105–2109. [PubMed: 14600506]

Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cognitive Brain Research. 2002; 13:17–26. [PubMed: 11867247]

Bestelmeyer PEG, Belin P, Grosbras M-H. Right temporal TMS impairs voice detection. Current Biology. 2011; 21:R838–R839. [PubMed: 22032183]

Binns C, Culling JF. The role of fundamental frequency contours in the perception of speech against interfering speech. Journal of the Acoustical Society of America. 2007; 122:1765–1776. [PubMed: 17927436]

Bozic M, Tyler LK, Ives DT, Randall B, Marslen-Wilson WD. Bihemispheric foundations for human speech comprehension. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:17439–17444. [PubMed: 20855587]

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Braun B, Dainora A, Ernestus M. An unfamiliar intonation contour slows down on-line speech comprehension. Language and Cognitive Processes. 2011; 26:350–375.

Carroll J, Tiaden S, Zeng FG. Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. Journal of the Acoustical Society of America. 2011; 130:2054–2062. [PubMed: 21973360]

Cruttenden, A. Intonation. Cambridge University Press; Cambridge, UK: 1986.

Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: A literature review. Language and Speech. 1997; 40:141–201. [PubMed: 9509577]

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. Journal of Neuroscience. 2003; 23:3423–3431. [PubMed: 12716950]

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. Human Brain Mapping. 1999; 7:89–97. [PubMed: 9950066]

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 2005; 25:1325–1335. [PubMed: 15850749]

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. Journal of Neuroscience. 2010; 30:7179–7186. [PubMed: 20505085]

Evans S, Kyong J, Rosen S, Golestani N, Warren JE, McGettigan C, et al. The pathways for intelligible speech: Multivariate and univariate perspectives. Cerebral Cortex. 2013 doi: dx.doi.org/doi:10.1093/cercor/bht083.

Gandour J, Tong Y, Wong D, Talavage T, Dzemidzic M, Xu Y, et al. Hemispheric roles in the perception of speech prosody. Neuroimage. 2004; 23:344–357. [PubMed: 15325382]

Gfeller K, Turner C, Oleson J, Zhang X, Gantz B, Froman R, et al. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. Ear and Hearing. 2007; 28:412–423. [PubMed: 17485990]

Greenwood DD. A cochlear frequency-position function for several species-29 years later. Journal of the Acoustical Society of America. 1990; 87:2592–2605. [PubMed: 2373794]

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, et al. “Sparse” temporal sampling in auditory fMRI. Human Brain Mapping. 1999; 7:213–223. [PubMed: 10194620]

Hart, JT.; Collier, R.; Cohen, A. A perceptual study of intonation. An experimental-phonetic approach to intonation. Cambridge University Press; Cambridge, UK: 1990.

Hickok G, Poeppel D. Opinion—The cortical organization of speech processing. Nature Reviews Neuroscience. 2007; 8:393–402.

Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain. 2000; 123:155–163. [PubMed: 10611129]

Kawahara H, Masuda-Katsuse I, de Cheveigné A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. Speech Communication. 1999; 27:187–207.

Kriegstein KV, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage. 2004; 22:948–955. [PubMed: 15193626]

Ladd, RD. Intonational phonology. Vol. 79. Cambridge University Press; Cambridge, UK: 1996.

McGettigan C, Eisner F, Agnew ZK, Manly T, Wisbey D, Scott SK. T'ain't what you say, it's the way that you say it—Left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. Journal of Cognitive Neuroscience. 2013 doi:10.1162/jocn_a_00427.

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK. An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. Journal of Cognitive Neuroscience. 2012; 24:636–652. [PubMed: 22066589]

McGettigan C, Faulkner A, Altarelli I, Obleser J, Baverstock H, Scott SK. Speech comprehension aided by multiple modalities: Behavioural and neural interactions. Neuropsychologia. 2012; 50:762–776. [PubMed: 22266262]

McGettigan C, Scott SK. Cortical asymmetries in speech perception: What's wrong, what's right and what's left? Trends in Cognitive Sciences. 2012; 16:269–276. [PubMed: 22521208]

Meister H, Landwehr M, Pyschny V, Grugel L, Walger M. Use of intonation contours for speech recognition in noise by cochlear implant recipients. Journal of the Acoustical Society of America. 2011; 129:E204–E209.

Meister H, Landwehr M, Pyschny V, Wagner P, Walger M. The perception of sentence stress in cochlear implant recipients. Ear and Hearing. 2011; 32:459–467. [PubMed: 21187749]

Meister H, Landwehr M, Pyschny V, Walger M, Wedel HV. The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. International Journal of Audiology. 2009; 48:38–48. [PubMed: 19173112]

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY. fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Human Brain Mapping. 2002; 17:73–88. [PubMed: 12353242]

Miller SE, Schlauch RS, Watson PJ. The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. Journal of the Acoustical Society of America. 2010; 128:435–443. [PubMed: 20649237]

Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. Journal of the Acoustical Society of America. 2012; 131:1307–1314. [PubMed: 22352504]

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, et al. Defining a left-lateralized response specific to intelligible speech using fMRI. Cerebral Cortex. 2003; 13:1362–1368. [PubMed: 14615301]

Nooteboom S. The prosody of speech: Melody and rhythm. The Handbook of Phonetic Sciences. 1997; 5:640–673.

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, et al. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex. 2010; 20:2486–2495. [PubMed: 20100898]

Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral Cortex. 2012 doi:10.1093/cercor/bhs118.

Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time.". Speech Communication. 2003; 41:245–255.

Rosen S, Wise RJS, Chadha S, Conway E-J, Scott SK. Hemispheric asymmetries in speech perception: Sense, nonsense and modulations. PloS One. 2011; 6 doi:10.1371/journal.pone.0024672.

Rubinstein JT. How cochlear implants encode speech. Current Opinion in Otolaryngology & Head and Neck Surgery. 2004; 12:444–448. [PubMed: 15377959]

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

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends in Neurosciences. 2003; 26:100–107. [PubMed: 12536133]

Scott SK, Rosen S, Lang H, Wise RJS. Neural correlates of intelligibility in speech investigated with noise vocoded speech-A positron emission tomography study. Journal of the Acoustical Society of America. 2006; 120:1075–1083. [PubMed: 16938993]

Seghier ML. Laterality index in functional MRI: Methodological issues [Review]. Magnetic Resonance Imaging. 2008; 26:594–601. [PubMed: 18158224]

Slotnick SD, Moo LR, Segal JB, Hart J. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cognitive Brain Research. 2003; 17:75–82. [PubMed: 12763194]

Sohoglu E, Peelle JE, Carlyon RP, Davis MH. Predictive top-down integration of prior knowledge during speech perception. Journal of Neuroscience. 2012; 32:8443–8453. [PubMed: 22723684]

Tong Y, Gandour J, Talavage T, Wong D, Dzemidzic M, Xu Y, et al. Neural circuitry underlying sentence-level linguistic prosody. Neuroimage. 2005; 28:417–428. [PubMed: 16006150]

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. Cognitive Brain Research. 2003; 17:48–55. [PubMed: 12763191]

von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. Journal of Cognitive Neuroscience. 2005; 17:367–376. [PubMed: 15813998]

Warren JE, Wise RJ, Warren JD. Sounds do-able: Auditory-motor transformations and the posterior temporal plane. Trends in Neurosciences. 2005; 28:636–643. [PubMed: 16216346]

Watson PJ, Schlauch RS. Fundamental frequency variation with an electrolarynx improves speech understanding: A case study. American Journal of Speech-Language Pathology. 2009; 18:162–167. [PubMed: 19106204]

Zatorre RJ, Baum SR. Musical melody and speech intonation: Singing a different tune? Plos Biology. 2012; 10 doi:10.1371/journal.pbio.1001372.

Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. Cerebral Cortex. 2001; 11:946–953. [PubMed: 11549617]

Europe PMC Funders Author Manuscripts

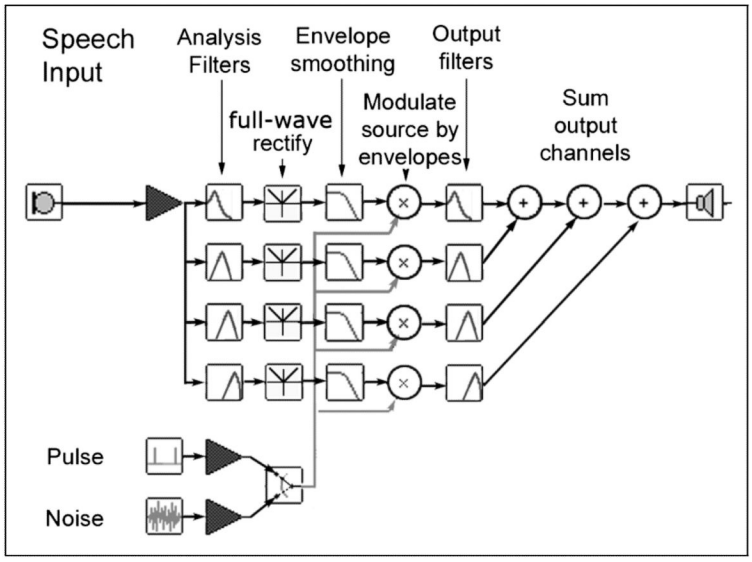

Figure 1. Schematic of the signal processing steps used to create stimuli for the current experiment.

See Methods for details.

Europe PMC Funders Author Manuscripts

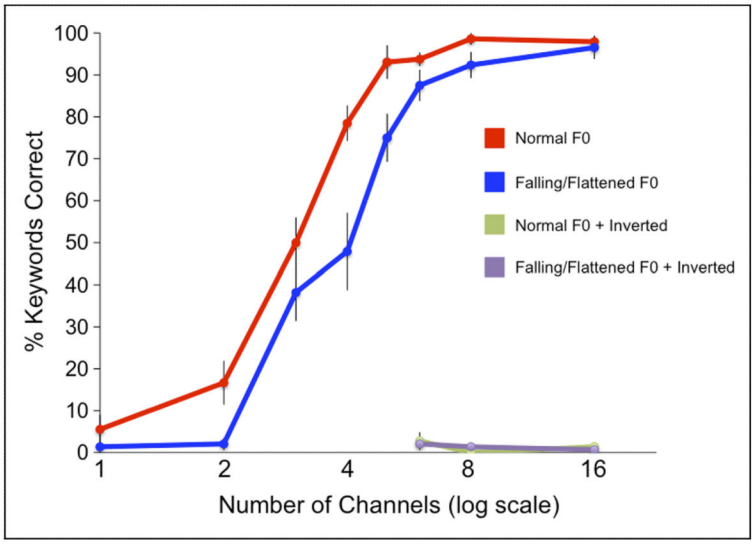

Figure 2.

Results of a behavioral pilot experiment showing the effects of F0 Excitation, Channels, and Inversion on the intelligibility of vocoded sentences. Error bars show 1 *SEM*.

Europe PMC Funders Author Manuscripts

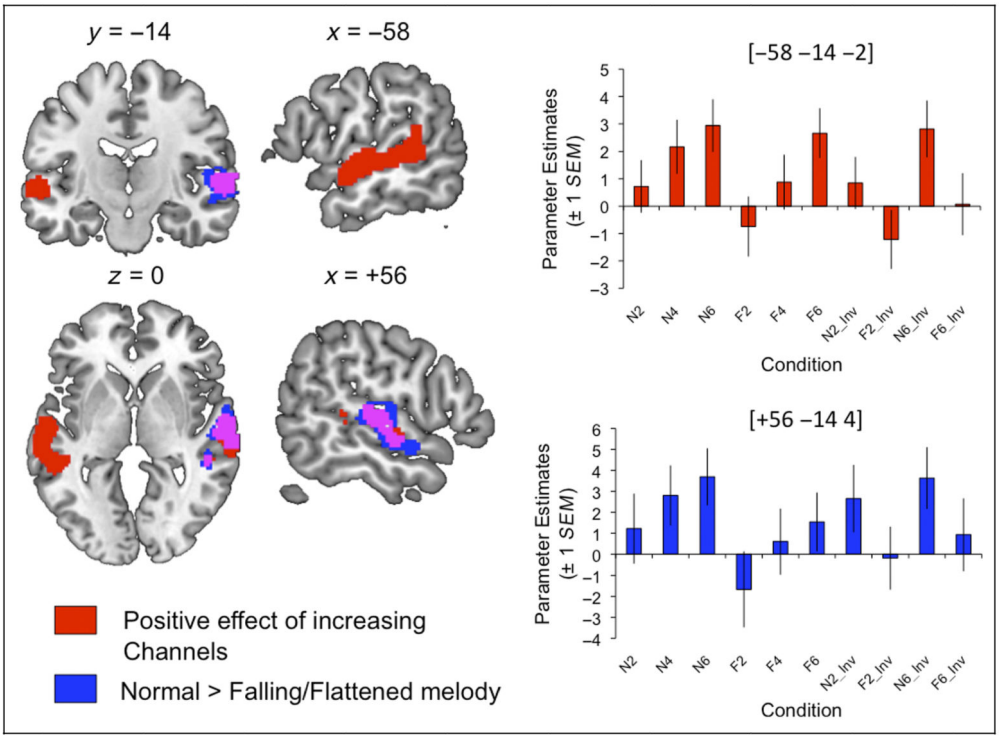

Figure 3. Regions showing increased activation for stimuli with greater spectrotemporal detail (Channels; shaded red) and natural F0 contour (compared with uninformative falling/flattened contours; shaded blue).

Overlap of the two effects is shown in violet. Plots show parameter estimates extracted from spherical ROIs built around the peak voxels (+1 *SEM*). Images are shown at a voxelwise threshold of $p < .001$ (uncorrected) and a cluster extent threshold of $p < .001$ (corrected; Slotnick et al., 2003). Numbers above slices indicate coordinates in Montreal Neurological Institute (MNI) stereotactic space.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

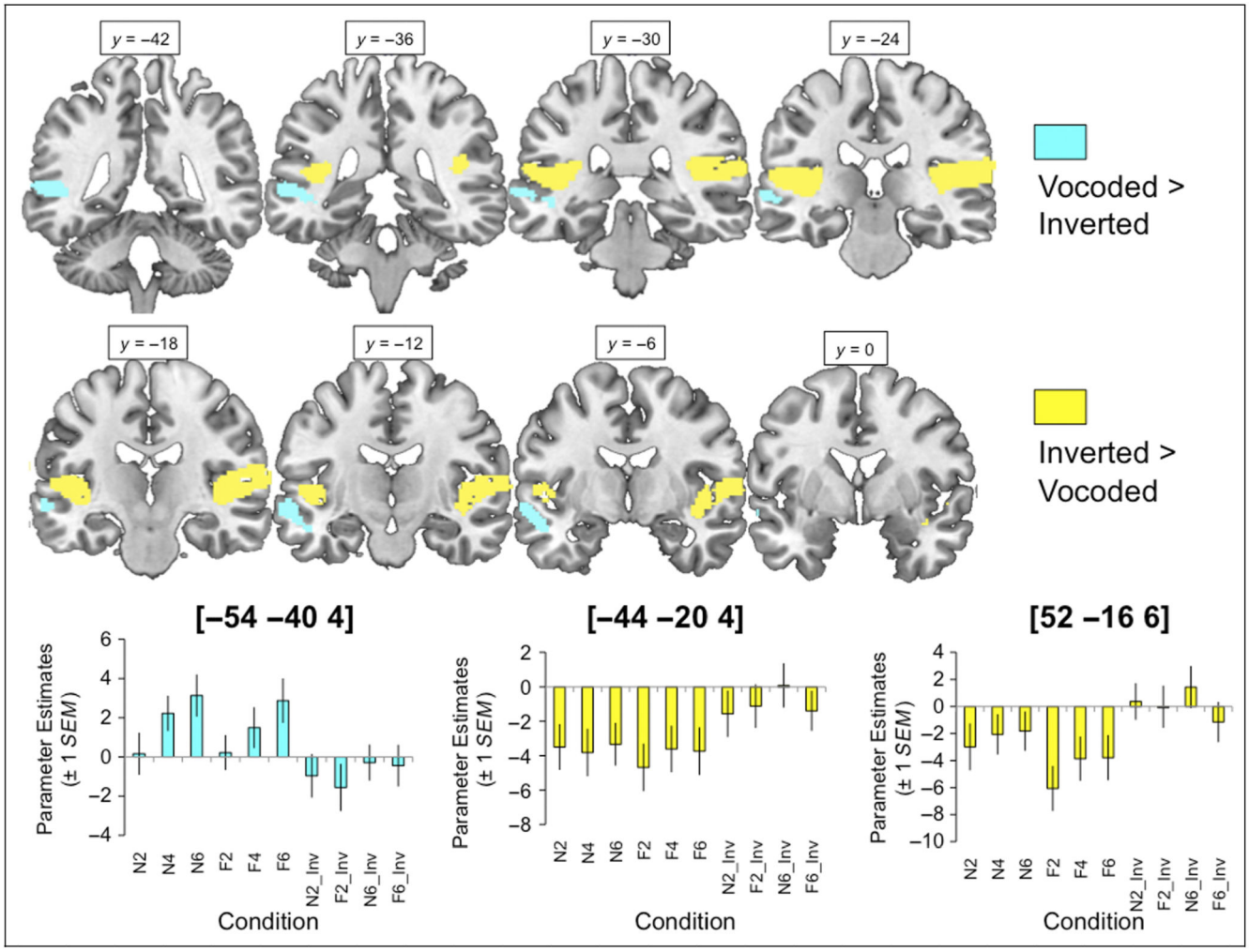

Figure 4. Comparison of vocoded and inverted conditions, where regions showing greater signal for vocoded (and partially intelligible) items are shaded with cyan, and those showing the opposite preference are shaded with yellow.

Plots show parameter estimates extracted from spherical ROIs built around the peak voxels (+1 *SEM*). Images are shown at a voxelwise threshold of $p < .001$ (uncorrected) and a cluster extent threshold of $p < .001$ (corrected; Slotnick et al., 2003). Numbers above slices indicate coordinates in MNI stereotactic space.

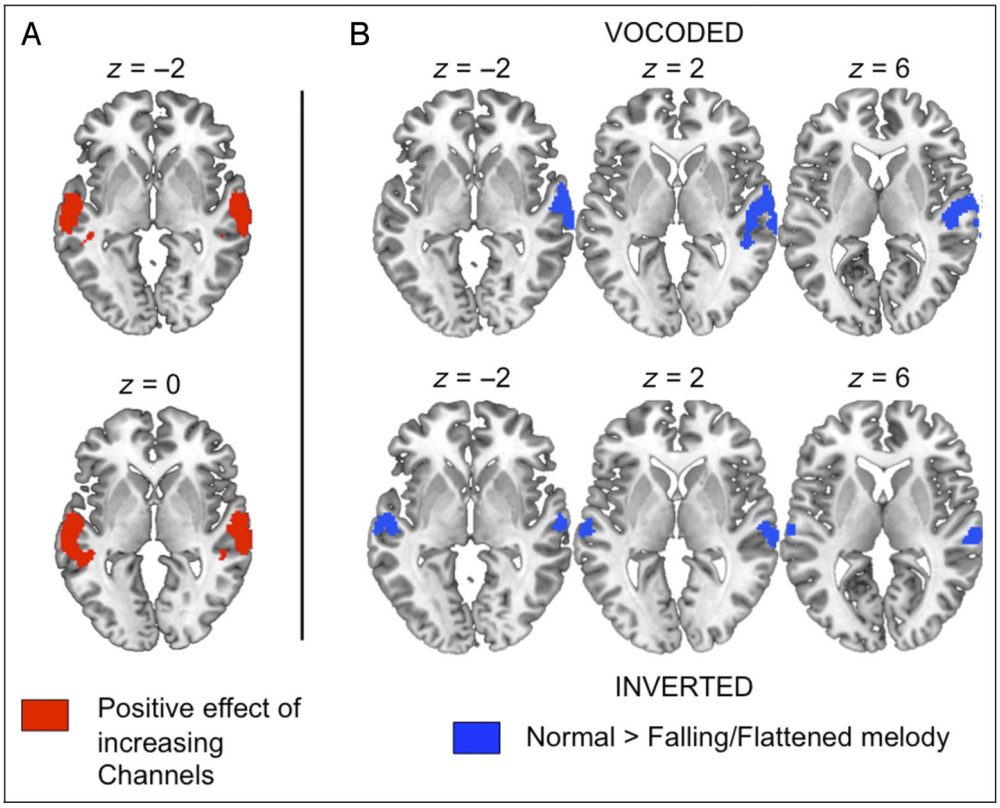

Figure 5. Results of factorial analyses of the effects of Channels and F0 Excitation within vocoded and inverted conditions separately.

Images show (A) the positive effect of Channels for vocoded sentences and (B) the positive effect of F0 Excitation (normal > falling) for vocoded and inverted sentences. Images are shown at a voxelwise threshold of $p < .001$ (uncorrected), and a cluster extent threshold of $p < .001$ (corrected; Slotnick et al., 2003). Numbers above slices indicate coordinates in MNI stereotactic space.

Europe PMC Funders Author Manuscripts

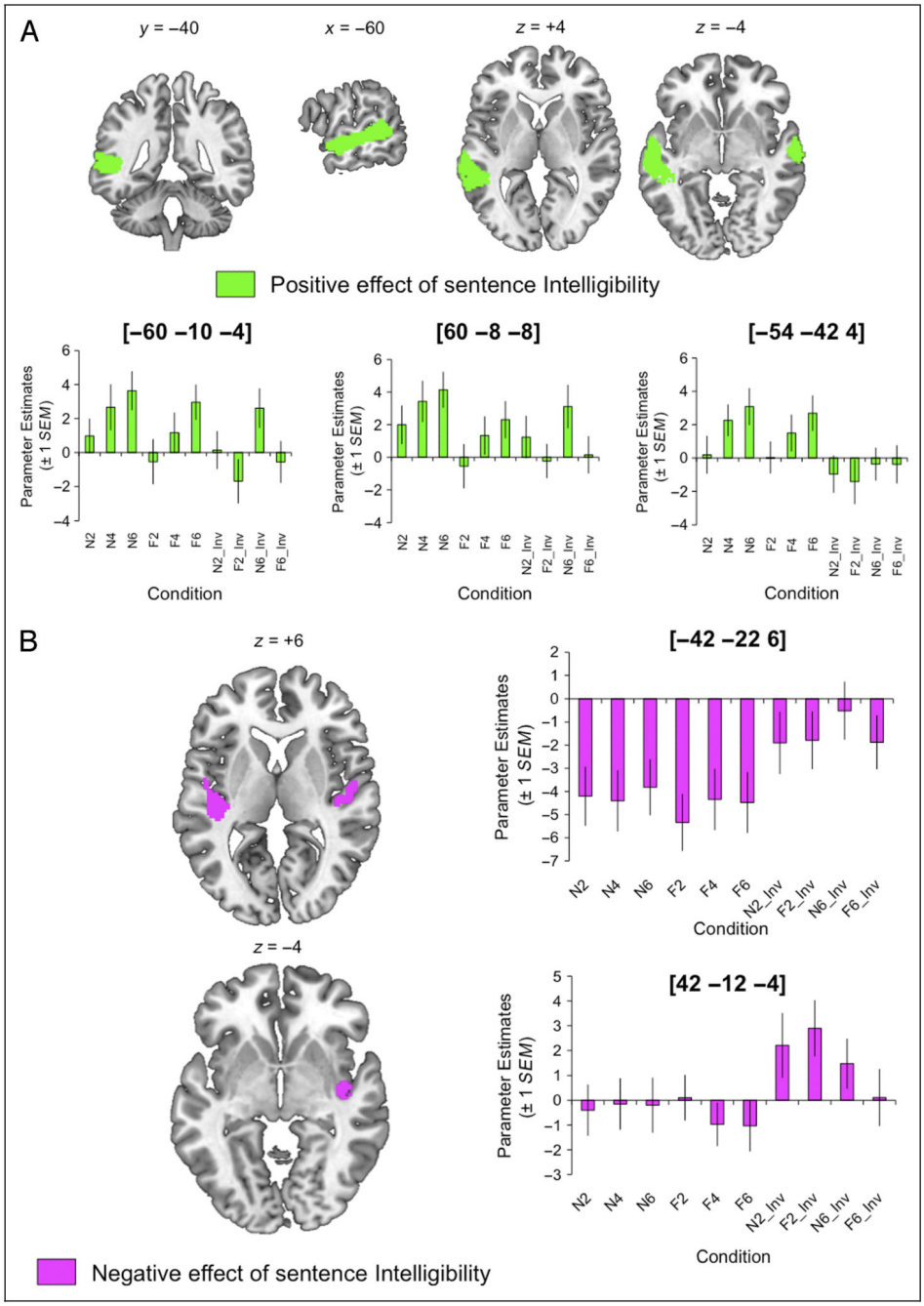

Figure 6. (A) Positive and (B) negative effects of sentence intelligibility across all conditions.
Plots show parameter estimates extracted from spherical ROIs built around the peak voxels (+1 *SEM*). Images are shown at a voxelwise threshold of $p < .001$ (uncorrected) and a cluster extent threshold of $p < .001$ (corrected; Slotnick et al., 2003). Numbers above slices indicate coordinates in MNI stereotactic space.

Table 1

Results of a Flexible Factorial ANOVA on the Vocoded Conditions Only, with Factors Channels (2, 4, 6) and F0 Excitation (Normal, Falling)

			Coordinate				
Contrast	*No. of Voxels*	*Region*	*x*	*y*	*z*	*F/T*	*z*
Main effect of channels (F test)	611	Left STG/STS	−58	−14	−2	15.16	4.60
			−50	−42	6	13.05	4.25
			−66	−22	2	12.76	4.20
	412	Right STG/STS	62	−8	−2	14.19	4.44
			66	−22	2	10.69	3.81
Main effect of excitation (F test)	998	Right STG/STS	56	−14	4	37.97	5.49
			62	−8	−4	34.34	5.25
			70	−18	2	22.68	4.33
Positive effect of channels (6 > 4 > 2)	1121	Left STG/STS	−58	−14	−2	5.43	5.03
			−50	−42	6	5.07	4.74
			−66	−22	2	5.02	4.70
	757	Right STG/STS	62	−8	−2	5.33	4.95
			66	−22	2	4.62	4.36
			52	−34	0	3.57	3.44
Positive effect of excitation (N > F)	557	Right STG/STS	56	−14	4	6.16	5.61
			62	−8	−4	5.85	5.38
			70	−18	2	4.75	4.48

Voxel height threshold $p < .001$ (uncorrected), cluster extent 68 voxels (corrected; Slotnick et al., 2003). Coordinates indicate the position of the peak voxel from each significant cluster in MNI stereotactic space.

Table 2

Results of a Full Factorial ANOVA with Factors Inversion (Vocoded, Inverted), Channels (2, 6), and F0 Excitation (Normal, Falling)

Contrast	*No. of Voxels*	*Region*	*Coordinate*			*F/T*	*z*
			x	*y*	*z*		
Main effect of inversion (F test)	1063	Left planum temporale/Heschl's gyrus	−44	−20	4	66.76	6.75
			−40	−30	8	58.71	6.43
			−52	−26	10	38.09	5.39
	1623	Right Heschl's gyrus/planum temporale	52	−16	6	61.27	6.53
			56	−24	12	47.71	5.92
			42	−28	14	42.34	5.63
	551	Left STS	−54	−40	4	30.42	4.89
			−62	−12	−6	25.96	4.55
			−56	−8	−12	19.44	3.97
Positive effect of inversion (vocoded > inverted) (*t* Test)	684	Left STS	−54	−40	4	5.52	5.02
			−52	−12	−6	5.10	4.69
			−56	−8	−12	4.41	4.13
Negative effect of inversion (inverted > vocoded) (*t* Test)	1157	Left planum temporale/Heschl's gyrus	−44	−20	4	8.17	6.85
			−40	−30	8	7.66	6.53
			−52	−26	10	6.17	5.51
	1791	Right Heschl's gyrus/planum temporale	52	−16	6	7.83	6.64
			56	−24	12	6.91	6.03
			42	−28	14	6.51	5.75

Voxel height threshold $p < .001$ (uncorrected), cluster extent 68 voxels (corrected; Slotnick et al., 2003). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space. See Methods for details.

Table 3

Results of a Flexible Factorial ANOVA on the Vocoded Conditions Only, with Factors Channels (2,6) and F0 Excitation (Normal, Falling)

			Coordinate				
Contrast	*No. of Voxels*	*Region*	*x*	*y*	*z*	*F/T*	*z*
Main effect of channels (F test)	844	Left STS/STG	−60	−14	−2	31.97	4.85
			−52	−42	8	29.59	4.70
			−66	−22	2	22.17	4.13
	601	Right STG/STS	62	−12	0	29.82	4.71
			68	−18	0	27.06	4.52
			66	−26	2	24.12	4.29
Main effect of excitation (F test)	719	Right STG	54	−14	6	34.48	5.01
			60	−12	0	31.08	4.80
			68	−22	−2	22.03	4.12
Positive effect of channels (6 > 2)	1076	Left STS/STG	−60	−14	−2	5.65	4.99
			−52	−42	8	5.44	4.84
			−66	−22	2	4.71	4.29
	780	Right STG/STS	62	−12	0	5.46	4.85
			68	−18	0	5.20	4.66
			66	−26	2	4.91	4.45
Positive effect of excitation (N > F)	965	Right STG/STS	54	−14	6	5.87	5.14
			60	−12	0	5.58	4.93
			68	−22	−2	4.69	4.28

Voxel height threshold $p < .001$ (uncorrected), cluster extent 20 voxels (corrected; Slotnick et al., 2003). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space.

Table 4

Results of a Flexible Factorial ANOVA on the Inverted Conditions Only, with Factors Channels (2,6) and F0 Excitation (Normal, Falling)

			Coordinate				
Contrast	***No. of Voxels***	***Region***	***x***	***y***	***z***	***F/T***	***z***
Main effect of excitation (F test)	151	Left STG/STS	−64	−20	2	23.46	4.24
			−54	−16	−4	19.16	3.87
	145	Right STG	68	−22	2	22.69	4.18
			60	−18	2	18.00	3.75
			58	−28	8	13.96	3.32
Positive effect of excitation (N > F)	214	Left STG/STS	−64	−20	2	4.84	4.39
			−54	−16	−4	4.38	4.03
	241	Right STG	68	−22	2	4.76	4.33
			60	−18	2	4.24	3.92
			58	−28	8	3.74	3.51

Voxel height threshold $p < .001$ (uncorrected), cluster extent 20 voxels (corrected; Slotnick et al., 2003). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space.

Table 5
Brain Regions Sensitive to Sentence Intelligibility

			Coordinate				
Contrast	***No. of Voxels***	***Region***	***x***	***y***	***z***	***F/T***	***z***
Main effect of intelligibility (F test)	1493	Left STS/temporal pole	−60	−10	−4	48.39	6.39
			−54	−42	4	47.84	6.36
			−66	−24	2	37.11	5.65
	332	Right STG	60	−8	−8	32.69	5.32
			64	−16	−2	24.92	4.67
	390	Left Heschl's gyrus/planum temporale	−42	−22	6	23.20	4.50
			−44	−30	8	23.18	4.50
	190	Right insula/Heschl's gyrus	42	−12	−4	17.14	3.86
			46	−18	6	15.05	3.61
			52	−8	6	14.49	3.54
	73	Right planum temporale/rolandic operculum	42	−28	16	15.83	3.71
			52	−26	16	13.43	3.40
Positive effect of intelligibility (*t* test)	1654	Left STS	−60	−10	−4	6.96	6.50
			−54	−42	4	6.92	6.46
			−66	−24	2	6.09	5.77
	395	Right STG/STS	60	−8	−8	5.72	5.45
			64	−16	−2	4.99	4.81
Negative effect of intelligibility (*t* test)	475	Left Heschl's gyrus/planum temporal	−42	−30	9	5.03	4.84
			−42	−21	3	4.96	4.78
			−30	−30	18	3.45	3.38
	465	Right insula/Heschl's gyrus/rolandic operculum	42	−12	−3	4.36	4.24
			42	−27	12	4.02	3.92
			51	−9	3	3.94	3.84

Voxel height threshold $p < .001$ (uncorrected), cluster extent 20 voxels (corrected; Slotnick et al., 2003). Coordinates indicate the position of the peak voxel from each significant cluster, in MNI stereotactic space.

Europe PMC Funders Author Manuscripts

Table 6

Results of Laterality Index Analyses (Using the LI Toolbox in SPM; Wilke & Schmithorst, 2006)

Contrast	*Trimmed Mean*	*Min*	*Max*	*Weighted Mean*
Positive effect of channels (2 < 4 < 6)	0.17	0.13	0.22	**0.23**
Normal > Flattened (3 × 2 ANOVA)	−0.27	−0.68	−0.066	− **0.57**
Positive effect of channels (2 < 6; 2 × 2 ANOVA with vocoded conditions)	0.13	0.1	0.16	0.16
Normal > Flattened (2 × 2 ANOVA with vocoded conditions)	−0.29	−0.65	−0.095	− **0.57**
Normal > Flattened (2 × 2 ANOVA with inverted conditions)	−0.019	−0.076	−0.016	0.0052
Vocoded > Inverted (2 × 2 × 2 ANOVA)	0.74	0.63	0.87	**0.84**
Inverted > Vocoded (2 × 2 × 2 ANOVA)	−0.12	−0.16	−0.032	−0.015
Positive effect of intelligibility (one-way ANOVA)	0.61	0.48	0.73	**0.71**
Negative effect of intelligibility (one-way ANOVA)	−0.013	−0.054	0.07	0.19