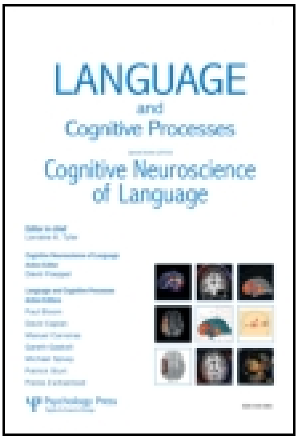

Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/plcp20

Brain regions recruited for the effortful comprehension of noise-vocoded words

Alexis G. Hervais-Adelman [a b], Robert P. Carlyon [b], Ingrid S. Johnsrude [c] & Matthew H. Davis [b]

[a] Functional Brain Mapping Laboratory, Department of Fundamental Neuroscience, University of Geneva Medical School, Geneva, Switzerland

[b] MRC Cognition and Brain Sciences Unit, Cambridge, UK

[c] Department of Psychology, Queen's University, Kingston, Ontario, Canada

Version of record first published: 12 Jul 2012

To cite this article: Alexis G. Hervais-Adelman, Robert P. Carlyon, Ingrid S. Johnsrude & Matthew H. Davis (2012): Brain regions recruited for the effortful comprehension of noise-vocoded words, Language and Cognitive Processes, 27:7-8, 1145-1166

To link to this article: http://dx.doi.org/10.1080/01690965.2012.662280

LANGUAGE AND COGNITIVE PROCESSES
2012, 27 (7/8), 1145–1166

Psychology Press
Taylor & Francis Group

Brain regions recruited for the effortful comprehension of noise-vocoded words

Alexis G. Hervais-Adelman[1,2], Robert P. Carlyon[2], Ingrid S. Johnsrude[3], and Matthew H. Davis[2]

[1]Functional Brain Mapping Laboratory, Department of Fundamental Neuroscience, University of Geneva Medical School, Geneva, Switzerland
[2]MRC Cognition and Brain Sciences Unit, Cambridge, UK
[3]Department of Psychology, Queen's University, Kingston, Ontario, Canada

We used functional magnetic resonance imaging (fMRI) to investigate the neural basis of comprehension and perceptual learning of artificially degraded [noise vocoded (NV)] speech. Fifteen participants were scanned while listening to 6-channel vocoded words, which are difficult for naïve listeners to comprehend, but can be readily learned with appropriate feedback presentations. During three test blocks, we compared responses to potentially intelligible NV words, incomprehensible distorted words and clear speech. Training sessions were interleaved with the test sessions and included paired presentation of clear then noise-vocoded words: a type of feedback that enhances perceptual learning. Listeners' comprehension of NV words improved significantly as a consequence of training. Listening to NV compared to clear speech activated left insula, and prefrontal and motor cortices. These areas, which are implicated in speech production, may play an active role in supporting the comprehension of degraded speech. Elevated activation in the precentral gyrus during paired clear-then-distorted presentations that enhance learning further suggests a role for articulatory representations of speech in perceptual learning of degraded speech.

Keywords: Vocoded speech; Neuroimaging; Motor system; Left inferior frontal cortex.

The networks of brain regions involved in speech perception have been extensively studied with functional brain imaging (see Davis et al., 2007; Hickok & Poeppel, 2004, 2007; Scott & Johnsrude, 2003). Under favourable listening conditions, spoken word perception engages multiple regions in the lateral temporal lobes: primarily the anterior superior and middle temporal gyri bilaterally, the left temporo-parietal

Correspondence should be addressed to Alexis Hervais-Adelman, Functional Brain Mapping Laboratory, Department of Fundamental Neuroscience, University of Geneva Medical School, Rue Michel Servet 1, CH-1206, Geneva, Switzerland. E-mail: alexis.adelman@unige.ch

This work was supported by a United Kingdom Medical Research Council PhD studentship to Alexis Hervais-Adelman, and by Medical Research Council funding of the Cognition and Brain Sciences Unit (MC_US_A060_0038) and by the Canada Research Chairs Programme (Ingrid S. Johnsrude). We would also like to thank two anonymous reviewers for their helpful comments and suggestions.

http://www.psypress.com/lcp http://dx.doi.org/10.1080/01690965.2012.662280

junction and left posterior middle and inferior temporal gyri and connected frontal regions (Saur et al., 2008).

However, these networks have been largely outlined for clearly presented, highly intelligible speech, which may not reflect our everyday experience of spoken language. The speech that we understand may be strongly accented, or degraded by signal compression or poor quality audio reproduction (e.g., telephones, Milchard & Cullington, 2004). Furthermore, it is frequently heard against a background of noise, interruptions, or other voices (e.g., G. A. Miller & Licklider, 1950; Warren, 1984), and distorted by room acoustics (e.g., A. J. Watkins, 2005), changes in speaking rate (J. L. Miller, Grosjean, & Lomanto, 1984), and accent (Ferreira, Henderson, Anes, & Weeks, 1996). Human speech perception is generally robust to almost all of these sorts of manipulations to a certain extent, although hearing-impaired and non-native listeners are adversely affected by all of these challenges (Bent & Bradlow, 2003; van Wijngaarden, Steeneken, & Houtgast, 2002).

In this article, we explore the neural correlates of the perception of speech under a form of artificial acoustic degradation (noise-vocoding), in particular focusing on additional processes that are engaged when listeners are exposed to speech signals that can only be comprehended with additional effort. Noise-vocoding (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) is a manipulation that degrades the spectral content of speech. It is used as a simulation of speech as transduced through cochlear implant processors, and is particularly useful in the study of speech-comprehension under acoustically-challenging conditions, as its difficulty can be easily manipulated. Normally-hearing participants exposed to noise-vocoded speech initially find it difficult to understand, although perception improves over time due to perceptual learning (Dahan & Mead, 2010; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008; Loebach, Pisoni, & Svirsky, 2010). We hypothesise that some of the additional processes engaged when comprehending this form of degraded speech are responsible for listeners' gradual improvements in perception. Neuroimaging investigations may help delineate systems that are responsible for this form of perceptual learning. This is of particular importance for the rehabilitation of cochlear implant users, whose initial experiences of their devices can be significantly enhanced if they receive appropriate training (Fu & Galvin, 2008; Stacey et al., 2010). A more complete understanding of the neural basis of the ability to understand simulations of cochlear-implant processed speech may help to elaborate and further enhance training strategies for newly implanted postlingually deafened individuals (see, e.g., those described by Loebach et al., 2010; Stacey et al., 2010; Stacey & Summerfield, 2007).

Functional imaging investigations of speech perception have shown that, under challenging listening conditions, different patterns of activation emerge for degraded compared to clear speech. Davis and Johnsrude (2003) compared neural responses to sentences degraded in a variety of ways with responses to clear speech and signal correlated noise. They observed additional activation for degraded speech in left inferior frontal and premotor regions, as well as in the superior temporal gyrus surrounding auditory cortex. However, these activations were observed in the context of a secondary task (rating the intelligibility of speech), and so might be a neural correlate of increased task difficulty, rather than perceptual effort per se. A study by Binder, Liebenthal, Possing, Medler, and Ward (2004) showed that inferior frontal activity was directly linked to difficulty of perceptual discriminations for speech sounds masked by varying levels of background noise. It would, therefore, be valuable to measure neural responses to speech, without a secondary task, to investigate whether additional activation reflects difficulties in perception or merely increased task demands.

Two further functional magnetic resonance imaging (fMRI) studies have explored changes in neural responses to noise-vocoded speech under conditions that include perceptual learning. Giraud et al. (2004) compared fMRI responses to a set of vocoded sentences before and after participants were taught to perceive them correctly. They found that left inferior frontal gyrus (Broca's area) responded significantly more to noise-vocoded speech after training, perhaps reflecting additional resources engaged in "auditory search" for comprehension. However, their pre- and posttraining comparisons did not include comparison of clear and degraded speech before and after training, and it is hence unclear whether the response to degraded speech changed or whether participants simply attended more closely to the degraded stimuli after training. Condition-specific differences in the degree to which participants attend to certain stimuli are particularly likely since this was a blocked design in which the same type of stimulus was presented for 30 seconds at a time.

A better-controlled fMRI study of training effects for vocoded sentences was reported by Eisner, McGettigan, Faulkner, Rosen, and Scott (2010). They presented listeners with potentially comprehensible noise-vocoded and incomprehensible (spectrally inverted) noise-vocoded sentences. They showed differences in activity evoked by these two forms of vocoded speech in the left inferior frontal and superior temporal gyri. The difference in inferior frontal activity between the potentially comprehensible and the incomprehensible stimuli was correlated (across participants) with the degree of improvement in report scores seen over the course of the experiment. However, from this result alone it is unclear whether the change in frontal activity is directly associated with perceptual learning—it may equally stem from the downstream effects of increased intelligibility (e.g., interpreting longer strings of words and the associated increases in semantic and syntactic processing). Those participants that perceive vocoded speech more successfully will have more opportunity to process sentence-level syntax and meaning, both of which have previously been associated with inferior frontal gyrus activity (e.g., Davis et al., 2007; Friederici, 2002; Friederici, Opitz, & von Cramon, 2000; Hagoort, 2005; Hagoort & Van Berkum, 2007; Rodd, Davis, & Johnsrude, 2005; Tettamanti et al., 2009). Eisner and colleagues (2010) also report a connectivity analysis, in which they find increased coupling between left inferior frontal gyrus and angular gyrus for learnable vs. unlearnable degraded sentences. They suggest this relates to the mapping of written feedback onto learnable, but not unlearnable, degraded sentences.

In this study, we used a contrast that assessed the neural basis of the effortful perception of degraded speech while excluding task effects and sentence level processes. Instead of using sentences (as employed by Davis & Johnsrude, 2003; Eisner et al., 2010; Giraud et al., 2004) we use single spoken words, in an event-related fMRI paradigm using a nonspeech target detection task. We are particularly interested in assessing additional activation associated with listening to vocoded speech compared to clear speech—a neural correlate of listening effort similar to that assessed by Davis and Johnsrude (2003). By comparing responses to noise-vocoded and clearly spoken words during monitoring for nonspeech targets, we can be confident that activation differences between the two conditions are not due to task performance, since the monitoring task is the same regardless of stimulus type. Furthermore, there is no requirement to engage in higher-level sentence processing or make an overt behavioural response on critical trials. One potential source of higher-level information does remain, however, and that is the phonological and lexical structure of the familiar English words that are presented.

A further goal of the present study is to investigate the neural changes that underpin listeners' improved comprehension of degraded speech after exposure, by comparing pre and posttraining functional responses to degraded speech. To achieve this, we employed a multi-stage design to (1) image the brain's responses to degraded speech; (2) monitor listeners' performance; (3) train them to understand noise-vocoded words to a reasonable level of proficiency within a single fMRI experiment before; and (4) re-examining cerebral responses to noise-vocoded words. By interleaving the training and assessment sessions with test sessions, we can assess functional changes to the brain's response to degraded speech as performance improves.

Recent behavioural studies (Davis et al., 2005; Hervais-Adelman et al., 2008) have demonstrated the importance of feedback for rapid perceptual learning of noise-vocoded speech. In the present study, we will assess neural activity associated with a feedback presentation that facilitates perceptual learning. This contrast can help further our understanding of the neural mechanisms driving perceptual learning of vocoded speech.

Although in the present article we focus on training using clear auditory feedback, which cannot be used to train postlingually deafened cochlear implant users, Davis and colleagues (2005) demonstrated that written feedback is as effective as auditory feedback, and the efficacy of written feedback has been demonstrated for rehabilitating recently implanted cochlear implant (CI) users (Stacey et al., 2010). The training method we choose to apply involves providing listeners with a clear instance of a word, followed by its noise-vocoded equivalent. It has been shown that prior access to the identity of a vocoded word produces a perceptual experience of enhanced clarity of the degraded token. This experience of perceptual "pop-out" has been linked to enhanced rates of perceptual learning in previous investigations into training on noise-vocoded speech (e.g., Davis et al., 2005; Hervais-Adelman et al., 2008). Although the previous studies used triplets in which the first stimulus was always a distorted probe, for the sake of brevity we elected to use only pairs of stimuli. We choose to compare paired presentations of clear-then-distorted (CD) words with distorted-then-clear (DC), clear-then-clear (CC), and distorted-then-distorted (DD) in order to seek any neural responses unique to the CD pairs, while controlling for the potential confounds of having pairs of stimuli differing in clarity, possible effects of priming and any potential effects of the ordering of the members of pairs. As we will review in a later section, existing evidence demonstrates a clear superiority of CD vs. DC pairs in supporting learning. This comparison of conditions that do and do not produce pop-out, therefore, reveals processes that are likely to contribute to perceptual learning. Although the relative effectiveness for perceptual learning of noise vocoded (NV) speech has not been previously assessed, this study lays the groundwork for further investigation of the potential neural drivers of perceptual learning in conditions that induce perceptual pop-out.

METHODS

Participants

Fifteen adults (10 female, aged 18–35 years, right-handed native speakers of British English, without hearing problems or dyslexia) took part.

Stimuli

Stimuli were 350 monosyllabic and 350 bisyllabic concrete nouns of intermediate frequency (1–75 occurrences/million), more than three letters in length, selected from the CELEX database (Baayen, Piepenbrook, & Gulikers, 1995). These words were recorded directly to.wav files by a male native speaker of Southern British English, at 16 bits and a 44.1 kHz sampling rate. Recordings were padded with silence (such that all sound files were 1 second long), noise or pulse-train vocoded (described below), and pre-emphasised to compensate for the uneven frequency response of the pneumatic headphone system in the MRI scanner.

Critical words in the test and training blocks were six-band noise-vocoded words (NV6 words) using the procedure described by Shannon and colleagues (1995) implemented in Matlab (The Mathworks, Natick, MA, USA). The words were filtered into six logarithmically spaced frequency bands from 50 to 5000 Hz (based on Greenwood, 1990). Pass bands were 3 dB down at 50, 200, 456, 889, 1626, 2876, and 5000 Hz with a 16 dB/octave roll-off. The amplitude envelope from each band was extracted by half-wave rectification and removal of pitch-synchronous oscillations above 30 Hz with a 2nd-order Butterworth filter. The resulting envelopes were multiplied with broadband noise which was then band-pass filtered in the same frequency ranges as the source. A similar procedure was used to create unintelligible, one-band noise-vocoded words (NV1 words), using an amplitude envelope extracted for one frequency band, between 50 and 5000Hz. Whereas NV6 words are potentially comprehensible, especially after a period of exposure with feedback (cf. the training methodology employed in Davis et al., 2005; and Hervais-Adelman et al., 2008), NV1 words remain entirely unintelligible, even after considerable exposure and training. Although NV1 words are less acoustically complex than NV6 words we chose this as a control condition since informal listening suggested that other, more complex stimuli such as “spectrally rotated” words can still contain phonological information and hence might evoke some attempt at understanding, even if such stimuli cannot be readily identified. The primary purpose of the NV1 stimuli was to provide a low-level, unintelligible baseline and assistance in interpreting the comparison of NV6 and clear words.

In addition to NV1 words, a set of 12 words were one-band vocoded using a 70 pulse-per second harmonic complex as the carrier signal, yielding unintelligible signals with a distinct, buzzy timbre (cf. Deeks & Carlyon, 2004; Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011). Listeners were asked to respond with a button press whenever they heard these buzzy sounds during scanning, ensuring that participants remained awake and attentive throughout the scanning sessions. Words were randomly allocated to different conditions in the test and training blocks for each participant, ensuring that stimulus-specific characteristics did not make a contribution to the effects observed over all listeners. Each word was presented only once to each participant, with the exception of the unintelligible words in the NV1 condition which were also randomly allocated to other conditions.

PROCEDURE

The experiment was divided into eight blocks (see Figure 1), with three types of task. Blocks 1, 4, and 7 were “fMRI test runs” in which listeners monitored for buzzy sound targets; Blocks 2, 5, and 8 were “behavioural test sessions”, in which listeners repeated words, and Blocks 3 and 6 were “training runs” in which listeners again performed the

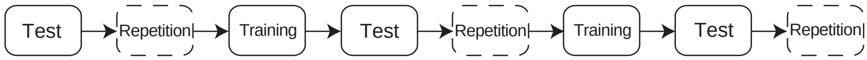

Figure 1. The order of the eight blocks of the experiment. Dashed blocks denote behavioural test sessions, solid blocks indicate fMRI scanning sessions.

buzzy-target detection task. Before beginning the experiment, participants were played examples of each of the kinds of stimuli that they were to be presented with, and the levels adjusted so that stimuli were audible and at a comfortable level. These were played until listeners were satisfied that they could tell the difference between NV1 and the buzzy target stimuli.

Scanning was carried out using a Siemens 3T Tim Trio MR system. Echo-planar imaging (EPI) volumes comprised 21 4-mm thick slices (interslice distance 1 mm, matrix size 64×64, in-plane resolution 3×3 mm, TR = 2,500 ms or 3,500 ms, respectively for test and training blocks, TA = 1,300 ms, TE = 30 ms, flip angle = 78°). The slices were transverse-oblique, angled away from the eyeballs to avoid ghost artefacts from eye movements. Acquisitions covered the majority of the brain, but missed the superior aspect of the parietal lobe in participants with larger brains. A T1-weighted structural scan was acquired for each subject using a three-dimensional MPRAGE sequence (TR = 2,250 ms; TE = 2.99 ms; flip angle = 9°; field of view = $256 \times 240 \times 160$ mm; matrix size = $256 \times 240 \times 160$ mm; spatial resolution = $1 \times 1 \times 1$ mm).

fMRI test blocks

These blocks were designed to measure the brain's response to clear speech, comprehensible noise-vocoded speech and incomprehensible noise-vocoded speech in naïve listeners (Block 1), after some training (Block 4), and after more training (Block 7). Listeners were instructed to listen to the stimuli presented over the headphones, monitor them for "buzzy" sounds (one-band pulse-train-vocoded words), and to press a button on the button box each time they heard one of these. A fast-sparse imaging design was used (cf. Orfanidou, Marslen-Wilson, & Davis, 2006) such that stimuli were presented in the 1.2 second silent intervals between 1.3 second scans, to minimise scanner-noise interference (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999). Each of the blocks consisted of 50 clear words, 50 NV1 words, 50 NV6 words, 50 silent trials and all 12 target stimuli, resulting in a total of 215 scans per block. Stimuli were presented in a fully randomised order (see Figure 2a for an illustration of the timeline of stimulus presentation and scanning). The words used as NV1 stimuli were re-used in other conditions, as they were incomprehensible and unrecognisable, thus a total of 300 different words were used over the three test blocks.

Behavioural test blocks

Comprehension of NV speech was assessed before and after the first training block (i.e., in Blocks 2 and 5) and at the end of the experiment (Block 8) by presenting listeners with the 50 NV6 words from the preceding fMRI test block in a newly randomised order and asking them to repeat the words if they were able to or to say "don't know" if they were unable to even guess the word's identity. Each word was preceded by a warning tone, and listeners had 5 seconds after each word in which to make a response. Participants were not scanned during these sessions, as head movements during speech production could have contaminated concurrent fMRI data.

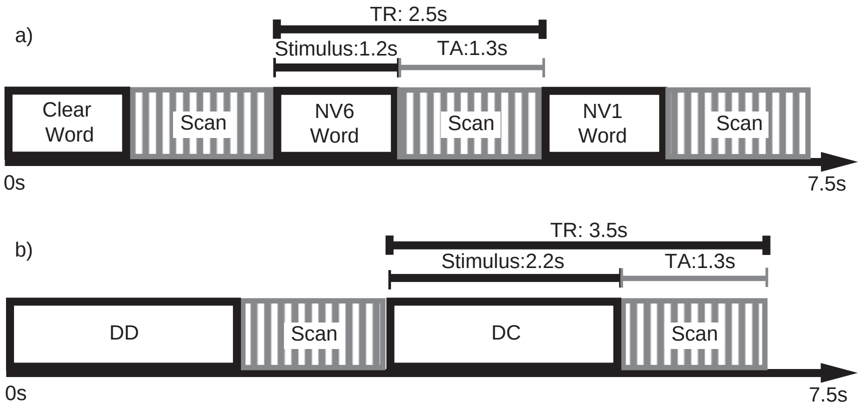

Figure 2. Schematic illustration of the presentation of stimuli and scanning in fMRI runs. (a) Shows a timeline for 3 stimuli in the fMRI Test blocks; (b) shows a timeline for 2 stimuli in the fMRI Training runs. Stimulus durations and hence silent intervals differ, but the acquisition time (TA) is identical in the two types of run.

fMRI training blocks

Blocks 3 and 6 were "fMRI training blocks" in which listeners heard pairs of clear and NV6 words in differing combinations—clear-clear (CC), degraded-degraded (DD), clear-degraded (CD), and degraded-clear (DC). Fifty pairs of each type were presented. A total of 266 scans were acquired, with a TR of 3.5 seconds. Stimulus pairs were presented in the 2.2 seconds period between 1.3 second scans: a slightly longer silent interval between successive scans was required for these blocks compared to the "test" blocks to permit the presentation of a pair of stimuli at a time.

Behavioural studies of the perception of vocoded sentences (Davis et al., 2005) have shown that CD feedback presentations provide more effective training than DC pairs or a single additional distorted presentation. For isolated words, the difference between CD and DC presentation is even more pronounced. Hervais-Adelman et al. (2008) showed only limited learning with DC feedback. In both these studies participants first heard a distorted instance of the stimulus, which they were asked to report (i.e., presentation was distorted-clear-distorted (DCD) or distorted-distorted-clear (DDC) with initial presentation used to provide behavioural scores). Since this initial presentation of a stimulus occurred in all conditions, the differential efficacy of the presentation conditions can only be due to the different feedback provided after report. Furthermore, in other work we have observed that exposure to a series of 180 noise-vocoded words without feedback produced no significant improvements in free-report scores (Hervais-Adelman, 2007). Therefore, we can assume that DD word pairs would also produce markedly less perceptual learning. These blocks served the dual function of training volunteers and enabling us to seek characteristic neural responses to effective (CD) over less effective (DC, DD, CC) training stimuli. To ensure attention, participants were asked to monitor for occasional pairs of buzzy target stimuli (12 pairs per session).

Instructions, stimulus presentation and response collection

Participants were informed before the beginning of the experiment of all the stages they would encounter—i.e., runs involving single or paired presentation of syllables, as well as periods in which they would be required to repeat aloud. Participants were also informed that the stimuli they would hear would be a series of comprehensible or incomprehensible English words and presented with examples of each of these, along with example target stimuli.

Auditory stimuli were presented through a pair of Etymotic Research (Etymotic Research Inc., Elk Grove Village, IL, USA) ER3A insert pneumatic earphones. Listeners were fitted with ear defenders over the earphones to attenuate the noise of the scanner.

The responses for the task of listening for target stimuli were made by pressing the index-finger button of a proprietary MR-compatible four-button box, using the right hand. Response times and accuracy were automatically recorded on computer. In the behavioural test runs, listeners' responses were spoken aloud and were made into an FOMRI (Fibre Optical Microphone for MRI Communication) dual-channel MR-compatible optical microphone (Opto-Acoustics Ltd, Or-Yehuda, Israel), and recorded directly to hard-disk. The accuracy of these responses was assessed later, and responses were scored correct if listeners had repeated the test word exactly.

Data preprocessing and analysis

Data were preprocessed and analysed using Statistical Parametric Mapping software (SPM8; Wellcome Department of Cognitive Neurology, London, UK). Prior to analysis, all images were corrected for motion by realigning them with respect to the first image. They were also "unwarped" to correct for geometric image distortions resulting from interactions between subject head movement and magnetic field inhomogeneities. The mean of the realigned and unwarped images was coregistered with the structural T1 volume, which was then spatially normalised to a standard template. The same spatial transformation was applied to the realigned EPI volumes. Finally, the normalised images were smoothed with a 10 mm full-width half-maximum (FWHM) Gaussian kernel suitable for random-effects analysis (Xiong et al., 2000). Analysis of the EPI images acquired from the test and training sessions was performed separately, since they had different TRs.

Test sessions

Data from each subject were entered into a general linear model using an event-related analysis procedure (Josephs & Henson, 1999). Six event types were modelled for each session: these were the three conditions (Clear, NV1, NV6), responses to target trials divided into hits and misses, and false alarms. Null events (silent trials) were left unmodelled (Josephs & Henson, 1999). Each event was modelled using the canonical haemodynamic response function (HRF) in SPM8. Our analysis focused primarily on responses to noise-vocoded and clear-speech trials without overt behavioural response. Movement parameters estimated at the realignment stage of preprocessing were added as regressors of no interest. A high-pass filter (cutoff 128 seconds) and AR1 correction for serial autocorrelation were applied.

Analysis of group data was achieved by entering contrasts of parameter estimates from single-subject models into random-effects analyses (T-tests), comparing differences of parameter estimates over subjects to zero. In addition to these analyses, an analysis of variance was conducted to test for a main effect of condition in the fMRI test runs as well as any condition-by-session interaction (cf. Henson & Penny, 2005).

Training sessions

The data from each subject were entered into a general linear model, using an event-related analysis procedure. Seven event types were modelled for each session:

CC, CD, DC, and DD stimuli alongside three categories of response to the buzz target stimuli (hit, miss, false alarm). Null events were left unmodelled as in the test sessions. Due to the longer duration of the stimuli and increased TR, model inspection revealed a comparatively poor fit for SPM's canonical HRF. We, therefore, modelled the events using a finite impulse response (FIR) model, with three time bins. This model enabled us to examine the activation attributable to each stimulus type over the subsequent 3 scans. Movement parameters were added as regressors of no interest, and a high-pass filter (cutoff 128 seconds) and AR1 correction for serial autocorrelation were applied. Group random-effects analyses were assembled from the contrasts of parameter estimates for the CC, CD, DC, and DD conditions, for each of the three time bins of the single-subject FIR model, collapsed over both training sessions. The data were then amenable to an analysis of variance in which the four conditions were modelled as two crossed factors, each with two levels (first stimulus clear or degraded; second stimulus clear or degraded).

RESULTS

Behavioural tests

All participants tested were successful in detecting the occasional buzz targets in the test and training scanning runs and withholding button presses for clear or noise-vocoded words (mean $d' = 3.59$, $SE = 0.023$). Word report scores from the repetition task showed a significant improvement in performance from Test 1 to Test 3, $F(2, 28) = 27.085$, $p < .001$, $\eta^2 = 0.659$, indicating that the training blocks produced effective learning of NV6 speech. These behavioural results are shown in Figure 3.

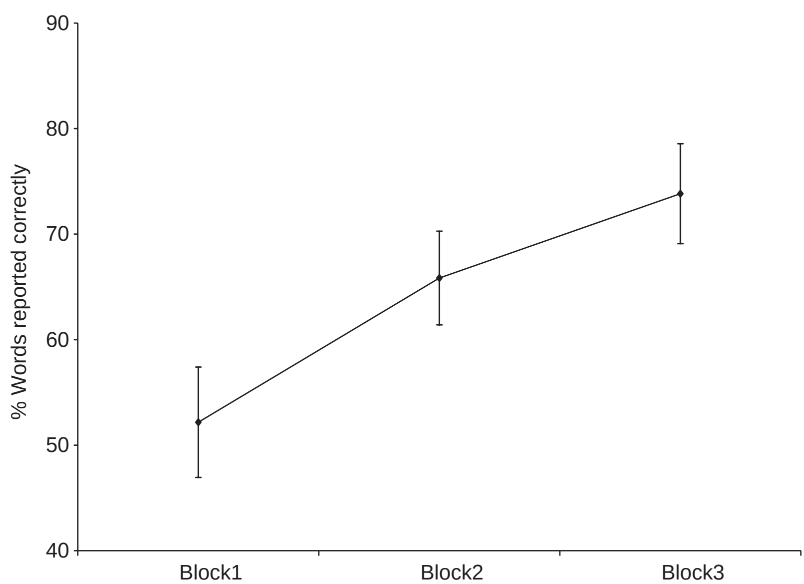

Figure 3. Behavioural data from the repetition blocks. Mean proportion of words reported correctly in each of the three test blocks. Error bars represent *SE* of the mean.

Test sessions

Our analysis of neural activity focused on two components of the network involved in speech perception: one network that is involved in comprehension of speech generally, which will be tested by comparing activations relating to Clear stimuli with activations related to NV1 stimuli (NV6 stimuli would also be expected to activate these areas to the extent that they are also comprehensible). The second network we expect to be activated additionally for NV6 compared to clear speech and can be thought of as being involved in the additional "effort" of processing degraded speech. All the results discussed below are significant at an uncorrected $p < .001$ at the voxel level, with a cluster extent threshold of 100 voxels, and a family-wise error corrected cluster-level significance of $p < .05$.

Comprehension network

We examined the contrast between clear and NV1 speech to locate brain regions involved in comprehending clear speech, while controlling for the presence and processing of a relatively complex acoustic signal. The contrast is presented in Figure 4. The peak activations are shown in Table 1. Significant activations are found extending along most of the length of the left superior temporal sulcus and left superior temporal gyrus, along the right superior temporal gyrus, in the left fusiform gyrus and the left hippocampus, extending to the posterior amygdala. The activations in these areas were not unique to the clear speech stimuli, as can be seen in the plots of parameter estimates shown in Figure 4. The response of these regions was elevated relative to NV1 stimuli for NV6 as well as for clear speech.

Effort network

Additional activation due to hearing potentially intelligible, degraded speech is shown by the contrast between NV6 and Clear speech. This showed additional activation for NV6 words in left precentral gyrus (motor cortex), as part of a cluster extending into left inferior frontal gyrus (frontal operculum), and the left and right anterior insulae.

To rule out the possibility that the observed activity could have been due to the button-pressing component of the task, we compared neural activation in trials when participants successfully identified target stimuli (by a button-press) with NV1 stimuli and applied this activation map (thresholded at uncorrected $p < .001$) as an exclusive mask to the analysis of NV6 vs. Clear stimuli. This mask included the right anterior insula region in which NV6 vs. Clear differences were observed. Differences in activation in this region will, therefore, not be discussed. The exclusively-masked results are presented in Figure 4 and Table 2. The plots of parameter estimates included in Figure 4 show that these areas respond to NV6 more than clear speech and NV1. This profile suggests a response that is enhanced during the processing of potentially comprehensible degraded speech.

Neural responses related to changes in perceptual report

To test for changes in activation over the three test sessions we submitted single-condition parameter estimates to a 3-by-3 analysis of variance (ANOVA) assessing the differential effect of the three listening conditions (Clear, NV1, and NV6 speech) and 3 scanning sessions (fMRI Test runs 1, 2, and 3) using methods described by Henson and Penny (2005). There was no significant main effect of session. The main effect of

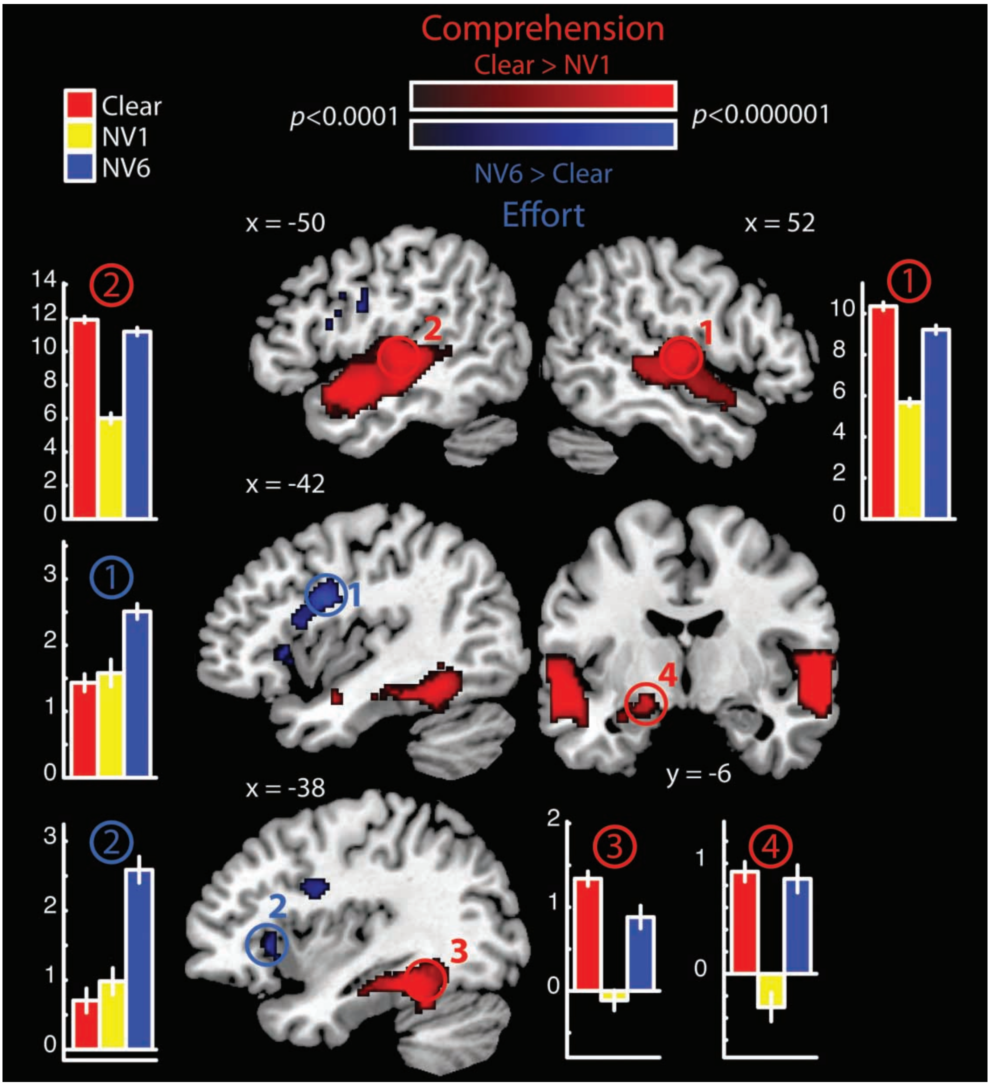

Figure 4. Comprehension (red scale) and effort (blue scale) networks displayed on sagittal sections of canonical brain. Coordinates of the plane of section of the slices are shown. The contrasts displayed are the positive second-level T-contrasts (on the canonical haemodynamic response function), exclusively masked by the activation map for button-presses, thresholded at uncorrected $p < .001$. Only voxels that are significant at uncorrected $p < .0001$ are shown, for clarity in this display. Colour scale represents uncorrected significance levels. Bar graphs show the parameter estimate for each condition at the peak voxels that reach whole-brain corrected significance in the circled regions (numerical labels correspond to the numbered peaks in Table 1 and 2), *y*-axes are parameter estimates of the fit of the canonical haemodynamic response plotted in arbitrary units. Error bars show *SE* of the mean, corrected for repeated-measures comparisons (after Loftus and Masson, 1994).

condition (Clear, NV6, NV1) reveals areas where different conditions produced significantly different activations (described above). The condition-by-session interaction was nonsignificant. Despite the highly reliable changes in report score observed over the three scanning runs, no brain area exhibited differential change in response to the three classes of stimuli (Clear, NV1, NV6) over the three test blocks of the experiment.

TABLE 1
Comprehension network. Peak voxels for Clear-NV1 contrast, second level group analysis. T-contrast (on canonical haemodynamic response function parameter estimates)

Brain region	*Peak voxel coordinates (x, y, z mm)*	*Z-score*	*Cluster size (voxels)*
****Right superior temporal sulcus [1]**	**64, −22, 0**	**6.19**	**1626**
	64, −6, −2	5.97	
	50, −30, 2	4.11	
****Left superior temporal Sulcus [2]**	**−60, −18, 0**	**6.01**	**2010**
	−54, 4, −12	5.40	
	−60, −38, 4	4.6	
****Left fusiform gyrus [3]**	**−38, −44, −14**	**5.13**	**462**
	−38, −24, −16	4.11	
***Left hippocampus [4]**	**−18, −6, −12**	**4.43**	**99**
	−30, −2, −18	4.14	
	−24, 6, −14	3.98	

Notes: All voxels are significant at uncorrected $p < .0001$. Activations marked with * are significant at the cluster level at FWE-corrected $p < .05$, those marked ** are significant at the voxel level at FWE-corrected $p < .05$. The table shows significant peaks and sub-peaks of clusters a minimum of 8 mm from each other. Numbers in square brackets refer to peaks whose activations are plotted in Figure 4 (marked in red). Bold entries denote the peak voxel of each cluster.

To ensure that the effects seen were not due to subjects becoming aware that they should attend more carefully to the NV6 stimuli, as they were later to be tested on them, we carried out an ANOVA comparing the magnitude of activation for NV6 and Clear conditions over the three test sessions, at the peak voxel of the NV6 > Clear contrast in the left precentral gyrus (MNI co-ordinates: –40, 4, 26). This test shows no significant main effect of session, $F(2, 28) = 2.025$, $p = .151$, no significant condition-by-session interaction, $F(2, 28) = 1.37$, $p = .271$, and a significant main effect of condition, $F(1, 28) = 61.60$, $p < .001$. We note therefore, that activation of left precentral gyrus for the contrast of NV6—Clear is apparent in the first scanning session before participants could have been aware that the NV6 items would be repeated in the behavioural assessment. The effect can thus not be attributed to the effects of deliberate attention being paid in order to enhance later recollection of stimuli.

To assess differences in activation that were related to changes in performance (i.e., the result of perceptual learning), activations due to NV6 stimuli in the final fMRI test

TABLE 2
Effort network. Peak voxels for NV6-Clear contrast, second level group analysis. T-Contrast (on canonical haemodynamic response function parameter estimates)

Brain region	*Peak voxel coordinates (x, y, z mm)*	*Z-score*	*Cluster size (voxels)*
****Left precentral gyrus/pars opercularis [1]**	**−40, 4, 26**	**4.78**	**232**
	−50, −2, 26		
	−42, 14, 16		
Left anterior insula [2]	**−38, 22,4**	**4.19**	**33**

Notes: All voxels are significant at uncorrected $p < .0001$. Activations marked with * are significant at the cluster level at FWE-corrected $p < .05$, those marked ** are significant at the voxel level at FWE-corrected $p < .05$. The table shows significant peaks and sub-peaks of clusters a minimum of 8 mm from each other. Numbers in square brackets refer to peaks whose activations are plotted in Figure 4 (marked in blue). Bold entries denote the peak voxel of each cluster.

session were contrasted with the activations elicited by NV6 stimuli in the first fMRI test session (Session 3 NV6 vs. Session 1 NV6). This produced no significant differences. Further analyses were carried out to seek any performance-related changes by using changes in performance as regressors for brain activity in individual participants. No brain areas were found in which activation was significantly correlated with changes in comprehension over the three test bocks, either for differences due to increased comprehension of NV6 stimuli (i.e., the NV6 minus NV1 contrast) or for listening effort (NV6 minus Clear speech) when an appropriate region of interest was used (defined by the main effect of condition, thresholded at uncorrected $p < .001$). Analyses in which correct/incorrect report for single trials of NV6 was included as a parametric modulator also yielded no significant results. Thus after detailed examination of the data, we found no changes in the fMRI response to the experimental stimuli at different points during perceptual learning of noise-vocoded speech. It may be that perceptual changes were insufficient to modulate neural responses or that our between-session design lacked sensitivity to detect these changes.

Training sessions

Our primary goal in analysis of fMRI responses during the training sessions was to assess additional neural responses associated with Clear-then-Degraded (CD) word pairs, since these have been shown to enhance perceptual retuning (Davis et al., 2005; Hervais-Adelman et al., 2008). To do this we tested for an interaction between first and second stimulus type, equivalent to the contrast (CD-DD)–(CC-DC). Significant interactions between first and second stimulus type (at FWE-corrected $p < .05$) are observed in the left precentral sulcus, extending onto the precentral gyrus and into the inferior frontal sulcus (Figure 5, Table 3). In order to rule out the possibility of this motor-region's differential response being related to the button-pressing task, we exclusively masked the activation map with the hits—NV1 contrast, as described in the analysis of the effort contrast. The precentral activation reported was nonoverlapping with this contrast. Since this interaction could arise due to a differential response to either CD or DC stimuli, we used post hoc analyses of these responses to explore the origin of this neural interaction, specifically looking for an increased response in CD (pop-out) trials compared to DC (non pop-out trials). As seen in the plot of the parameter estimates for each condition in each time bin of the FIR model, we see a specific increase in neural activity for CD stimuli in the left precentral sulcus in the third time bin, i.e., approximately 7 seconds after the offset of the pair of stimuli. This is confirmed by t-tests comparing CD–DC for each of the peak voxels in each time-bin; this contrast is only significant in the third time-bin in [$T(14) = 2.60$, $p < .05$, sidak-corrected for multiple comparisons], and the effect is significantly larger in time-bin 3 compared to time-bin 2. $T(14) = 2.11$, $p < .05$. Although the exact timing of the neural response cannot be determined from fMRI data with such a long TR, these data show a late neural response specific to CD presentation. The magnitude of perceptual improvement in individual participants' word-report scores did not predict activity in this region.

DISCUSSION

In our study, listening to both clear speech and potentially comprehensible noise-vocoded speech produced bilateral temporal-lobe activity extending from anterior to

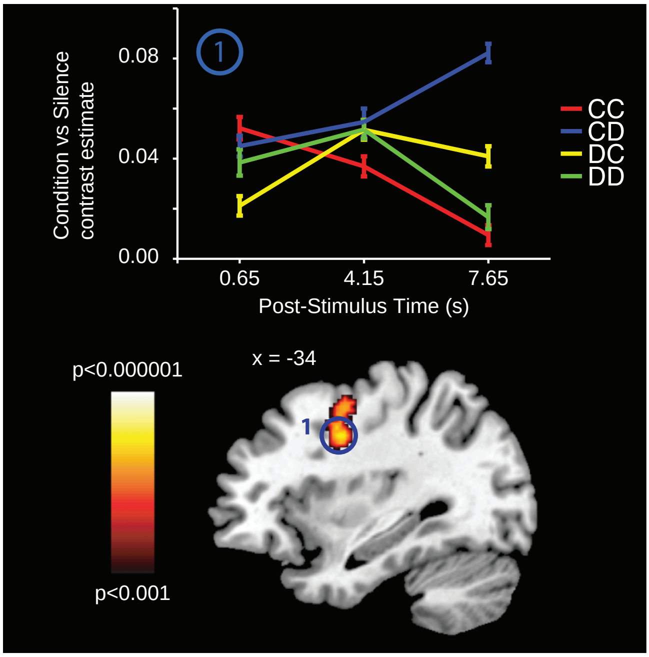

Figure 5. Location of regions showing a significant interaction between clear/distorted first and second stimuli [interaction contrast: (CD–DD)–(CC–DC) where CC = clear-clear; CD = clear-distorted; DC = distorted-clear; DD = distorted-distorted pairs]. Activations projected on a single slice of a canonical single-subject brain, for uncorrected $p < .001$, inclusively masked by the activation map of the effects of interest in the test sessions, thresholded at uncorrected $p < .001$. Significance of activations can be found in Table 3. Line plots show the parameter estimates of the conditions at a peak voxel in (1) the inferior frontal sulcus over the three analysed time bins of the FIR model. Error bars represent *SE* of the mean, corrected for between-subjects comparisons (Loftus & Masson, 1994).

posterior regions of the superior and middle temporal gyri in comparison with unintelligible NV1 stimuli, similarly to existing neuroimaging studies comparing listening to speech and nonspeech (Mummery, Ashburner, Scott, & Wise, 1999; Rodd et al., 2005; see also Scott & Johnsrude, 2003). Clear speech also activated the left fusiform gyrus and left Hippocampus more than NV1, consistent with existing studies of word comprehension (e.g., Davis & Johnsrude, 2007). The comparison of intelligible, clear spoken words contrasted with nonspeech noises highlights multiple temporal regions that contribute to speech comprehension under favourable listening conditions. Given that, in everyday life, the speech we hear is sometimes not clearly spoken or is heard in the presence of background noise or other forms of degradation, an important focus of the present research was to examine the neural correlates of increased listening effort associated with comprehension of degraded spoken words.

TABLE 3
Training trials, second-level group ANOVA analysis

Brain region	*Peak voxel coordinates (x, y, z mm)*	*Z-score*	*Cluster size (voxels)*
***Left precentral sulcus [1]**	**−34, 0, 38**	**4.18**	**187**

Notes: Interaction between first and second stimulus type, significant at uncorrected $p < .001$. Activations marked with * are significant at the cluster level at FWE-corrected $p < .05$. Bold entries denote the peak voxel of each cluster.

The majority of functional imaging studies have assessed neural responses to intelligible speech by comparison with less intelligible control or baseline conditions (e.g., Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000). Such studies do not consider the possibility of additional neural processes associated with the effortful perception of distorted or degraded speech stimuli.

Compared with listening to clear speech, we saw additional activity for potentially intelligible noise-vocoded words in a region extending from the precentral gyrus into the left frontal operculum. The pattern of activation suggests that in addition to recruiting the same temporal-lobe processing pathways as clear speech, potentially comprehensible noise-vocoded speech recruits alternative processing pathways involving premotor and prefrontal regions in the left hemisphere. The *pars opercularis* has been associated with nonword processing and phonological working memory (e.g., Poldrack et al., 1999). We have previously suggested a role for phonological short-term memory in degraded speech perception and learning (Davis et al., 2005; Hervais-Adelman et al., 2008). Eisner and colleagues (2010) found a similar region of left inferior frontal gyrus to be implicated in the perception of degraded sentences, and further that the response of this region to degraded words correlated, between participants, with perceptual improvements over the course of the study. The left inferior frontal gyrus has also been associated with many nonlinguistic functions such as task switching, (e.g., Derrfuss, Brass, von Cramon, Lohmann, & Amunts, 2009), working memory and pain perception (e.g., Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), and hence evidently does not subserve a purely speech-motor function. However, increased activation in response to NV6 vs. clear speech is not confined to the left inferior frontal gyrus and extends into more posterior and medial regions that have been more specifically associated with articulatory functions.

The regions of precentral gyrus revealed by this contrast are situated in the inferior portion of the gyrus, proximal to and ventral to the junction of the inferior frontal sulcus and the precentral gyrus, consistent with the localisation of tongue- and mouth-related regions of cortex found in many motor-cortical mapping studies (e.g., Grafton, Woods, Mazziotta, & Phelps, 1991; Lotze et al., 2000; Penfield & Boldrey, 1937). Given this, we think it is reasonable to suggest that the activation of the precentral gyrus reported here is in a region that contributes to the control of speech articulation. However, we acknowledge that in the absence of an explicit articulatory movement localiser task this conclusion must remain tentative. Nonetheless, we believe that the cortical regions activated here are constituents of the dorsal auditory pathway, and the activations observed likely reflect a well established contribution of this pathway to mapping heard speech onto articulatory representations of speech (Hickok & Poeppel, 2007; Saur et al., 2008; Scott & Johnsrude, 2003). We note again that these regions do not overlap with those involved in the button-press responses of the buzz-detection task.

The engagement of the left anterior insula by NV6 stimuli is greater than for either clear or NV1 stimuli. Although the anterior insula is often associated with the processing of disgust or other emotionally valent stimuli (Schienle et al., 2002), it has also been shown to have very strong connections with the inferior frontal gyri (Jabbi & Keysers, 2008) and to contribute to speech production (Ackermann & Riecker, 2004; Borovsky, Saygin, Bates, & Dronkers, 2007). For instance, it has been suggested that the insulae play a role in the control of vocal tract musculature (Ackermann & Riecker, 2004Ackermann & Riecker, 2010; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000), and neuropsychological evidence associates lesions to the anterior insula with speech production deficits (Dronkers, 1996). At the level of speech reception the anterior insula has been implicated in auditory processing deficits contributing to dyslexia (Steinbrink, Ackermann, Lachmann, & Riecker, 2009). We take this evidence to suggest a possible speech-motor role for the insula, bolstering the case for the involvement of a speech-motor circuit in the processing of challenging but comprehensible degraded speech stimuli.

Unlike previous studies, we can rule out the possibility that these precentral and inferior frontal activations are associated with performing overt tasks on degraded speech. For instance, Davis and Johnsrude (2003) found that regions of the left inferior frontal cortex and adjacent premotor and motor regions were activated when listeners rated the intelligibility of degraded speech of various types and various degrees. Although they suggested that the left inferior frontal and precentral gyri play a role in compensation for degraded speech input, task-specific contributions (i.e., rating speech intelligibility) could not be ruled out. Consistent with this, Binder and colleagues (2004) observed an increased response in the inferior frontal lobe and anterior insula for decisions made on spoken syllables presented in background noise. Increased difficulty of decision processes is therefore a plausible explanation of elevated frontal responses for degraded compared to clear speech in both these studies. No such decision-based explanation is likely for the present study. Participants were engaged in a passive monitoring task (listening for rare buzzy target stimuli) and made no explicit response in trials involved in the critical contrast. Despite the absence of an explicit task to be performed on either Clear or NV6 words, however, we see substantial additional activation of prefrontal and premotor regions for degraded speech—the conclusion we draw is that these regions are indeed involved in effortful perception of degraded speech.

Previous studies have shown similar activation of prefrontal and premotor regions in response to degraded speech. Giraud et al. (2004) reported additional posttraining activation for noise-vocoded speech in bilateral insula and inferior frontal gyrus. However, the critical comparison in this study was between 30-second-long blocks perceived as speech or nonspeech and hence differential responses might result from slow fluctuations in the degree to which participants attend to different stimuli. A similar observation of increased activation in ventral premotor cortex was seen during adaptation to time-compressed sentences (Adank & Devlin, 2010), whereas Eisner and colleagues (2010) showed activation in inferior frontal gyrus associated with adaptation to noise-vocoded speech. However, in all these studies the critical contrast may activate frontal regions because of their contribution to syntactic and semantic processing of whole sentences (cf. Friederici, 2002; Friederici, Meyer, & von Cramon, 2000; Hagoort & Van Berkum, 2007; Heim, 2005; Rodd et al., 2005; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006). As comprehension of degraded sentence improves, the amount of sentence-level processing occurring will also increase, this alone could explain the enhanced activation observed by Eisner and colleagues (2010) and by Giraud and colleagues (2004) in these frontal regions. No such explanation is plausible in the present event-related fMRI study. The single words that we presented will not engage sentence level processes even when they are fully intelligible. We, therefore, propose that the prefrontal, insular, and premotor activation that we observe is linked to processes recruited for the perception of degraded speech. We note that two other recent studies have similarly reported activation of precentral gyrus during perception of isolated words (Osnes, Hugdahl, & Specht, 2010; Tremblay & Small, 2011).

Motor-cortical responses to degraded speech

The motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000) proposes that the recognition of phonological units in speech is achieved by inferring the articulatory gestures of the speaker, and hence involves the recruitment of motoric processes. This theory is currently experiencing a revival of interest in response to mounting evidence from functional imaging (e.g., Pulvermuller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004) and Transcranial Magnetic Stimulation (TMS) (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; K. E. Watkins, Strafella, & Paus, 2003) that motor regions are activated during speech perception. A study by K. E. Watkins and Paus (2004) combined TMS potentiation of motor activation with concurrent positron emission tomography (PET) imaging. It demonstrated that changes in the excitability of speech-motor areas in volunteers listening to speech were significantly correlated with the magnitude of speech-evoked activity in the posterior part of the left inferior frontal gyrus (Broca's area) and in inferior parietal regions; suggesting that these regions comprise a functional network; similar inferior frontal regions were activated in the present contrast of effortful versus effortless comprehension of spoken words.

Despite evidence for the activation of motor regions in speech perception, controversy remains concerning whether or not this activation implicates motor regions as a necessary component of the speech perception system (see Lotto, Hickok, & Holt, 2009; Scott, McGettigan, & Eisner, 2009). It has been suggested that such motor activity is artefactual, and is not observed when comparisons are made with appropriately matched nonspeech stimuli, or not significant at a whole-brain corrected level (Scott et al., 2009). Such criticisms cannot readily be applied to the results reported here. Indeed, the subtraction that reveals whole-brain corrected motor involvement involves additional activation for degraded compared to clear speech. A second criticism—that activation of motor regions is task-dependent and observed in situations that require subvocal rehearsal such as phonemic segmentation (Lotto et al., 2009)—is also addressed by the present study. The critical comparison that activated motor regions here involved two conditions, both of which were perceived as speech, in the context of a passive monitoring task that did not require any form of overt or covert vocal response. We therefore argue that activation of motor regions in the present study automatically accompanies the perception of degraded speech, and is not an additional process that participants opt to perform due to the nature of a task. Thus, unlike in older instantiations of the motor theory that posited that all speech perception was via motoric representations, we hypothesise that such representations are only important when the input is degraded and purely acoustic forms of representation are perhaps insufficient to yield optimal intelligibility.

Since functional imaging data are by nature correlational, we cannot state with any certainty that motor recruitment plays a necessary functional role in speech perception. Such data can only come from studies of brain injured patients with impaired speech production (cf. Moineau, Dronkers, & Bates, 2005; Utman, Blumstein, & Sullivan, 2001), or from studies in which TMS is used to transiently interfere with neural processes in motor regions (D'Ausilio, 2007; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Mottonen & Watkins, 2009). We note with interest, however, that all these studies show modulation of perception of speech that has been degraded—either by artificial speech manipulations (D'Ausilio, 2007; Meister et al., 2007; Moineau et al., 2005) or by the use of phonemically ambiguous materials

(Mottonen & Watkins, 2009; Utman et al., 2001). These studies, therefore, converge with our fMRI work in suggesting that motor responses to speech are more likely of functional significance in listening situations in which the perception of speech is challenged by noise or phonetic ambiguity.

The exact function of premotor regions and anterior insula in speech perception remains to be established in future studies. However, on the basis of the present data, we propose that the effortful perception of noise-vocoded speech, and perhaps other forms of degraded speech also, is assisted by networks involving regions of the dorsal auditory pathway that decode speech using nonacoustic, articulatory templates. This accords well with the suggestion made by Davis and Johnsrude (2007), Iacoboni (2008), Poeppel and Monahan (2010) and others that these regions provide an internal simulation that helps to match degraded speech input to internal templates derived from a prototypical motor pattern. Just such a feedback process has been invoked to explain the perceptual learning of noise-vocoded speech (Hervais-Adelman et al., 2008), and the current functional imaging study therefore provides initial activation evidence for brain regions that contribute to this form of perceptual feedback.

Further evidence from the present study for a role of the dorsal auditory pathway in perceptual retuning comes from our observation that left precentral gyrus is specifically engaged by CD word pairs compared to otherwise matched presentations. The distorted item in these CD stimuli produce a distinct perceptual experience that speech is perceptually more clear ("pop-out") that has been linked in a number of behavioural studies to enhanced perceptual learning (Davis et al., 2005; Hervais-Adelman et al., 2008). A recent study by Wild, Davis, and Johnsrude (in press) examined the neural correlates of perceptual pop-out with vocoded speech, using printed feedback (cf. Davis et al., 2005) to induce the experience. They showed that conditions inducing pop-out produced significantly greater activation in the left precentral gyrus and precentral sulcus than conditions that did not, consistent with the present results.

One putative mechanism for enhanced learning using CD stimuli arises from hearing degraded speech when the identity of the degraded speech token is known. We propose that when hearing a degraded speech stimulus that has been preceded by presentation of equivalent clear speech listeners can generate a training or error signal that permits modifications of preceding levels of processing so as to produce more optimal perception of degraded input subsequently. An anonymous reviewer has pointed out that this argument is based on the association between pop-out and learning, based on previous work (Davis et al., 2005; Hervais-Adelman et al., 2008). At present, data do not exist to conclusively demonstrate a causal relationship between the magnitude of pop-out and more effective perceptual learning with the stimuli used in this study. In our work, we show a neural correlate of this form of supervisory feedback in the late onset of additional activation in the precentral gyrus. This result is consistent with our proposal that motor regions generate an error signal, which is used to direct retuning of lower levels of the speech perception system (Davis & Johnsrude, 2007). The absence of any direct link between the magnitude of this response to CD stimuli and the observed behavioural improvement in perception is disappointing. However, we are hopeful that future research on the timing and connectivity of auditory–motor interactions might reveal a more direct association between neural generators of pop-out and perceptual learning.

CONCLUSION

The data presented earlier demonstrate that the human speech perception system recruits frontal regions (including the precentral gyrus and left anterior insula) as well as temporal lobe regions during comprehension of degraded speech. We propose that this involvement stems from the recruitment of articulatory representations of speech sounds or words, and that these can be used to help to identify the content of degraded speech. Furthermore, paired presentations that produce "pop-out" and enhance perceptual learning also increase activity in some of the same regions (precentral gyrus). Thus, we further propose that learning to comprehend degraded speech may be mediated by supervisory influences from the dorsal auditory pathway. Taken together, this evidence provides support for the view that perception of degraded speech can engage alternative, nonacoustic, representations of speech. We acknowledge that this remains a controversial view (Lotto et al., 2009; Scott et al., 2009). However, we believe that our data and other evidence reviewed here shows recruitment of precentral gyrus during speech perception and hence it is possible that links between auditory and articulatory representations may play some functional role in speech perception. Although older instantiations of the motor theory of speech perception made claims about the obligatory involvement of motor structures whenever speech is perceived, we suggest that our data are consistent with a more limited role, with these regions becoming involved specifically when speech is hard to understand. Further evidence, both from functional imaging, TMS and neuropsychology will be required, however, to show that the regions implicated here are necessarily recruited because of their motoric contribution. Nevertheless, this idea, that motor representations of speech can be used to enhance perception of degraded speech, merits further investigation for other forms of degraded speech and speech heard in adverse conditions.

REFERENCES

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain Lang*, *89*(2), 320–328.

Ackermann, H., & Riecker, A. (2010). The contribution (s) of the insula to speech production: A review of the clinical and functional imaging literature. *Brain Structure and Function*, *214*(5–6), 419–433.

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage*, *49*(1), 1124–1132.

Baayen, R. H., Piepenbrook, R., & Gulikers, L. (1995). The CELEX Lexical Database. Retrieved from http://www.ru.nl/celex/subsecs/section_psy.html

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, *114*(3), 1600–1610.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*(3), 295–301.

Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia*, *45*(11), 2525–2533.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*(Pt 5), 1193–1201.

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 704–728.

D'Ausilio, A. (2007). The role of the mirror system in mapping complex sounds into actions. *Journal of Neuroscience*, *27*(22), 5847.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F., et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences of USA*, *104*(41), 16032–16037.

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*(8), 3423–3431.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241.

Deeks, J. M., & Carlyon, R. P. (2004). Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation. *Journal of the Acoustical Society of America*, *115*(4), 1736–1746.

Derrfuss, J., Brass, M., von Cramon, D. Y., Lohmann, G., & Amunts, K. (2009). Neural activations at the junction of the inferior frontal sulcus and the inferior precentral sulcus: Interindividual variability, reliability, and association with sulcal morphology. *Human Brain Mapping*, *30*(1), 299–311.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*(6605), 159–161.

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*(2), 89–97.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*(21), 7179–7186.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Short communication: Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*, 399–402.

Ferreira, F., Henderson, J. M., Anes, M. D., & Weeks, P. A. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 324–335.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Languages*, *74*(2), 289–300.

Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. *Cerebral Cortex*, *10*(7), 698–705.

Fu, Q. J., & Galvin, J. J. III. (2008). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research*, *242*(1–2), 198–208.

Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, *14*(3), 247–255.

Grafton, S. T., Woods, R. P., Mazziotta, J. C., & Phelps, M. E. (1991). Somatotopic mapping of the primary motor cortex in humans: Activation studies with cerebral blood flow and positron emission tomography. *Journal of Neurophysiology*, *66*(3), 735–743.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species – 29 years later. *Journal of the Acoustical Society of America*, *87*(6), 2592–2605.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423.

Hagoort, P., & Van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). Sparse temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223.

Heim, S. (2005). The structure and dynamics of normal language processing: Insights from neuroimaging. *Acta Neurobiologiae Experimentalis*, *65*(1), 95–116.

Henson, R. N., & Penny, W. D. (2005). *ANOVAs and SPM*. London: Wellcome Department of Imaging Neuroscience, Institute of Cognitive Neuroscience.

Hervais-Adelman, A. (2007). *The perceptual learning of noise-vocoded speech*. Cambridge: University of Cambridge.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual Learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460–474.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 283–295.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology-Paris*, *102*(1–3), 31–34.

Jabbi, M., & Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion*, *8*(6), 775–780.

Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *354*(1387), 1215–1228.

Liberman, A. M., Cooper, E., Shankweiler, D., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(1), 224.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Lotto, A., Hickok, G., & Holt, L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110–114.

Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., & Grodd, W. (2000). fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage*, *11*(5 Pt 1), 473–481.

Meister, I., Wilson, S., Deblieck, C., Wu, A., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692–1696.

Milchard, A. J., & Cullington, H. E. (2004). An investigation into the effect of limiting the frequency bandwidth of speech on speech recognition in adult cochlear implant users. *International Journal of Audiology*, *43*(6), 356–362.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, *22*, 167.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*(4), 215–225.

Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech, Language, and Hearing Research*, *48*(4), 884–896.

Mottonen, R., & Watkins, K. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, *29*(31), 9819.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America*, *106*(1), 449–457.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13*(12), 1362–1368.

Orfanidou, E., Marslen-Wilson, W. M., & Davis, M. H. (2006). Neural response suppression predicts repetition priming of spoken words and pseudowords. *Journal of Cognitive Neuroscience*, *18*(8), 1237–1252.

Osnes, B., Hugdahl, K., & Specht, K. (2010). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage*, *54*(3), 2437–2445.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389.

Poeppel, D., & Monahan, P. J. (2010). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26*(7), 1–17.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, *10*(1), 15–35.

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of USA*, *103*(20), 7865–7870.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport*, *11*(9), 1997–2000.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*(8), 1261–1269.

Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., & Abel, S. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*(46), 18035–18040.

Schienle, A., Stark, R., Walter, B., Blecker, C., Ott, U., Kirsch, P., et al. (2002). The insula is not specifically involved in disgust processing: An fMRI study. *Neuroreport*, *13*(16), 2023–2026.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends and Neuroscience*, *26*(2), 100–107.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Stacey, P. C., Raine, C. H., O'Donoghue, G. M., Tapper, L., Twomey, T., & Summerfield, A. Q. (2010). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology*, *49*(5), 347–356.

Stacey, P. C., & Summerfield, Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America*, *121*(5), 2923–2935.

Steinbrink, C., Ackermann, H., Lachmann, T., & Riecker, A. (2009). Contribution of the anterior insula to temporal auditory processing deficits in developmental dyslexia. *Human Brain mapping*, *30*(8), 2401–2411.

Tettamanti, M., Rotondi, I., Perani, D., Scotti, G., Fazio, F., Cappa, S. F., et al. (2009). Syntax without language: Neurobiological evidence for cross-domain syntactic computations. *Cortex*, *45*(7), 825–838.

Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *Neuroimage*, *57*(4), 1561–1571.

Utman, J. A., Blumstein, S. E., & Sullivan, K. (2001). Mapping from sound to meaning: Reduced lexical activation in Broca's aphasics. *Brain and Language*, *79*(3), 444–472.

van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, *111*, 1906.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, *96*(2), 371–383.

Watkins, A. J. (2005). Perceptual compensation for effects of reverberation in speech identification. *Journal of the Acoustical Society of America*, *118*(1), 249–262.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience*, *16*(6), 978–987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994.

Wild, C. J., Davis, M. H., & Johnsrude, I. S. (in press). Human auditory cortex is sensitive to the perceived clarity of speech. *Neuroimage*.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702.

Xiong, J., Rao, S., Jerabek, P., Zamarripa, F., Woldorff, M., Lancaster, J., et al. (2000). Intersubject variability in cortical activations during a complex language task. *Neuroimage*, *12*(3), 326–339.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, *32*(4), 1826–1836.