# The Continuity Illusion Does Not Depend on Attentional State: fMRI Evidence from Illusory Vowels

**Antje Heinrich**[1,2], **Robert P. Carlyon**[1],
**Matthew H. Davis**[1], **and Ingrid S. Johnsrude**[2]

## Abstract

We investigate whether the neural correlates of the continuity illusion, as measured using fMRI, are modulated by attention. As we have shown previously, when two formants of a synthetic vowel are presented in an alternating pattern, the vowel can be identified if the gaps in each formant are filled with bursts of plausible masking noise, causing the illusory percept of a continuous vowel ("illusion" condition). When the formant-to-noise ratio is increased so that noise no longer plausibly masks the formants, the formants are heard as interrupted ("illusion break" condition) and vowels are not identifiable. A region of the left middle temporal gyrus (MTG) is sensitive both to intact synthetic vowels (two formants present simultaneously) and to illusion stimuli, compared to illusion Break stimuli. Here, we compared these conditions in the presence and absence of attention. We examined fMRI signal for different sound types under three attentional conditions: full attention to the vowels; attention to a visual distracter; or attention to an auditory distracter. Crucially, although a robust main effect of attentional state was observed in many regions, the effect of attention did not differ systematically for the illusory vowels compared to either intact vowels or to the illusion break stimuli in the left STG/MTG vowel-sensitive region. This result suggests that illusory continuity of vowels is an obligatory perceptual process, and operates independently of attentional state. An additional finding was that the sensitivity of primary auditory cortex to the number of sound onsets in the stimulus was modulated by attention.

## INTRODUCTION

In many everyday situations, the sound we wish to listen to—such as someone's voice—is often masked to some extent by other sounds that are present in the environment. The ability of listeners to, nevertheless, maintain a coherent perceptual representation of the target sound is reflected in a phenomenon known as the continuity illusion: When a portion of a target sound is replaced by another sound that may plausibly have masked it, the target is heard as continuous, not interrupted.

In order to generate a continuous percept of an interrupted stimulus, listeners often draw on higher-level knowledge of the structure of the sensory input. For instance, in the phoneme restoration effect, a section of speech that is masked by noise is perceived as being an instance of the most plausible missing phoneme. Hence, in the context /parədʌɪ/ listeners hear a noise-masked phoneme /?/ as being [s] to make the word "paradise." One natural interpretation of this finding is that higher-level lexical representations act "top–down" to reinstate missing elements of the acoustic input (Shahin, Bishop, & Miller, 2009; Samuel, 1981). One previous study of the neural basis of the phoneme restoration effect has associated prefrontal activation in response to illusory continuity with top–down repair (Shahin et al., 2009). This study demonstrates that regions supporting higher-level processes are more activated during the perception of interrupted word stimuli than pseudowords, whereas lower-level auditory areas do not distinguish between real and pseudowords. However, to what extent the continuity illusion relies on attentional top–down processes remains an open question. In the present work, we address this question by using an established neural correlate of the continuity illusion (Heinrich, Carlyon, Davis, & Johnsrude, 2008) to assess whether the illusion operates when a listener's attention is diverted and higher-level processes are unavailable (e.g., Sabri et al., 2008). In this way, we can test the extent to which it is appropriate to model the continuity illusion using "bottom–up" neural processes (Husain, Lozito, Ulloa, & Horwitz, 2005; Beauvois & Meddis, 1991), or whether, as often proposed for speech perception more generally, top–down processes play a critical role (e.g., Davis & Johnsrude, 2007; McClelland, Mirman, & Holt, 2006).

In a previous study (Heinrich et al., 2008), we used functional magnetic resonance imaging (fMRI) to obtain a neural correlate of the continuity illusion as it pertains to vowel sounds. Listeners are poor at identifying two-formant vowels in which the formants alternate in time, but they improve substantially when the silent gaps in each formant frequency region are filled by noise that could plausibly mask the formants. In this case, listeners hear an illusory continuous vowel ("illusion" stimuli; Carlyon,

[1] MRC Cognition and Brain Sciences Unit, Cambridge, UK,
[2] Queen's University, Kingston, Canada

Deeks, Norris, & Butterfield, 2002). Performance deteriorates again when the synthetic formant levels are increased so that they are no longer plausibly masked by the noise, causing the illusion to break down ("illusion break" stimuli). These latter two conditions are closely matched acoustically, but differ in terms of the degree to which they yield a vowel percept. Our initial fMRI study (Heinrich et al., 2008) demonstrated a neural correlate of the illusion: Activation in vowel-sensitive regions of the brain (bilaterally in middle temporal gyrus/superior temporal sulcus [MTG/STS]) was observed not only for intact but also for illusory vowels. In both cases, activation was greater than in nonspeech conditions such as illusion break (Figure 1).

Heinrich et al.'s (2008) results allowed us to conclude that the continuity illusion is complete at a stage of processing at or below that relying on vowel-sensitive cortex in the superior temporal gyrus (STG)/MTG. Perhaps more importantly, by providing an objective neural marker for the continuity illusion, these results provide a tool that allows us to determine whether the continuity illusion is dependent on attention. The use of such an objective measure allows one to overcome a problem inherent to behavioral approaches, namely, that of having to ask the participant about some aspect of the stimulus that they are meant to be ignoring. A previous study suggests that attention may not be crucial for the continuity illusion. Using EEG, Micheyl et al. (2003) recorded a mismatch negativity (MMN) response to a tone that deviated from a sequence of more common standards only by virtue of illusory continuity. Because an MMN is only observed when the response to the deviants and standards differ, Micheyl et al.'s results provided an independent marker of the illusion. However, as they pointed out, although their subjects were instructed to ignore the sounds and watch a silent movie, attention was not manipulated systematically, and so their conclusions were limited to the statement that the continuity illusion is partly complete for sounds that are outside the *focus* of attention. Here, we measure the neural response to illusory and to intact vowels both when subjects attend to those stimuli, and when they are required to perform a demanding task on other auditory or visual stimuli. If top–down processes and focused attention are critical for the perception of illusory continuity, then the neural response to illusory vowels should be reduced or absent in the absence of attention directed to the stimuli. We include both auditory and visual distractor stimuli because, if attention does affect the continuity illusion, it could do so either in a modality-specific way, as has been observed, for example, in divided attention tasks (e.g., Duncan, Martens, & Ward, 1997), or in a modality-independent manner, as has been observed for the effects of attention on auditory streaming (Carlyon, Plack, Fantini, & Cusack, 2003).

When studying the effects of attention on perceptual organization, it is important to distinguish between the effect of that influence on the resulting neural activity, and more general changes in neural activation that result from attending to a sound. For example, the results of several fMRI studies have shown that, in general, paying attention to a sound results in enhanced activity in periauditory regions of the STG (e.g., Sabri et al., 2008; Rinne et al., 2007; Petkov et al., 2004; Hugdahl, Thomsen, Ersland, Rimol, &

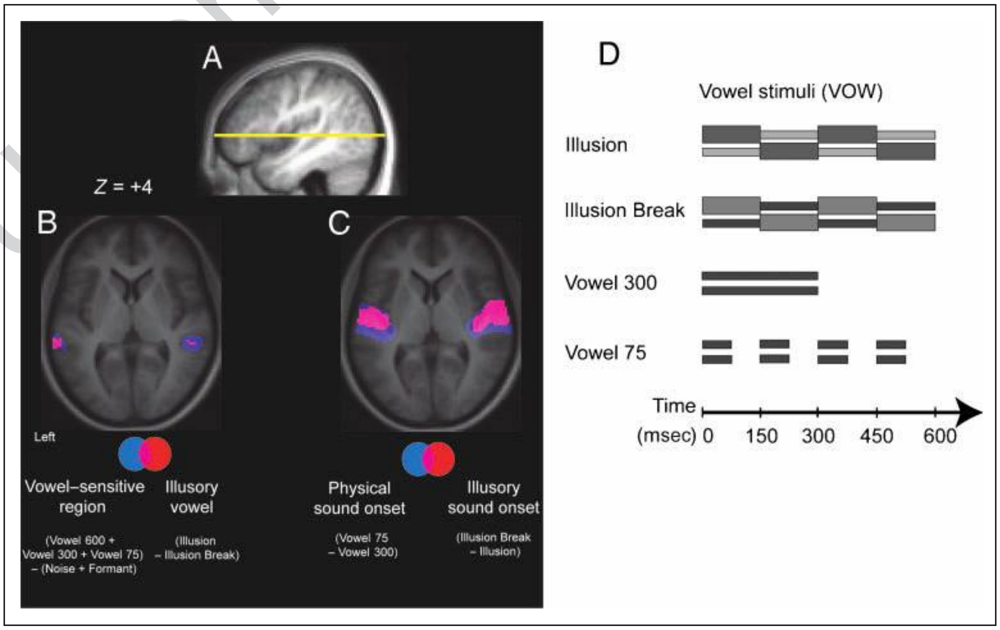

**Figure 1.** Schematic representation of selected conditions and results of Heinrich et al. (2008). (A) Location of horizontal slices B and C. (B) Areas showing significant activation in response to intact vowels compared to nonspeech stimuli (blue) ([Vowel 600 + Vowel 300 + Vowel 75] − [Isolated Formants + Isolated Noises]) and illusory vowels (red) (illusion–illusion break) in Heinrich et al. An overlap in activation of these conditions is displayed in magenta. (C) Areas showing significant activation in response to an increased number of physical sound onsets (blue) (Vowel 75–Vowel 300) and to a change in the number of perceived sound onsets due to the continuity illusion breaking down (illusory sound onset effect in red) (illusion break–illusion). An overlap in activation of these conditions is shown in magenta. The activations in this and subsequent images, unless otherwise specified, are shown superimposed on the mean T1 structural of all 22 participants of the current study, thresholded at $p < .05$ (FDR). (D) Schematic spectrograms of the four vowel conditions from the previous study that are used here. See text for further details.

Niemi, 2003; Jäncke, Mirzazade, & Shah, 1999; Pugh et al., 1996). Similarly, paying attention to visual stimuli in a audiovisual selective attention task increases activation in occipital cortex over conditions where the same stimuli were present but not attended (Mozolic et al., 2008; Sabri et al., 2008). Here we study whether the activation of the vowel-sensitive areas, identified by Heinrich et al. (2008), is greater when subjects attend to illusory vowels than when they attend to a competing auditory or visual stimulus. It is expected that paying attention to speech sounds will enhance activity for these sounds—for intact and illusory vowels alike. The main interest of the current study, however, lies in investigating the differential effects that the attentional manipulation might have on intact and illusory vowels, that is, whether the change in activation for different attention conditions is greater than that observed for intact (nonillusory) vowels. Such an interaction would indicate that the illusion is specifically susceptible to disruption when attention is diverted. Furthermore, we compare the effects of attention on vowel-sensitive regions to those in areas near primary auditory cortex, where, in the previous study (Heinrich et al., 2008), the neural response depended on the number of perceptual onsets in the sound, rather than on whether the sounds were perceived as vowels.

## METHODS

### Participants

Twenty-four adults between the ages of 18 and 39 years (mean age = 24.4 years, *SD* = 6.1, 11 men) took part. Two participants were subsequently excluded due to excessive head movement or equipment malfunction (one participant each), leaving a total of 22 datasets for statistical analysis. All participants were right-handed, fluent speakers of English, and reported no neurological disease or hearing loss. The study was cleared by the Local Research Ethics Committee at Addenbrooke's Hospital in Cambridge. Each volunteer provided written consent prior to participating and was paid £25.00 for their time.

### Stimuli

Stimuli on each trial were 8.4 sec long, and consisted of three concurrent stimulus sequences: one sequence included sounds from one of the four possible "vowel" stimulus types shown in Figure 1D; one sequence was an auditory distracter in a different frequency range to the vowel stimuli (AUD); and one sequence was a visual distracter (VIS). Depending on the attention condition, participants performed a target detection task on one of the three sequences. All three distracters were physically present on all trials.

#### *Vowel-condition (VOW) Sound Types*

Each stimulus was constructed from a 600-msec harmonic complex with energy at the first 100 harmonics. Synthetically generated formants were created by passing the harmonic complex through a second-order infinite impulse response (IIR) filter (Rabiner & Schafer, 1978) with a center frequency at the formant value and the 3-dB bandwidth of the filter fixed at 90 Hz. The following four vowels were generated: [ɑ] as in "far," [ε] as in "head," [ɔ] as in "ford," and [ɜ] as in "heard." Each vowel was synthesized five times on different fundamental frequencies (f0s): 115, 122, 137, 145, 150 Hz. We created four stimulus types, as shown in Figure 1D. Two of these were "chequerboard" stimuli (Howard-Jones & Rosen, 1993) in which formants alternated with noise that was either intense enough to plausibly mask a formant ("illusion"), or not ("illusion break"), and two types were intact, two-formant vowels ("Vowel 300"; "Vowel 75") (Heinrich et al., 2008; Carlyon et al., 2002). Speech-likeness ratings and vowel identification performance for the stimuli used here are given in Heinrich et al. (2008).

In the *illusion* condition, the first (f1) and second (f2) formants of 150-msec duration alternated for a total of four times and a total duration of 600 msec. The alternation started with f1 and ended with f2. The silent gap in each frequency region after the switch was filled with filtered noise in such a way that a band-pass noise complemented f1 and low-pass noise complemented f2. The filter cutoff was placed at the geometric mean between the two formant frequencies of each vowel and had a 96-dB/octave roll-off. This filter cutoff defined a low-frequency region for the lower formant or (low-pass) noise, and a high-frequency region for the upper formant or (band-pass) noise. The upper cutoff for the high-frequency region was set to 22,050 Hz, identical to the upper cutoff of the harmonic complex on which the vowel sounds were based. The formant-to-noise ratio (FNR) was calculated as the ratio of the root-mean-square (rms) value of each noise band over the rms value of each formant. For f1, FNR was calculated with reference to the respective low-pass noise band; for f2 it was calculated with reference to the respective band-pass noise. The FNR was set to −20 dB because extensive pilot testing in connection with Heinrich et al. (2008) had shown that an FNR of −20 dB produces a noise that is intense enough to plausibly mask the formants and thus induce an illusory vowel percept.

The *illusion-break* condition was identical to the illusion condition, except that the FNR was raised to +15 dB. This was done because the continuity illusion is not evoked if the noise is not loud enough to plausibly mask the formants. However, an increase in formant level leads to an increase in the overall level of the sound if the noise level is not simultaneously decreased. As a change in overall sound level may introduce confounding activation, we sought to keep overall sound level similar between illusion and illusion-break conditions. Therefore, the +15 dB FNR was achieved by increasing the level of f1 and f2 while simultaneously decreasing the level of the noise. As described in more detail by Heinrich et al. (2008), this resulted in the formant level being about 23 dB higher, and the noise level about 12 dB lower, than in the illusion condition. As a consequence, the illusory vowel percept did not emerge, rather, the stimulus was perceived as a series of pitched sounds alternating with noise and, therefore, sounded less speech-like than the illusion condition. Given the extensive piloting and the successful use of the illusion and illusion-break stimuli in Heinrich et al. (2008), we did not obtain another rating measure for the speech-likeness of those stimuli from the current group of participants.

In the *Vowel 300* condition, both formants were gated on and off simultaneously for 300 msec, followed by a 300-msec silence. The resulting percept was that of a vowel (Heinrich et al., 2008).

In the *Vowel 75* condition, both formants were presented simultaneously for 75 msec, followed by a 75-msec silence. This pattern was repeated four times. The overall duration of each vowel was the same as in the Vowel 300 condition, but with four sound onsets rather than one. Again, the resulting percept was that of a vowel (Heinrich et al., 2008).

Each segment of sound was gated on and off with a 5-msec raised cosine ramp. Stimuli were low-pass filtered at 4 kHz, with a 145-dB/oct roll-off. Low-pass filtering the vowel stimuli enabled us to present the auditory distracter sounds in the frequency region above 4 kHz, without any overlap of sound energy with the vowel sounds.

Finally, VOW stimuli were assembled into 8.4-sec sequences of 10 stimuli. Each sequence only contained stimuli from one of the four vowel conditions (Vowel 300, Vowel 75, illusion, illusion break). Sounds were combined so that vowel identity and f0 changed between successive items. The ISI between sounds was variable (mean = 240 msec, *SD* = 41.23 msec), as was the period of silence at the beginning and end of each 8.4-sec sound sequence (mean = 120 msec, *SD* = 18.86 msec). Twenty-four unique sequences were generated in each of the four vowel conditions for a total of 96 VOW sequences.

Three sequences out of the 24 of each VOW type (i.e., 1 in 8) contained two consecutive tokens attenuated by 9 dB. These served as target stimuli when participants were instructed to attend to the vowel sequences (see the first trial of Figure 2). Targets were placed at either the third and fourth, the fifth and sixth, or the seventh and eighth tokens of the sequence.

### *Auditory Distraction Stimuli (AUD)*

When instructed to attend to the auditory distracters, participants monitored this sequence of noise bursts for occasional targets that differed from the other noise bursts on the basis of their amplitude envelope (see the second trial of Figure 2). To generate the noises, a broadband white noise signal of 400-msec duration was passed through a 4- to 5-kHz brick wall band-pass filter that resulted in a 60-dB SPL dropoff from full scale within 100 Hz of the nominal cutoff frequency on either side. As a consequence, there was minimal overlap in sound energy between the vowel and auditory distracter stimuli. Over the 400-msec duration, the noise stimuli changed their amplitude in one of two ways: "Approach" noises increased in amplitude over 380 msec linearly from zero to full-scale amplitude followed by a 20-msec linear amplitude decrease back to zero (and sounded like they were getting closer); "depart" noises—which were the targets—conformed to the opposite pattern with a 20-msec linear rise and a 380-msec linear fall.

Sequences of 12 noise bursts, each 8.4 sec long, were created. The ISI between consecutive noise bursts was jittered between 220 and 380 msec (mean = 300 msec, *SD* = 49 msec), and each sequence began with 200 msec of silence and ended with 100 msec of silence. The majority of noise bursts had the "approach" temporal envelope. Twelve of the 96 noise sequences (i.e., 1 in 8) contained targets ("depart" noises). Six sequences contained only one target, whereas the others contained two, with at least one of these coming in the second half of the sequence. Note that the different presentation rates of the approach/depart noise bursts and of the vowel sounds meant that their onsets and offsets were asynchronous, thereby facilitating perceptual segregation.

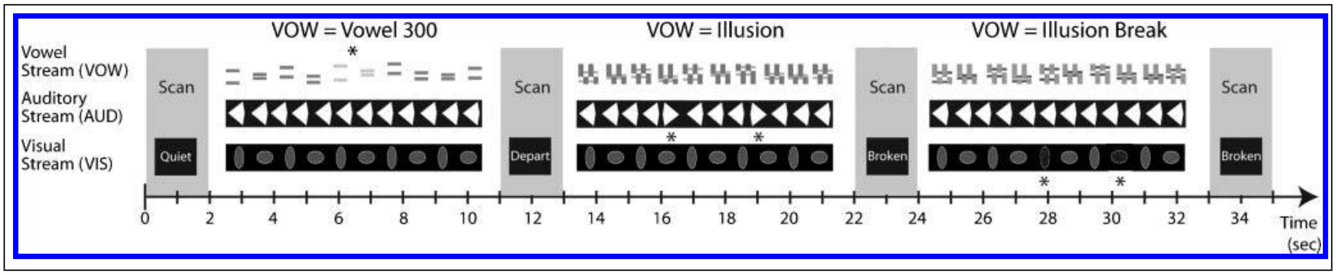

**Figure 2.** Schematic diagram of the imaging protocol. Stimuli were presented in the 9-sec period between successive 2-sec whole-brain volume acquisitions (TR = 11 sec). On every trial, three stimulus streams were presented concurrently; these were VOW (with four different possible sound types), AUD (auditory distracter task), and VIS (visual distracter task). During the scan (volume acquisition) at the end of a trial, a cue was displayed that directed the focus of attention on the next trial. The position of targets in each attended stream is indicated by asterisks. (In reality, targets were only present on 1 out of 8 trials, on average). Subjects pressed a key whenever targets in the attended stream were detected. Targets were never present in an unattended stream.

*Visual Distraction Stimuli (VIS)*

The visual stimuli consisted of a series of cross-hatched white ellipses presented on a black background; these stimuli were similar to those used in the study by Carlyon et al. (2003), in which they were shown to be sufficiently engaging to demonstrate an effect of attention on auditory streaming. When instructed to attend to this sequence, participants monitored the ellipses for occasional targets with broken lines instead of a cross-hatched fill pattern (see Trial 3 of Figure 2). The orientation of the ellipses alternated every 150 msec between 0 and 90 degrees. The total duration of the visual sequence was 8.4 sec. In 12 of 96 sequences (i.e., 1 in 8), one or two of the frames were replaced with broken-line targets. Six of these target sequences contained one target, and six contained two. Targets were placed randomly in the sequence to minimize predictability.

*Method of Combining Stimuli*

Each VOW sequence was combined with an AUD sequence at an equal sound intensity by first calculating the rms level of the VOW sequence and then adjusting the rms level of the AUD sequence to the same level. All sequences of one specific VOW type had very similar rms levels, however, they did vary slightly among VOW types. The overall presentation level of the combined VOW and AUD sequences varied between 62 and 65 dB. For two listeners, the presentation level was increased by another 5 dB to ensure a comfortable listening level. Each sound sequence was combined with a VIS sequence—thus stimuli of all three types (VOW, AUD, VIS) were present on all trials (Figure 2). Targets, when present, were only present in one sequence—the one to which listeners were instructed to attend.

Visual stimuli were displayed by a Christie video projector on an MRI-compatible screen at the head of the scanner bore, which was then projected onto a mirror situated approximately 90 mm from the eyes. The full-screen display subtended a visual angle of 16.7°. Auditory stimuli were presented diotically using Nordic Neuro Labs (NNL) electrostatic headphones which have a relatively flat frequency response up to at least 8 kHz, and which passively attenuate background (i.e., imaging system) noise by approximately 30 dB.

## Procedure

We used a sparse-imaging technique (Edmister, Talavage, Ledden, & Weisskopf, 1999; Hall et al., 1999), in which stimuli were presented in the silent intervals between successive scans. In each trial, a 300-msec silent pause was followed by an 8.4-sec stimulus sequence. This was followed by another 300-msec pause and then a 2-sec image acquisition period, for a total repeat time (TR) of 11 sec (Figure 2). A visual cue word for the next trial was presented in the 2-sec image acquisition period of the preceding trial. The cue told participants to which stimulus sequence they were expected to direct their attention in the upcoming trial. The cue word “quiet” instructed participants to listen to the VOW stream, and monitor for two consecutive attenuated stimuli. The cue word “depart” directed listeners’ attention to the auditory distracter stream (AUD) in order to detect departing noises among a sequence of approaching noises. Lastly, the cue word “broken” instructed listeners to direct their attention to the visual stream (VIS) and to look for ellipses with broken lines. Listeners were instructed to press a button (with their right index finger) each time they perceived a target. In the case of vowel targets, participants were instructed to press the button only once in response to perceiving two consecutive attenuated vowels. For noise and visual stimuli, participants were asked to indicate the occurrence of every target in the sequence. Only one out of eight (12.5%) of the attended sequences contained targets, and targets were never present in unattended sequences.

Twenty-four trials of each of 12 stimulus conditions (i.e., the four VOW stimulus types crossed with the three attentional domains) plus 24 silent trials were presented in six blocks of 52 trials each, with four trials of each condition in each block. Targets were present in six trials in each block, two for each of the three different attentional domains. During the four rest trials (cued by the word “rest”) in each block, participants were instructed to relax their eyes and hands while keeping still, and to wait for the next trial.

We used a dynamic stochastic design (Henson & Penny, 2005) in which the probability of occurrence of a particular attention condition varied in a sinusoidal fashion over time. As a consequence, attention could be focused on one particular domain (VOW, AUD, or VIS), for between one and six consecutive trials. Clusters of two and three trials of the same attention condition were most common, followed by single trials. Clusters of four, five, and six trials of the same attention condition were rare. The number of each type of trial cluster was identical across attention conditions for all participants; only the order in which they were presented varied because the six blocks of trials were presented in counterbalanced order over subjects (every order was presented to two participants).

Before the scanning session, each participant received a 15-min practice session in which he or she was familiarized with all types of stimuli and all types of target and nontarget sequences. Each participant received training with the experimental target-detection procedure (including trials from all 12 different conditions) for a minimum of 30 trials and until they felt comfortable with the task.

Imaging data were acquired using a Magnetom Trio (Siemens, Munich, Germany) 3-Tesla MRI system with a head gradient coil. Three hundred twenty-four echo-planar imaging volumes were acquired in six 10-min scanning runs for each of the 22 participants. Each volume consisted of 32 slices (slice order: interleaved; resolution:

3 × 3 × 3 mm; interslice gap: 25%; field view: 192 × 192 mm; matrix size: 64 × 64; TE: 30 msec; TR: 11 sec, TA: 2 sec). Acquisition was transverse oblique, angled to avoid the eyeballs and to cover as much of the brain as possible. When whole-brain coverage was not possible (i.e., for large males), the very top of the parietal lobe was omitted. At the beginning of each run, two dummy scans were acquired to allow for a stable level of magnetization before data collection commenced. A high-resolution T1-weighted structural anatomical image was also acquired on every subject.

### Analysis of fMRI Data

Data were processed and analyzed using Statistical Parametric Mapping (SPM5; Wellcome Trust Centre for Neuroimaging, London, UK; www.fil.ion.ucl.ac.uk/spm/). Each subject's functional time-series was first aligned to the first image of the first run, and the structural image was coregistered to the mean functional image. Spatial normalization of the functional images was accomplished by first normalizing the structural image using the combined segmentation/normalization procedures in SPM5 and the default ICBM 152 tissue probabilistic maps in MNI space as reference templates, and then applying these normalization parameters to functional images. Functional images were spatially smoothed using a Gaussian kernel with a full width at half maximum of 8 mm. Movement parameters were entered as separate regressors in the design matrix so that variance due to scan-to-scan movement could be modeled. Due to the long TR used in this sparse-imaging study, no correction for serial autocorrelation in the time series was necessary, nor was any high-pass filter applied.

A design matrix incorporating one column for each of 12 experimental conditions (4 vowel stimulus types × 3 attention domains) was constructed for each participant. Three additional columns were included to code correctly identified targets (hits), missed targets (misses), and misidentified nontargets (false alarms). Because targets were only presented in the attended sequence, these extra columns prevent any difference in the physical stimuli of the target sequences from contributing to our results. Realignment parameters and a dummy variable coding the six sessions were included as variables of no interest. Fixed-effects analyses were conducted on each listener's data. The parameter estimates, derived from the least-mean-squares fit of these models, were entered into second-level group analyses in which $t$ values were calculated for each voxel, treating intersubject variability as a random effect. For whole-brain analyses of main effects, we only report activation clusters with more than 100 voxels that pass a false discovery rate (FDR; Genovese, Lazar, & Nichols, 2002), corrected threshold of $p < .05$. Where exceptions occur, it is explicitly noted.

## RESULTS

### Target Detection

Participants had little difficulty detecting the rare targets in the stimulus blocks. The values of $d'$ were high overall (VOW = 3.08, AUD = 3.11, VIS = 3.30), and did not differ among the three attention domains [$F(2, 42) = 2.35$, $p = .11$].

### Imaging Results

*Replication of Heinrich et al.'s (2008) Findings*

The current study is a replication and extension of Heinrich et al. (2008) with a new group of participants. In that first study, we showed that both intact and illusory vowels activated a common "speech-sensitive region" in temporal cortex, bilaterally (Figure 1B). Additionally, we observed that a region of auditory cortex, bilaterally, was sensitive to the number of sound onsets in stimuli, exhibiting greater activity when more onsets were present (Figure 1C). We first attempted to replicate these earlier observations, before examining whether activity in vowel-sensitive regions and sound-onset sensitive regions is modulated by attentional state. To identify vowel-sensitive and onset-sensitive regions, we averaged over all three attention conditions in the current study.

*Activation in response to intact vowels.* The contrast between intact vowels and illusion break stimuli ([Vowel 75 + Vowel 300] − 2 * illusion break) revealed activation in the left and right MTG (Figure 3A and Table 1). The Euclidean distances between the peaks in the original study (Heinrich et al., 2008) [LH: −68 −32 4; RH: 54 −36 4] and in this replication were 8 mm in the left and 17 mm in the right hemisphere.

*Activation in response to illusory vowels.* The contrast (illusion–illusion break) revealed activity across a broad region of the STG bilaterally, as well as in the right middle frontal gyrus, supplementary motor area (SMA), and inferior parietal lobule (Figure 3B and Table 1).

The closest illusory-vowel peak to the vowel-sensitive areas observed in the present study was within 13 mm in the left hemisphere and 8 mm in the right hemisphere. This comparison is nonorthogonal because both contrasts share one condition (illusion break), but, reassuringly, illusory-vowel activity in this study is also close to vowel-sensitive areas observed in the independent dataset of Heinrich et al. (2008). In the left hemisphere, one of the peaks [−64 −36 14] was only 11 mm from the peak for intact vowels in the previous study, whereas in the right hemisphere, the closest peak [52, −24, 8] was 13 mm away. We have therefore successfully replicated the results of the previous study (Heinrich et al., 2008) and have observed activation in response to illusory vowels bilaterally in vowel-sensitive regions.

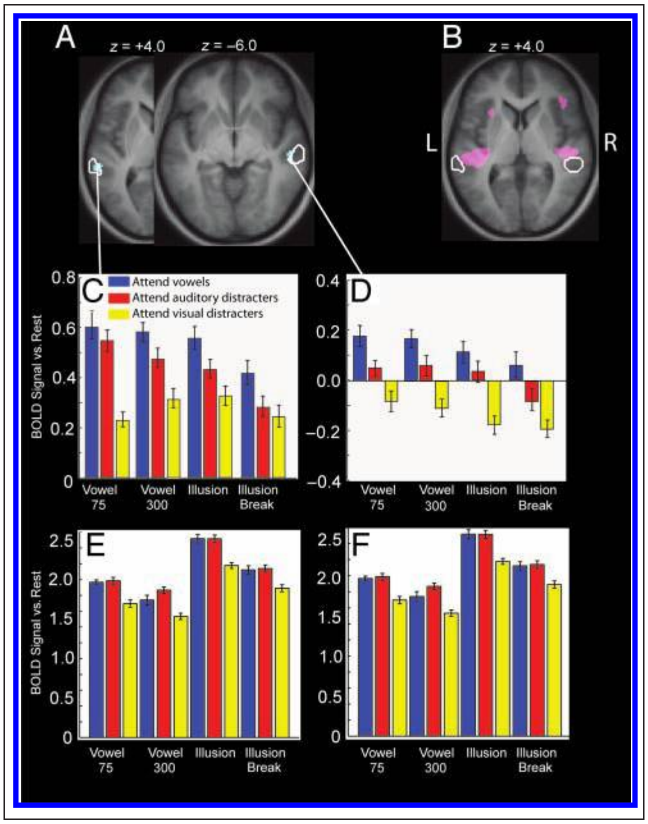

**Figure 3.** Areas showing significant activation in response to intact (A; in cyan) and illusory (B; in magenta) vowels in the current study. The white contours depict the outline of the activation to intact vowels in Heinrich et al. (2008) (thresholded at $p$ < .05, FDR). (A) The current study replicates the activation in the STG and the MTG observed in Heinrich et al. for intact vowels. (B) The observed activation to illusory vowels is slightly more anterior, medial, and superior, compared to the activation for intact vowels observed by Heinrich et al. (C) Parameter estimates for the 12 conditions from the peak voxel from the intact vowels–illusion break contrast in the left hemisphere: stereotaxic coordinates (−60, −32, 2). (D) Parameter estimates for the 12 conditions from the peak voxel from the intact vowels–illusion break contrast in the right hemisphere: stereotaxic coordinates (54, −22, −6). (E) Parameter estimates for the 12 conditions from the peak voxel from the illusion–illusion break contrast in the left hemisphere: stereotaxic coordinates (−44, −32, 8). (F) Parameter estimates for the 12 conditions from the peak voxel from the illusion–illusion break contrast in the right hemisphere: stereotaxic coordinates (52, −24, −8).

*Response to sound onsets*. We used two different contrasts to examine sensitivity to the number of sound onsets. In the (Vowel 75–Vowel 300) contrast, there are four physical sound onsets in the Vowel 75 condition versus one in the Vowel 300 condition, but vowels are perceived in both. In the (illusion break–illusion) contrast, the number of physical sound onsets is the same in both conditions, but stimuli are perceived as nonvowels in the illusion break condition (and the onsets are more salient), whereas they are perceived as more speech-like in the illusion condition (and onsets are less salient).

In both contrasts, the current study replicated the previous findings ([LH: −50 −16 6], [RH: 58 −14 2]) (Heinrich et al., 2008) of peak activation in auditory regions bilaterally (Figure 4 and Table 2). These peaks are within 12 mm of the peaks observed in the homologous contrasts in the previous study.

### *Effect of Attentional State on Patterns of Activity*

Having first replicated the results of Heinrich et al. (2008) regarding sensitivity to intact and illusory vowels, and to sound onsets, we now assess the effects of attentional state, and examine whether attentional state modulates activity in vowel and onset-sensitive regions. Accordingly, a within-subjects ANOVA with two factors (4 levels of sound type: Vowel 75, Vowel 300, illusion, illusion break; 3 levels of attention condition: VOW, AUD, VIS) was set up in SPM5 (Henson & Penny, 2005). This permits us to examine the main effect of attention condition as well as the interaction between attention condition and sound type. The overall main effect of sound type is not considered because the section on Replication of Heinrich et al.'s (2008) Findings already covered the interesting constituent effects.

*Main effect of attention*. A main effect of attention in a brain region would indicate that shifting attention among VOW, AUD, and VIS stimuli affected the amount of activation in that region. As noted above, all stimulus types were present on every trial, hence, differences in activation cannot be attributed to differences in sensory stimulation.

The SPM–ANOVA revealed strong effects of attentional state in a number of areas, including occipital areas, SMA, and temporal lobe, all bilaterally, and in the right inferior parietal lobule, left posterior angular gyrus, and left superior frontal gyrus (Figure 5 and Table 3). To identify the simple effects underlying the main effect in each region, we extracted the mean activation values for the main voxel coordinates (in bold font in Table 3) for each of the three attention conditions (compared to rest) in each subject, and conducted pairwise comparisons (Sidak-corrected). In the right STG and right inferior parietal lobule, attention to VOW stimuli led to more activation than attention to either AUD or to VIS stimuli. In the bilateral STG and left SMA, attention to either auditory stimulus type (VOW or AUD) led to more activation than attention to visual stimuli (VIS). In the right middle occipital gyrus, attention to VIS stimuli led to more activation than attention to VOW or AUD stimuli, and this was also true of a left superior frontal region. Finally, in the left inferior parietal lobule, attention to VOW stimuli elicited less activity than attention to VIS or AUD stimuli. We also used SPM to calculate the corresponding simple effects, which we used to mask the main effect (see Figure 5). The effects are consistent with the pairwise comparisons on peak voxels. In sum, attentional state significantly modulated patterns of brain activity, but did it do so in a way that was different for intact vowels, illusory vowels, and illusion break stimuli?

*Interaction between sound type and attention condition*. An analysis of the interaction between the attention and

**Table 1.** Coordinates and Statistics for Activation Peaks Resulting from the Contrasts Intact Vowels–Illusion Break ([Vowel 75 + Vowel 300] − 2 * Illusion Break), and Illusion–Illusion Break, Collapsed across All Attention Conditions

| *Contrast* | *No. of Voxels* | *T* | *p(FDR)* | *Coordinates (x y z)* | *Area* |
|---|---|---|---|---|---|
| Intact vowels–illusion break | **83** | **7.35** | **.015** | **−60 −32 2** | **L MTG** |
| | **29** | **5.89** | **.024** | **54 −22 −6** | **R MTG** |
| Illusion–illusion break | **1487** | **9.78** | **<.001** | **−44 −32 8** | **L Planum temporale** |
| | | 7.65 | <.001 | −32 −32 16 | L HG |
| | | 7.54 | <.001 | | |
| | | | | −64 −36 14 | L Posterior STG |
| | **871** | **8.37** | **<.001** | **52 −24 8** | **R STG** |
| | | 5.65 | .002 | 34 −26 14 | R HG |
| | | 5.45 | .002 | 56 −14 −4 | R STS |
| | **808** | **6.25** | **.001** | **52 26 34** | **R Middle frontal gyrus** |
| | | 5.10 | .004 | 44 4 50 | R Precentral gyrus |
| | | 5.02 | .005 | 42 10 36 | R Inferior frontal gyrus (pars opercularis) |
| | **335** | **5.79** | **.002** | **2 10 60** | **R SMA** |
| | | 5.49 | .002 | 6 16 56 | R SMA |
| | | 4.62 | .009 | 6 32 46 | R Superior frontal gyrus |
| | **479** | **5.64** | **.002** | **38 −52 38** | **R Inferior parietal lobule** |

We report voxels that are significant at $p < .05$ (using the FDR correction for multiple comparisons), and within clusters of more than 100 contiguous voxels. The only exception is the intact vowel–illusion break contrasts, which showed a very concise activation pattern. HG = Heschl's gyrus; L = left; MTG = middle temporal gyrus; R = right; SMA = supplementary motor area; STG = superior temporal gyrus; STS = superior temporal sulcus.

sound type factors highlights how attention condition modulates activation for intact vowels, illusory vowels, and illusion break stimuli, and is the heart of the current study. We start by investigating the overall interaction (4 sound types × 3 attention conditions), and then, in subsequent sections, by investigating attentional modulation on activation revealed by more targeted, hypothesis-relevant, contrasts.

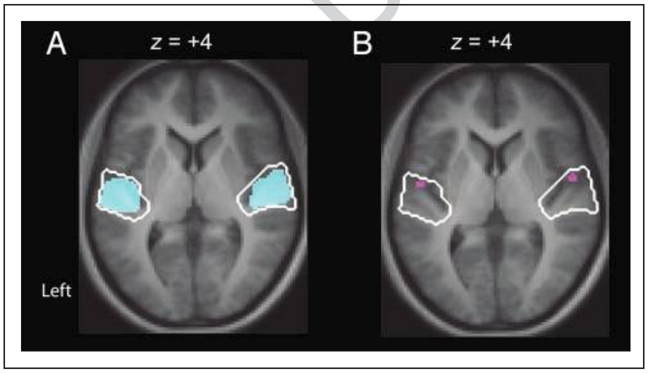

**Figure 4.** The activation to physical (A; in cyan) and illusory (B; in magenta) sound onsets replicates the previous findings (Heinrich et al., 2008) very closely and reveals activation in the STG close to Heschl's gyrus, bilaterally. The white contours depict the outline of the activation for physical sound onsets (Vowel 75–Vowel 300) in Heinrich et al. (2008) (thresholded at $p < .05$, FDR).

In the whole-brain analysis, we observed a significant interaction between attention condition and sound type in three clusters, one cluster of 70 voxels, with the chief peak in the left precentral gyrus, one cluster of 8 voxels in the left supramarginal gyrus, and one cluster of 9 voxels in the right SMA (Table 4), well away from temporal cortex that is the focus of interest in this study. Importantly, none of these regions were particularly sensitive to intact or illusory vowels, as determined by Heinrich et al. (2008), and as replicated here.

In order to characterize the observed interaction effects, we conducted Sidak-corrected post hoc pairwise comparisons on the signal extracted from the peak voxels of the three significant clusters (Figure 6C–E). The only consistent effect of attention across all three regions was that activity for the Vowel 300 [and illusion] stimuli was greater than for the other two sound types in the auditory distracter condition, but not in the other two attention conditions.

In a further attempt to observe an interaction between attention condition and sound type, we assessed the interaction in a sphere of interest of radius 20 mm around the vowel-sensitive region [LH: −60 −32 2; RH: 54 −22 −6], at a liberal threshold of $p < .001$, uncorrected for multiple comparisons. This liberal threshold (Figure 6A and B) revealed one small cluster of 30 voxels [$F(6, 231) = 4.59$] at the lateral edge of Heschl's gyrus [−62 −26 14], 11 and

**Table 2.** Coordinates and Statistics for Activation Peaks for the Two Contrasts Assessing Sensitivity to Sound Onsets

| *Contrast* | *No. of Voxels* | *T* | *p(FDR)* | *Coordinates (x y z)* | *Area* |
|---|---|---|---|---|---|
| Vowel 75–Vowel 300 | **1387** | **10.43** | **<.001** | **−58 −18 6** | **L STG** |
| | | 8.08 | <.001 | −36 −30 12 | L HG |
| | **1357** | **9.34** | **<.001** | **44 −24 10** | **R HG** |
| | | 9.23 | <.001 | 48 −16 8 | R HG |
| Illusion break–illusion | **154** | **5.97** | **.043** | **−52 −6 4** | **L STG** |
| | **209** | **6.32** | **.043** | **54 −2 0** | **R STG** |

We report voxels that are significant at $p < .05$ (using the FDR correction for multiple comparisons), and are within clusters of more than 100 contiguous voxels. HG = Heschl's gyrus; L = left; R = right; STG = superior temporal gyrus.

10 mm away from the vowel-sensitive [−60 −32 2] and illusion-sensitive [−64 −36 14] peaks of the current study, respectively. However, this region was clearly not selective for vowels, as activation was at least as large for the illusion break as for the intact vowel stimuli (Figure 6B). Furthermore, the interaction here did not reflect a greater effect of attention on responses to illusory than to intact vowels, but, instead, was due to attention modulating the response to the Vowel 300 stimuli differently to that of other stimuli.

Because we have a specific hypothesis about attentional modulation in vowel-sensitive regions, we now conduct analyses targeted at these regions. Accordingly, we first examine the effect of attention in the vowel-sensitive peak voxels (Figure 3A and Table 1) in the left and right hemispheres. Although activation in these vowel-sensitive voxels is strongest when attention is paid to intact vowels, and weakest when participants attend to the visual stimuli, these effects are similar for all four sound types, particularly in the left hemisphere (Figure 3C and D). Repeated measures ANOVAs (with 3 levels of attention and 4 levels of sound type) on signal extracted from the two peak voxels (right and left hemisphere) revealed main effects of attention [LH: $F(2, 42) = 13.96, p < .001$; RH: $F(2, 42) = 17.47, p < .001$] and sound type [LH: $F(3, 63) = 11.35, p < .001$; RH: $F(3, 63) = 8.20, p < .001$], but no interaction [$F(6, 126) = 1$] for either the right or left peak voxel.

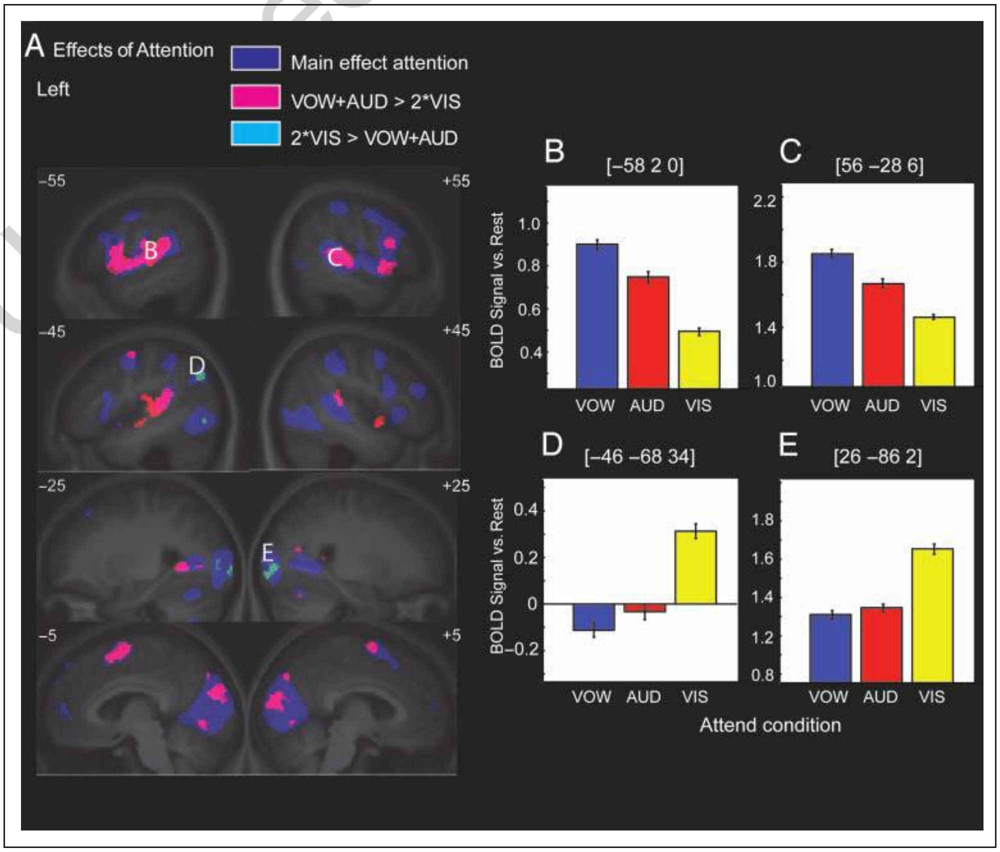

**Figure 5.** (A) Main effect of attention in the left and right hemispheres, shown at $p < .05$, FDR-corrected for multiple comparisons. Blue indicates where the $F$-statistic for the interaction contrast at the group level was statistically significant. Magenta and cyan regions explain the interaction in terms of simple effects: Magenta indicates regions where attention to auditory stimuli (VOW and AUD) resulted in greater activity than attention to visual stimuli. Cyan indicates regions where visual stimuli (VIS) resulted in greater activity than attention to auditory stimuli (VOW and AUD). B–E depict parameter estimates for the two strongest peak foci observed for the simple effects VOW + AUD > 2 * VIS (B, C) and 2 * VIS > VOW + AUD (D, E). Simple effects are collapsed over sound type.

**Table 3.** Coordinates and Statistics for the Foci of the Main Effect of Attention Condition

| *No. of Voxels* | *F* | *p(FDR)* | *Coordinates (x y z)* | *Area* | *Simple Effects* |
|---|---|---|---|---|---|
| **16,396** | **83.25** | **<.001** | **32 −92 2** | **R Middle occipital G** | **VOW = AUD**<br>**VIS >> AUD**<br>**VIS >> VOW** |
| | 77.58 | <.001 | 2 −86 10 | R Cuneus | |
| | 73.29 | <.001 | −28 −88 4 | L Middle occipital G | |
| **1329** | **58.44** | **<.001** | **−4 0 64** | **L SMA** | **VOW = AUD**<br>**VOW >> VIS**<br>**AUD >> VIS** |
| | 42.50 | <.001 | 6 2 64 | R SMA | |
| | 12.60 | <.001 | 6 22 48 | R SMA | |
| **8918** | **52.08** | **<.001** | **62 −24 2** | **R STG** | **VOW >> AUD**<br>**VOW >> VIS**<br>**AUD >> VIS** |
| | 49.49 | <.001 | 50 −36 4 | R MTG | |
| | 45.18 | <.001 | 54 0 46 | R Precentral G | |
| **7371** | **40.01** | **<.001** | **−66 −44 12** | **L Posterior STG** | **VOW = AUD**<br>**VOW >> VIS**<br>**AUD >> VIS** |
| | 36.34 | <.001 | −52 −6 48 | L Precentral G | |
| | 36.26 | <.001 | −62 −16 2 | L STG | |
| **1428** | **30.74** | **<.001** | **48 −36 50** | **R Inferior parietal lobule** | **VOW >> AUD**<br>**VOW >> VIS**<br>**AUD = VIS** |
| | 8.18 | .002 | 32 −50 50 | R Superior parietal lobule | |
| **579** | **17.85** | **<.001** | **−44 −70 36** | **L Inferior parietal lobule** | **VOW << AUD**<br>**VOW << VIS**<br>**AUD = VIS** |
| **1685** | **16.69** | **<.001** | **−12 42 50** | **L Superior frontal G** | **VOW = AUD**<br>**VOW << VIS**<br>**AUD << VIS** |
| | 14.49 | <.001 | −10 58 36 | L Superior frontal G | |
| | 12.93 | <.001 | 0 56 8 | Superior frontal G (medial) | |

The last column shows the results of post hoc pairwise comparisons (Sidak-corrected) conducted at the peak voxel, which explain the main effect at that voxel in terms of simple effects. We report voxels that are significant at $p < .05$ (using the FDR correction for multiple comparisons), and are within clusters of more than 100 contiguous voxels. G = gyrus; HG = Heschl's gyrus; L = left; MTG = middle temporal gyrus; R = right; SMA = supplementary motor area; STG = superior temporal gyrus.

=: not significantly different, $p > .05$; > : significantly different at $p < .05$; >>: significantly different at $p < .01$.

The pattern of results remained unchanged when the illusion break condition was excluded from the calculations. Moreover, the same pattern of results was obtained when sound types and attention conditions were compared in the peak voxels of the illusion−illusion break contrast (Figure 3E and F).

In the next section, we look for differential effects of attention specifically on activity elicited by illusory and intact vowels.

**Table 4.** Coordinates and Statistics for the Foci of the Interaction between Attention Condition and Sound Type

| | | | | | *Simple Effects* | |
|---|---|---|---|---|---|---|
| *No. of Voxels* | *F* | *p* | *Coordinates (x y z)* | *Area* | *Attention Condition* | *Sound Type* |
| 70 | 6.52 | .043 | −40 −24 60 | L Precentral G | AUD | Vowel 300 > Break |
| | | | | | | Illusion > Break |
| | | | | | VIS | Break > Vowel 75 |
| 8 | 5.96 | .043 | −58 −20 40 | L Supramarginal G | AUD | Vowel 300 > Vowel75 |
| | | | | | | Vowel300 >> Break |
| | | | | | VIS | Break >> Vowel300 |
| 9 | 5.92 | .043 | 6 16 52 | R SMA | VOW | Illusion >> Vowel75 |
| | | | | | | Illusion >> Vowel300 |
| | | | | | | Illusion >> Break |
| | | | | | AUD | Vowel300 >> Break |

Due to reduced statistical power of interaction effects, we report voxel clusters of any size that are significant at $p < .05$ (using the FDR correction for multiple comparisons). The last two columns present the results of post hoc pairwise comparisons (Sidak) that explain the interaction in that voxel in terms of simple effects. G = gyrus; SMA = supplementary motor area.

>: significantly different at $p < .05$; >>: significantly different at $p < .01$.

### *Testing for Effects of Attention on Specific Contrasts*

To this point, we have been assessing interaction effects evident across all sound types. However, we have specific hypotheses regarding particular sound types. In this section, we examine attentional effects on activity revealed by specific contrasts: (1) intact vowels–illusion break, (2) illusion–illusion break, and (3) physical (Vowel 75–Vowel 300) and perceived (illusion break–illusion) sound onsets. We computed one-way repeated measures ANOVAs in SPM5 with three levels of attention (VOW, AUD, VIS) for each of the contrasts of interest, and examined activation: (1) across the whole brain, and (2) in spherical regions of interest of 20-mm radius around vowel-sensitive peak voxels revealed in the current study (Figure 3A and Table 1).

*Effect of attention on activity in response to intact and illusory vowels.* If the perception of illusory vowels depends on attention being focused on the illusion stimuli themselves, then we would expect activation in vowel-sensitive regions to be greater for the illusion–illusion break contrast in the VOW condition than in the AUD or VIS attention conditions. In contrast, the activation in such regions for intact vowels compared to illusion break should be less susceptible to attention. In fact, no effect of attention was observed for the illusion–illusion break contrast, either at a whole-brain corrected level of significance or at an uncorrected level within regions of interest (20-mm spheres) around vowel-sensitive foci. No effect of attention was observed for the intact vowels versus illusion break contrast at a whole-brain corrected level. Twenty-millimeter spherical regions of interest around vowel-sensitive foci did reveal an effect of attention (Table 5). However, activity in these regions was not greater for intact vowels when attention was on these stimuli, compared to when it was focused on either type of distracter; in fact, greater activity was observed when attention was directed to the auditory distracters (AUD) compared to when it was directed to either VOW or VIS stimuli.

*Effect of attention on the activation in response to sound onsets.* Repeated measures ANOVA on the two contrasts highlighting responses to sound onset (Vowel 75–Vowel 300 and illusion break–illusion) revealed no interaction (i.e., no attentional modulation of activity evoked by different sound types) at a whole-brain corrected level of significance. Similarly, when we searched within 20-mm spherical regions of interest around the peaks of the illusion break–illusion contrast, no significant effects were obtained. However, when we searched in similar regions of interest around the peaks of the Vowel 75−Vowel 300 contrast, we observed an effect of attention in the left postcentral gyrus and the right STG. Sidak-corrected pairwise comparisons revealed that in both STG regions, the difference in activity between Vowel 75 stimuli and Vowel 300 stimuli was greater when attention was focused on them (VOW) than when it was focused on auditory distracters (Table 6). Attention focused on visual distracters led to activation in between VOW and AUD conditions.

## DISCUSSION

### Summary of Present Results and Comparison to Previous Research

We replicated previous results (Heinrich et al., 2008) and showed that stimuli perceived as vowels produced greater

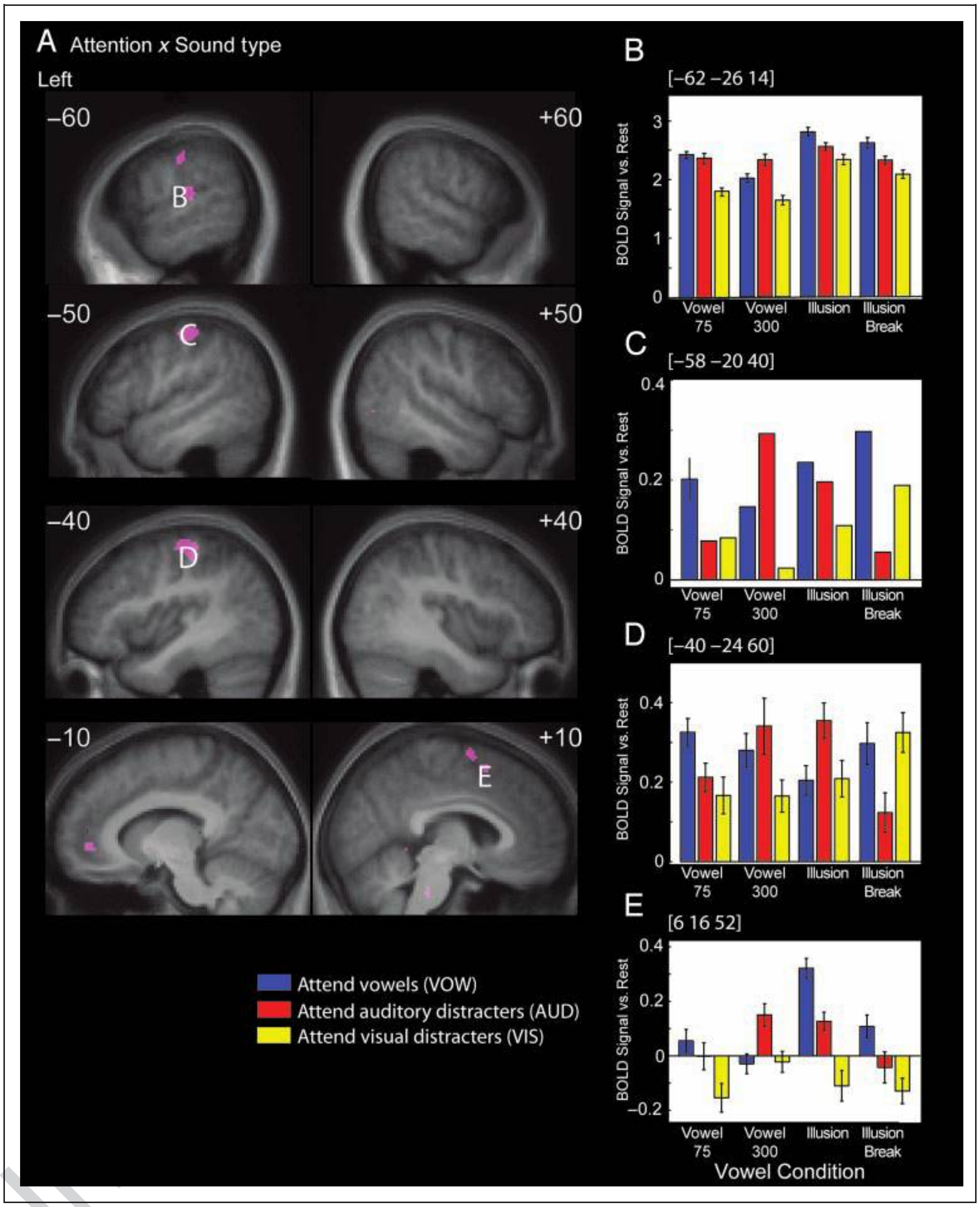

**Figure 6.** Interaction between attention condition and sound type. (A) Activation is shown at a liberal threshold of $p < .001$, uncorrected for multiple comparisons. At the more conservative FDR-corrected threshold of $p < .05$, three clusters are significant (C, D, and E). The parameter estimates across the 12 conditions for the peak voxels of these three clusters are shown in C–E. Only one small region in the vicinity of the vowel-sensitive region is not significant at the conservative threshold, but is significant at the liberal threshold within a 20-mm search volume around the peak vowel-sensitive voxel (B). Parameter estimates across the 12 conditions for this voxel are shown in B.

**Table 5.** Coordinates and Statistics for the Foci of the Effect of Attention Condition on Intact Vowels ([Vowel 75 + Vowel 300] − 2 * Illusion Break), within a Spherical Volume (Radius 20 mm) Centered on Vowel-sensitive Peak Voxels from the Present Study

| *No. of Voxels* | *F* | *p(Uncorr)* | *Coordinates (x y z)* | *Area* | *Simple Effects* |
|---|---|---|---|---|---|
| 5 | 8.70 | .001 | −60 −24 8 | L STG | AUD > VOW |
| | | | | | AUD >> VIS |
| | | | | | VOW = VIS |
| 49 | 10.61 | <.001 | 60 −22 10 | R STG | AUD > VOW |
| | | | | | AUD >> VIS |
| | | | | | VOW = VIS |

The last column shows the results of post hoc pairwise comparisons among the attention conditions for activity in that voxel (Sidak-corrected), detailing direction and strength of difference. L = left; R = right; STG = superior temporal gyrus.

=: not significantly different, $p > .05$; >: significantly different at $p < .05$; >>: significantly different at $p < .01$.

**Table 6.** Coordinates and Statistics for the Foci of the Effect of Attention Condition on the Onset Contrast Vowel 75–Vowel 300, within a Spherical Volume (Radius 20 mm) Centered on Onset-sensitive Peak Voxels of the Current Study

| *No. of Voxels* | *F* | *p(Uncorr)* | *Coordinates (x y z)* | *Area* | *Simple Effects* |
|---|---|---|---|---|---|
| 5 | 9.58 | <.001 | −66 −18 20 | L Postcentral gyrus | VOW = VIS<br>VOW >> AUD<br>VIS >> AUD |
| 18 | 9.49 | <.001 | 56 −34 14 | R Posterior STG | VOW > VIS<br>VOW >> AUD<br>VIS = AUD |

All values are significant at $p < .001$, uncorrected for multiple comparisons. The last column shows the results of post hoc pairwise comparisons among the attention conditions for activity in that voxel (Sidak-corrected), detailing direction and strength of difference. L = left; R = right; STG = superior temporal gyrus.

=: not significantly different, $p > .05$; >: significantly different at $p < .05$; >>: significantly different at $p < .01$.

activation in the left and right STS/MTG than stimuli not perceived as vowels. This was true not only for intact vowels but also for sounds that were not physically complete and depended on the continuity illusion to be perceived as vowels. We also replicated the finding that sounds with more perceptible onsets activated auditory regions more than sounds with fewer onsets. These findings provide the background to further contrasts that assess whether attention plays a specific role in the perception critical for auditory scene analysis such as perception of sound onsets and illusory continuity.

We found strong effects of attention in many areas across the brain, particularly in auditory and visual sensory areas. This is in accord with a number of imaging studies (e.g., Sabri et al., 2008; Wong, Uppunda, Parrish, & Dhar, 2008; Johnson & Zatorre, 2005; Petkov et al., 2004; O'Leary et al., 1997). In the left and right STG, near Heschl's gyrus (the macroanatomical landmark for primary auditory cortex), attention to the auditory stimuli (either the auditory distracters or the intact vowels, illusion and illusion break stimuli) yielded greater activity than attention to the visual distracters. In contrast, in occipital regions, attention to visual distracters yielded greater activity than did attention to either auditory sequence. These main effects confirm that our attention manipulation produced robust effects on neural activity, validating our further analyses to assess stimulus-specific effects of attention in auditory perception.

Despite the robust and predictable effects of attentional state, responses in vowel-sensitive areas, although modulated by attention, were affected similarly for illusory as for intact vowels. Some interactions between sound type and attentional state were observed, but these were not in regions sensitive specifically to vowel stimuli, nor were they of the form that would be expected if the continuity illusion depended on attention. We therefore conclude that the continuity illusion required to perceive alternating formants and noise as vowels does not depend on allocation of attention to the auditory stimulus. We will explore the implications of this finding for auditory scene analysis in a later section.

Attentional state, however, did influence activity reflecting number of sound onsets in the left postcentral gyrus and right STG. In the right STG, sensitivity to the number of sound onsets was greater when attention was focused on VOW stimuli than when it was focused on a distracter stimulus (AUD or VIS). In the left postcentral gyrus, sensitivity to the number of sound onsets was higher in VOW and VIS attention conditions than in AUD. These results replicate and extend previous findings by Rinne et al. (2005) and Mayer, Franco, Canive, and Harrington (2009) in auditory cortices. For instance, Rinne et al. presented harmonic sound complexes at rates between 0.5 and 4 Hz and asked listeners to focus their attention either on the sounds or on a visual distracter. Limiting their examination to bilateral superior temporal cortex, they showed an interaction between the presentation rate and attention, such that higher presentation rates led to more activation particularly when attention was focused on the sound onset stimuli themselves. Our study replicated the main effect of presentation rate and the interaction between presentation rate and attentional focus for the right hemisphere: In the left hemisphere, however, the onset rate effect was equally strong whether attention was focused on the sound onset stimuli or on visual distracters. Mayer et al. (2009), who showed similar results, offered an intriguing explanation for why attention to visual distracters may lead to activation in auditory cortices: They argued for an obligatory recruitment of auditory networks in tasks with a significant temporal component, regardless of their modality. In addition, they suggested that attending to visual stimuli while ignoring auditory information is effortful and recruits additional neuronal resources.

In our work, we included two conditions in which participants directed their attention away from the vowel stimuli and toward auditory or visual distracters. In comparing these two conditions, we found that directing attention to the auditory distracters (AUD) minimized differences in activation based on the number of sound onsets in both hemispheres. This may have happened because in the AUD condition, the distracter is also a multiple-onset stimulus. If activation to the attended identical multiple-onset stimulus is stronger than the activation to the unattended Vowel 75 and Vowel 300 stimuli, then one would not expect to see sound onset differences.

### Attention, the Continuity Illusion, and Auditory Scene Analysis

The auditory system is often faced with the task of processing complex sounds, such as speech, in the presence of one or more interfering sources. It must therefore deal with instances where portions of the target speech are inaudible, correctly assign individual frequency components to the correct source, and track the target source over time. The processes by which the brain solves this problem have been collectively termed "Auditory Scene Analysis (ASA)" (Bregman, 1990). Early research into ASA focused on, and successfully identified, the stimulus parameters governing phenomena such as the auditory grouping of simultaneous sounds (Darwin & Carlyon, 1995), auditory streaming (Moore & Gockel, 2002), and the continuity illusion (Warren, 1999). More recently, researchers have started to investigate the neural bases of ASA, and how it interacts with other, more cognitive, functions such as attention. Much of this effort has been directed toward the study of auditory streaming, with neural correlates in humans being identified both in auditory cortex and the parietal lobe using fMRI (Wilson, Melcher, Micheyl, Gutschalk, & Oxenham, 2007; Cusack, 2005), and electrophysiologically using EEG (Sussman, Ritter, & Vaughan, 1999). At the same time, a number of behavioral studies have shown that the temporal course of auditory streaming can be dramatically altered by a demanding auditory, visual, or cognitive task (Cusack, Deeks, Aikman, & Carlyon, 2004; Carlyon et al., 2003; Carlyon, Cusack, Foxton, & Robertson, 2001).

The present study adds to existing evidence, from both animal and human studies, that correlates of the continuity illusion can be observed in periauditory areas (Heinrich et al., 2008; Petkov, O'Connor, & Sutter, 2007). More importantly, it shows that, in contrast to the build-up of auditory streaming, the illusion does not depend strongly on the direction of attention, and that its neural correlates can be observed even when the subject is performing a demanding task in either the auditory or visual modality. Whereas top–down neural processes have often been inferred during some forms of continuity illusion (for instance, in the phoneme restoration effect; see Shahin et al., 2009; Samuel, 1981), our results indicate that the perception of illusory vowels in alternating formant/noise stimuli is independent of attention, and thus, probably occurs in the absence of frontally mediated top–down processes. The robustness of the neural signature of the continuity illusion to attentional modulation strongly suggests that bottom–up mechanisms alone are sufficient to account for it, at least for the vowel stimuli used here.

Another aspect of ASA that may not depend on attention is the processing of mistuning. Alain and colleagues have presented EEG evidence that a negative deflection produced by mistuning one component from a harmonic complex is of similar amplitude when subjects are required to detect that mistuning and when they are instructed to ignore it and to watch a silent movie (Alain, Arnott, & Picton, 2001). When combined with such previous research, our results begin to map out the relationships among different aspects of ASA and how they are mediated by cognitive processes. It will be interesting to extend this method to study the influences of attention on the use of other cues important for perceptual organization, such as onset asynchrony and the binaural precedence effect. A more complete understanding of the ways in which attentional state influences perceptual organization is critical if we are to develop an account of the brain's ability to process sounds in challenging environments.

Reprint requests should be sent to Antje Heinrich, Department of Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK, or via e-mail: ah540@cam.ac.uk.

## REFERENCES

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1072–1089.

Beauvois, M. W., & Meddis, R. (1991). A computer model of auditory stream segregation. *Quarterly Journal of Experimental Psychology, 43A,* 517–542.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 115–127.

Carlyon, R. P., Deeks, J., Norris, D., & Butterfield, S. (2002). The continuity illusion and vowel identification. *Acta Acustica United with Acustica, 88,* 408–415.

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32,* 1393–1402.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience, 17,* 641–651.

Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 643–656.

Darwin, C. J., & Carlyon, R. P. (1995). Auditory grouping. In B. C. J. Moore (Ed.), *Hearing* (pp. 387–424). Orlando, FL: Academic Press.

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

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature, 387,* 808–810.

Edmister, W., Talavage, T., Ledden, P., & Weisskopf, R. (1999). Improved auditory cortex imaging using clustered volume acquisition. *Human Brain Mapping, 7,* 89–97.

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

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

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 20,* 1737–1752.

Henson, R. N. A., & Penny, W. D. (2005). *ANOVAs and SPM*. Technical report. Wellcome Department of Imaging Neuroscience.

Howard-Jones, P. A., & Rosen, S. (1993). Uncomodulated glimpsing in checkerboard noise. *Journal of the Acoustical Society of America, 93,* 2915–2922.

Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M., & Niemi, J. (2003). The effects of attention on speech perception: An fMRI study. *Brain and Language, 85,* 37–48.

Husain, F. T., Lozito, T. P., Ulloa, A., & Horwitz, B. (2005). Investigating the neural basis of the auditory continuity illusion. *Journal of Cognitive Neuroscience, 17,* 1275–1292.

Jäncke, L., Mirzazade, S., & Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266,* 125–128.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex, 15,* 1605–1620.

Mayer, A. R., Franco, A. R., Canive, J., & Harrington, D. L. (2009). The effects of stimulus modality and frequency of stimulus presentation on cross-modal distraction. *Cerebral Cortex, 19,* 993–1007.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10,* 363–369.

Micheyl, C., Carlyon, R. P., Shtyrov, Y., Hauk, O., Dodson, T., & Pullvermuller, F. (2003). The neurophysiological basis of the auditory continuity illusion: A mismatch negativity study. *Journal of Cognitive Neuroscience, 15,* 747–758.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88,* 320–333.

Mozolic, J. L., Joyner, D., Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Maldjian, J. A., et al. (2008). Cross-modal deactivations during modality-specific selective attention. *BMC Neurology, 8*.

O'Leary, D. S., Andreasen, N. C., Hurtig, R. R., Torres, I. J., Flashman, L. A., Kesler, M. L., et al. (1997). Auditory and visual attention assessed with PET. *Human Brain Mapping, 5,* 422–436.

Petkov, C., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attention modulation of human auditory cortex. *Nature Neuroscience, 7,* 658–663.

Petkov, C., O'Connor, K. N., & Sutter, M. L. (2007). Encoding of illusory continuity in primary auditory cortex. *Neuron, 54,* 153–165.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., et al. (1996). Auditory selective attention: An fMRI study. *Neuroimage, 4,* 159–173.

Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice-Hall.

Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jaaskelainen, I. P., Sams, M., et al. (2005). Modulation of auditory cortex activation by sound presentation rate and attention. *Human Brain Mapping, 26,* 94–99.

Rinne, T., Stecker, G. C., Kang, X., Yund, E. W., Herron, T. J., & Woods, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *NeuroReport, 18,* 1311–1314.

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage, 39,* 1444–1456.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage, 44,* 1133–1143.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related potentials. *Psychophysiology, 36,* 22–34.

Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge, UK: Cambridge University Press.

Wilson, E. C., Melcher, J. R., Micheyl, C., Gutschalk, A., & Oxenham, A. J. (2007). Cortical fMRI activation to sequences of tones alternating in frequency: Relationship to perceived rate and streaming. *Journal of Neurophysiology, 97,* 2230–2238.

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *Journal of Speech, Language, and Hearing Research, 51,* 1026–1041.