Illusory Vowels Resulting from Perceptual Continuity: A Functional Magnetic Resonance Imaging Study

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

Abstract

■ We used functional magnetic resonance imaging to study the neural processing of vowels whose perception depends on the continuity illusion. Participants heard sequences of two-formant vowels under a number of listening conditions. In the ''vowel conditions,'' both formants were always present simultaneously and the stimuli were perceived as speech-like. Contrasted with a range of nonspeech sounds, these vowels elicited activity in the posterior middle temporal gyrus (MTG) and superior temporal sulcus (STS). When the two formants alternated in time, the ''speech-likeness'' of the sounds was reduced. It could be partially restored by filling the silent gaps in each formant with bands of noise (the ''Illusion'' condition) because the noise induced an illusion of continuity in each formant region, causing the two formants to be *perceived* as simultaneous. However, this manipulation was only effective at low formant-to-noise ratios (FNRs). When the FNR was increased, the illusion broke down (the ''illusion-break'' condition). Activation in vowel-sensitive regions of the MTG was greater in the illusion than in the illusion-break condition, consistent with the perception of Illusion stimuli as vowels. Activity in Heschl's gyri, the approximate location of the primary auditory cortex, showed the opposite pattern, and may depend instead on the number of perceptual onsets in a sound. Our results demonstrate that speech-sensitive regions of the MTG are sensitive not to the physical characteristics of the stimulus but to the perception of the stimulus as speech, and also provide an anatomically distinct, objective physiological correlate of the continuity illusion in human listeners. ■

INTRODUCTION

Following one sound source (i.e., a particular talker), despite interference from other sources, is a feat we all accomplish every day. When the frequency components of sounds from different sources are interleaved across the frequency spectrum, the intruding sound may partially or completely mask the sound of interest for short periods. When this happens, if the duration of the masking is sufficiently brief, the brain can complete the partially masked sound so that it is heard as a coherent whole. This ''continuity illusion'' is an instance of the Gestalt principle of perceptual closure. It is typically investigated by briefly interrupting a sound, and filling the ensuing silent period with an ''inducing'' sound.

Over the last few decades, the continuity illusion has been extensively studied in behavioral experiments using a variety of stimuli, including pure and modulated tones, tone glides, and speech (Carlyon, Micheyl, Deeks, & Moore, 2002; Warren, Wrightson, & Puretz, 1988; Ciocca & Bregman, 1987; Powers & Wilcox, 1977; Houtgast, 1972; Warren, Obusek, & Ackroff, 1972; Vicario, 1960; Miller & Licklider, 1950). The results of all these studies have led to a firm understanding of the stimulus parameters that do and do not lead to a percept of illusory continuity. For example, an important prerequisite for the illusion to occur is that the physical characteristics of the inducer are such that, if the interrupted sound had remained on, then the inducer would have masked it. Furthermore, evidence that the illusory percept can either hinder or help performance in forced-choice tasks has shown that it is a genuine perceptual phenomenon, rather than simply reflecting participants' reports of what they think they ''should have'' heard (Carlyon, Micheyl, Deeks, & Moore, 2004; Carlyon, Deeks, Norris, & Butterfield, 2002; Plack & White, 2000) However, evidence concerning the neural basis of the illusion has only recently become available (Petkov, O'Connor, & Sutter, 2007; Micheyl et al., 2003; Sugita, 1997).

Petkov et al. (2007) measured the responses of macaque A1 neurons to steady tones, to noise bursts, and to interrupted tones where the silent gap could optionally be filled by a noise. In this latter case, they found that the response of about half the neurons to the noise was more similar to that elicited by an uninterrupted tone than it was to that produced by an isolated noise burst. Evidence that the continuity illusion can occur at an early stage of cortical auditory processing in humans comes from a study by Micheyl et al. (2003). Using pure-tone stimuli, they showed that the continuity illusion was at least partially complete at the stage

[1]MRC Cognition & Brain Sciences Unit, Cambridge, UK, [2]Queens University, Kingston, Ontario, Canada

of processing at which the mismatch negativity (MMN) potential is generated. In this case, the MMN occurs with a latency of under 200 msec, indicating that the illusion is generated within 200 msec of the onset of the inducing tone. They further tentatively located the source generator of the illusion somewhere in the auditory cortex, probably close to the primary auditory cortex (PAC).

Here, we search for a neural correlate of the illusion using functional magnetic resonance imaging (fMRI), and with a paradigm inspired by a behavioral study reported by Carlyon, Deeks, et al. (2002) and Carlyon, Micheyl, et al. (2002). They presented listeners with pairs of synthetic, steady-state formants, which, in one condition, alternated on and off every 100–200 msec (Figure 1B). In this condition, listeners perceived a sequence of alternating ''buzzes,'' and vowel identification was poor. However, when the silent gaps in each formant region were filled with a burst of noise (Figure 1C), the sounds were perceived as vowels and could be easily identified. The most likely explanation for these results is that, when the noise was a plausible masker, the formant segments in each frequency region were perceptually completed, changing the qualitative percept of the sound from a nonspeech, filtered harmonic complex to a speech-like vowel. Note that the performance advantage observed here probably occurred at a prephonetic level (where formants were being integrated into a vowel percept), rather than from the noise filling in gaps in (presumably) already preformed phonemes. Carlyon and colleagues concluded that the neural mechanisms responsible for the continuity illusion ''feed into'' those mechanisms that integrate formants across frequency, and which are necessary for the identification of vowels.

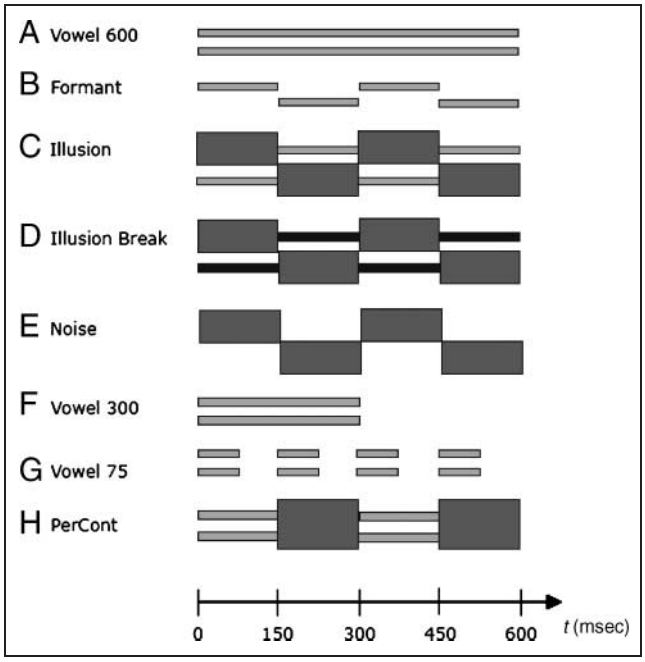

Figure 1. Schematic spectrogram of the eight sound conditions used in the study. f1 and f2 are each indicated depicted by narrow gray rectangles. Each formant was created by amplifying a small number of harmonics in the narrow range of the formant frequency. High- and low-pass noises are indicated by wider rectangles. For clarity, only one stimulus is depicted for each condition. For details of the conditions, see text.

One aim of the present study is to test this hypothesis and to examine whether the continuity illusion is present by the time information reaches cortical stages engaged in speech-specific processing, and contributes directly to these stages of processing. If so, then vowels that are perceived only by virtue of the continuity illusion should elicit activity at these speech-specific cortical stages. A number of recent studies have shown preferential activation in the posterior superior temporal gyrus (STG) for vowels and other speech sounds compared to nonspeech stimuli (e.g., Uppenkamp, Johnsrude, Norris, Marslen-Wilson, & Patterson, 2006; Jancke, Wustenberg, Scheich, & Heinze, 2002; Giraud & Price, 2001). An important challenge is to determine whether this activation is driven purely by the physical characteristics of speech, or whether it reflects the *perception* of speech and can be influenced by factors other than the representation of the stimulus in the peripheral auditory system. To distinguish between these alternatives, we compare activation in the two conditions illustrated in Figure 1C and D. Whereas in the illusion condition (Figure 1C) the formant-to-noise ratio (FNR) is sufficiently low to induce the continuity illusion, this is not the case in the ''illusion-break'' condition (Figure 1D), where the formants are too intense to be plausibly masked by the noise. Hence, if a brain region receives input from mechanisms responsible for the continuity illusion, *and* responds preferentially to speech-like stimuli, then we would predict more activation in the illusion condition than in the illusion-break condition. Crucially, we predict the opposite pattern of results in areas earlier in the auditory pathway (Heschl's gyrus [HG]), which do not distinguish between speech and nonspeech sounds. This prediction arises from evidence that HG activation depends on the number of sound onsets over time, which will be greater in the illusion-break than in the illusion condition. For example, in a preliminary report, Cusack, Carlyon, Johnsrude, and Epstein (2001) showed greater activation in HG for tones presented to the two ears when there was a small onset disparity between them, resulting in the percept of two sounds, than when they started synchronously, leading to the percept of a single sound. More recently, Harms, Guinan, Sigalovsky, and Melcher (2005) and Harms and Melcher (2002) have shown greater activation in HG for bursts of white noise presented with fast (8–10 Hz) onset rates, compared to slower rates, whereas the same pattern was not observed in the STG. Other studies have also shown relatively greater activation in HG to sounds with high onset rates, with a different pattern of results observed in regions closer to the STG (Herdener et al.,

2007; Hart, Palmer, & Hall, 2003). Another test of this prediction can be made by comparing activation for vowel stimuli that are turned on and off quickly (Condition ''Vowel 75''; Figure 1G) compared to more slowly (''Vowel 300''; Figure 1F).

METHODS

Participants

Nineteen adults between the ages of 19 and 40 years (mean = 26.9 years, *SD* = 6.9 years; 6 women) took part in the study. All participants were right-handed, fluent speakers of English, and reported no neurological disease or hearing loss.

The study was approved by the Addenbrooke Hospital's (Cambridge) Local Research Ethics Committee and written informed consent was obtained from all volunteers, who were each reimbursed £25.00 for their time.

Stimuli

The sounds were based on 600-msec harmonic complexes with energy at the first 100 harmonics of one of a set of fundamental frequencies (f0, see below), and were synthesized with a sampling rate of 44.1 kHz. Formants were generated by passing the harmonic complex through a second-order infinite impulse response (IIR) filter (Rabiner & Schafer, 1978) with the center frequency at the formant value and the 3-dB bandwidth of the filter fixed at 90 Hz. We chose two-formant vowels because they can evoke a vowel percept while allowing for a great deal of experimental control at the same time (Carlyon, Deeks, et al., 2002; Carlyon, Micheyl, et al., 2002; Akeroyd & Summerfield, 2000). The four vowels used in the study were: / ɑː / as in ''far,'' / ɛː / as in ''head,'' / ɔː / as in ''ford,'' and / ɜː / as in ''heard.'' The formant frequencies were loosely based on Peterson and Barney (1952) (/ ɑː / and / ɔː /), Ladefoged (1967) (/ ɛː /), and Wells (1962) (/ ɜː /). The vowels (see Table 1) were chosen from a larger set used in speech-rating pilot studies because they gave the largest difference in speech-likeness ratings when formants were played simultaneously (perceived as speech) and alternating (perceived as less speech-like). Because only four different vowels were used, we roved the f0 of each vowel to create five different exemplars of a particular vowel (see Table 1). Following Smith, Patterson, Turner, Kawahara, and Irino's (2005) observation that a proportionate increase of the formant and fundamental frequencies of a vowel can keep its identity stable but changes speaker identity, we increased f1 and f2 in proportion to the increase in f0, and fine-adjusted formant frequencies such that the ''color'' of the vowel was as similar as possible across all five tokens (Table 1).

The study included eight conditions whose stimuli differed with respect to whether formants were presented simultaneously or alternating, and how the spectral gaps in alternating-formant stimuli were filled. Sounds were concatenated into 8.4-sec sequences of 10 sounds. Each sequence only contained sounds from one condition. Sounds in a sequence were combined in such a way that neither the vowel identity nor f0 was identical in any two successive sounds. A variable ISI (mean = 240 msec, *SD* = 41.23 msec) was interposed between sounds; moreover, the sound sequence started and ended with a period of silence of variable duration (mean = 120 msec, *SD* = 18.86 msec). Twenty-four unique sound sequences were generated for each of the eight sound conditions. Schematic spectrograms of the eight stimulus conditions are presented in Figure 1. These conditions were:

(a) **Vowel 600:** Both formants were gated on and off together for 600 msec.
(b) **Formant:** The f1 and f2 alternated for a total of four times and a total duration of 600 msec starting with f1

Table 1. Vowel Identity (Vowel ID), Fundamental Frequency (f0), First (f1) and Second (f2) Formant Frequencies, and Low-pass (LP)/High-pass (HP) Cutoff Frequencies of the Complementary Noise Band of All Vowel Sounds Used in the Study

Vowel Identification	*f0*	*f1*	*f2*	*LP/HP Cutoff*
/ah/	115	730	1090	892
	122	750	1120	916
	137	780	1165	953
	145	810	1209	990
	150	830	1239	1014
/eh/	115	790	1697	1158
	122	810	1740	1187
	137	850	1826	1246
	145	870	1869	1275
	150	910	1955	1334
/or/	115	570	850	696
	122	590	880	720
	137	610	910	745
	145	650	970	794
	150	680	1014	830
/er/	115	560	1480	910
	122	580	1531	942
	137	600	1584	975
	145	610	1610	991
	150	630	1663	1024

and ending with f2. The switch between the two formants occurred every 150 msec.

(c) **Illusion:** Same as formant except that the silent gaps in each frequency region left by the missing complementary formant were filled with high-pass (complementing f1) or low-pass (complementing f2) filtered white noise. The filter cutoff, placed at the geometric mean between the two formant frequencies of each vowel, had a 96-dB/octave roll-off. Pilot testing showed that an FNR of −20 dB (when combining formants and noise) was sufficient to evoke the continuity illusion; listeners perceived the sound as a vowel embedded in noise bursts.

(d) **Illusion Break:** same as illusion, except that formant and noise were combined with an FNR of +15 dB. Because the FNR was higher than in the illusion condition, the sounds were not perceived as continuous vowels but rather as alternating formants and noise.

(e) **Noise:** Band-pass noise bursts of alternating high and low frequency: This condition was the same as the illusion and illusion-break conditions but without the formants (and thus with unfilled spectral gaps).

(f) **Vowel 300:** Both formants were gated on and off together for 300 msec followed by a 300-msec silence. This condition was intended as comparison for total sound duration in the formant condition.

(g) **Vowel 75:** Both formants were presented simultaneously for 75 msec followed by 75 msec of silence; this pattern was repeated four times. This condition has the same number of sound onsets as the formant, illusion-break, and noise conditions, but sounded more speech-like.

(h) **PerCont** (Perceived continuous): 150 msec of simultaneous formant presentation was followed by 150 msec of combined low- and high-pass noise presentation. The condition was created to provide a comparison for the illusion condition, as the perception of the vowel in a background noise is similar in both conditions, yet in the PerCont condition the illusion only contributes to the perception of continuity, not to the perception of the vowel itself.

Each segment of sound was gated on and off with a 5-msec raised cosine ramp. At the end of the generation process, all stimuli were scaled to the maximum and low-pass filtered below 1 kHz with a 12-dB-per-octave roll-off, as is typical for natural speech.

For those parts of the experiment performed outside of the scanner, stimuli were presented diotically over Sennheiser HD250 headphones. In the scanner sessions, stimuli were presented diotically using Etymotic ER-3 pneumatic tube phones with 4.5-m-long air-conduction tubes and insert ear pieces. To minimize the level of scanner noise, listeners also wore ear defenders. Because the tube headphones had a steep frequency roll-off above 800 and below 300 Hz, we used a digital filter to compensate for the transfer function and flatten the output of the headphones. However, because our filter was only able to compensate fully up to 5200 Hz, we also low-pass filtered all stimuli at 4 kHz (145 dB/oct roll-off) to make sure that no high-frequency artifacts were introduced during the compensation process. This low-pass filter (but not the digital compensation) was also applied to the stimuli presented outside of the scanner.

The overall levels of the stimuli (in dB SPL) for each condition, averaged across all vowels, are shown in the last column of Table 2. It can be seen that this level varied by a maximum of 6.1 dB across conditions. The levels of the ''formant'' and ''noise'' components in each condition are shown in the first two columns. Because, in the illusion and PerCont conditions, the noise was more intense than the vowel, and because the overall levels were similar across conditions, the formant levels in these conditions were substantially lower than in the ''vowel alone'' condition. It is also worth noting that one of our two main predictions is that activity on speech areas will be greater in the illusion than in the illusion-break condition. As Table 2 shows, the formant level in the illusion condition is actually *less* than that in the illusion-break condition, so, if our prediction is confirmed, this will be so despite lower formant energy in the condition showing greater activation. Our other prediction is that primary areas will show more activation in the illusion-break condition; as the last column of Table 2 shows, this would occur despite the overall rms level being slightly lower in the illusion-break than in the illusion condition.

Procedure

Each volunteer completed a number of tasks in the course of the study. The order of the tasks was identical

Table 2. Levels in dB SPL for the Various Conditions of the Experiment

	Formant	*Noise*	*Total*
Vowel 600	70.6		70.6
Formant	72.5		72.5
Illusion	48.4	68.4	68.4
Illusion break	71.3	56.3	71.4
Noise		68.5	68.5
Vowel 300	70.6		70.6
Vowel 75	70.6		70.6
PerCont	46.3	66.3	66.3

The level of the harmonic (''formant'') and noise components are given in the first two columns, with the total level being shown in the third column. In the imaging part of the experiment, these levels were reported to be too soft by two participants, and were therefore increased by 5 dB.

for all participants. The experimental session began with a behavioral rating task, followed by vowel identification training and training on the target-detection task used in the imaging session. Volunteers were then scanned and finally completed a vowel identification task. All tasks other than the target-detection task (the ''scanner'' task) were conducted in a single-walled sound-attenuated booth, where sounds were played through a PC and presented over HD 250 Linear II headphones at approximately 80 dB SPL.

Rating Task

The rating task required participants to listen to a subset of the sequences used in the later scanning experiment and to rate each sequence for its speech-likeness on a scale between 1 (*not speech-like at all*) and 7 (*very speech-like*). Before the rating task, participants listened to ''anchor sounds'' that were examples of particularly speech-like (recording of naturally spoken vowels) and non-speech-like sounds. The non-speech-like anchor sounds were ''musical rain'' stimuli used by Uppenkamp et al. (2006). These sounds were sets of four ''damped sinusoids'' in which the four carrier frequencies and the start times of the sinusoids within a cycle were randomized. The stimuli sound like a rapid splatter of overlapping tone pips with a rain-like quality (see Uppenkamp et al., 2006 for further details). The first four volunteers rated six sequences per condition; the remaining 15 participants rated 12 sequences per condition. The sequences were presented in random order. Participants responded to the presentation of each sequence by pressing one of seven buttons labeled ''1'' to ''7'' on the computer keyboard. No feedback was provided. The test took about 15 min to complete.

Vowel Identification Training

To ensure that they attached the correct label to each vowel, listeners completed a brief vowel identification training session. In this task, they listened to a 600-msec example of each of the 20 vowel sounds (4 different vowels generated with 5 different fundamental frequencies). Two seconds after the vowel was played, the correct label of the vowel was displayed on the computer screen. Subsequently, the vowel sound was played again while the label remained on the screen. No response was required.

Target-detection Training

During the imaging session, listeners were asked to perform a simple detection task to keep them alert and focused on the vowel and nonvowel stimuli. The task required participants to press a button with their right index finger each time they detected two consecutive tokens in a sequence that were softer than the sounds surrounding them (these stimuli were attenuated by 9 dB). Listeners were asked to press a button as soon as they detected the attenuated sounds.

Participants were trained on the detection task, outside of the scanner, to ensure that they could reliably detect the quieter stimuli. During the training session, participants were presented with one sample sequence per condition and listened to it twice: first with all 10 sounds at full intensity and then with two adjacent sounds attenuated by 9 dB.

Imaging Session

Imaging data were acquired using a Magnetom Trio (Siemens, Munich, Germany) 3-T MRI system with a head gradient coil. For 16 out of 19 participants, 224 echo-planar imaging (EPI) volumes were acquired over four 10-min scanning runs (56 volumes per run). For an additional three participants, only three scanning runs were conducted, and thus, only 168 volumes were collected. Each volume consisted of 32 slices (slice order: interleaved; slice thickness: 3 mm; resolution: 3 × 3 mm; interslice gap: 25%; field of view: 192 × 192 mm; matrix size: 64 × 64; echo time: 30 msec; acquisition time: 2 sec). Acquisition was transverse oblique, angled to avoid the eyeballs and to cover the whole brain except for, in a few cases, the top of the parietal lobe. The temporal lobe, the area in which activation was most likely to occur in the current study, was fully covered in all cases.

We used a sparse-imaging technique (Edmister, Talavage, Ledden, & Weisskoff, 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 sound sequence, a further 300 msec pause, and then a 2-sec data acquisition. The total repetition time of a trial was 11 sec. At the beginning of each run, two dummy scans were acquired to allow for a stable level of magnetization before data collection commenced. The remaining 54 trials per run included six sequences from each of eight sound conditions, plus six rest trials. Conditions were presented in pseudorandom order within each run, with not more than two successive sequences from the same condition. Sixteen unique orders of presentation were used; the presentation order for the first three volunteers was repeated for the last three participants. Six sound sequences in each run contained targets for the attenuation-detection task, hence, included two consecutive tokens attenuated by 9 dB. Participants were told at the beginning of the imaging session that they would hear the same sound sequences as in the behavioral rating experiment, and were asked to perform the target-detection task. For rest trials, listeners were asked to relax their hands and eyes while keeping still, and to wait for the experimental task to continue.

Vowel Identification Task

After the imaging session, volunteers participated in a short vowel identification task. They listened to all 160 sounds used in the study (five versions of each of four vowels in eight different conditions) and made a four-alternative forced-choice decision for each sound as to which vowel the sound represented. This task was included to obtain an additional measure of the differences among sounds in speech and nonspeech conditions.

Analysis of fMRI Data

Data were processed and analyzed using Statistical Parametric Mapping (SPM5; Wellcome Department of Cognitive Neurology, London, UK, www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included within-subject alignment of the blood oxygenation level-dependent (BOLD) time series to the first image of the first run, coregistration of the mean BOLD image with the structural image, and normalization of the structural image to the MNI average brain using the combined segmentation/normalization procedure in SPM5. Normalization parameters were then applied to BOLD images which were spatially smoothed using a Gaussian kernel with a full width at half maximum of 10 mm. Movement parameters were entered as separate regressors in the design matrix so that variance due to scan-to-scan movements could be modeled. Due to the long TR used in this sparse-imaging fMRI study, no correction for serial autocorrelation was necessary, nor was any high-pass filter applied to the fMRI data. The order of conditions was pseudo-randomly set within each scanning session with all transitions between conditions being approximately equally probable. This procedure is sufficient to ensure that low-frequency noise in the fMRI time series does not confound activation results.

A separate design matrix, that included eight columns for the sound conditions, was constructed for each participant. Three additional columns were included to code whether a sequence contained a target (quieter sound), whether the listener correctly identified it (hit), and when a listener misidentified a nontarget as target (false alarm). Realignment parameters and a dummy variable coding the four sessions (for 16 participants) or three sessions (for three participants) were included as covariates 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 variation as a random effect. For whole-brain analyses, we report peak voxels that pass both uncorrected ($p < .001$) and whole-brain corrected (false discovery rate [FDR], $p < .05$; Genovese, Lazar, & Nichols, 2002; Benjamini & Hochberg, 1995) thresholds in clusters with more than 10 contiguous voxels. By combining an uncorrected and corrected threshold, we limit the effect of the adaptive FDR threshold that can otherwise produce a statistical threshold more lenient than $p < .001$, uncorrected (if a large number of voxels are active). For masked contrasts we report peak voxels that survive FDR, $p < .05$, correction within the appropriate search volume. The peak voxels were localized using the anatomical automatic labeling (AAL) map of the MNI canonical brain (Tzourio-Mazoyer et al., 2002).

RESULTS

Behavioral Results

Rating Task

The results of the rating task are presented in Table 3 with conditions arranged in descending order of speech-likeness rating. Mean ratings, even in the ''vowel'' conditions, were below the maximum of 7, probably because the comparison stimulus for speech-like sounds was a sequence of naturally spoken vowels. In contrast, the vowels used in the experiment were all synthetic two-formant steady-state vowels that even in their complete form (both formants present at the same time) differed somewhat from natural speech.

The crucial comparison in the task is the speech-likeness of illusion and illusion-break conditions. A repeated measures analysis of variance (ANOVA) including all eight conditions (Vowel 600, Vowel 300, Vowel 75, PerCont, Illusion, Illusion Break, Formant, Noise) revealed a main effect of condition [$F(7, 126) = 115.31$, adjusted $p < .001$]. Subsequently, we compared conditions with adjacent vowel ratings by performing seven

Table 3. Mean and Standard Error of the Mean (*SEM*) for Speech-likeness Ratings of All Eight Experimental Conditions

	Speech-likeness Rating		
	Mean	*SE*	*t*
Vowel 600	5.89	0.13	
			0.257
Vowel 300	5.86	0.14	
			3.808*
PerCont	5.05	0.19	
			1.363
Vowel 75	4.61	0.17	
			2.513
Illusion	3.87	0.17	
			6.016*
Illusion Break	2.16	0.19	
			1.487
Formant	1.96	0.15	
			4.752*
Noise	1.20	0.13	

The *SEM* was adjusted to remove between-subject variance according to Loftus and Masson (1994). Asterisks represent significant differences between rating values, which were adjusted for multiple comparisons (7) using the multiplicative Bonferroni correction (Sidak, 1967) at $t(18) = 3.02$.

*$p < .05$.

t tests using a variant of the Bonferroni correction (Sidak, 1967) to adjust for multiple comparisons. This correction corresponds to a t value of 3.02 for a threshold of $p = .05$. The speech ratings for the Vowel 600 and Vowel 300 conditions did not differ statistically. Ratings were higher in the Vowel 300 condition than in PerCont, but did not differ between the PerCont and Vowel 75 conditions, and between the Vowel 75 and illusion conditions. Crucially, ratings were higher in the illusion than in the illusion-break condition. Illusion-break and formant conditions did not differ, but ratings in the noise condition were significantly lower. Thus, we were successful in our attempt to elicit the vowel illusion as evidenced by the fact that the speech ratings for the illusion condition did not differ from those for the intact vowel condition (Vowel 75), but were significantly higher than the illusion-break condition.

Attenuation Detection Task

Participants were able to correctly detect the attenuated sounds in a sequence, as illustrated by a d' value of 3.21 ($SD = .35$) for all participants averaged over all scanning runs.

Vowel Identification

The results of the vowel identification task, performed after the scanning session, are shown in Table 4, with each condition shown in order of descending performance. This ordering was almost identical to that obtained in the rating task (Table 3), but the results were somewhat more variable. A repeated measures ANOVA across the eight sound conditions yielded a main effect of condition [$F(7, 126) = 59.65$, $p < .001$]. Performing seven Bonferroni-adjusted t tests (Sidak, 1967) revealed that only one pair of adjacent test scores was significantly different: Vowel 75 and illusion. Hence, unlike the study of Carlyon, Deeks, et al. (2002) and Carlyon, Micheyl, et al. (2002), we were unable to demonstrate superior identification performance in the illusion than in the formant condition. Although there were several differences between the stimuli and the procedure used in the two studies, one potentially important difference is that the noise levels used here were higher (relative to the formants) than those employed by Carlyon et al. As they pointed out, the beneficial effects of adding noise may be counteracted by partial masking of each formant from the noise burst in the complementary frequency region, rendering it more difficult to accurately identify each formant frequency. This effect would have been stronger at the higher noise levels used here, but may not have affected the overall impression of ''speech-likeness.''

Table 4. Mean and *SEM* for Vowel Identification Performance of All Eight Experimental Conditions

Vowel Identification	*Mean*	*SE*	*t*
Vowel 600	0.74	0.02	
			0.00
PerCont	0.74	0.02	
			0.43
Vowel 300	0.73	0.02	
			1.68
Vowel 75	0.67	0.03	
			6.20*
Illusion	0.40	0.02	
			0.36
Formant	0.38	0.02	
			1.90
Illusion break	0.34	0.03	
			2.07
Noise	0.27	0.03	

The *SEM* was adjusted to remove between-subject variance according to Loftus and Masson (1994). Asterisks represent significant differences between rating values adjusted for multiple comparisions (7) using the multiplicative Bonferroni correction (Sidak, 1967) at $t(18) = 3.02$.

Imaging Results

Activation in Response to Sound Compared to Rest

Cortical activation in response to all eight conditions compared to silence yielded bilateral activation in two large clusters in the lateral STG; although these clusters include HG (the morphological marker of the PAC on the superior temporal plane), they were centered somewhat more inferiorly and laterally. Contrasts of each of the eight conditions with silence revealed robust bilateral temporal activation. This activation extended anterior and posterior to HG along the STG and the middle temporal gyrus (MTG).

Our two nonspeech sound conditions (formant and noise) together, compared to silence, revealed a significant activation in the STG, paracentral lobule, and precentral gyrus in the right hemisphere; and the MTG and postcentral gyrus in the left hemisphere (see Table 5 and Figure 2A). This contrast was used to define a generic, sound-responsive region of interest (ROI) for analyses in subsequent sections.

Activation in Response to Vowel Sounds

We contrasted the three conditions that could be most clearly characterized as only containing vowels (Vowel 600, Vowel 300, Vowel 75) with the two sound conditions that would not be expected to produce a speech percept (formant and noise). This speech contrast reveals activation in the posterior MTG bilaterally (see Table 5 and Figure 2B), and defines a speech ROI within which we look for significant activity in the Illusion versus Illusion-Break contrast.

Perception of Illusorily Continuous Vowels

To investigate the neural correlates of the subjective perception of speech, we compared the pattern of neural

Table 5. Peak Voxels Activated for Sound (Formant + Noise vs. Rest), Speech (Vowel 600 + Vowel 300 + Vowel 75 vs. Formant + Noise), and Illusion Compared to Illusion Break Stimuli (Masked with Speech Contrast Activation), as well as for the Activation Associated with Positive Speech-likeness Ratings over All Sound Conditions

	Coordinates					
Contrast	*x*	*y*	*z*	*t*	*No. of Voxels in Cluster*	*Area*
Sound ROI (Formant + Noise)	**64**	**−24**	**6**	**15.49**	**4756**	**R STG**
	50	−26	8	13.29		R STG
	46	−18	4	13.12		R HG
	−60	**−14**	**0**	**14.44**	**5361**	**L MTG**
	−40	−30	12	13.55		L Planum
	−54	−20	6	11.941		L lat HG
	8	**−28**	**64**	**7.85**	**371**	**R paracentral lobule**
	−8	−34	62	5.44		L paracentral lobule
	−12	−24	64	4.90		L paracentral lobule
	54	**0**	**38**	**4.99**	**208**	**R precentral G**
	52	−10	42	3.82		R precentral G
	34	**−20**	**54**	**4.94**	**131**	**R precentral G**
	−54	**−14**	**42**	**4.21**	**122**	**L postcentral G**
	−50	−10	50	3.92		L postcentral G
	2	**10**	**20**	**4.01**	**13**	**Corpus Callosum**
	−26	**28**	**0**	**4.01**	**19**	**L ant Insula**
	22	**26**	**8**	**3.89**	**13**	**Caudate**
Speech ROI Vowel − Sound: (Vowel 600 + Vowel 300 + Vowel 75) − (Formant + Noise)	**−68**	**−32**	**4**	**6.89**	**150**	**L post MTG**
	54	**−36**	**4**	**6.22**	**438**	**R STS**
Illusion − Illusion Break (using speech ROI)	**−68**	**−36**	**8**	**3.65**	**12**	**L post MTG**
	54	**−38**	**4**	**2.76**	**12**	**R STS**
Speech-likeness (pos. correlation)	**−68**	**−36**	**4**	**11.01**	**320**	**L MTG**
	58	**−32**	**0**	**5.57**	**422**	**R MTG**
	50	−34	4	5.45		R STS

For whole-brain analyses, we report peak voxels that are $p < .05$ (FDR) and $p < .001$, uncorrected within clusters of more than 10 contiguous voxels. For ROI analyses, we report peak voxels that are $p < .05$, FDR corrected.

R = right; L = left; ant = anterior; post = posterior; STG = superior temporal gyrus; MTG = middle temporal gyrus; STS = superior temporal sulcus; HG = Heschl's gyrus; G = gyrus; LB = lobule.

activation in illusion and illusion-break conditions within the areas responsive to vowel sounds (speech ROI; see Table 5). The contrast between them revealed significant foci of greater activation in the left posterior MTG and the right STS (−68 −36 8; 54 −38 4) (see Table 5 and Figure 2D). When the same contrast was computed without restricting the region of interest, no activation peak reached the predefined threshold. Moreover, we wished to compare activation between PerCont and illusion conditions. As stated earlier, the PerCont condition provided a control for the illusion condition because both led to the perception of a vowel in a background of noise bursts. On the other hand, PerCont vowels were rated as more speech-like and were more intelligible than illusory vowels (compare Tables 3 and 4). If we found activation differences between the two conditions, they could reflect either one or both of these differences. However, direct comparison of the activation of these conditions revealed no activation differences that reached the defined threshold.

We also examined whether illusion and illusion-break conditions engage in distinct processes. One way to

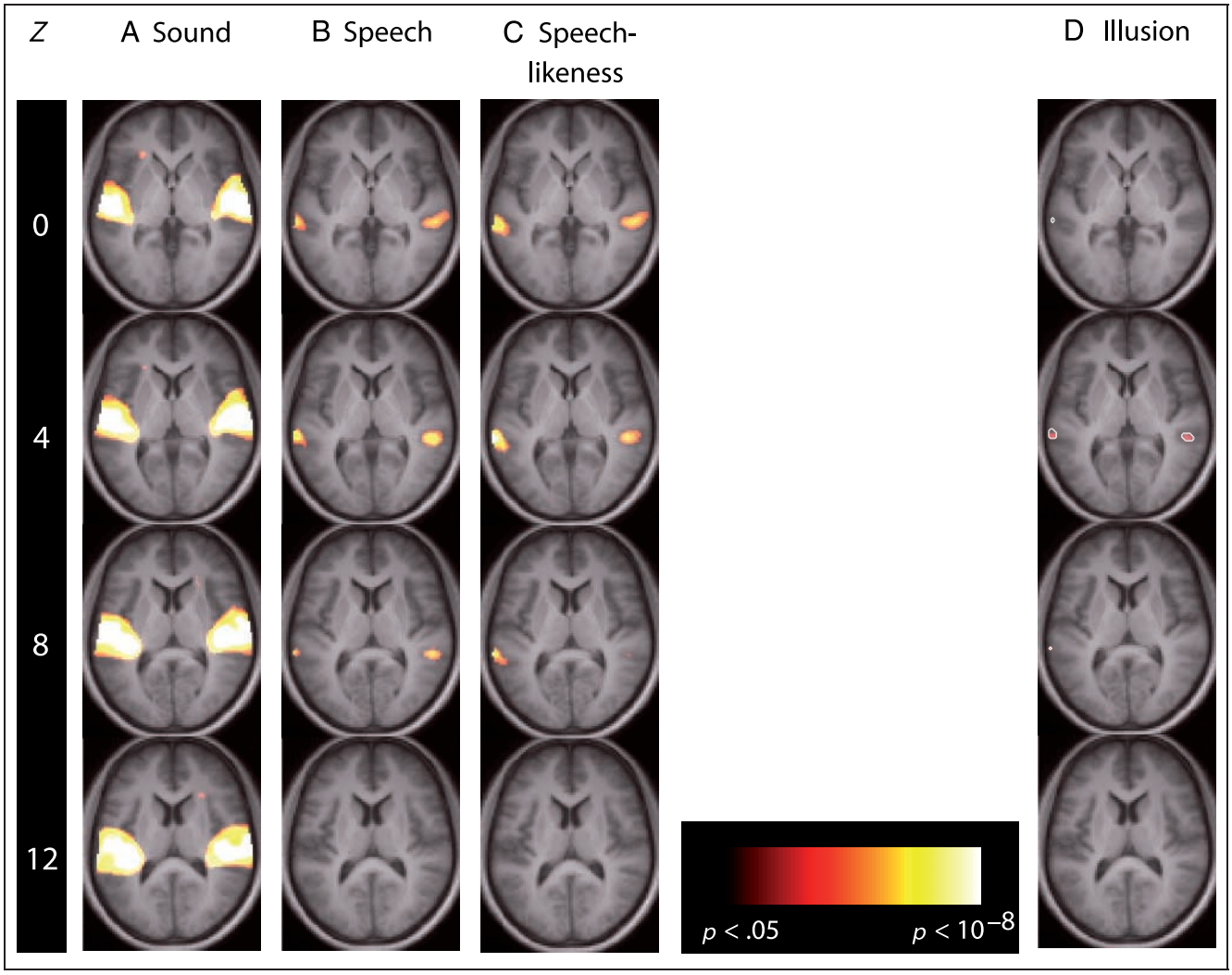

Figure 2. Areas showing significant activation in response to (A) sound [(Formant + Noise) − Rest]; (B) speech [(Vowel 600 + Vowel 300 + Vowel 75) − (Formant + Noise)]; (C) a positive correlation with speech-likeness ratings; and (D) Illusion − Illusion Break (masked with activation in response to speech). Activations are superimposed on the mean T1 structural of all 19 participants and thresholded at $p < .001$, uncorrected for multiple comparisons. Axial slices through the auditory cortex are shown. General sounds (A) activate large areas of the superior temporal gyrus (STG) as well as the middle temporal gyrus (MTG), whereas activation to speech sounds (B) is restricted to a small area in the MTG. Using speech-likeness ratings across the eight sound conditions for each listener as individual contrast weights (C), yields very similar activation similar to that for foci as the categorical speech-versus-nonspeech contrast, that is, a small area of activation in the posterior MTG. Within the speech area (shown as a white contour line), illusion stimuli evoke greater activation than illusion-break stimuli (D) (see Table 5 for details).

address this question is to measure the magnitude of neural activation for these two conditions at different sites in the brain. Following Henson (2006), we argue that a significant Condition-by-Brain area interaction provides evidence for regional specialization, and therefore, supports the notion of separate underlying mechanisms in the processing of illusion and illusion-break stimuli. We compared signal magnitude in these two conditions in 5-mm spheres at brain sites that we identified before as being either particularly sensitive to speech sounds or that responded to nonspeech harmonic complexes in independent contrasts that involved neither the illusion nor illusion-break conditions. One sphere was centered around the left-hemisphere peak voxels of the speech ROI (−68 −32 4), the other around the peak voxel of the sound ROI (−60 −14 0) (see Table 5). Subsequently, we used the mean signal estimates from both spheres in a repeated measures ANOVA with two conditions (illusion, illusion break) and two locations (speech and sound ROI). The ANOVA revealed a main effect of location [$F(1, 18) = 10.48$, $p < .005$], with significantly greater activation in the sound ROI than in the speech ROI, and, more importantly, an interaction between location and condition [$F(1, 18) = 23.42$, $p < .001$]. Post hoc t tests revealed that, in the sphere around the speech ROI, the signal elicited by the illusion condition was larger than the signal elicited by the illusion-break condition [$t(18) = 2.16$, $p = .045$]. In the general sound ROI, on the other hand, the illusion-break condition elicited a larger signal than the

illusion condition [$t(18) = -3.16$, $p = .005$; see Table 6 and Figure 3].

We were also curious to know whether we could replicate the Condition-by-Area interaction when contrasting activation in the speech ROI as determined by the current study with activation in the PAC as estimated in previous studies. We were interested in this comparison because, in a hierarchical model of sound processing, the PAC is assumed to comprise cognitively lower-level auditory areas that serve functions different from higher-cognitive areas in the STS/MTG. This hierarchical model was developed based on anatomical evidence from macaque monkeys which indicates that the auditory cortex is organized into at least three separate zones (core, belt, and parabelt) each comprising multiple cortical fields (Hackett & Kaas, 2004). The hierarchical connectivity of these zones suggests at least three discrete stages of processing (Hackett & Kaas, 2003; Kaas, Hackett, & Tramo, 1999) with the PAC (core) being the lowest of them. By contrasting signal strength of illusory speech-like sounds (illusion) with acoustically very similar nonspeech sounds (illusion break) in the core (primary) auditory cortex and speech areas (speech ROI), we can perhaps identify functionally distinguishable regions in the auditory cortex. To do this, we first estimated the coordinates of HG by averaging across four separate estimates of the center of either HG (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penhune, Zatorre, MacDonald, & Evans, 1996) or cytoarchitectonically defined the PAC (Morosan et al., 2001; Rademacher et al., 2001). All four sets of coordinates are published in Patterson et al. (2002, Table 1). Our estimated coordinates were (−45 −19 7) and (48 −15 7). We then repeated the previously described ANOVA, but with signal strength measured in 5-mm spheres around the left peak voxel of the speech ROI (the same MTG ROI used before) and the left PAC. The two conditions (illusion, illusion break) by two locations (MTG: −68 −32 4; PAC: −45 −19 7) repeated measures ANOVA yielded a main effect of location [$F(1, 18) = 29.20$, $p < .001$] with the PAC generally responding more to sound stimulation than the vowel area in the MTG. Crucially, the interaction between condition and area was once more significant [$F(1, 18) = 28.94$, $p < .001$]. As mentioned above, in the speech ROI, illusion stimuli elicited greater activation than illusion-break stimuli. Conversely, in the area around the PAC, illusion-break stimuli elicited significantly greater activity [$t(18) = -3.63$, $p = .002$]. Taken together, these two analyses demonstrate that an area that is particularly sensitive to speech responds to the illusion condition with increased activation, whereas areas sensitive to all sounds show greater activation in response to the illusion-break condition. The interactions did not occur simply because one of the two areas was inactive in one of the two conditions: Table 6 reports means and standard error of the mean signal in each condition in all tested areas.

Neural Correlates of Speech-likeness Ratings

As an additional test of the idea that posterior MTG and STS regions are important for speech perception, we looked for regions in which the BOLD signal correlated with speech-likeness ratings, within subjects. We did this by zero-mean normalizing the speech-likeness ratings across the eight conditions for each listener, and using these as contrast weights on the eight stimulus columns of the design matrix. The results of these listener-specific fixed-effects contrasts were then combined in a random-effects analysis. We particularly looked at areas in which BOLD signal correlated positively with the speech ratings. Two regions of significant positive correlation were observed in the left and right hemisphere posterior MTG. In the left hemisphere, this activation was within 4 mm of the region that was sensitive to vowels compared to nonspeech sounds and that was active in the illusion compared to the illusion-break condition (Table 5 and Figure 2C).

Number of Sound Onsets

In the Introduction we argued that, based on previous research, we might expect to find greater activation in primary auditory areas for illusion-break than illusion stimuli because the former contains more audible sound onsets. In order to test this notion, we computed two contrasts between conditions that differed in the number of sound onsets: Vowel 75 versus Vowel 300, and

Table 6. Mean (*SEM*) and *t* and *p* Values of the Signal in Illusion and Illusion Break Conditions in 5-mm Spheres around the Vowel Area, Sound Area, and the Center of the Auditory Cortex ($n = 19$ in All Conditions)

	Illusion	*Illusion Break*	*t*	*p*
Speech area (−68 −32 4)	0.72 (0.03)	0.60 (0.05)	2.16	.045
General Sound area (−60 −14 0)	1.08 (0.04)	1.25 (0.04)	−3.16	.005
Heschl's gyrus (−45 −19 7)	1.43 (0.04)	1.61 (0.04)	−3.63	.002

The *SEM* is calculated with the between-subject variance removed as appropriate for repeated-measures comparisons (Loftus & Masson, 1994).

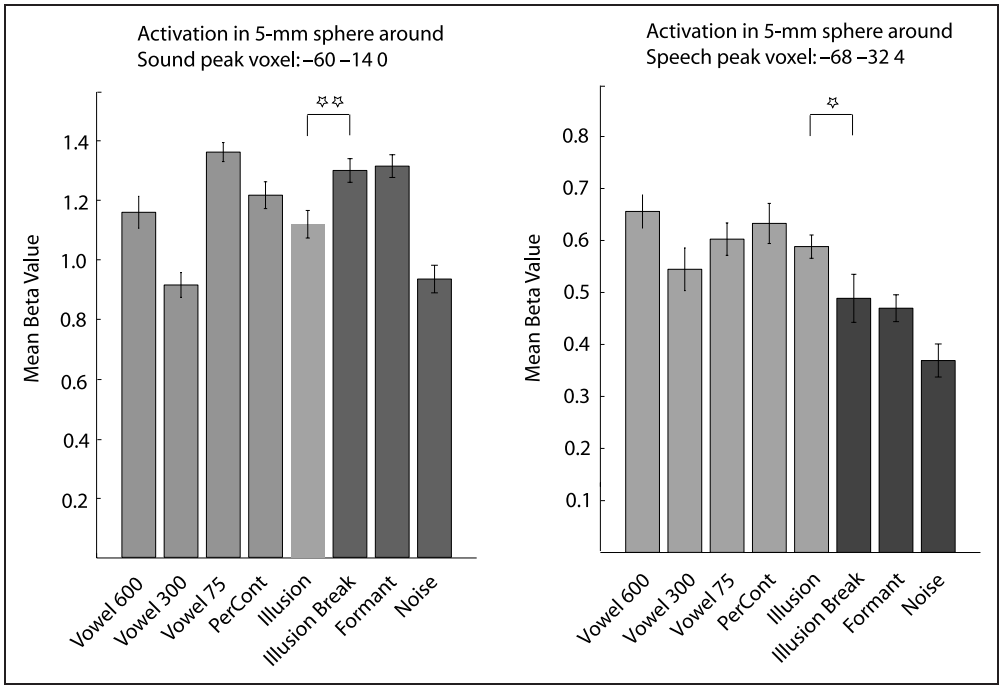

Figure 3. Graphs show BOLD signal increases activation for all eight sound conditions compared to rest in 5-mm spheres around speech and sound peak voxels. Error bars indicate the *SEM* after between-subject variability has been removed (Loftus & Masson, 1994). Activation differences between illusion and illusion-break conditions are significant in both plots (*$p < .05$; **$p < .001$). Light gray shading indicates conditions in which sounds were perceived as speech-like, dark gray shading indicates conditions in which sounds were perceived as significantly less speech-like.

Illusion Break versus Illusion. When comparisons were whole-brain corrected, only the former contrast yielded a significant result. We then used the thresholded region of activation in this former contrast as an ROI within which to look for significant activation to sound onsets in the Illusion Break versus Illusion comparison.

As can be seen in Table 7 and Figure 4, both contrasts reveal more activity in auditory regions of the STG for sounds with more onsets. Despite the fact that the search volume for the second contrast is fairly large, both contrasts show a maximum in activity in the lateral part of the STG and the dorsal part of the STS, slightly lateral and posterior to HG with peaks in the two contrasts differing only by 8.9 and 6.3 mm in the left and right hemispheres, respectively.

DISCUSSION

We used fMRI to measure neural activity in response to vowels, illusory vowels, and stimuli that were perceived as less speech-like. Compared to stimuli that were not perceived as speech, sounds perceived as vowels produced greater activation in the posterior MTG and the STS. Importantly, this was true even for sounds that were not physically complete vowels and depended on the continuity illusion to be perceived as speech-like,

Table 7. Peak Voxels Activated for Two Sound Onset Contrasts: Vowel 75 versus Vowel 300, and Illusion Break versus Illusion (Masked with the Vowel 75 − Vowel 300 Contrast)

	Coordinates					
Contrast	*x*	*y*	*z*	*t*	*No. of Voxels in Cluster*	*Area*
Vowel 75 − Vowel 300	**−50**	**−16**	**6**	**17.12**	**1962**	**L STG**
	58	**−14**	**2**	**13.67**	**2328**	**R STG**
	68	*−18*	*10*	*8.40*		*STG*
	62	*−24*	*10*	*7.40*		*STG*
Illusion Break − Illusion (using Vowel 75 − Vowel 300 as ROI)	**−58**	**−12**	**6**	**5.14**	**198**	**L STG**
	−48	*−12*	*2*	*5.13*		*L STG*
	56	**−8**	**2**	**5.80**	**381**	**R STG**
	42	*−18*	*8*	*3.68*		*HG*

For whole-brain analyses, we report peak voxels that are $p < .05$ (FDR) and $p < .001$, uncorrected within clusters of more than 10 contiguous voxels. For ROI analyses, we report peak voxels that are $p < .05$, FDR corrected. R = right; L = left; STG = superior temporal gyrus.

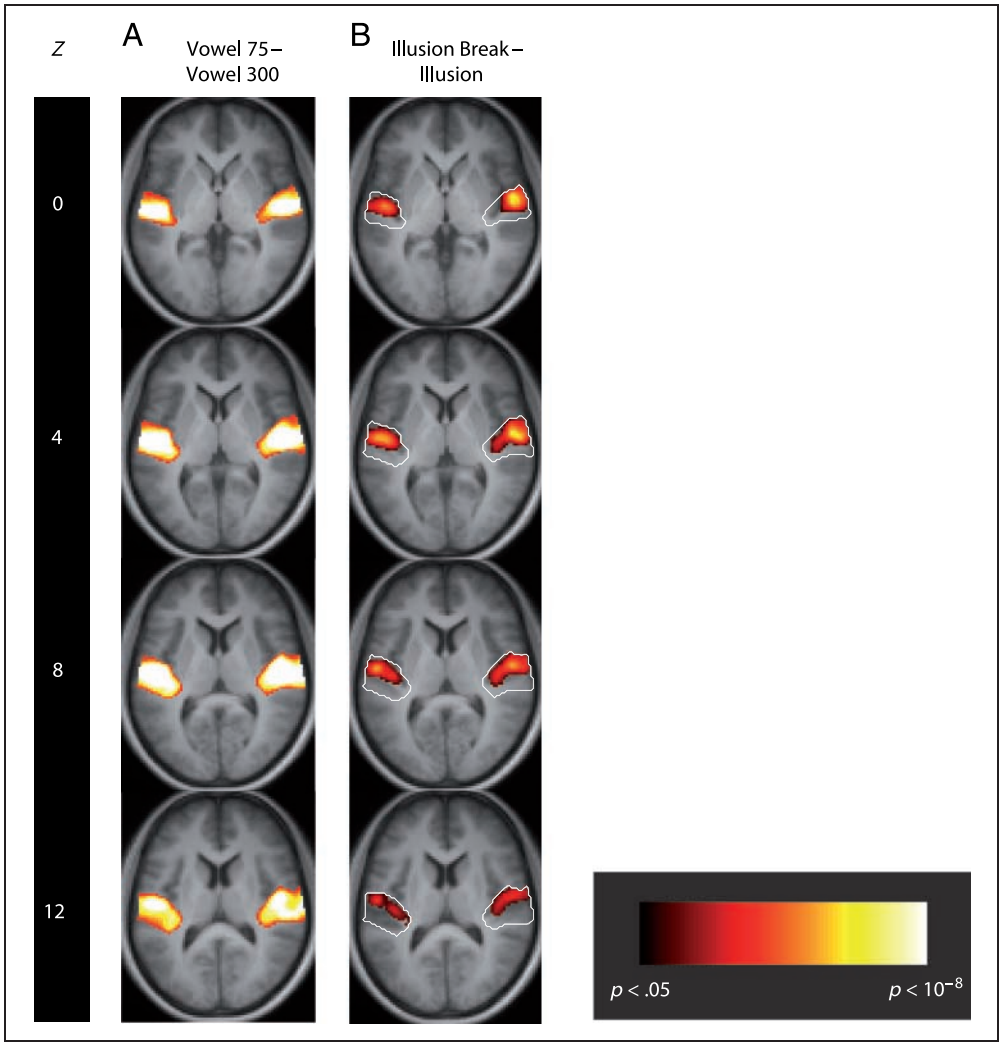

Figure 4. Areas showing significant activation related to an increase in number of sound onsets. (A) Vowel 75 and Vowel 300 are equated for overall sound energy and perceptual quality (both are perceived as speech), but Vowel 75 stimuli have four times as many onsets and evoke more activation in the superior temporal gyrus (STG), close to Heschl's gyrus, bilaterally. Within that area (shown as a white line) illusion-break stimuli also lead to greater cortical activation than illusion stimuli (B). As before, activations are superimposed on the mean T1 structural of all 19 participants and thresholded at $p < .001$, uncorrected for multiple comparisons.

indicating that speech-sensitive activation in the MTG reflects the *perception* of speech, and not necessarily its physical characteristics.

In contrast, sounds that were perceived as less speech-like produced greater activation than a silent baseline in the lateral STG. This activation included HG but was focused lateral to it. Whereas the illusion stimuli produced greater activation than the illusion-break stimuli in the MTG, the opposite was true in this lateral STG region. This interaction indicates that the *higher* activity for illusion stimuli in the MTG cannot be due to one stimulus being generally more effective at eliciting neural activity than another stimulus.

Finally, stimuli with more perceptual or physical onsets produced greater activity in the STG and in the dorsal part of the STS, suggesting a neural correlate of onset-sensitive processes in brain regions lateral to the PAC.

Sound Onset

As described in the Results section, stimuli with more audible onsets and offsets produced greater activation in regions close to but not in the PAC. According to previous studies (Herdener et al., 2007; Hart et al., 2003; Harms & Melcher, 2002), the locus of peak activation can depend on the actual rate of stimulation and can shift from HG for high rates of stimulation to the STG lateral and posterior of HG for lower rates. For instance, Hart et al. (2003) compared sustained harmonic complexes and tones to complexes with sound onset rates of 5 Hz. They found that regions extending lateral and posterior to HG and ventrally toward the STS in both hemispheres were particularly sensitive to the 5-Hz amplitude modulation over sustained sounds. Using white noise, Giraud and colleagues found a slight topographic gradient to the activation in response to the number of sound onsets with the medial geniculate body showing greatest activation to presentation rates of around 16 Hz, HG to sound onsets of around 8 Hz, and regions lateral and posterior to HG to onsets of 4 Hz.

The most straightforward comparison between our results and those obtained previously concerns the greater activation observed for the Vowel 75 than for

the Vowel 300 condition. Averaged over the 600-msec stimulus, the Vowel 75 condition had an onset rate of 6.7 Hz, whereas the stimulus was presented only once in the Vowel 300 condition. This result is therefore consistent with the previous finding that sound onsets between 5 and 10 Hz produce greatest activation in HG and areas immediately lateral to it. Note, however, that in no region did we see greater activation in the Vowel 300 than in the Vowel 75 condition.

It is less straightforward to express the stimuli in the illusion and illusion-break conditions in terms of onset rates because the perceived onsets occurred at different times in different frequency regions. However, in each frequency region, there were more perceived onsets in the break condition, where the onset of both the noise and the formant could be heard, than in the illusion condition. The finding of greater activation in the break condition is therefore consistent with Cusack et al.'s (2001) finding of greater activation for sounds with more onsets, but the results are hard to compare directly to those of studies investigating the effects of presentation rate. The most important implication of the results of this comparison is that the greater activation produced by the illusion stimuli in the MTG cannot be due to those particular stimuli producing uniformly more neural excitation.

Speech-likeness

Our vowel stimuli (Vowel 75, Vowel 300, and Vowel 600) elicit activity in the MTG; this result is consistent with a number of recent imaging studies of speech perception (Uppenkamp et al., 2006; Ashtari et al., 2004; Jancke et al., 2002; Giraud & Price, 2001; Vouloumanos, Kiehl, Werker, & Liddle, 2001). Table 8 lists the coordinates of peak voxels of those studies in which simple phonetic stimuli (vowels, consonant–vowel syllables) were compared to acoustically matched nonspeech sound conditions. We calculated the Euclidean distance (ED) between the peak voxel of the relevant contrast in those studies and the peak voxel activated in the current study. The average distance of peaks observed in these studies from the MTG peak that we observe was 7 mm (range 4–13 mm). Giraud and Price (2001) subtracted the activation produced by environmental sounds and noise from that produced by words and syllables, and found peak activation in the ventral posterior STS within 7 mm of the current study. Ashtari et al. (2004) measured activation in response to speech-specific phoneme and non-speech-specific frequency processing. Activation specific to word pairs occurred in a variety of locations along the anterior–posterior plane of the STS and the MTG with the most pronounced activity within 6 mm of the speech-specific peak activation of the current study. Jancke et al. (2002) and Vouloumanos et al. (2001) both compared rather basic speech stimuli (CV syllables and CVC nonwords) to nonspeech control stimuli and found activation that was very similar in location (within 4 mm) to the speech activation observed in the present study. Lastly, Uppenkamp et al. (2006) used synthetic vowels and compared them to acoustically closely matched musical rain sounds and also found activation specific to vowels in the left MTG.

Comparing these recent studies to the current investigation, it appears that there is good evidence that the activation we demonstrated in the MTG/STS is related to the perception of our intact vowels as speech sounds. Moreover, despite the fact that the illusion stimuli were acoustically incomplete and needed an acoustic illusion to enable listeners to hear the formants simultaneously and integrate them to a speech sound, they activated the same area in the MTG, suggesting that this area is more sensitive to the *perception* of a sound than its acoustical complexity.

Table 8. Descriptions, Left Hemisphere Coordinates, and Euclidian Distances of Recent Studies Investigating Speech Perception Compared to the Current Study

		Coordinates				
Authors	*Contrast*	*x*	*y*	*z*	*Area*	*Euclidian Distance in mm*
Present study	Vowels − (Formant + Noise)	−68	−32	4	L MTG	
Ashtari et al. (2004)	Phonetic detail (phonemes − tones)	−63	−36	3	L MTG	6
Giraud and Price (2001)	Phonetic detail (words and syllables)	−70	−38	6	L STG	7
Jancke et al. (2002)	CV − (tones + noise)	−64	−32	4	L STS	4
		−64	−28	8	L STS/PT	7
Uppenkamp et al. (2006)	Speech − nonspeech	−66	−20	0	L STS	13
Vouloumanos et al. (2001)	Phonetic detail (nonwords − SWS comparison)	−64	−32	4	L STG	4

CV = consonant-vowel syllable; MTG = middle temporal gyrus; STG = superior temporal gyrus; STS = superior temporal sulcus; SWS = sine-wave speech; PT = planum temporale.

Perceptual Learning Studies

A number of studies have exploited learning in order to contrast acoustically identical stimuli that are perceived as nonspeech by naïve listeners, but perceived as intelligible speech by trained listeners (Mottonen et al., 2006; Meyer et al., 2005; Liebenthal, Binder, Piorkowski, & Remez, 2003). These studies used sine-wave speech and scanned listeners first before training when stimuli were unintelligible, and then again after a period of training. Unfortunately, the results of these studies are not entirely consistent. When intelligible stimuli were contrasted with unintelligible stimuli, the studies led by Mottonen et al. (2006) and by Meyer et al. (2005) revealed activation in the left posterior STS/MTG, whereas Liebenthal et al. (2003) did not report any areas of increased activation for conditions of increased intelligibility for the sine-wave stimuli, but instead showed deactivation peaks in HG bilaterally and in the posterior STG 23 mm away from our speech peak voxel (−51, −30, +19). The differences in results between the studies may be caused either by the fact that only a minority of listeners in the Liebenthal et al. study perceived the stimuli as speech, even after training, or that the experimental task did not require participants to make use of their linguistic knowledge to perform well.

The Continuity Illusion

Although the physical parameters of stimuli necessary to elicit the continuity illusion have been well established over the last few decades (Warren, 1999), its neural basis has, until recently, remained elusive. The results presented here provide an objective, physiological correlate of the illusion in human subjects. Furthermore, the fact that subjects were monitoring the stimuli for a temporary decrease in intensity means that this correlate can be observed even when subjects' attention is focused on a feature of the stimuli that is independent of the generation of the illusion. In this regard, the present results are consistent with Micheyl et al.'s (2003) electroencephalogram finding that the illusion can influence the size of the MMN, even though subjects were instructed to ignore the sounds completely and to watch a silent movie. However, it is important to note that in neither of the two studies was attention manipulated explicitly. Hence, although one may conclude that the illusion can be measured without requiring subjects to respond to the relevant feature of the stimuli, this does not mean that it is completely unaffected by attention. It remains possible that, if we had required subjects to perform a demanding competing task, activation of speech-related areas may have been reduced in the illusion condition. Physiological measures of the illusion, such as that reported here, pave the way for further studies in which the effects of attention are probed fully by requiring subjects to perform a range of different competing tasks. We plan to pursue this opportunity in future experiments.

A further advantage of physiological measures such as ours is that they may allow one to understand the relationship between the continuity illusion and other neural processes. For example, the present results support the conclusion of Carlyon, Deeks, et al. (2002) and Carlyon, Micheyl, et al. (2002) that the illusion is already neurally represented at the stage where multiple formants are integrated into a speech percept, and, more importantly, shows that the integration of both physically and perceptually simultaneous formants have similar effects at the level of the MTG. Furthermore, combined with other physiological and behavioral measures, our findings help impose limits on the stages of processing at which the illusion can occur. For example, a consideration of the response properties of the auditory nerve (''AN'') reveals that the illusion is unlikely to be reflected at this stage of processing. This is because the AN responds instantaneously to sounds, whereas the illusion necessarily operates over a larger time scale: When a noise interrupts a sound such as a steady tone, a modulated tone, or a glide, the listener hears the appropriate sound continue, even though the AN activity during the noise would be the same for all three stimuli. At a cortical level, the evidence to date is less clear. Recent findings by Petkov et al. (2007), obtained in macaques, suggests a correlate of the illusion in A1, whereas the only physiological correlate of the illusion that has been reported in humans showed that it may only be partially complete at the stage of MMN generation (Micheyl et al., 2003), which presumably occurs after A1. The present results contribute to this rather sparse set of evidence by showing that the illusion occurs prior to the MTG. In addition, the fact that the presence, but not the phase, of frequency modulation is preserved during the illusion (Carlyon, Deeks, et al., 2002; Carlyon, Micheyl, et al., 2002) suggests that it operates at a stage where higher-order features of sounds have been extracted, and where some fine-grained acoustic information has already been discarded (Carlyon, Deeks, et al., 2002; Carlyon, Micheyl, et al., 2002). Finally, the fact that sound occurring *after* the inducer can affect the strength of the illusion (e.g., Ciocca & Bregman, 1987) indicates that it may be constrained to occur at a stage where information has already been integrated over tens of milliseconds. The challenge for future physiological studies is to constrain these explanations further by obtaining measurements at subcortical levels.

Acknowledgments

This research was funded by the Canadian Institutes of Health Research (Operating Grant MGP-69046) and the Medical Research Council. We thank the volunteers and the radiographers for their help with data collection (UK).

Reprint requests should be sent to Antje Heinrich, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, UK, or via e-mail: antje.heinrich@mrc-cbu.cam.ac.uk.

REFERENCES

Akeroyd, M. A., & Summerfield, A. Q. (2000). Integration of monaural and binaural evidence of vowel formants. *Journal of the Acoustical Society of America, 107,* 3394–3406.

Ashtari, M., Lencz, T., Zuffante, P., Bilder, R., Clarke, T., Diamond, A., et al. (2004). Left middle temporal gyrus activation during a phonemic discrimination task. *NeuroReport, 15,* 389–393.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate—A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, Methodological, 57,* 289–300.

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., Micheyl, C., Deeks, J. M., & Moore, B. C. J. (2002). FM phase and the continuity illusion. *Journal of the Acoustical Society of America, 111,* 2468.

Carlyon, R. P., Micheyl, C., Deeks, J. M., & Moore, B. C. J. (2004). Auditory processing of real and illusory changes in frequency modulation (FM) phase. *Journal of the Acoustical Society of America, 116,* 3629–3639.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics, 42,* 476–484.

Cusack, R., Carlyon, R. P., Johnsrude, I. S., & Epstein, R. (2001). Functional interaction between the left and right auditory pathways demonstrated using fMRI. Paper presented at the Society for Neuroscience Meeting, San Diego.

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,* 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.

Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience, 13,* 754–765.

Hackett, T. A., & Kaas, J. H. (2003). *Auditory processing in the primate brain* (Vol. 3). Hoboken, NJ: Wiley.

Hackett, T. A., & Kaas, J. H. (2004). *Auditory cortex in primates: Functional subdivisions and processing streams* (Vol. 14). Cambridge: MIT Press.

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,* 213–223.

Harms, M. P., Guinan, J. J., Sigalovsky, I. S., & Melcher, J. R. (2005). Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. *Journal of Neurophysiology, 93,* 210–222.

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology, 88,* 1433–1450.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex, 13,* 773–781.

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences, 10,* 64–69.

Herdener, M., Esposito, F., Di Salle, F., Lehmann, C., Bach, D. R., Scheffler, K., et al. (2007). BOLD correlates of edge detection in human auditory cortex. *Neuroimage, 36,* 194–201.

Houtgast, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America, 51,* 1885–1894.

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

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex (Vol. 9, p. 164). *Current Opinion in Neurobiology, 9,* 500.

Ladefoged, P. (1967). *Three areas of experimental phonetics*. Oxford: Oxford University Press.

Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience, 15,* 549–558.

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

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

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.

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America, 22,* 167–173.

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

Mottonen, R., Calvert, G. A., Jaaskelainen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage, 30,* 563–569.

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

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex, 6,* 661–672.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

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

Plack, C. J., & White, L. J. (2000). Perceived continuity and pitch perception. *Journal of the Acoustical Society of America, 108,* 1162–1169.

Powers, G. L., & Wilcox, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America, 61,* 195–199.

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

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage, 13,* 669–683.
Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62,* 623–633.
Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America, 117,* 305–318.
Sugita, Y. (1997). Neuronal correlates of auditory induction in the cat cortex. *NeuroReport, 8,* 1155–1159.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.
Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., & Patterson, R. D. (2006). Locating the initial stages of speech–sound processing in human temporal cortex. *Neuroimage, 31,* 1284–1296.
Vicario, G. (1960). L'effetto tunnel acustico. *Rivista di Psicologia, 54,* 41–52.
Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13,* 994–1005.
Warren, R. M. (1999). *Auditory perception: A new analysis and synthesis*. Cambridge: Cambridge University Press.
Warren, R. M., Obusek, C. J., & Ackroff, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science, 176,* 1149–1151.
Warren, R. M., Wrightson, J. M., & Puretz, J. (1988). Illusory continuity of tonal and infratonal periodic sounds. *Journal of the Acoustical Society of America, 84,* 1338–1342.
Wells, J. C. (1962). *A study of the formants of the pure vowels of British English*. Unpublished MA thesis.