Behavioral/Systems/Cognitive

How Silent Is Silent Reading? Intracerebral Evidence for Top-Down Activation of Temporal Voice Areas during Reading

Marcela Perrone-Bertolotti,[1,2] **Jan Kujala,**[1,2] **Juan R. Vidal,**[1,2] **Carlos M. Hamame,**[3] **Tomas Ossandon,**[4] **Olivier Bertrand,**[1,2] **Lorella Minotti,**[5] **Philippe Kahane,**[5] **Karim Jerbi,**[1,2] **and Jean-Philippe Lachaux**[1,2]

[1]INSERM U1028-CNRS UMR5292, Brain Dynamics and Cognition Team, Lyon Neuroscience Research Center, F-69500 Lyon-Bron, France; [2]University Claude Bernard, F-69000 Lyon, France; [3]Laboratoire de Psychologie Cognitive, CNRS (UMR 6146), Aix-Marseille Université 13003 Marseille Cedex, 3 France; [4]Departamento de Psiquiatría, Facultad de Medicina y Centro Interdisciplinario de Neurociencia, Pontificia Universidad Católica de Chile, CL-8330024 Santiago, Chile; and [5]CHU Grenoble and Department of Neurology, INSERM U704, F-38043 Grenoble, France

As you might experience it while reading this sentence, silent reading often involves an imagery speech component: we can hear our own "inner voice" pronouncing words mentally. Recent functional magnetic resonance imaging studies have associated that component with increased metabolic activity in the auditory cortex, including voice-selective areas. It remains to be determined, however, whether this activation arises automatically from early bottom-up visual inputs or whether it depends on late top-down control processes modulated by task demands. To answer this question, we collaborated with four epileptic human patients recorded with intracranial electrodes in the auditory cortex for therapeutic purposes, and measured high-frequency (50–150 Hz) "gamma" activity as a proxy of population level spiking activity. Temporal voice-selective areas (TVAs) were identified with an auditory localizer task and monitored as participants viewed words flashed on screen. We compared neural responses depending on whether words were attended or ignored and found a significant increase of neural activity in response to words, strongly enhanced by attention. In one of the patients, we could record that response at 800 ms in TVAs, but also at 700 ms in the primary auditory cortex and at 300 ms in the ventral occipital temporal cortex. Furthermore, single-trial analysis revealed a considerable jitter between activation peaks in visual and auditory cortices. Altogether, our results demonstrate that the multimodal mental experience of reading is in fact a heterogeneous complex of asynchronous neural responses, and that auditory and visual modalities often process distinct temporal frames of our environment at the same time.

Introduction

When children learn to read, they learn to associate written symbols with spoken sounds, until the association is so well trained that it occurs effortlessly. Around that time, they become able to read silently. Basic introspection suggests that adults continue to pronounce written text covertly using auditory verbal imagery (AVI; Jäncke and Shah, 2004). This intuition has been confirmed by several psychophysics experiments (Alexander and Nygaard, 2008).

Until recently, AVI had received little interest from experimental psychology and cognitive neuroscience (Kurby et al., 2009; Yao et al., 2011). Yet, it is one of the most striking examples of cross talk between sensory modalities, as written words produce a vivid auditory experience almost effortlessly. It is therefore a privileged situation to understand multisensory integration and its underlying functional connectivity. It also offers a unique window into the actual reading experience of "what it is like" to read. AVI is one of its most salient components and should be included in any neurocognitive model of reading. Finally, few would contest that most of our waking time is spent talking to ourselves covertly, and some diseases, like schizophrenia or depression, are often characterized by a lack of control over AVI (Linden et al., 2011). Silent reading is an efficient way to elicit AVI experimentally with precise control over its timing, and understand its neural bases.

The neurophysiological investigation of AVI has been focused on the auditory cortex, because that region is active during several forms of auditory imagery, including hallucinations (Lennox et al., 1999), silent lip reading (Calvert et al., 1997), and music imagery (Halpern and Zatorre, 1999). One might therefore assume that if silent reading generates AVI, it should activate the auditory cortex concurrently. This hypothesis was recently validated with functional magnetic resonance imaging (fMRI) by Yao et al. (2011), who showed a strong activation of auditory voice-selective areas, or temporal voice areas (TVAs; Belin et al., 2000).

One remaining question, unresolved by fMRI, is whether TVA activation is mostly early and bottom-up, or controlled by late

Table 1. Talairach coordinates of sites of interest

Patient	Site	X	Y	Z
P1	T8	58	0	−12
P2	T7	57	−16	−6
P3	U'9	−60	−15	4
P4	T8	61	−11	−5
P2	U2	37	−28	8
P2	F'2	−25	−56	−8

Coordinates are expressed in millimeters in the Talairach coordinate system (Talairach and Tournoux, 1988).

top-down processes (Alexander and Nygaard, 2008) and therefore modulated by attention or cognitive strategy. We addressed this question with direct intracerebral electroencephalographic (EEG) recordings of TVA in epileptic patients. We analyzed high-frequency "gamma" activity (HFA), between 50 and 150 Hz, because it is now considered as a reliable physiological marker of neuronal activity at the population level, and a general index of cortical processing during cognition (Lachaux et al., 2012). Lower frequencies were not considered in this study, either because of their less precise association with fine cognitive processes (α and β; Jerbi et al., 2009; Crone et al., 2011) or simply because their frequency range was incompatible with the timing of the experimental design (theta and delta).We found that the neural response to written words in TVA comes between 400 and 800 ms after the visual response, with a highly variable delay, and is strongly modulated by attention. We therefore demonstrate that the rich and seemingly coherent audiovisual experience of reading arises from a heterogeneous amalgam of asynchronous neural responses.

Materials and Methods

Participants. Intracranial recordings were obtained from four neurosurgical patients with intractable epilepsy (three female, mean age: 28.75 +/− 10.04 years) at the Epilepsy Department of the Grenoble Neurological Hospital (Grenoble, France). All patients were stereotactically implanted with multilead EEG depth electrodes. All electrode data exhibiting pathological waveforms were discarded from the present study. This was achieved in collaboration with the medical staff and was based on visual inspection of the recordings and by systematically excluding data from any electrode site that was a posteriori found to be located within the seizure onset zone. All participants provided written informed consent, and the experimental procedures were approved by the Institutional Review Board and by the National French Science Ethical Committee.

Electrode implantation. Eleven to 15 semirigid, multilead electrodes were stereotactically implanted in each patient. The stereotactic electroencephalography (SEEG) electrodes used have a diameter of 0.8 mm and, depending on the target structure, consist of 10–15 contact leads 2 mm wide and 1.5 mm apart (DIXI Medical Instruments). All electrode contacts were identified on a postimplantation MRI showing the electrodes, and coregistered on a pre-implantation MRI. MNI and Talairach) coordinates were computed using the SPM (http://www.fil.ion.ucl.ac.uk/spm/) toolbox (see Table 1).

Intracranial recordings. Intracranial recordings were conducted using a video-SEEG monitoring system (Micromed), which allowed the simultaneous data recording from 128 depth EEG electrode sites. The data were bandpass filtered on-line from 0.1 to 200 Hz and sampled at 512 Hz in all patients. At the time of acquisition data are recorded using a reference electrode located in white matter, and each electrode trace is subsequently re-referenced with respect to its direct neighbor (bipolar derivations). This bipolar montage has a number of advantages over common referencing. It helps eliminate signal artifacts common to adjacent electrode contacts (such as the 50 Hz mains artifact or distant physiological artifacts) and achieves a high local specificity by canceling out effects of distant sources that spread equally to both adjacent sites through volume conduction. The spatial resolution achieved by the bi-

polar SEEG is on the order of 3 mm (Kahane et al., 2003; Lachaux et al., 2003; Jerbi et al., 2009). Both spatial resolution and spatial sampling achieved with SEEG differ slightly from that obtained with subdural grid electrocorticography (Jerbi et al., 2009).

Time–frequency analysis, gamma power, and envelope computations. The frequency band of interest, between 50 and 150 Hz, was defined from preliminary time-frequency analysis of the SEEG data using wavelets (Tallon-Baudry et al., 1997) (performed with in-house software package for electrophysiological signal analysis) (Aguera et al., 2011) and from previous studies by our group (Jerbi et al., 2009).

Raw data were transformed into gamma-band amplitude (GBA) time series with the following procedure (Ossandón et al., 2011). Step 1: continuous SEEG signals were first bandpass filtered in multiple successive 10 Hz wide frequency bands (e.g., 10 bands from 50–60 Hz to 140–150 Hz) using a zero phase shift noncausal finite impulse filter with 0.5 Hz roll-off. Step 2: next, for each bandpass filtered signal we computed the envelope using standard Hilbert transform (Le Van Quyen, 2001). The obtained envelope is downsampled to a sampling rate of 64 Hz (i.e., one time sample every 15,625 ms). Step 3: for each band this envelope signal (i.e., time-varying amplitude) was divided by its mean across the entire recording session and multiplied by 100. This yields instantaneous envelope values expressed in percentage (%) of the mean. Step 4: finally, the envelope signals computed for each consecutive frequency band (the 10 bands of 10 Hz intervals between 50 and 150 Hz) were averaged together to provide one single time series (the GBA) across the entire session. By construction, the mean value of that time series across the recording session is equal to 100. Note that computing the Hilbert envelopes in 10 Hz sub-bands and normalizing them individually before averaging over the broadband interval allows us to account for a bias toward the lower frequencies of the interval that would otherwise occur due to the 1/f drop-off in amplitude.

Auditory functional localizer. During the auditory localizer task (AUDI), participants listened to sounds of several categories with the instruction that three of them would be presented again at the end of the task, together with three novel sounds and that they should be able to detect previously played items. There were three speech and speech-like categories, including sentences told by a computerized voice in a language familiar to the participant (French) or unfamiliar (Suomi), and reversed speech, originally in French (the same sentences as the "French" category, played backwards). These categories were compared with nonspeech-like human sounds (coughing and yawning), music, environmental sounds, and animal sounds. In the text, these categories will be referred to as French, Suomi, Revers, Cough, Yawn, Music, Envir, and Animal. Some of the sounds (Cough, Yawn, Envir, and Animal) were concatenations of shorter stimuli of the functional localizer of the TVAs (Belin et al., 2000; Pernet et al., 2007).

Participants were instructed to close their eyes while listening to three sounds of each category, with a duration of 12 s each, along with three 12 s intervals with no stimulation, serving as a baseline (Silence). Consecutive sounds were separated by a 3 s silent interval. The sequence was pseudorandom, to ensure that two sounds of the same category did not follow each other. Sound sequences were delivered binaurally via earphones via the Presentation stimulus delivery software (Neurobehavioral Systems). All acoustic signals had the same intensity, quantified by their root mean square, set at a comfortable sound level.

Reading task. The reading task (Read) was adapted from Nobre et al. (1998). In each block of the experiment, participants were presented with two intermixed stories, shown word by word at a rapid rate. One of the stories was written in gray (on a black screen) and the other in white. Consecutive words with the same color formed a meaningful and simple short story in French. Participants were instructed to read the gray story to report it at the end of the block, while ignoring the white one. Each block comprised 400 words, with 200 gray words (Attend condition) and 200 white words (Ignore condition) for the two stories. The time sequence of colors within the 400 words series was randomized, so that participants could not predict whether the subsequent word was to be attended or not; however, the randomization was constrained to forbid series of more than three consecutive words with the same color. After the block, participants were asked questions about the attended text, which could not have been answered from general knowledge. The ex-
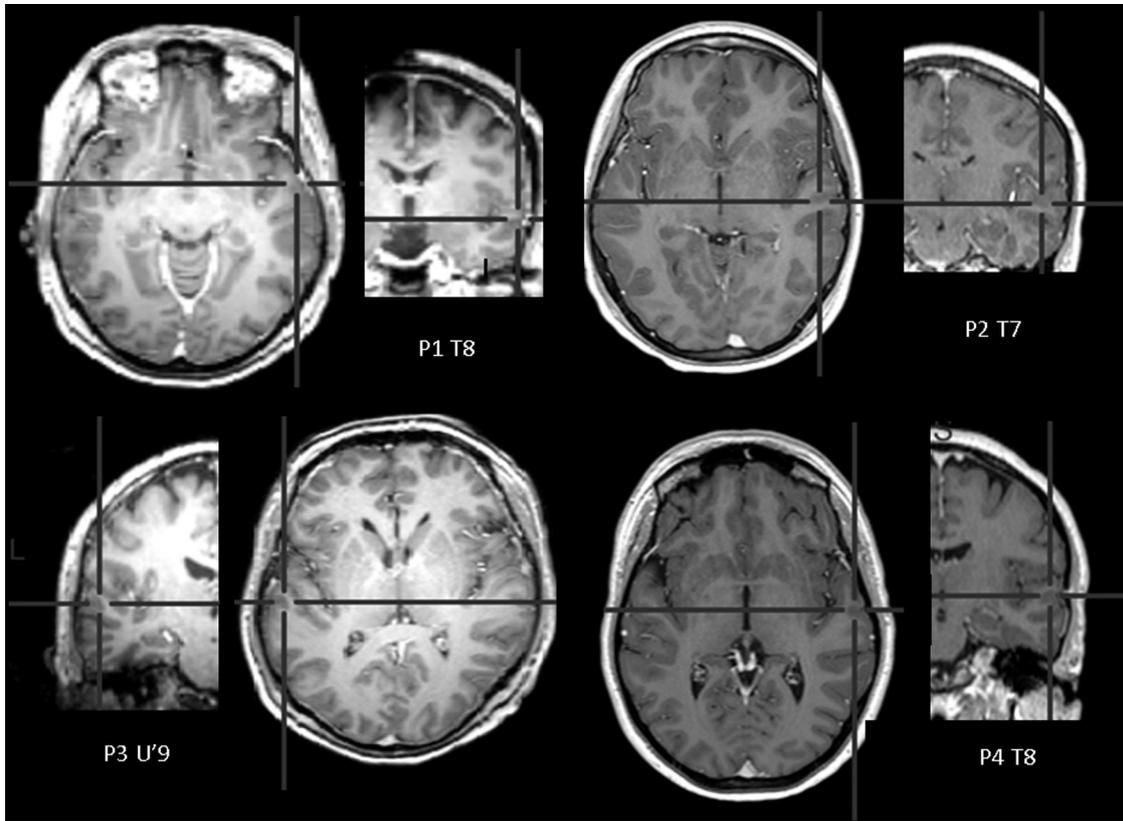
Figure 1. Anatomical location of recording sites. Black crosshairs indicate site location on coronal and axial views of individual MRI.

perimental procedure took place in patients' hospital rooms. Stimuli were presented to the participants on a 17 inch computer screen at a 20 cm viewing distance and the average word subtended 28 of visual angle. Words appeared singly for 100 ms every 700 ms.

Statistical analysis. Statistical analyses were performed on high-frequency activity time series, HFA, computed as above. In the AUDI task, for each sound (12 s duration), we selected 20 nonoverlapping 500 ms windows covering latencies between 1500 ms after stimulus onset to 500 ms before stimulus offset (windows within the first 1500 ms were excluded to avoid the transient response to the stimulus). The mean HFA was measured in each window to provide 20 values per stimulus, and a total of 60 values per category (since there were three sounds for each category). We used a nonparametric Kruskal–Wallis test to detect an effect of stimulus category on those values (60 HFA values × 8 categories, that is, excluding the Silence category). We used a Bonferroni procedure to correct for multiple comparisons. In the Read task the effect of silent reading on TVA sites was evaluated by comparing HFA across consecutive, nonoverlapping 100 ms windows covering a 1 s interval following word presentation (Kruskal–Wallis test, corrected for multiple comparisons with a Bonferroni procedure). This comparison was done separately for attended and ignored words. We reasoned that if a TVA site responds to word presentation, then a significant HFA modulation should be observed when the participant is processing that stimulus,
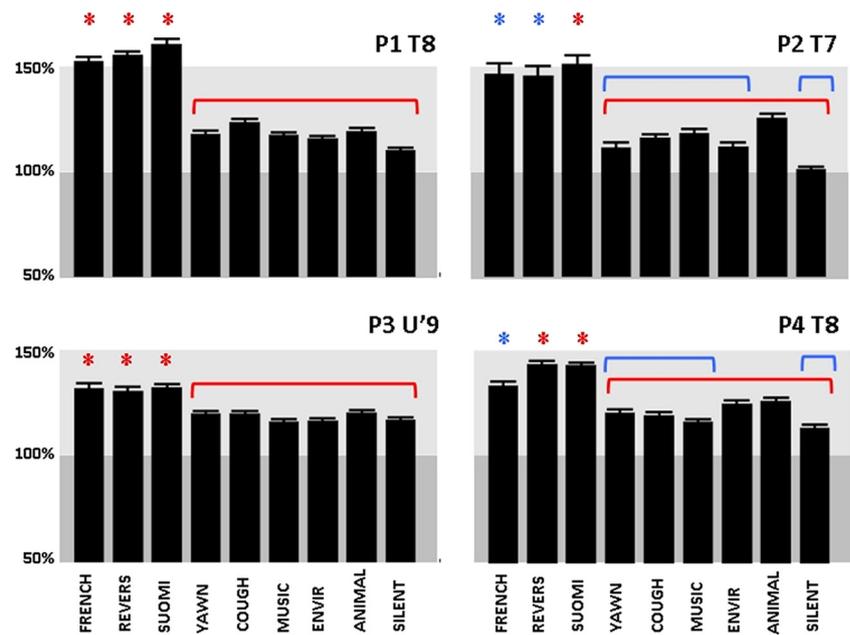

Figure 2. Selectivity to speech in the auditory cortex. Bar plots show the mean HFA (±SEM) measured while participants listened to sounds of several categories (averaged across 60 nonoverlapping 500 ms windows during sounds, see Materials and Methods). HFA is expressed in percentage of the average amplitude measured across the entire recording session. Values with a colored asterisk are significant higher than values indicated by a bracket of the same color (Kruskal–Wallis, *post hoc* pairwise comparison). Note that in P2 and P4 the French sentence did not elicit a response significantly stronger than to animal sounds, but the sites were nevertheless included in the study because of the strong response to other speech-like stimuli (Revers and Suomi).

that is, within the first second following stimulus onset. Another approach would have been to compare HFA before and after stimulus onset, as most studies usually do, but the interstimulus interval was too short to define a neutral baseline before stimulus onset.
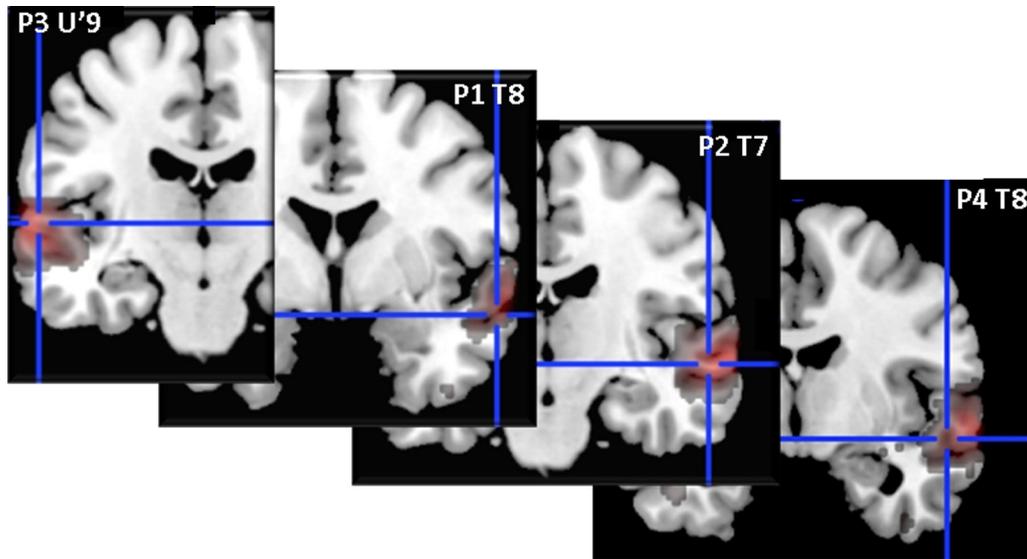
Figure 3. Correspondence between recording sites and TVAs. For each site, blue crosshairs indicate its normalized MNI coordinates projected upon the single-subject MRI. Red overlay shows the probabilistic map of TVA, as defined by Belin et al. (2000). (TVA map available for download at http://vnl.psy.gla.ac.uk/resources.php.)
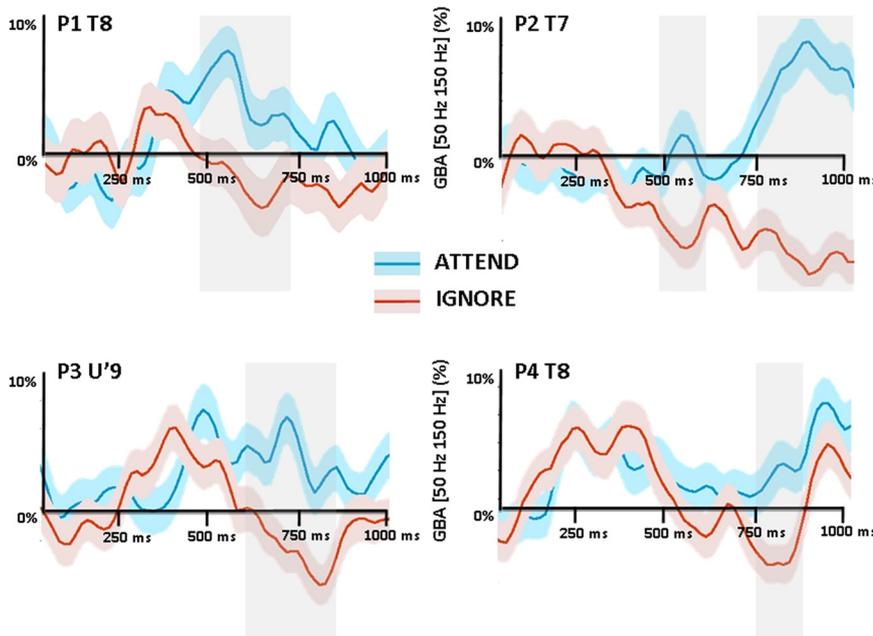

Figure 4. Response to written words in TVAs. For each site, plots display HFA modulations (50 –150 Hz) averaged across trials in the two attention conditions (±SEM): Attend (blue) and Ignore (red). Amplitude is expressed in percentage of the average amplitude measured across the entire recording session. Shaded areas indicate time windows during which HFA is significantly higher in the Attend condition. Words were presented at 0 ms for 100 ms. Note that in that dataset, the probability for the preceding word to be attended or ignored is the same (50%).

Directed interactions. To quantify how attentional modulation affects the amount of influence brain regions exert on each other, Granger Causality (Granger, 1980; Geweke, 1982) was estimated between selected recording sites over the visual and auditory cortices. The analysis was conducted using a MATLAB Toolbox developed by Seth (2010). In the analysis, Granger Causality was computed across trials separately in the Attend and Ignore conditions; the computations were performed using partially overlapping 200 sample-long windows, with the window centers spanning the range from 100 ms prestimuli to 700 ms poststimuli. Each time window was then detrended, the ensemble mean was subtracted, and first-order differencing was applied. The stationarity of the processed data was tested by examining the autocorrelation functions in each time window for both conditions and recording sites (Seth, 2010); no violations of covariance stationarity were

detected. As neither the Akaike (1974) nor Bayesian information (Seth, 2005) criterion yielded conclusively an optimal model order, the lowest order that led to 80% of the data to be captured by the autoregressive model was used. Subsequently, the Granger Causality terms were computed, and the significance of the difference of influence terms was determined using permutation testing (200 permutations).

Results
Behavioral data
In both tasks, behavioral data analysis was mostly qualitative, simply to ensure that participants had understood tasks instructions. It was the case for all patients. In particular, after the Reading task, debriefing sessions clearly indicated that they had read the target story, captured its global meaning, and were able to tell the correct sequence of events.

Selective responses to speech in the auditory cortex
The auditory localizer task revealed four voice-selective sites, one in each patient (Fig. 1). All sites were characterized by a stronger increase of HFA (50–150 Hz) after speech-like stimuli than after other stimulus categories such as music or environmental sounds (Kruskal–Wallis, corrected $p < 0.05$; Fig. 2). In each patient, voice-selective sites were located in the superior temporal gyrus, one in the left hemisphere (P1) and three in the right hemisphere (P2, P3, and P4). All of them were within the TVAs defined by Pernet et al. (2007) (Fig. 3). Interestingly, the strongest responses were not obtained for French sentences, but for meaningless speech stimuli, Revers and Suomi, as already reported by Lachaux et al. (2007).

Responses to written words in TVAs
In three of the four TVA sites (in P1, P2, and P4), word presentation induced a significant HFA modulation in the Attend

Table 2. Demographic (sex, age, manual lateralization, index of manual lateralization, laterality index of language) and verbal neuropsychological performance (verbal comprehension index, WAIS-IIII, and reading scores assessed by means of Stroop test) data of patients P1–P4

Patient	Sex	Age	ML	EIML	LIL	ICV	Reading (Stroop)
P1	F	39	R	100%	LH	107 (68 ct)	44.38 s (−0.73 ET)
P2	F	16	R	NA	NA	77 (6 ct)	70 s (−2 ET)
P3	H	34	R	90%	LH	114 (82 ct)	37.33 s (−0.44 ET)
P4	F	26	R	100%	LH	96 (39 ct)	61 s (−0.27 ET)

Manual lateralization, ML; R, right-handed; LH, left-handed; Edinburgh Handedness Inventory (Oldfield, 1971), EIMI; laterality index of language lateralization, LIL; verbal comprehension index, ICV; information not available, NA.
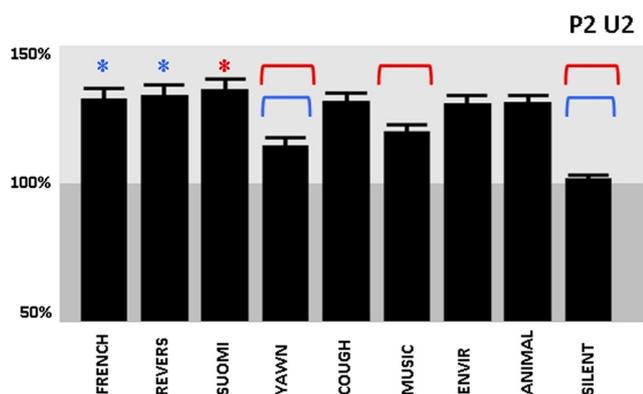

Figure 5. Response to sounds in the primary auditory cortex. Bar plots show the mean HFA (±SEM) measured while participants listened to sounds of several categories (averaged across 60 nonoverlapping 500 ms windows during sounds, see Materials and Methods). Values with a colored asterisk are significant higher than values indicated by a bracket of the same color (Kruskal–Wallis, *post hoc* pairwise comparison). HFA is expressed in percentage of the average amplitude measured across the entire recording session. Note that speech-like and nonspeech-like sounds elicit equally strong responses.

condition (Kruskal–Wallis, corrected $p < 0.05$; Fig. 4). In the remaining patient (P3), the modulation was significant only in the Ignore condition. However, the response to attended words was stronger than to ignored words in all sites (Attend > Ignore; Kruskal–-Wallis, corrected $p < 0.05$), in at least one 100 ms window. The effect of attention was late and variable in time, starting 400 ms after word presentation at the earliest (P1 and P2), but later than 500 ms in the remaining sites (P3 and P4). This effect of attention occurred independently of participants' reading skills (see Table 2); although P1 and P2 had different reading skills (better for P1 than P2), the effect of attention was observed in the same temporal window for both patients.

A wavering sequence of activation from visual to auditory cortex

In one patient (P2), two additional electrodes were located in the primary auditory cortex (Fig. 5) and in the ventral occipital temporal cortex (VOTC). This provided a unique opportunity to compare the timing of neural responses to written words across visual and auditory regions. As shown in Figure 6, there was a clear propagation peaking first in the left VOTC (F′2, 300 ms), then in the right primary auditory cortex (U2, 700 ms), and finally in TVA (T7, 800 ms). This is a clear indication that auditory and visual areas do not react conjointly to provide a multimodal representation of the stimulus. That "representation" is in fact a heterogeneous collection of asynchronous activities, with no binding mechanism to assemble them. Further, it is also clear that primary auditory areas also participate in the auditory representation of the written word, and sometimes before TVA, which

means that the auditory representation itself is not homogeneous, even within a single modality in the auditory cortex. In addition, single-trial display of visual and auditory responses revealed that the lag between visual and auditory responses is far from constant (Fig. 7). In this figure, the response of the VTOC (F′2) and primary auditory cortex (U2) can be seen for several consecutive attended words, with a clear time separation varying between 450 and 750 ms. To quantify this qualitative observation, we measured the peak latency of the response to individual words in the attend condition, in F′2 and U2. We selected only trials with a clearly visible response in each of the two sites (i.e., an excellent signal-to-noise ratio), as defined by a peak value exceeding the mean of the signal by at least two SDs (the mean and SD of HFA over the entire experimental session for that channel). Thirty-six trials were selected, and the mean peak latencies were, respectively, 324 ms in F′2 (SD = 63 ms) and 681 ms in U2 (SD, 164 ms), significantly later in the auditory cortex than in the visual cortex (sign test; $p < 10$–9). When considering all visible peaks to increase the number of trials (not necessarily visible in both U2 and F′2 in the same trial), we found similar latencies: 333 ms in F′2 (SD, 100 ms; 83 trials) and 674 ms in U2 (SD, 174 ms; 87 trials). Mean latency in T7 was 651 ms (SD, 219 ms; 54 trials); however, there were not enough common peaks in T7 and U2 or F′2 to perform a meaningful comparison of their latencies (<20 peaks).

There was also a clear difference between the Attend and Ignore conditions in the directed interactions between the VOTC and primary auditory cortex (Fig. 8). For the Attend condition, there was significantly ($p < 0.01$) more influence from VOTC to primary auditory cortex than vice versa in two consecutive time windows, centered at ~400 and 500 ms, whereas no dominant direction of influence was observed in the Ignore condition.

Finally, it is also clear from Figures 6 and 7 that the response of the auditory cortex often occurs during the display of the next word, which means that auditory and visual areas are active simultaneously while processing different temporal frames of the participant's environment. We tested for a possible interference between those responses by computing the correlation between GBA in T7 and F′2 at the peak latency in F′2, between 200 and 400 ms. Since the response peak in T7 to the previous stimulus occurs around that latency, we reasoned that both processes might compete for resources and interfere. Accordingly, we found that average GBA values over the 200 ms:400 ms time window for F′2 and T7 are negatively correlated (Spearman $r = -0.1887$, $p = 0.00014$). However, one might argue that Attend stimuli are slightly more often preceded by an Ignore stimulus, and therefore that strong F′2 responses should be often simultaneous with a weak T7 response (to the preceding stimulus). We corrected for that possible bias by considering only Attend stimuli preceded by an Attend stimulus and found that the correlation was in fact more negative ($r = -0.2167$); however, it failed to reach significance level because of a smaller number of trials ($p = 0.06$).

Discussion

We used direct electrophysiological recordings from the auditory cortex to show that silent reading activates voice-selective regions of the auditory cortex, or TVAs. The neural response involves high-frequency activity, already shown to be exquisitely sensitive to auditory stimuli, including speech (Crone et al., 2001; Towle et al., 2008; Chang et al., 2010). This result is reminiscent of an earlier study by our group, reporting that a region in the left anterior superior temporal gyrus (STG) was active during a silent reading task (Mainy et al., 2008), in line with an earlier suggestion that the left STG might comprise an auditory analog of the word
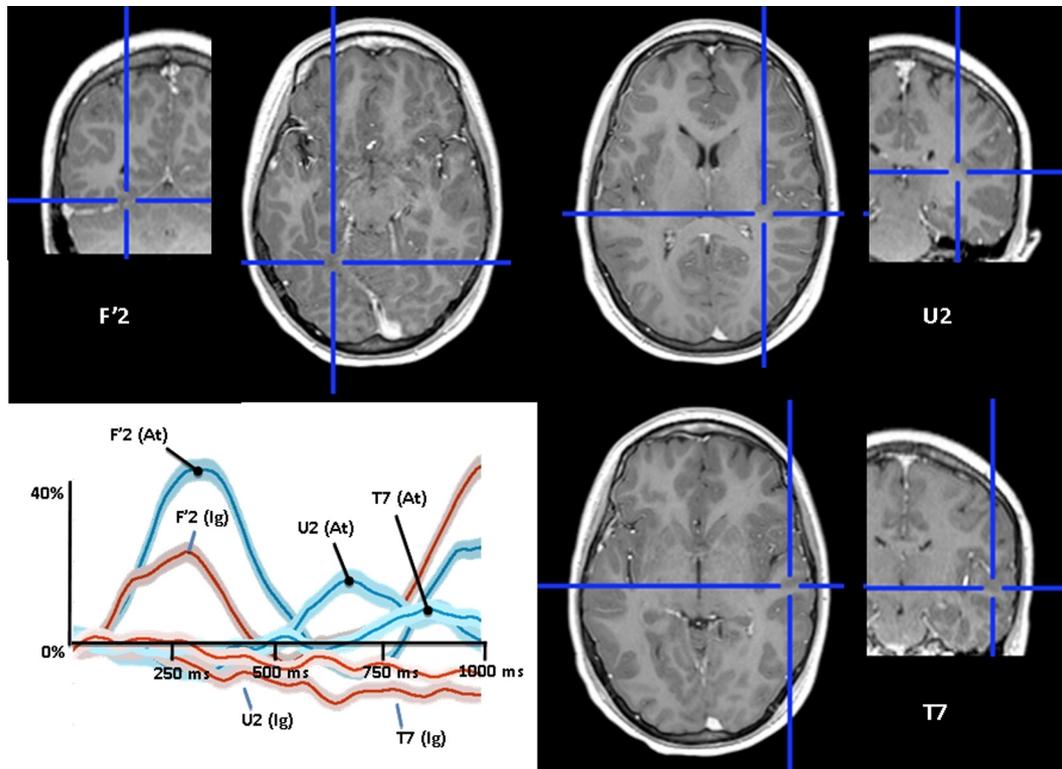
Figure 6. Neural response to written words in the visual and auditory cortex. The three sites were recorded in the same patient, P2, in the left VOTC (F'2), the primary auditory cortex (U2), and in a TVA, in the STG (T7), respectively. Plots display HFA modulations (50 –150 Hz) averaged across trials in the two attention conditions ($\pm$SEM): Attend (blue) and Ignore (red), for each site. Amplitude is expressed in percentage of the average amplitude measured across the entire recording session. Words were presented at 0 ms for 100 ms. Note that in that dataset, the probability for the preceding word to be attended (Attend) or ignored (Ignore) is the same (50%). However, the upcoming word shown at 700 ms has a higher probability to be of the opposite condition. This explains the second peak of activation ~1000 ms in T7 (red curve).
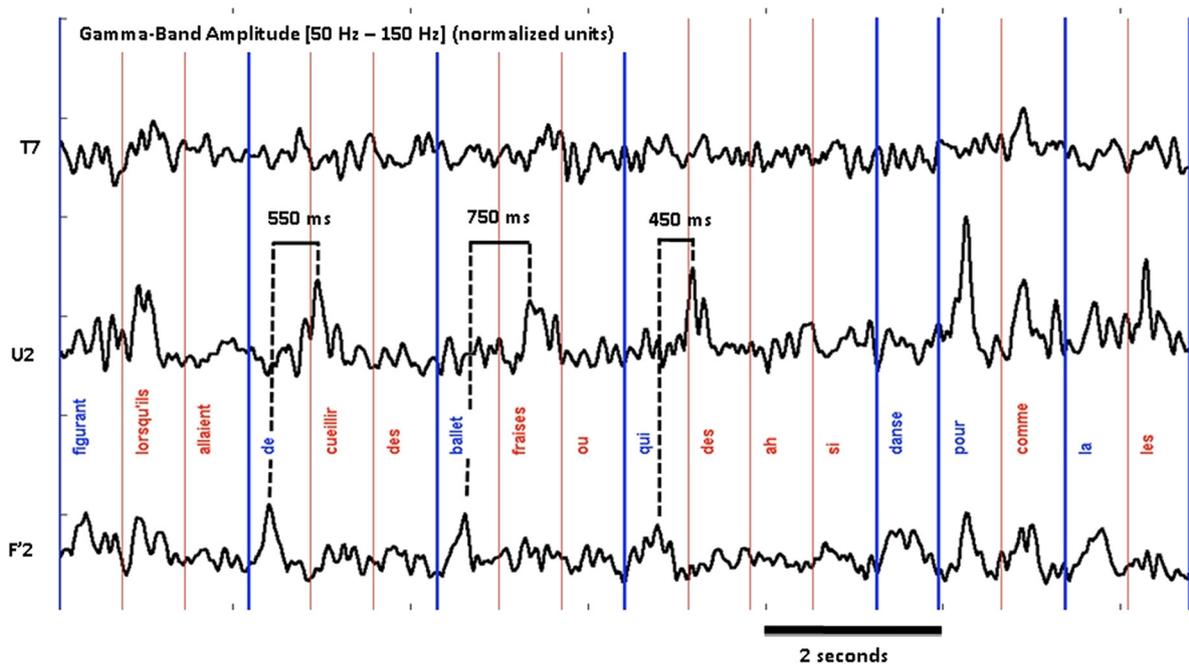

Figure 7. Single-trial responses to written words in visual and auditory cortex. Plots are HFA fluctuations (50 –150 Hz) during the reading task. Vertical lines indicate word onsets in the Attend (blue) and Ignore (red) conditions. Individual activation peaks can be seen in VOTC (F'2) and the primary auditory cortex (U2). Note that the lag between visual and auditory peaks is not constant.

form area (Cohen et al., 2004). We confirm that finding and extend it to both hemispheres. In addition, we show that this activity is strongly modulated by attention: the response is strong and sustained only if participants read words attentively.

Our results confirm a recent fMRI study by Yao et al. (2011) showing that TVA is active during silent reading, especially when participants read direct speech. One specificity of that study, emphasized by the authors, was that the auditory cortex was active
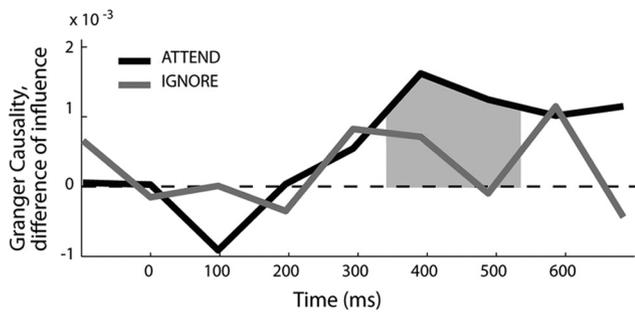
Figure 8. Directed interactions between VOTC and the primary auditory cortex. Difference between the Granger Causality terms quantifying the directed interactions from VOTC to primary auditory cortex and vice versa for the Attend (black) and Ignore (gray) condition; positive values indicate more influence from VOTC to primary auditory cortex. Time windows showing significant ($p < 0.01$) differences of influence are marked by the shaded area. Time windows showing significant ($p < 0.01$) differences of influence are shaded in gray (indicating significance in the Attend condition).

even though participants were not explicitly encouraged to use auditory imagery. Understandably, most studies so far have used task designs emphasizing covert pronunciation, such as rhyme and phoneme detection tasks (Cousin et al., 2007, Khateb et al., 2007; Perrone-Bertolotti et al., 2011) to maximize the imagery component. Our task did not emphasize mental pronunciation and therefore confirms the conclusion of Yao et al. (2011): reading spontaneously elicits auditory processing in the absence of any auditory stimulation (Linden et al., 2011), which might indicate that readers spontaneously use auditory images.

Possible role of the inner voice during reading
In that sense, our study also contradicts a recent claim that readers are unlikely to experience an auditory image of a written text, unless the text corresponds to sentences spoken by people whose voice is familiar (Kurby et al., 2009). We show that participants produce inner voice even when reading narrative with no identified speaker. Our results are more compatible with an alternative interpretation of that group, that inner voice processes are also, and in fact mostly, modulated by attention. This modulation was also suggested by Alexander and Nygaard (2008) to explain why readers tend to use auditory imagery more often when reading difficult texts.

In line with this interpretation, recent studies have proposed that reading might rely more on phonological processes as texts become more difficult to read. Phonological activation would be more active as linguistic complexity increases, or in nonproficient readers (Jobard et al., 2011), and triggered by top-down attentional processes. This suggestion is compatible with our observation that the strongest TVA activation was found in patient P2, who was not a proficient reader. It would also explain why AVI is active while reading sentences, but not while reading isolated words (Price, 2000; Jobard et al., 2007): AVI might facilitate the processing of prosodic information and the active integration of consecutive words into a sentence, using verbal working memory.

Indeed, the superior temporal sulcus and superior temporal gyrus (STG), which support AVI, play an important role in several phonological processes, such as the integration of letters into speech sounds (Vigneau et al., 2006; van Atteveldt et al., 2009), the identification of phonological word forms (Shalom and Poeppel, 2008), and the analysis of speech sounds (Hickok, 2009; Grimaldi, 2012). Furthermore, the STG has been shown to react specifically to readable stimuli, i.e., to words and pseudowords,

but not to, e.g., consonant strings (Vartiainen et al., 2011). TVA might then interact with frontal regions to support the phonological component of verbal working memory during reading, through articulatory rehearsal processes (Baddeley and Logie, 1992; Yvert et al., 2012). Those processes would allow an update and refresh of the phonological material stored in working memory (Baddeley and Logie, 1992; Smith et al., 1995). Auditory verbal Imagery would thus facilitate verbal working memory, as suggested by several authors (Smith et al., 1995; Sato et al., 2004), to process the sentence as a whole, and not as a collection of unrelated pieces. Simultaneously, the activation of phonological representations in TVA would produce the vivid experience of the inner voice (Abramson and Goldinger, 1997).

How automatic is the inner voice during sentence reading?
Our results directly relate to a long-standing debate as to whether expert readers automatically access phonological representations when reading, since it is difficult to think of a phonological representation that would not include an auditory imagery component (Alexander and Nygaard, 2008). In the three most proficient readers (P1, P3, and P4), unattended words triggered the same initial activation peak as attended words ~300 ms, followed by a fast decline, which might suggest an automatic activation of TVA in expert readers. This assumption is further supported by a recent event-related potential study (Froyen et al., 2009) showing that early and automatic letter–speech sound processing only occurs in advanced readers (Blomert and Froyen, 2010). One neural possible explanation is that learning to read might strengthen the connectivity between visual and auditory areas (Booth et al., 2008; Richardson et al., 2011) based on Hebbian plasticity: both regions would be repeatedly coactivated because of repeated associations between visual and auditory inputs during the learning period (the written word and the auditory percept of one's own voice while reading overtly). With practice, this connectivity would allow for a direct activation of the auditory cortex by visual inputs through the visual cortex, in the absence of overt speech, very much like an automatic stimulus-response association.

However, we also show that if it is the case, this connectivity can still be modulated by top-down control. Sustained inner voice activation is not an automatic process occurring systematically in response to any written word. It is clearly enhanced when participants read attentively (to understand and memorize sentences) and minimized when words are not processed attentively. Provided that high-frequency neural activity can be measured with sufficient specificity, as in patient P2, for instance, it can even be proposed as an on-line measure of attention: the engagement of attention during reading can be followed in time on a word-by-word basis (as well as covert speech, consequently). One remaining question is whether the reduced response to unattended words results from active inhibition or from the absence of phasic activation. This second interpretation is supported by several intracranial EEG studies showing that high-frequency activity in sensory cortices comprise two consecutive components: a transient peak determined by the stimulus, followed by a sustained component determined by task condition, that is, whether the stimulus contains task-relevant information (Jung et al., 2008; Ossandón et al., 2011; Bastin et al., 2012). This interpretation is supported also by the findings of the present study, showing that directed interactions from the visual to auditory cortex are significantly stronger than vice versa only during attentive reading. Therefore, words might trigger an automatic TVA re-

sponse (at least in proficient readers), which is sustained by top-down attentional processes when the word is task-relevant.

Early auditory cortex first, TVA second

Jäncke and Shah (2004) had shown that in a population trained to imagine syllables in response to flicker lights, AVI generates bilateral hemodynamic responses in TVA. However, they also suggested that during auditory imagery, the auditory system activates in reverse order, from higher order to lower order areas, and that the primary auditory cortex is not active (Bunzeck et al., 2005, for nonverbal complex stimuli). Our results qualify these two conclusions: we argue that the extent to which auditory areas activate during AVI depends on the strategy used by the subject. In patient P2, the primary auditory cortex was clearly active during AVI, and, most importantly, before higher order voice-selective areas. However, this sequence was driven by a visual stimulus in our case, and we cannot exclude that the primary auditory cortex might be driven by the visual cortex (Booth et al., 2008); in fact, our Granger Causality results support this assumption. Still, one intriguing finding of our study is that the activation of the auditory cortex can follow the response of the visual cortex by >500 ms, which means that in natural reading conditions paced by an average of four saccades per second, the auditory cortex might react with a lag of two saccades, that is, two words in most instances.

What remains to be determined is whether the brief activation of the auditory cortex after unattended words is sufficient to lead to a conscious experience of an inner voice. Indeed, showing that a voice-selective region of the auditory cortex is active is not undisputable evidence that the reader actually hears a voice. The conscious experience might arise only for longer durations, for instance, or when that activation broadcasts to a wider brain network (Baars, 2002). This will remain the main limitation of all studies in that field, until new experiments reveal the neural correlates of auditory consciousness.

References

Abramson M, Goldinger SD (1997) What the reader's eye tells the mind's ear: silent reading activates inner speech. Percept Psychophys 59:1059–1068. CrossRef Medline

Aguera P, Jerbi K, Caclin A, Bertrand O (2011) ELAN: a software package for analysis and visualization of MEG, EEG, and LFP signals. Comput Intell Neurosci 158970.

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723. CrossRef

Alexander JD, Nygaard LC (2008) Reading voices and hearing text: talker-specific auditory imagery in reading. J Exp Psychol Hum Percept Perform 34:446–459. CrossRef Medline

Baars BJ (2002) The conscious access hypothesis: origins and recent evidence. Trends Cogn Sci 6:47–52. CrossRef Medline

Baddeley AD, Logie R (1992) Auditory imagery and working memory. In: Auditory imagery (Reisberg D, ed.), pp 179–197. Hillsdale, NJ: Lawrence Erlbaum.

Bastin J, Committeri G, Kahane P, Galati G, Minotti L, Lachaux JP, Berthoz A (2012) Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. Hum Brain Mapp. Advance online publication. Retrieved Jan 30, 2012. doi:10.1002/hbm.21515. CrossRef

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312. CrossRef Medline

Blomert L, Froyen D (2010) Multi-sensory learning and learning to read. Int J Psychophysiol 77:195–204. CrossRef Medline

Booth JR, Mehdiratta N, Burman DD, Bitan T (2008) Developmental increases in effective connectivity to brain regions involved in phonological processing during tasks with orthographic demands. Brain Res 1189:78–89. CrossRef Medline

Bunzeck N, Wuestenberg T, Lutz K, Heinze HJ, Jancke L (2005) Scanning silence: mental imagery of complex sounds. Neuroimage 26:1119–1127. CrossRef Medline

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK (1997) Activation of auditory cortex during silent lipreading. Science 276:593–596. CrossRef Medline

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT (2010) Categorical speech representation in human superior temporal gyrus. Nat Neurosci 13:1428–1432. CrossRef Medline

Cohen L, Jobert A, Le Bihan D, Dehaene S (2004) Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage 23:1256–1270. CrossRef Medline

Cousin E, Peyrin C, Pichat C, Lamalle L, Le Bas JF, Baciu M (2007) Functional MRI approach for assessing hemispheric predominance of regions activated by a phonological and a semantic task. Eur J Radiol 63:274–285. CrossRef Medline

Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. Clin Neurophysiol 112:565–582. CrossRef Medline

Crone NE, Korzeniewska A, Franaszczuk PJ (2011) Cortical gamma responses: searching high and low. Int J Psychophysiol 79:9–15. CrossRef Medline

Froyen DJ, Bonte ML, van Atteveldt N, Blomert L (2009) The long road to automation: neurocognitive development of letter-speech sound processing. J Cogn Neurosci 21:567–580. CrossRef Medline

Geweke J (1982) Measurement of linear dependence and feedback between multiple time series. J Am Stat Assoc 77:304–313. CrossRef

Granger CWJ (1980) Testing for causality: a personal viewpoint. J Econ Dyn Control 2:329–352. CrossRef

Grimaldi M (2012) Toward a neural theory of language: old issues and new perspectives. J Neurolinguistics 25:304–327. CrossRef

Halpern AR, Zatorre RJ (1999) When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cereb Cortex 9:697–704. CrossRef Medline

Hickok G (2009) The functional neuroanatomy of language. Phys Life Rev 6:121–143. CrossRef Medline

Jäncke L, Shah NJ (2004) 'Hearing' syllables by 'seeing' visual stimuli. Eur J Neurosci 19:2603–2608. CrossRef Medline

Jerbi K, Ossandón T, Hamamé, CM, Senova S, Dalal SS, Jung J (2009) Task-related gamma-band dynamics from an intracerebral perspective: review and implications for surface EEG and MEG. Hum Brain Mapp 30:1758–1771. CrossRef Medline

Jobard G, Vigneau M, Simon G, Tzourio-Mazoyer N (2011) The weight of skill: interindividual variability of reading related brain activation patterns in fluent readers. J Neurolinguistics 24:113–132. CrossRef

Jung J, Mainy N, Kahane P, Minotti L, Hoffmann D, Bertrand O, Lachaux JP (2008) The neural bases of attentive reading. Hum Brain Mapp 29:1193–1206. CrossRef Medline

Kahane P, Minotti L, Hoffmann D, Lachaux J-P, Ryvlin P (2003) Invasive EEG in the definition of the seizure onset zone: depth electrodes. In: Handbook of Clinical Neurophysiology, Vol 3 (Rosenow F, Lüders HO, eds) pp 109–133. Amsterdam: Elsevier.

Khateb A, Pegna AJ, Landis T, Michel CM, Brunet D, Seghier ML, Annoni JM (2007) Rhyme processing in the brain: an ERP mapping study. Int J Psychophysiol 63:240–250. CrossRef Medline

Kurby CA, Magliano JP, Rapp DN (2009) Those voices in your head: activation of auditory images during reading. Cognition 112:457–461. CrossRef Medline

Lachaux JP, Rudrauf D, Kahane P (2003) Intracranial EEG and human brain mapping. J Physiol (Paris) 97:613–628. CrossRef

Lachaux JP, Jerbi K, Bertrand O, Minotti L, Hoffmann D, Schoendorff B, Kahane P (2007) A blueprint for real-time functional mapping via human intracranial recordings. PLoS One 2:e1094. CrossRef Medline

Lachaux JP, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301. CrossRef Medline

Le Van Quyen M, Foucher J, Lachaux JP, Rodriguez E, Lutz A, Martinerie J, Varela FJ (2001) Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. J Neurosci Methods 111:83–98. CrossRef Medline

Lennox, BR, Bert, S, Park, G, Jones, PB, Morris, PG (1999) Spatial and tem-

poral mapping of neural activity associated with auditory hallucinations. Lancet 353.

Linden DE, Thornton K, Kuswanto CN, Johnston SJ, van de Ven V, Jackson MC (2011) The brain's voices: comparing nonclinical auditory hallucinations and imagery. Cereb Cortex 21:330–337. CrossRef Medline

Mainy N, Jung J, Baciu M, Kahane P, Schoendorff B, Minotti L, Hoffmann D, Bertrand O, Lachaux JP (2008) Cortical dynamics of word recognition. Hum Brain Mapp 29:1215–1230. CrossRef Medline

Nobre AC, Allison T, McCarthy G (1998) Modulation of human extrastriate visual processing by selective attention to colours and words. Brain 121:1357–1368. CrossRef Medline

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113. CrossRef Medline

Ossandón T, Jerbi K, Vidal JR, Bayle DJ, Henaff MA, Jung J (2011) Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance. J Neurosci 31:14521–14530. CrossRef Medline

Pernet C, Charest I, Bélizaire G, Zatorre RJ, Belin P (2007) The temporal voice areas: spatial characterization and variability. 13th International Conference on Functional Mapping of the Human Brain, Chicago, IL. Neuroimage 36 [Supp 1].

Perrone-Bertolotti M, Pichat C, Le Bas JF, Baciu A, Baciu M (2011) Functional MRI evidence for modulation of cerebral activity by grapheme-to-phoneme conversion in French, and by the variable of gender. J Neurolinguistics 24:507–520. CrossRef

Price C (2000) The anatomy of language: contributions from functional neuroimaging. J Anat 197:335–359. CrossRef Medline

Richardson FM, Seghier ML, Leff AP, Thomas MS, Price CJ (2011) Multiple routes from occipital to temporal cortices during reading. J Neurosci 31:8239–8247. CrossRef Medline

Sato M, Baciu M, Loevenbruck H, Schwartz JL, Cathiard MA, Segebarth C, Abry C (2004) Multistable representation of speech forms: a functional MRI study of verbal transformation. Neuroimage 23:1143–1151. CrossRef Medline

Seth AK (2005) Causal connectivity of evolved neural networks during behavior. Network 16:35–54. CrossRef Medline

Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. J Neurosci Methods 186:262–273. CrossRef Medline

Shalom DB, Poeppel D (2008) Functional anatomic models of language: assembling the pieces. Neuroscientist 14:119–127. Medline

Smith JD, Wilson M, Reisberg D (1995) The role of subvocalization in auditory imagery. Neuropsychologia 33:1433–1454. CrossRef Medline

Talairach J, Tournoux P (1988) Co-Planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme.

Tallon-Baudry C, Bertrand O, Delpuech C, Permier J (1997) Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. J Neurosci 17:722–734. Medline

Towle VL, Yoon HA, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire JP, Kohrman MH (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain 131:2013–2027. CrossRef Medline

van Atteveldt N, Roebroeck A, Goebel R (2009) Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. Hear Res 258:152–164. CrossRef

Vartiainen J, Liljeström M, Koskinen M, Renvall H, Salmelin R (2011) Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. J Neurosci 31:1048–1058. CrossRef Medline

Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, et al. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. Neuroimage 30:1414–1432. CrossRef Medline

Yao B, Belin P, Scheepers C (2011) Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. J Cogn Neurosci 23:3146–3152. CrossRef Medline

Yvert G, Perrone-Bertolotti M, Baciu M, David O (2012) Dynamic causal modeling of spatiotemporal integration of phonological and semantic processes: an electroencephalographic study. J Neurosci 32:4297–4306. CrossRef Medline