Cerebral Cortex January 2006;16:115–123
doi:10.1093/cercor/bhi091
Advance Access publication April 13, 2005

Time Course of Top-down and Bottom-up Influences on Syllable Processing in the Auditory Cortex

Milene Bonte[1,2], Tiina Parviainen[2], Kaisa Hytönen[2] and Riitta Salmelin[2]

[1]Department of Cognitive Neuroscience, Faculty of Psychology, University of Maastricht, PO Box 616, 6200 MD, Maastricht, The Netherlands and [2]Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, FIN-02015, HUT, Espoo, Finland

In speech perception, extraction of meaning from complex streams of sounds is surprisingly fast and efficient. By tracking the neural time course of syllable processing with magnetoencephalography we show that this continuous construction of meaning-based representations is aided by both top-down (context-based) expectations and bottom-up (acoustic–phonetic) cues in the speech signal. Syllables elicited a sustained response at 200–600 ms (N400m) which became most similar to that evoked by words when the expectation for meaningful speech was increased by presenting the syllables among words and sentences or using sentence-initial syllables. This word-like cortical processing of meaningless syllables emerged at the build-up of the N400m response, 200–300 ms after speech onset, during the transition from perceptual to lexical–semantic analysis. These findings show that the efficiency of meaning-based analysis of speech is subserved by a cortical system finely tuned to lexically relevant acoustic–phonetic and contextual cues.

Keywords: language, MEG, N400, speech comprehension

Introduction

The brain system underlying spoken language comprehension is exposed to a continuously changing stream of speech sounds from which meaning must be extracted. Linguistic context and acoustic–phonetic features of the speech signal provide top-down and bottom-up cues that may be used to predict incoming information, enabling the fast and automatic recognition of up to 150 words per minute in healthy adults. The present study utilizes the high temporal resolution of magnetoencephalography (MEG) to investigate the influence of lexical–semantic context (top-down information) and acoustic–phonetic cues (bottom-up information) on the time course of the neural processing of natural speech.

MEG and electroencephalography (EEG) methods have been crucial in characterizing the time course of neural systems involved in different aspects of speech processing, including those related to the processing of acoustic–phonetic (Eggermont and Ponton, 2002), phonological and semantic information (Kutas and Schmitt, 2003). MEG studies have shown that acoustic–phonetic features of speech modulate activity in non-primary auditory cortex from 50–100 ms onwards, as reflected in a robust response that emerges 100 ms after sound onset and is usually referred to as the N100/N100m (Kuriki and Murase, 1989; Poeppel *et al.*, 1996; Obleser *et al.*, 2004; Parviainen *et al.*, 2005). Converging evidence from hemodynamic brain imaging studies suggests that the neural processing of acoustic–phonetic features specifically involves posterior superior temporal areas in the left hemisphere (Hickok and Poeppel, 2000; Scott and Johnsrude, 2003). The onset of language specific phonetic–phonological analysis has been estimated at ~100–200 ms. In this time window, an MEG/EEG response associated with mnemonic functions of the auditory association cortex, i.e. the mismatch negativity (MMN), indicates access to phonological categories (Phillips *et al.*, 2000; Vihla *et al.*, 2000), and distinct processing of native versus nonnative phonetic contrasts (Näätänen *et al.*, 1997; Cheour *et al.,* 1998; Winkler *et al.*, 1999). This is the approximate level to which cortical analysis proceeds when the stimuli are small sets of synthetic vowels or consonant–vowel (CV) syllables presented in passive paradigms or using simple perceptual tasks, as has typically been the case in these previous studies of early perceptual aspects of speech processing.

Words and word-like speech stimuli further evoke a sustained activation that starts at ~200 ms after stimulus onset, reaches a maximum at ~400 ms and lasts until 600–800 ms (Kutas and Federmeier, 2000). MEG reports based on equivalent current dipole (ECD) modeling associate this so-called N400/N400m response (Kutas and Hillyard, 1980) with activation of the superior temporal cortex in the immediate vicinity of the auditory cortex (Helenius *et al.*, 2002; Kujala *et al.*, 2004). Distributed source modeling of MEG data (Marinkovic *et al.*, 2003) suggests that neural activity underlying the N400 response may additionally extend into (left) anterior temporal and frontal areas.

The N400 response probably reflects multiple processes, ranging from phonological analysis to lexical access and semantic processing. Facilitating factors like semantic or phonological priming tend to reduce the N400 amplitude (Van Petten *et al.*, 1999; Dumay *et al.*, 2001; Helenius *et al.*, 2002; Perrin and Garcia-Larrea, 2003; Bonte and Blomert, 2004). A review of previous studies suggests a natural division at ~350–400 ms, around the N400 maximum, reflecting a gradual shift from predominantly phonological to predominantly semantic processing. Evidence for initial phonological analysis and lexical access in the onset window of the N400 comes from studies using a variant of the classical N400 sentence paradigm where the final word is semantically wrong (Kutas and Hillyard, 1980) but shares its initial phonemes with the expected word. This initial phonological congruency results in a delayed onset of the semantic N400 effect (Van Petten *et al.*, 1999; Helenius *et al.*, 2002) or in a separate event-related response around 200–350 ms, preceding the N400 (Connolly and Phillips, 1994; Hagoort and Brown, 2000; Van den Brink *et al.*, 2001). Furthermore, recent studies on the neural time course of spoken word recognition in dyslexics, who experience difficulties in phonological processing, have demonstrated specific abnormalities at 100–300 ms, prior to and during the onset of the N400, but not in later N400 windows (Helenius *et al.*, 2002; Bonte and

Blomert, 2004). These later windows probably reflect further lexical processing. For example, the latency of the N400 response maximum has been shown to depend on the time point at which the acoustic signal can only represent one particular word, i.e. the latency is delayed when this recognition point occurs later within the word (O'Rourke and Holcomb, 2002).

In the present MEG study we examined in detail the cortical processing of syllables in the time window from 200 to 350 ms, when the utterance is processed as speech but its lexical–semantic content has not yet been established. Our stimuli of interest were natural Finnish CV syllables, potentially meaningful, but only if followed by further speech input. We assumed that increasing the expectation for a meaningful utterance would result in an increasingly word-like sustained N400m response to these syllables. In natural speech, expectation may be built both by the linguistic context and by subtle acoustic–phonetic cues in the utterances. Here, we studied the influence of linguistic context (top-down) by varying the probability that the syllable was part of a meaningful utterance, i.e. CV syllables were presented in two different contexts, together with complete sentences and sentence-initial words (*context*) and as a separate sequence of syllables only (*isolation*). We studied the influence of acoustic–phonetic cues (bottom-up) by comparing CV syllables pronounced separately (*syl*) and CV syllables cut from the beginning of complete sentences (*sylsent*) (Fig. 1). These two types of syllables contained different acoustic–phonetic cues that may signal the absence or presence of subsequent speech input.

Materials and Methods

Subjects

Ten healthy Finnish-speaking members of laboratory personnel (five females; 23–29 years old, mean 25.5 years) took part in the study. Nine subjects were right-handed, one ambidextrous. None of the subjects had a history of hearing loss or neurological abnormalities. Informed consent was obtained from all subjects, in agreement with the prior approval of the Helsinki and Uusimaa Ethics Committee.

Stimuli

Stimuli were 31 meaningless Finnish CV syllables, e.g. 'ki', 62 monomorphemic bisyllabic words (CVCV) starting with the same set of syllables, e.g. 'kivi' ('stone'), and 124 sentences starting with the same set of words, e.g. 'kivi putoaa maahan' ('The stone falls on the ground'). In the total stimulus set, two words shared the same initial CV syllable and two sentences shared the same initial word (see Fig. 1). In this way, subjects would not expect one specific word or sentence upon hearing a syllable or word. We used two types of syllable stimuli. Words and one type of syllable stimuli (*sylsent*) were cut from the speech signal of the 124 sentences, resulting in two utterances of each word (2 × 62) and four utterances of each syllable (4 × 31). The second type of syllable stimuli (*syl*) consisted of the same 31 syllables pronounced separately (again, four utterances each).

The stimuli were spoken by a male native Finnish speaker and recorded at a sampling rate of 44.01 kHz on a DAT recorder in an anechoic chamber (Acoustics Laboratory, Helsinki University of Technology). The digitized stimuli were D/A converted with a 16-bit resolution, bandpass filtered (80 Hz to 10.5 kHz) and resampled at 22.05 kHz. We used a speech waveform editor (PRAAT 4.0: Boersma and Weenink, 2002) to determine acoustic onsets and offsets of syllables, words and sentences. For syllables and words cut from sentences, amplitude over the final 10 ms was tapered to zero to avoid acoustic transients (clicks) that would be created by a sharp cut-off. The overall sound intensity level was numerically equated across stimuli to generate equal rms values. The mean ± SD acoustical duration of syllables pronounced separately (*syl*) was 212 ± 45 ms, sentence-initial syllables (*sylsent*) 129 ± 45 ms, words 275 ± 48 ms and sentences 1490 ± 171 ms. Stimulus length differed significantly between the two syllable types ($P < 0.001$), i.e. *sylsent* were on average 83 ms shorter than *syl*. Furthermore, mean pitch was significantly higher for *sylsent* = 143 ± 53 Hz than for *syl* = 124 ± 13 Hz ($P < 0.001$).

Experimental Design and Procedure

Bottom-up effects on syllable processing were studied by comparing the syllables pronounced separately (*syl*) versus sentence-initial syllables (*sylsent*). Top-down effects were investigated by comparing the processing of these syllables in two types of experimental blocks: in one block both syllable types were presented together with words and sentences (*context*) and in two blocks each syllable type was presented in isolation (*isolation*). In the three experimental blocks, stimuli were presented in a pseudo-random order, i.e. two consecutive stimuli were not allowed to form an existing word or sentence and there had to be at least five intervening stimuli between the repetition of identical syllables or words. In order to maintain a stable attention level across the experimental blocks, the subject's task was to repeat the previous stimulus (which could be any of the stimulus types) whenever they heard a beep signal (1 kHz tone). The beep signal occurred in 6.5% of the trials (about once every 16 stimuli). Subject's responses were monitored on-line by the experimenter. All subjects correctly repeated the stimuli. The stimulus following a beep signal was not included in the analysis. Stimuli were presented binaurally at a comfortable listening level. The interstimulus interval (ISI) was 2 s for two consecutive

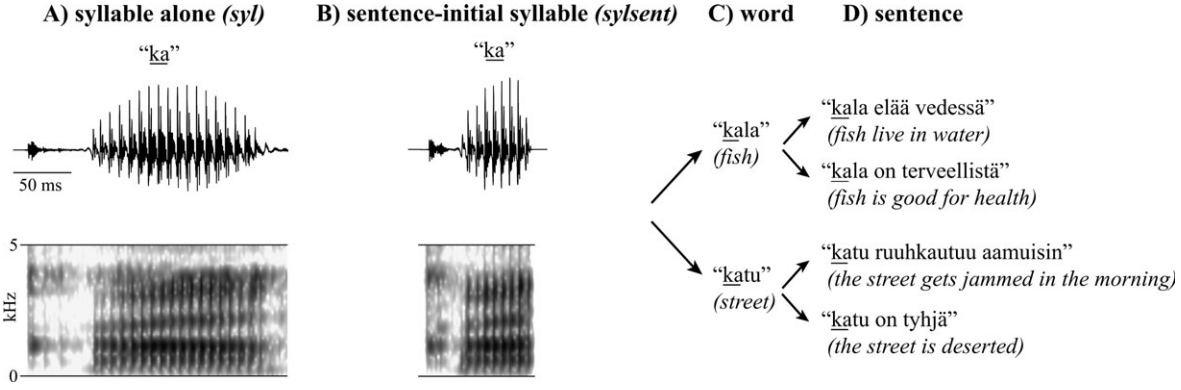

Figure 1. Stimuli used in the study. (*A*) Syllables pronounced separately (*syl*), (*B*) sentence-initial syllables (*sylsent*), (*C*) words and (*D*) sentences. In the complete stimulus set, two words shared the same initial syllable and two sentences the same initial word. Examples of waveforms (upper rows) and spectrograms (lower rows) are given for each syllable type.

experimental stimuli and 4 s after a beep signal. The total measurement time was ~40 min (20 min context block and 2 × 7 min isolation blocks). Subjects were given a short break every 5–8 min.

MEG Recording and Data Analysis

MEG recordings were conducted in a magnetically shielded room using a Vectorview™ whole-head system (Neuromag Ltd, Helsinki, Finland). The device contains 102 triple sensor elements composed of two orthogonal planar gradiometers and one magnetometer. The signals were bandpass filtered at 0.03–200 Hz and digitized at 600 Hz. The raw data was stored for off-line analysis. MEG signals were averaged on-line across trials, over an interval ranging from 200 ms before until 800 ms after stimulus onset. During the measurement, horizontal and vertical eye movements were monitored and trials with MEG or EOG signal amplitude exceeding 3000 fT/cm or ±150 μV, respectively, were discarded. At least 100 artifact-free trials were collected for each stimulus category. The averaged MEG responses were baseline corrected to the 200 ms interval immediately preceding the stimulus onset and low-pass filtered at 40 Hz.

To obtain an initial overview of the results, we calculated areal mean signals of (i) four gradiometer pairs over the left temporal lobe and (ii) four gradiometer pairs over the right temporal lobe that showed the strongest response. We first computed vector sums by squaring the MEG signals of each gradiometer pair, summing these signals together and then calculating the square root of this sum. The areal mean signals were computed by averaging these vector sums for each area of interest (left and right temporal lobe). The areal mean signals were computed from 400 ms before to 2000 ms after stimulus onset, individually for each subject. Finally, we calculated overall group averages. Because of the way the sensor-level areal mean signals are calculated (square root of sum of squared signals), they always have a positive value (>0).

The main analysis involved an individual estimation of the time course of neural activity in distinct brain areas using Equivalent Current Dipole (ECD) analysis (Hämäläinen *et al.*, 1993). The ECD analysis was performed up to 800 ms after stimulus onset. An ECD represents the mean location and strength of activation in a given brain area and the orientation of current flow therein. Dipoles were localized individually for each subject using a subset of planar gradiometers that ideally covered the distinct magnetic field patterns. After ECDs had been localized they were included into a multidipole model and, keeping their orientation fixed, their amplitudes were allowed to be adjusted to achieve maximum explanation for the measured whole-head data. All dipoles included in the model could be localized reliably, with goodness-of-fit values exceeding 80–90%. The final models were composed of 2–5 ECDs (mean = 4). In each individual, the same set of ECDs accounted for the pattern of auditory cortical activation evoked by all stimulus categories up to and including the N400m window. The ECDs explaining the field patterns around 100 ms (N100m) and 400 ms (N400m) were very similar both in location (mean Euclidean distance = 8 mm) and orientation (mean difference in orientation = 8°). In order to prevent spurious interactions between these two ECDs, both of these source areas were represented by a single ECD (at N400m) in the multidipole model.

The location of the ECDs was defined in head coordinates that were set by the nasion and two reference points anterior to the ear canals: the *x*-axis is directed from the left (negative) to the right (positive) preauricular point, the *y*-axis towards the nasion and the *z*-axis towards the vertex. Prior to the MEG measurement, the locations of four Head Position Indicator (HPI) coils attached to the subject's head were measured with a three-dimensional digitizer (Polhemus, Colchester, VT). Before each MEG session, the HPI coils were briefly energized to determine their location with respect to the MEG helmet.

For visualization purposes, the MR images of the individual subjects' brains were transformed into that of one representative subject (elastic transformation: Schormann *et al.*, 1996; Woods *et al.*, 1998). The individual ECDs were transformed accordingly to display the sources in a common coordinate system.

Strength and timing of the activation in the source areas as represented by the time course of the ECDs (source waveforms) were analyzed using repeated-measures analysis of variance (ANOVA), with linguistic context (*context* versus *isolation*) and syllable type (*sylsent* versus *syl*) as within-subject factors. Hemispheric differences were tested using hemisphere (left versus right) as an additional within-subject factor; this comparison was justified by first verifying that the distance of the ECDs from the centre of the head (and the MEG sensors) did not differ between the hemispheres. Estimation of activation strength included maximum activation and area under the ascending and descending slope of the N400m response. Area measures were calculated individually for each subject and separately for each condition. In the left hemisphere, the ascending window was defined as the time window between the latency at which the ascending slope reached 25% of the maximum activation (mean 260 ms) and the latency at the maximum. In the right hemisphere, where the N400m often started directly at the N100m response, the ascending window was defined as the time window between 200 ms and the latency at the maximum. In both hemispheres, the descending window was defined as the time window between the latency at the maximum and the latency at which the descending slope reached 25% of the maximum amplitude. Latency values included latency at the maximum activation, onset latency (latency at 25% of the maximum at the ascending slope), and the latencies at 50% of the maximum activation at the ascending and the descending slopes of the N400m.

Results

Overall Effects: Areal Mean Signals

Figure 2 displays the overall time course of MEG signals averaged across subjects. Both at the group level and in the

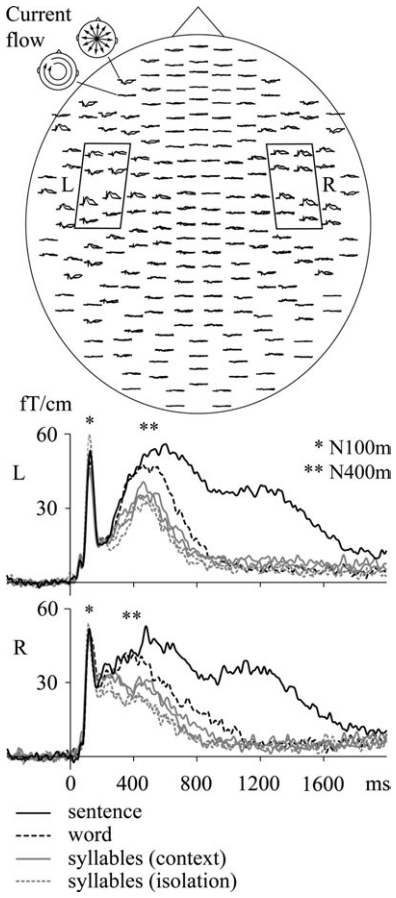

Figure 2. Grand average waveforms and areal mean signals of all 10 subjects. Areal mean signals were calculated for two sets of four MEG gradiometer pairs, indicated by parallelograms, over the left (L) and right (R) temporal lobes. Syllables include both syllable types (*syl* and *sylsent*) presented in the context of words and sentences (*context*) and as a separate sequence of syllables only (*isolation*).

individual subjects, activity concentrated over the left and right temporal lobes. The areal mean signals in these regions of interest are plotted in the lower panel of Figure 2. All stimuli elicited a clear and comparable N100m followed by a sustained response that reached the maximum ~450 ms after stimulus onset. Sentences elicited a strong sustained activity, including a second maximum at 1100–1200 ms, which returned to the baseline at ~1750 ms after stimulus onset, i.e. on average 250 ms after sentence ending.

Visual inspection of the areal mean signals elicited by the different stimuli suggested specific response characteristics and time windows of interest for further statistical testing on the time courses of activation at the source level. First, the largest experimental effects seemed to occur between 200 and 800 ms, in the time window of the sustained response. Second, the sustained response elicited by syllables, words and sentences seemed to show differences in both activation strength and timing. As expected, the sustained response was smallest and of shortest duration for syllables and was increased for words and sentences. Third, visual inspection suggested that the different stimulus conditions resulted in dissimilar patterns in the ascending versus descending slopes of the sustained response. Fourth, in the right hemisphere the sustained activity often started directly at the N100m response, rendering the definition of an N400m response less straightforward than in the left hemisphere. The areal mean signals thus suggest hemispheric differences in stimulus processing.

Overall Effects: Field Patterns and Dipole Models

Figure 3 shows a typical sequence of activation elicited by words in the left and right hemispheres, in one subject. There were clear dipolar field patterns at ~100, ~200 and ~400 ms. A similar sequence of MEG activity was obtained for syllables and sentences. The bilateral dipolar fields at ~100 ms (N100m) were characterized by a downward orientation of current flow perpendicular to the Sylvian fissure in both hemispheres, similarly in all subjects. ECD analysis indicated that this signal was generated by activation immediately posterior to Heschl's gyrus.

In the left hemisphere, the N100m was followed by a field pattern at ~200 ms that typically reflected a strong posterior temporal source with the current flow oriented anteriorly and inferiorly, almost perpendicular to the direction of current flow in the N100m time window. Occasionally, the field pattern also suggested presence of a weaker inferior frontal component in this same time window, with an anterior–superior direction of current flow. In the right hemisphere, the field patterns showed more inter-subject variability in this time window but most often indicated an anterior temporal source. It was possible to localize an ECD at ~200 ms in seven subjects in the left hemisphere and in nine subjects in the right hemisphere. A relatively large inter-subject variability in the location and orientation of these ECDs suggests that they reflected activity of a widespread network of brain areas. The corresponding source waveforms did not differentiate between stimulus conditions and are therefore not included in the further statistical analysis of top-down and bottom-up effects on syllable processing.

During the sustained activity peaking at ~400 ms, all subjects showed clear bilateral field patterns with the current flow downward perpendicular to the Sylvian fissure (Fig. 3). Figure 4 depicts the locations of the corresponding ECDs in all 10

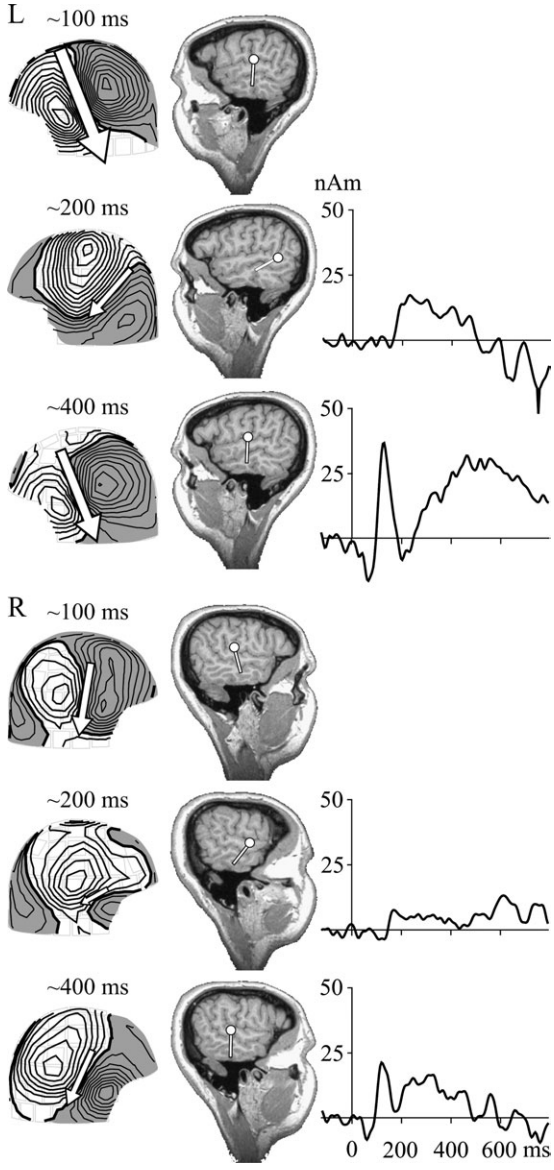

Figure 3. Typical MEG field patterns, equivalent current dipole (ECD) localization and corresponding source waveforms in response to words (single subject). All subjects showed similar field patterns and dipole localization in the N400m window. The sources of N100m and N400m were very similar in location and orientation. The N400m source thus accounted for most of the activity in the earlier N100m window. An ECD could also be reliably determined around 200 ms in seven subjects in the left hemisphere (L) and in nine subjects in the right hemisphere (R). The localization of this source showed large inter-subject variability, especially in the right hemisphere.

subjects. The sources clustered around the left and right posterior superior temporal gyrus, on average 3 mm medially to the sources of the N100m [left $t(9) = 3.9$, $P < 0.005$; right $t(9) = 2.4$, $P < 0.05$] but with no systematic differences along the anterior–posterior and superior–inferior axes. The orientation of current flow in the N100m and N400m time windows was essentially identical. Accordingly, the N400m sources also explained a major part of the N100m activity (see Fig. 3 and Materials and Methods). The location, orientation and time course of the sources of the sustained response suggested that it corresponded to the N400m reported in earlier MEG studies of semantic processing (e.g. Helenius *et al.*, 2002). The N400m source waveforms for syllables, words and sentences (Fig. 4)

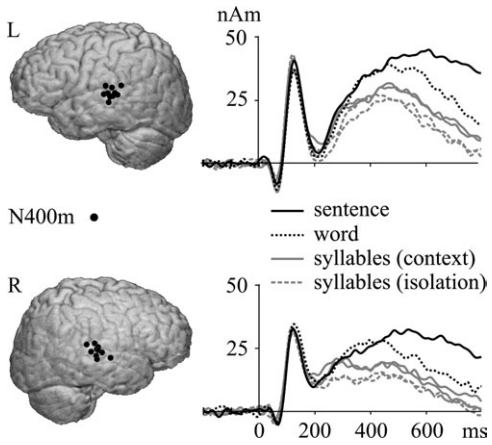

Figure 4. Locations and mean time course of the N400m sources. Black dots indicate the individual sources for all 10 subjects in the left (L) and right (R) posterior superior temporal areas. Syllables include both syllable types (*syl* and *sylsent*) presented in the context of words and sentences (*context*) and as a separate sequence of syllables only (*isolation*).

showed a pattern that was comparable to that of the areal mean signals over the left and right temporal lobes (Fig. 2).

Top-down Effects on Syllable Processing: Context versus Isolation

Figure 5*A* depicts the mean time course of activation in the left and right N100m/N400m source area when syllables cut from sentences (*sylsent*) and uttered separately (*syl*) were presented in the context of words and sentences (*context*) and when they were presented in isolation (*isolation*). The activation elicited by words is included for reference. Figure 5*B* shows the mean difference of the source waveforms (*context – isolation*) for each syllable type, indicating that there was stronger activation to syllables presented in context than in isolation starting at ~200 ms and reaching the maximum at ~280 ms, similarly in both hemispheres. In the left hemisphere, this enhancement was preceded by an opposite influence of context at ~100 ms, i.e. a weaker response to syllables when presented in context than isolation.

The N100m peak activation strength and latency were tested with a 2 (linguistic context) × 2 (syllable type) × 2 (hemisphere) repeated-measures ANOVA. Syllables evoked a weaker N100m response in context than isolation in the left hemisphere, but in the right hemisphere the responses were equal (Fig. 6*A*), as indicated by a significant context-by-hemisphere interaction [$F(1,9) = 6.2$, $P < 0.05$] and main effect of context in the left hemisphere [$F(1,9) = 8.2$, $P < 0.025$]. The peak latency of the N100m (Fig. 6*A*), for *sylsent* stimuli in the right hemisphere, was ~10 ms later in *context* than in *isolation* blocks [context × syllable type, $F(1,9) = 6.0$, $P < 0.05$, post-hoc *t*-test for *sylsent*: *context* versus *isolation* $t(9) = 2.9$, $P < 0.025$].

The activation strength and latency at the N400m response maximum were determined in both hemispheres and tested with a 2 (linguistic context) × 2 (syllable type) × 2 (hemisphere) repeated-measures ANOVA. In the left hemisphere, where the N400m response was clearly separate from the preceding N100m response, it was also possible to collect the latencies at the onset, and at 50% of the peak level on the ascending (+50%) and descending (–50%) slopes; the data were tested with a 2 (linguistic context) × 2 (syllable type) repeated-

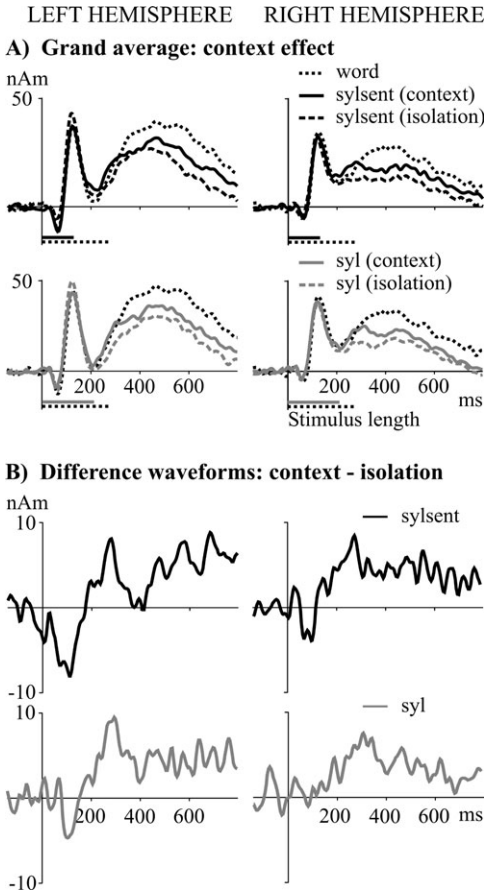

Figure 5. Top-down context effects on the mean time course of activation in the left and right N100m/N400m source area. (*A*) Grand average source waveforms of sentence-initial syllables (*sylsent*) and syllables uttered separately (*syl*) when presented in the context of words and sentences (*context*) and in isolation (*isolation*). The activation elicited by words is included for reference. The horizontal bars below each graph indicate the mean length of the stimuli. (*B*) Mean difference of the source waveforms (*context – isolation*) for each syllable type.

measures ANOVA. As depicted in Figure 6*B*, syllables evoked stronger N400m activity in the left than right hemisphere [main effect of hemisphere, $F(1,9) = 5.5$, $P < 0.05$]. Syllables also showed a stronger N400m response in *context* versus *isolation* blocks [main effect of linguistic context, $F(1,9) = 10.3$, $P < 0.025$]. This context effect was significant in the right hemisphere [$F(1,9) = 12.6$, $P < 0.01$], with a similar trend in the left hemisphere [$F(1,9) = 3.7$, $P < 0.10$].

As for the timing (Fig. 6*B*), the N400m response in the left hemisphere started earlier and lasted longer for syllables presented in *context* than *isolation* blocks [main effect of context: onset latency, $F(1,9) = 20.2$, $P < 0.005$; latency at 50% of the ascending flank, $F(1,9) = 4.9$, $P = 0.05$, and at 50% of the descending flank, $F(1,9) = 11.0$, $P < 0.01$]. Context did not influence the N400m peak latency in either hemisphere. The N400m response tended to reach the peak later in the left than right hemisphere [$F(1,9) = 3.3$, $P = 0.10$].

We also tested the ascending and descending slopes of the N400m separately by calculating the area under each flank in the left and right hemisphere (activation times duration; see Materials and Methods). Figure 8 (context and isolation columns) illustrates that syllables evoked a stronger response when presented in context than in isolation, in both time

windows and in both hemispheres [main effect of context: ascending window, $F(1,9) = 11.3$, $P < 0.01$; descending window, $F(1,9) = 17.9$, $P < 0.005$].

Interestingly, in the linguistic context, but not in isolation, the MEG signal elicited by both types of syllables seemed to follow the ascending flank of the N400m elicited by words until ~300 ms (Fig. 5*A*). In order to test this observation, we calculated the slope of the MEG signal at 200–300 ms for each individual subject. A one-way ANOVA (three levels: words, *context*, *isolation*) in the left hemisphere revealed a significant difference between conditions [$F(1,9) = 3.3$, $P < 0.05$]. Post-hoc *t*-tests showed that the slope of the activity elicited by words was significantly different from those of syllables in isolation ($P < 0.05$) but not from those of syllables in context ($P = 0.703$). A similar difference in the right hemisphere did not reach significance because of larger variability.

In sum, top-down influence of linguistic context led to a stronger bilateral N400m with an additional effect on N400m latency in the left hemisphere, i.e. an earlier N400m onset and a longer duration. Strikingly, the MEG signal elicited by syllables followed the signal elicited by words along the ascending flank of the N400m only when syllables were presented in the linguistic context and not in isolation.

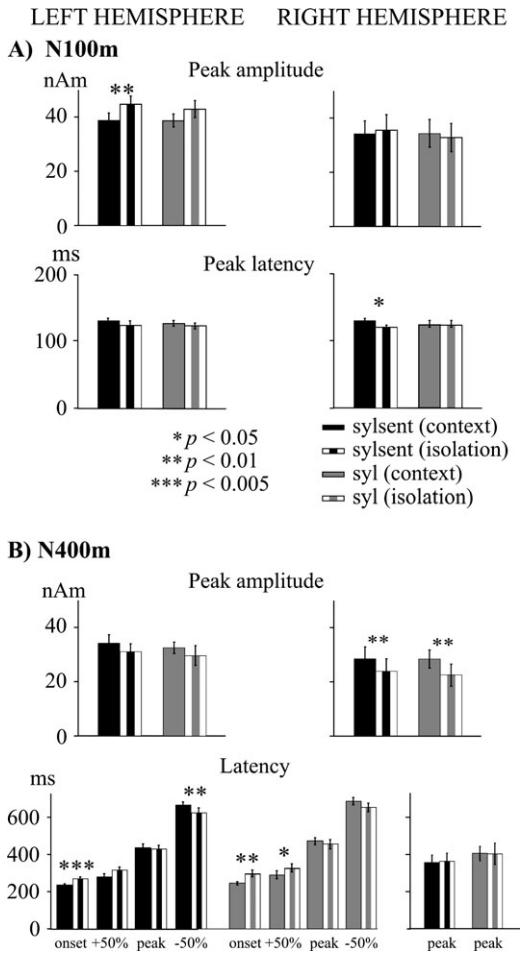

Figure 6. Activation strength and latency characteristics of the N100m and N400m. Mean amplitude and latency values of the (*A*) N100m and (*B*) N400m are given for both syllable types (*sylsent* and *syl*) in *context* and *isolation* blocks in the left and right hemisphere. Error bars represent SEM. Asterisks indicate significant differences between conditions (post-hoc *t*-comparisons).

Bottom-up Effects on Syllable Processing: Sylsent versus Syl

Figure 7*A* illustrates the influence of bottom-up information by comparing the time course of the N100m/N400m source waveforms elicited by sentence-initial syllables (*sylsent*) versus syllables pronounced separately (*syl*). The activation elicited by words is included for reference. Figure 7*B* shows the mean difference of the source waveforms (*sylsent-syl*) for the context and isolation blocks, again concentrating around 200–300 ms but clearly weaker than for the top-down influence (cf. Fig. 5).

The N100m activation strength showed no effects of syllable type (Fig. 6*A*, black versus grey bars). As for timing, the right-hemisphere N100m was ~6 ms earlier to *sylsent* than *syl* stimuli but only when they were presented in isolation [context-by-syllable type interaction, $F(1,9) = 6.0$, $P < 0.05$, post-hoc *t*-test for *sylsent* versus *syl*, $t(9) = 3.7$, $P = 0.005$]. The maximum strength of the N400m response (Fig. 6*B*, black versus grey bars) did not differ between the syllable types. However, the peak latency of the N400m was ~29 ms shorter for *sylsent* than *syl*, regardless of the context [main effect of syllable type, $F(1,9) = 5.3$, $P < 0.05$]. This effect was significant in the left hemisphere

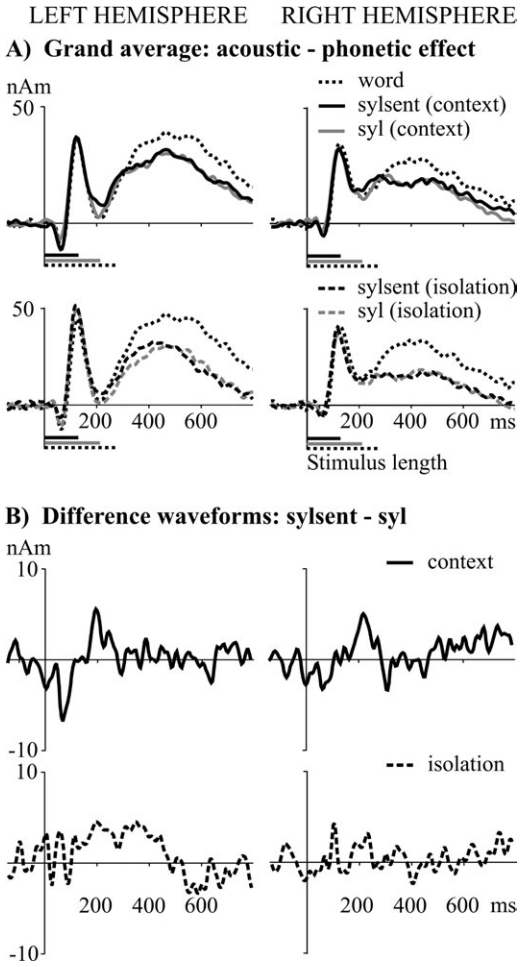

Figure 7. Bottom-up effects of acoustic–phonetic cues on the mean time course of activation in the left and right N100m/N400m source area. (*A*) Grand average source waveforms of sentence-initial syllables (*sylsent*) and syllables uttered separately (*syl*) in *context* and *isolation* blocks. The activation elicited by words is included for reference. The horizontal bars below each graph indicate the mean length of the stimuli. (*B*) Mean difference of the source waveforms (*sylsent – syl*) for the *context* and *isolation* blocks.

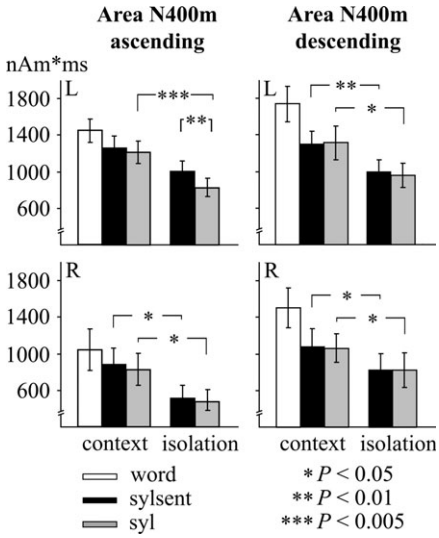

Figure 8. Mean area under the ascending and descending parts of the N400m source waveform. Mean area (activation strength multiplied by time) was determined separately in the left (L) and right (R) hemisphere. Error bars represent standard error of the mean. Asterisks indicate significant differences between conditions (post-hoc *t*-comparisons).

[$F(1,9) = 6.0$, $P < 0.05$] and showed a trend in the same direction in the right hemisphere [$F(1,9) = 3.3$, $P = 0.10$].

The separate test of area under the ascending and descending slopes of the N400m response (Fig. 8, black versus grey bars) showed that, in the left hemisphere, *sylsent* elicited a significantly stronger response than *syl* in the ascending window of the N400m, whereas no significant effect of syllable type was found in the right hemisphere [main effect of syllable type, $F(1,9) = 5.1$, $P < 0.05$; hemisphere-by-context-by-syllable type interaction, $F(1,9) = 4.1$, $P = 0.07$; main effects of syllable type in the left hemisphere, $F(1,9) = 5.2$, $P < 0.05$, and in the right hemisphere, $F(1,9) = 1.1$, NS). The presence of both bottom-up and top-down factors, i.e. *sylsent* in context, led to the largest, most word-like response, whereas the absence of both factors, i.e. *syl* in isolation, led to the smallest, least word-like response. When syllables were presented in the linguistic context the bottom-up effect appeared to be partly occluded by the stronger top-down effect. Responses to *sylsent* were indeed significantly larger than responses to *syl* when they were presented in isolation [$t(9) = 3.4$, $P < 0.01$] but not when presented in the linguistic context [$t(9) = 0.6$, NS].

In sum, bottom-up information of acoustic–phonetic cues in *sylsent* specifically increased neural activity during the build-up of the N400m response in the left hemisphere. Furthermore, this bottom-up modulation was strongest when syllables were presented in isolation, that is, without the additional presence of the top-down influence of linguistic context.

Discussion

We investigated neural correlates of speech comprehension by tracking the time course of MEG activation during attentive processing of potentially meaningful syllables as compared with that of words and sentences. Words evoked the expected sequence of cortical activity, characterized by two prominent MEG responses in bilateral superior temporal areas: a transient activity around 100 ms (N100m) and a sustained activity at 200–600 ms (N400m), reflecting the progression from perceptual to lexical–semantic analysis (Helenius *et al.*, 2002; Marinkovic *et al.*, 2003). Syllables did not only elicit a clear N100m response (Kuriki and Murase, 1989; Poeppel *et al.*, 1996; Obleser *et al.*, 2003; Parviainen *et al.*, 2005), but also a sustained N400m activity, albeit with a relatively reduced amplitude. The striking finding that all syllables evoked an N400m response suggests that syllable processing may proceed at a relatively abstract linguistic level when using a large number of different natural speech syllables and an active target repetition task. Crucially, in the experimental conditions in which we increased the expectation for meaningful speech [i.e. by manipulating lexical–semantic context (*context* versus *isolation*) and acoustic–phonetic cues (*sylsent* versus *syl*)] the N400m response evoked by syllables became more similar to the N400m evoked by words. This word-like neural processing of meaningless syllables was most prominent in the ascending window of the N400m at 200–300 ms, thus supporting the view that this interval is crucial for cognitive processes at the interface of phonological and lexical–semantic analysis (Hagoort and Brown, 2000).

The abstract linguistic nature of the sustained neural activity evoked by our syllables is further illustrated by the similarity in the timing of the N400m response to sentence-initial syllables (*sylsent*) and syllables pronounced separately (*syl*). Although stimulus length was ~83 ms shorter for *sylsent* than *syl*, the only significant latency difference was a ~29 ms earlier N400m peak latency for *sylsent* than *syl* in the left hemisphere. Tone and vowel stimuli with a duration of 150–200 ms or longer have been reported to elicit stimulus-locked sustained activity lasting for the duration of the stimulus (Hari *et al.*, 1989; Eulitz *et al.*, 1995). The absence of such stimulus-locked differences in the present study supports the interpretation that syllable processing proceeded beyond this type of low-level perceptual analysis of physical stimulus characteristics.

Whereas cortical activity was clearly bilateral throughout the entire time window, the N400m response was somewhat stronger in the left hemisphere. Previous MEG studies have shown a strong left lateralization of the visual N400m response (Helenius *et al.*, 1998; Halgren *et al.*, 2002; Marinkovic *et al.*, 2003), and a relatively small leftward bias for the auditory N400m (Helenius *et al.*, 2002; Marinkovic *et al.*, 2003; Kujala *et al.*, 2004). Our N400m results corroborate this leftward bias, and suggest an important contribution of left temporal areas during access of meaning-based representations (Scott and Johnsrude, 2003).

The present study specifically investigated top-down and bottom-up effects on the neural processing of natural speech syllables. In the following paragraphs we will discuss these effects in turn. As for the top-down effects, so far, only a few studies have investigated the effects of context on the processing of speech sounds. Most of these studies have compared speech processing during passive listening versus active tasks (Poeppel *et al.*, 1996; Noesselt *et al.*, 2003; Vihla and Salmelin, 2003) or during the performance of different types of experimental tasks (Szymanski *et al.*, 1999; Obleser *et al.*, 2004). As a consequence, these studies examined context effects which were directly related to distinct attentional demands of the experimental tasks employed. In contrast, we examined context effects on syllable processing independent of task demands as subjects performed the same general stimulus repetition task in all experimental blocks. This allowed us to highlight that the mere presence of words and sentences substantially changes

the cortical processing of meaningless syllables. Most strikingly, only in this linguistic context, the MEG signal evoked by syllables followed the signal of words along the ascending flank of the N400m until ~300 ms. This suggests that when the probability for meaningful speech is high, even meaningless syllables can trigger processes similar to those used in the construction of meaning-based representations.

Besides this specific context effect during the build-up of the N400m response, our results also indicated a more general effect of linguistic context. Syllables evoked a bilaterally enhanced N400m response when presented in the linguistic context versus isolation. Importantly, this enhancement of the N400m response was not preceded by a similar enhancement in earlier time windows. In fact there was even some evidence for an opposite modulation of the N100m in the left hemisphere, i.e. an amplitude reduction in context versus isolation. A context-dependent shift in the balance between neural activation underlying perceptual and higher-level cognitive processes may underlie these findings. For example, in a lexical–semantic context, brain areas engaged in meaning-based analysis of speech may show a generally enhanced level of activation (Noesselt *et al.*, 2003). Moreover, our N100m findings imply that context-based expectations may modulate the processing of speech already at an early perceptual level. Correspondingly, a recent EEG study reported similarly early N100 reductions to auditory speech stimuli in predictive cross-modal (auditory–visual) contexts (Van Wassenhove *et al.*, 2005).

Interestingly, the bottom-up cues as present in sentence-initial syllables (*sylsent*) led to specifically enhanced N400m activity in the left hemisphere. This suggests that left superior temporal areas may be tuned to acoustic–phonetic cues that are relevant for lexical access. Thereby the present observations extend previous fMRI findings which have associated these areas with the prelexical processing of phonetic cues and features of phonological significance in the perceiver's native language (Jäncke *et al.*, 2002; Jacquemot *et al.*, 2003; Gandour *et al.*, 2004). Additionally, our results identify a specific time window in which these bottom-up cues may modulate the neural processing of speech, i.e. ~200–350 ms after speech onset, when phonetic–phonological processes access lexical–semantic representations.

Which acoustic–phonetic cues underlie the present bottom-up effects? The two syllable types differed in several physical characteristics, with two prominent differences being a significantly shorter duration and a higher pitch for *sylsent* than *syl*. Psycholinguistic studies have shown that duration and pitch may represent lexically relevant prosodic cues (Davis *et al.*, 2002; Salverda *et al.*, 2003). During sentence processing, for example, syllables with longer durations have been found to bias lexical interpretations towards monosyllabic words (e.g. *ham*) rather than bisyllabic words (e.g. *hamster*) (Salverda *et al.*, 2003). Moreover, recent findings have indicated that the neural processing of these word-level prosodic cues specifically involves the left posterior superior temporal gyrus (Brechmann and Scheich, 2005; Gandour *et al.*, 2004). Thus, subtle cues related to segmental duration and/or pitch may have contributed to the present bottom-up effects.

Furthermore, due to the continuity of articulatory gestures during the production of speech, natural speech sounds are often co-articulated. Our *sylsent* stimuli were CV syllables cut from sentence-initial CVCV words (see Materials and Methods). We specifically checked that it was not possible to predict the identity of the subsequent consonant. Nevertheless, the final portion of the vowels probably contained subtle cues that anticipated articulation of the following consonant. A rich literature on speech co-articulation has shown that listeners use such anticipatory cues in a maximally efficient way to obtain the earliest possible recognition of spoken words (e.g. Warren and Marslen-Wilson, 1987; Marslen-Wilson and Warren, 1994; Dahan and Tanenhaus, 2004).

The observed left N400m enhancement to *sylsent* stimuli in the present study indicates that implicit knowledge about cues of anticipatory coarticulation and/or lexically relevant prosodic cues may automatically trigger neural processes involved in the access of meaning-based representations. Such an efficient use of bottom-up cues in natural speech may rely on predictive coding (see also Van Wassenhove *et al.*, 2005). That is, based on prior knowledge about, for example, phonological or semantic regularities, the speech-processing system may build on-line predictions of auditory signals which constrain their subsequent perceptual and/or cognitive processing.

Conclusion

The processing of meaningless syllables typically does not proceed beyond prelexical perceptual analysis. In contrast, our study reveals that acoustic–phonetic cues and the presence of a lexical–semantic context trigger word-like activation in the posterior superior temporal cortex. Most importantly, our findings indicate that the cortical system subserving meaning-based analysis of speech exploits predictive bottom-up cues in natural speech and context-induced expectation, thereby suggesting a neural basis for the efficiency and the adaptive nature of speech comprehension.

Notes

This study was supported by the Ter Meulen Fund of the Royal Netherlands Academy of Arts and Sciences, the Foundation 'De Drie Lichten' in the Netherlands, the Sigrid Juselius Foundation, Finnish Cultural Foundation and the Centre of Excellence Programme 2000–2005 of the Academy of Finland. We thank Sakari Arvela and Antti Kemppinen for reading the stimuli on tape, Poju Antsalo for valuable help with stimulus recordings and Mike Seppä for help in brain coordinate transformations.

Address correspondence to Milene Bonte, Department of Cognitive Neuroscience, Faculty of Psychology, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands. Email: m.bonte@psychology.unimaas.nl.

References

Boersma P, Weenink D (2002) Praat 4.0: a system for doing phonetics with the computer [computer software]. Amsterdam: Universiteit van Amsterdam.

Bonte ML, Blomert L (2004) Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. Cogn Brain Res 21:360–376.

Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb Cortex (in press).

Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R (1998) Development of language-specific phoneme representations in the infant brain. Nat Neurosci 1:351–353.

Connolly JF, Phillips NA (1994) Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. J Cogn Neurosci 6:256–266.

Dahan D, Tanenhaus MK (2004) Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of

verb-based thematic constraints. J Exp Psychol Learn Mem Cogn 30:498–513.
Davis MH, Marslen-Wilson WD, Gaskell M (2002) Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. J Exp Psychol Hum Percept Perform 28:218–244.
Dumay N, Benraiss A, Barriol B, Colin C, Radeau M, Besson M (2001) Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. J Cogn Neurosci 13:121–143.
Eggermont JJ, Ponton CW (2002) The neurophysiology of auditory perception: from single units to evoked potentials. Audiol Neurootol 7:71–99.
Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T (1995) Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. J Neurosci 15:2748–2755.
Gandour J, Tong Y, Wong D, Talavage T, Dzemidzic M, Xu Y, Li X, Lowe M (2004) Hemispheric roles in the perception of speech prosody. Neuroimage 23:344–357.
Hagoort P, Brown CM (2000) ERP effects of listening to speech: semantic ERP effects. Neuropsychologia 38:1518–1530.
Halgren E, Dhond RP, Christensen N, Van Petten C, Marinkovic K, Lewine JD, Dale AM (2002) N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. Neuroimage 17:1101–1116.
Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65:413–497.
Hari R, Hämäläinen M, Kaukoranta E, Mäkelä J, Joutsiniemi SL, Tiihonen J (1989) Selective listening modifies activity of the human auditory cortex. Exp Brain Res 74:463–470.
Helenius P, Salmelin R, Service E, Connolly JF (1998) Distinct time courses of word and context comprehension in the left temporal cortex. Brain 121:1133–1142.
Helenius P, Salmelin R, Service E, Connolly JF, Leinonen S, Lyytinen H (2002) Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. J Neurosci 22:2936–2944.
Hickok G, Poeppel D (2000) Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 4:131–138.
Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J Neurosci 23:9541–9546.
Jäncke L, Wustenberg T, Scheich H, Heinze HJ (2002) Phonetic perception and the temporal cortex. Neuroimage 15:733–746.
Kujala A, Alho K, Service E, Ilmoniemi RJ, Connolly JF (2004) Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. Cogn Brain Res 21:106–113.
Kuriki S, Murase M (1989) Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds. Exp Brain Res 77:127–134.
Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci 4:463–470.
Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205.
Kutas M, Schmitt BM (2003) Language in microvolts. In: Mind, brain, and language: multidisciplinary perspectives (Banich MT, Mack MA, eds), pp. 171–209. Mahwah, NJ: Lawrence Erlbaum.
Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal dynamics of modality-specific and supramodal word processing. Neuron 38:487–497.
Marslen-Wilson W, Warren P (1994) Levels of perceptual representation and process in lexical access: words, phonemes, and features. Psychol Rev 101:653–675.
Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432–434.
Noesselt T, Shah NJ, Jäncke L (2003) Top-down and bottom-up modulation of language related areas — an fMRI study. BMC Neurosci 4:13.
Obleser J, Lahiri A, Eulitz C (2003) Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. Neuroimage 20:1839–1847.
Obleser J, Elbert T, Eulitz C (2004) Attentional influences on functional mapping of speech sounds in human auditory cortex. BMC Neurosci 5:24.
O'Rourke TB, Holcomb PJ (2002) Electrophysiological evidence for the efficiency of spoken word processing. Biol Psychol 60:121–150.
Parviainen T, Helenius P, Salmelin R (2005) Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. Cereb Cortex (in press).
Perrin F, Garcia-Larrea L (2003) Modulation of the N400 potential during auditory phonological/semantic interaction. Cogn Brain Res 17:36–47.
Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T (2000) Auditory cortex accesses phonological categories: an MEG mismatch study. J Cogn Neurosci 12:1038–1055.
Poeppel D, Yellin E, Phillips C, Roberts TP, Rowley HA, Wexler K, Marantz A (1996) Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. Cogn Brain Res 4:231–242.
Salverda AP, Dahan D, McQueen JM (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition 90:51–89.
Schormann T, Henn S, Zilles K (1996) A new approach to fast elastic alignment with applications to human brains. Lect Notes Comput Sci 1131:337–342.
Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100–107.
Szymanski MD, Yund EW, Woods DL (1999) Human brain specialization for phonetic attention. Neuroreport 10:1605–1608.
Van den Brink D, Brown CM, Hagoort P (2001) Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. J Cogn Neurosci 13:967–985.
Van Petten CK, Coulson S, Rubin S, Plante E, Parks M (1999) Time course of word identification and semantic integration in spoken language. J Exp Psychol Learn Mem Cogn 25:394–417.
Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 102:1181–1186.
Vihla M, Salmelin R (2003) Hemispheric balance in processing attended and non-attended vowels and complex tones. Cogn Brain Res 16:167–173.
Vihla M, Lounasmaa OV, Salmelin R (2000) Cortical processing of change detection: dissociation between natural vowels and two-frequency complex tones. Proc Natl Acad Sci USA 97:10590–10594.
Warren P, Marslen-Wilson WD (1987) Continuous uptake of acoustic cues in spoken word recognition. Percept Psychophys 41:262–275.
Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, Aaltonen O, Raimo I, Alho K, Lang H, Iivonen A, Näätänen R (1999) Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cogn Brain Res 7:357–369.
Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC (1998) Automated image registration: II. Intersubject validation of linear and nonlinear models. J Comput Assist Tomogr 22:153–165.