Publication I

Johanna Uusvuori, Tiina Parviainen, Marianne Inkinen, and Riitta Salmelin. 2008. Spatiotemporal interaction between sound form and meaning during spoken word perception. Cerebral Cortex, volume 18, number 2, pages 456-466.

Spatiotemporal Interaction between Sound Form and Meaning during Spoken Word Perception

Johanna Uusvuori, Tiina Parviainen, Marianne Inkinen and Riitta Salmelin

Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, Espoo 02015 TKK, Finland

Cortical dynamics of spoken word perception is not well understood. The possible interplay between analysis of sound form and meaning, in particular, remains elusive. We used magnetoencephalography to study cortical manifestation of phonological and semantic priming. Ten subjects listened to lists of 4 words. The first 3 words set a semantic or phonological context, and the list-final word was congruent or incongruent with this context. Attenuation of activation by priming during the first 3 words and increase of activation to semantic or phonological mismatch in the list-final word provided converging evidence: The superior temporal cortex bilaterally was involved in both analysis of sound form and meaning but the role of each hemisphere varied over time. Sensitivity to sound form was observed at ~100 ms after word onset, followed by sensitivity to semantic aspects from ~250 ms onwards, in the left hemisphere. From ~450 ms onwards, the picture was changed, with semantic effects now present bilaterally, accompanied by a subtle late effect of sound form in the right hemisphere. Present MEG data provide a detailed spatiotemporal account of neural mechanisms during speech perception that may underlie characterizations obtained with other neuroimaging methods less sensitive in temporal or spatial domain.

Keywords: language, MEG, N100, N400, speech

Introduction

Analysis of spoken words is thought to proceed via acoustic and phonetic processing to extraction of phonological and semantic information (see Hickok and Poeppel 2004 for review). The manifestation of these processes at the neural level and, in particular, the possible interplay between phonological and semantic processing remains elusive. Here, we utilize the combined temporal and spatial resolution of whole-head magnetoencephalography (MEG) to address cortical dynamics of analysis of sound form (whether acoustic, phonetic, or phonological) and meaning in spoken language perception.

An isolated spoken word typically evokes the following sequence of cortical activation: A robust response at ~100 ms after the word onset, usually referred to as the N100m (N100 in electroencephalography [EEG] literature), is generated mainly in the planum temporale, immediately posterior to the primary auditory cortex (Lütkenhöner and Steinsträter 1998). The N100m response is not specific to words but it is evoked by any sound onset, offset, or a change in the sound (Hari 1990). However, the N100m response has been shown to differ between simple speech and acoustically matched nonspeech sounds (Tiitinen et al. 1999; Vihla and Salmelin 2003) specifically in the left hemisphere (Parviainen et al. 2005), indicating that neural processing is sensitive to acoustic signal of speech already in this time window.

At ~150-200 ms after the word onset, the cortical activation is strongly reduced. An experimental paradigm in which infrequent deviant stimuli interrupt a sequence of frequent standard stimuli ("oddball paradigm") can be used to focus on this time window. The infrequent stimuli evoke a so-called mismatch field (MMF) which is the MEG counterpart of the mismatch negativity (MMN), originally detected using EEG (Sams et al. 1985). The oddball paradigm has been used to demonstrate that the supratemporal auditory cortex is sensitive to the phonological structure of speech sounds by ~150 ms after stimulus onset (Näätänen et al. 1997; Phillips et al. 2000; Vihla et al. 2000).

At 200-800 ms after the word onset, a sustained response, usually referred to as the N400m, is recorded over the temporal areas (Helenius et al. 2002; Bonte et al. 2006). When the active cortical patches are represented by a set of focal Equivalent Current Dipoles (ECDs) the N400m response is found to be generated in the posterior superior temporal cortex, in the vicinity of the auditory cortex (Helenius et al. 2002; Kujala et al. 2004; Biermann-Ruben et al. 2005; Bonte et al. 2006). An MEG experiment using distributed source modeling technique (Marinkovic et al. 2003) has suggested that the activity underlying the N400m response additionally extends to frontal and anterior temporal areas.

The N400m (N400 in EEG literature) response is affected by semantic manipulation and, therefore, thought to reflect semantic analysis. When subjects listen to sentences that end with a semantically congruent or incongruent word the N400/m is remarkably subdued to the semantically congruent final words and significantly stronger to the incongruent final words (e.g., Connolly and Phillips 1994; Hagoort and Brown 2000; Helenius et al. 2002). This semantic priming effect occurs similarly for word pairs (e.g., Radeau et al. 1998; Perrin and Garcia-Larrea 2003). Phonological manipulation also influences neural activation in the N400/m time window, at 200-800 ms after stimulus onset. Experiments using sentences with final words that are semantically congruent but (phonologically) unexpected have specifically suggested presence of a separate response at 200-350 ms that would reflect analysis of phonological congruity, seemingly independent of any semantic processing (phonological MMN, PMN; Connolly and Phillips 1994; D'Arcy et al. 2004; Kujala et al. 2004). In word-pair experiments phonological priming shows as reduced activation, although the effects are weaker and more variable than for semantic priming (Praamstra and Stegeman 1993; Praamstra et al. 1994; Radeau et al. 1998; Dumay et al. 2001; Perrin and Garcia-Larrea 2003). The time window of the priming effect depends on whether the prime and target words share initial phonemes (alliteration, 250-450 ms; Praamstra et al. 1994) or final phonemes (rhyming, 300-1000 ms; Praamstra and Stegeman 1993; Praamstra et al. 1994;

Radeau et al. 1998; Dumay et al. 2001; Perrin and Garcia-Larrea 2003). Attenuation of the N400 response is detected when there are at least 2–3 common phonemes (rime or syllable) (Dumay et al. 2001). For a large part, however, the N400/m response clearly reflects semantic analysis as it is overall significantly stronger to semantically wrong sentence-final words (regardless of their phonological agreement with the expected word) than to acoustically/phonologically unexpected words that are semantically congruent with the preceding context (Helenius et al. 2002).

Based on reaction time experiments, it has been suggested that semantic priming effects can be explained by automatic spreading of activation and/or conscious strategic mechanisms (Posner and Snyder 1975a, 1975b). Activation of the prime word could preactivate the target due to overlapping neural representations in the mental lexicon (spreading-activation theory of semantic processing, Collins and Loftus 1975; distributed models, e.g., Gaskell and Marslen-Wilson 1997). In the context of neurophysiological responses, when the target word has been preactivated by the preceding prime word(s) it can be thought to generate weaker neuronal signal than when heard in isolation (Kutas and Federmeier 2000). However, there is plenty of evidence also for more controlled postlexical mechanisms accounting for priming effects (for review, see Hill et al. 2002). In neuroimaging experiments using auditory sentences the amplitude of the N400 has been suggested to reflect the ease of integration of the word to a larger context (Connolly and Phillips 1994; Hagoort and Brown 2000; Van den Brink et al. 2001). In these studies, lexical selection is suggested to occur earlier, at ~200–300 ms, and to be reflected in a separate response (PMN in Connolly and Phillips 1994; N200 in Hagoort and Brown 2000; Van den Brink et al. 2001).

Behavioral evidence of phonological priming effects is inconsistent. With alliterating word pairs either facilitation (e.g., Slowiaczek et al. 1987), interference (Radeau et al. 1989), or no effect at all (Slowiaczek and Pisoni 1986) has been observed. With rhyming word pairs, facilitation is usually detected (e.g., Slowiaczek et al. 1987). When both behavioral and electrophysiological responses were recorded in the same experiment, rhyming words showed facilitation in both measures whereas alliterating word pairs showed only an electrophysiological priming effect (Praamstra et al. 1994). It was suggested that the behavioral and electrophysiological effects reflect the same or closely related processes, and that the priming effects would be due to preactivation of some of the phonological components of the target by the prime word but this effect would be masked by the effect of lexical competition in the alliterating condition, resulting in disappearance of the behavioral effect (Praamstra et al. 1994).

Taken together, the existing experimental evidence suggests that semantic and phonological analysis may be reflected in temporally and spatially overlapping cortical activation. To investigate the neural representation of these processes in a systematic manner they need to be manipulated independently. The sentence paradigm that is frequently used to study semantic processing is not optimal for this purpose as the sentence creates an expectation for both the meaning and sound form of the final word. Phonological and semantic aspects are better dissociated when the stimuli are word pairs. However, a single prime word creates only a weak expectation of the target word. In the current experiment, we attempted to both separate the semantic and phonological priming effects and build a strong

context for the target word. Therefore, we used lists of 4 words instead of word pairs or sentences. The first 3 words of each list were semantically or phonologically related. They set the framework with which the last word either agreed or disagreed (see, e.g., Dehaene-Lambertz et al. 2000 for a similar type of paradigm in the context of MMN). Consequently, the stimuli were divided into 4 categories according to the list-final word: 1) semantically related, 2) semantically unrelated, 3) phonologically related, and 4) phonologically unrelated. To evaluate the effect of build-up of semantic versus phonological expectation we focused on activation during the first 3 words. The effect of semantic versus phonological mismatch was characterized by comparing the responses to 4 types of list-final words.

Methods

Subjects

Ten right-handed Finnish-speaking subjects (5 females and 5 males; 20–29 years, mean 25 years) participated in the experiment. None of the subjects reported 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 and Experimental Design

Stimuli were lists of 4 spoken words. The first 3 words had either a related meaning or began with the same 2 phonemes. The final word of the list either matched the framework set by the first 3 words or differed from it semantically or phonologically. The lists thus fell into 4 categories, according to the type of the final word: 1) semantically related, 2) semantically unrelated, 3) phonologically related, and 4) phonologically unrelated. Figure 1 gives an example of each list type. There were 87 word lists per category, and 348 lists in total, composed of 1392 different words. The subject's task was to press a button when s/he detected a word list in which one word appeared twice (detection rate on average 85%). The probability of these additional target lists was 6% (20 lists). The responses to the target lists were not included in the analysis. The words were bisyllabic 4- to 5-letter common Finnish nouns beginning with a consonant, chosen from a Finnish corpus (Laine and Virtanen 1999). Due to transparency of the Finnish language the number of phonemes equals the number of letters. Semantic lists were created by selecting from the corpus word groups that had a clearly related meaning. In semantic word lists the first phoneme of every word differed from that of every other word more than by voicing only. Most semantic lists included both concrete and abstract words but in case all the prime words were concrete/abstract the list-final word was also concrete/abstract. There was a 1-s interval between the onsets of the successive words in a list and a 2-s interval between the onset of the list-final word and the initial word of the next list.

The words were spoken by a male native Finnish speaker and recorded at a sampling rate of 48 kHz on a DAT recorder in an anechoic chamber (Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology). The sound intensity level of the stimuli was equated within 3 dB using a speech waveform editor (Praat; Boersma and Weenink 2002). The mean duration of the words was 546 ms (range 317 to 739 ms).

The word-list stimuli were presented in a pseudorandomized order, that is, neither phonological nor semantic lists appeared more than 3 times in a row. Words were presented binaurally at a comfortable listening level. The total measurement time was about 35 min (4 × 8-min blocks with short breaks in between).

MEG Recording and Data Analysis

MEG recordings were conducted in a magnetically shielded room with a Vectorview™ whole-head MEG device (Elekta Neuromag Ltd., Helsinki, Finland). The system contains 102 triple sensor elements composed of 2 orthogonal planar gradiometers and one magnetometer. The gradiometers detect the maximum signal directly above an active cortical area. The signals were band-pass filtered at 0.03–200 Hz and digitized at 600 Hz. During the measurement, subjects listened to the
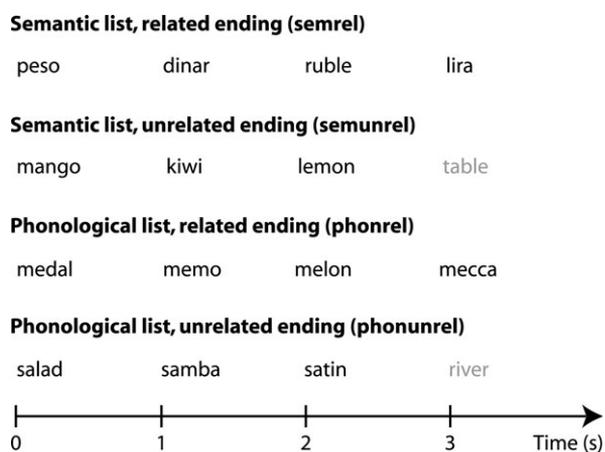
Semantic list, related ending (semrel)

peso   dinar   ruble   lira

Semantic list, unrelated ending (semunrel)

mango   kiwi   lemon   table

Phonological list, related ending (phonrel)

medal   memo   melon   mecca

Phonological list, unrelated ending (phonunrel)

salad   samba   satin   river

0   1   2   3   Time (s)

Figure 1. Examples of the word-list stimuli used in the experiment. In addition to the 4 list types, there were target lists in which one word appeared twice (probability 6%; not shown). The actual stimuli were in Finnish.

word lists, eyes closed. Horizontal and vertical eye movements were monitored (electro-oculogram, EOG).

The MEG data were averaged off-line across trials from –0.2 to 4.2 s relative to the onset of the list-initial word. The averaged MEG responses were baseline corrected to the 200-ms interval immediately preceding the list onset and low-pass filtered at 40 Hz. Trials with MEG amplitude exceeding 3000 fT/cm during the presentation of the list-final word were discarded automatically, thereafter, trials during which major disturbances remained (~10%) were removed from the data manually. In addition, artifactual slow, approximately 0.1 Hz shifts were detected in the data of 3 subjects. The disturbances were removed by applying a high-pass filter to the nonaveraged data of these individuals (center 0.15 Hz, width 0.1 Hz). The data were also averaged off-line with respect to eye movements (EOG). Principal component analysis was performed on this average, and the component that contained the artifacts due to eye movements was removed from the data (Uusitalo and Ilmoniemi 1997).

At least 60 (on average 78) artifact-free trials were obtained for each list type. For source analysis, the signal-to-noise ratio was enhanced by additionally averaging the responses to the list-initial word across the 4 word lists (all-average). This procedure increased the number of trials to at least 261 (on average 311 trials). Furthermore, for investigation of the responses to the second and third words, responses were averaged for the 2 semantic lists (sem-average) and for the 2 phonological lists (phon-average), resulting in at least 124 trials (on average 155).

To obtain an initial overview of the data, areal mean signals were calculated over 7 areas of interest: left and right frontal, temporal, and oc-cipital areas, and the parietal area. We first computed vector sums of each gradiometer pair by squaring the MEG signals, summing them together, and calculating the square root of this sum. The areal mean signals were computed by averaging these vector sums for each area of interest, individually for each subject. Finally, the areal mean signals were averaged across subjects. 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 data of each individual were analyzed further in 2 ways, using focal sources in ECD analysis (Hämäläinen et al. 1993) and a distributed model by computing Minimum Current Estimates (MCEs) (Uutela et al. 1999).

ECD Analysis

An ECD represents the center of an active cortical patch, and the mean orientation and strength of electric current in that area (Hämäläinen et al. 1993). ECDs were localized individually in each subject. The whole-head magnetic field patterns were visually inspected for dipolar field patterns, and the ECDs were identified one by one at time points at which each field pattern was clearest. ECDs were determined using a subset of planar gradiometers that covered the spatially and/or

temporally distinct magnetic field patterns. Thereafter, time courses of activation in those brain areas were obtained by including all ECDs simultaneously in a multidipole model: The locations and orientations of the ECDs were kept fixed, whereas their amplitudes were allowed to vary to best account for the measured data in each condition. The final multidipole models typically accounted for above 85% (at least 75%) of the total magnetic field variance at the peak of activation in each condition. In each subject, the cortical activation evoked by all words in the 4 types of lists was well represented by the same set of ECDs. The final models were composed of 4–7 ECDs (mean 5).

The location of the ECDs was defined in the head coordinate system that was set by the nasion and 2 reference points anterior to the ear canals. Prior to the MEG recording, 4 Head Position Indicator (HPI) coils were attached to the subject's head and their locations were measured with a 3D digitizer (Polhemus, Colchester, VT). At the beginning of the recording, the HPI coils were briefly energized to determine their location with respect to the MEG helmet. For visualization and comparison of the sources between subjects the ECDs were trans-formed to a standard brain (Roland and Zilles 1996) using elastic transformation (Schormann et al. 1996; Woods et al. 1998).

Strength and timing of activation in the source areas were repre-sented by the time courses of the ECDs (source waveforms). The statistical analysis was performed separately in 3 time windows: 75–160, 160–250, and 250–1000 ms after the onset of each word. The first time window contained the maximum of the N100m response in all subjects. The second time window contained the local minimum at ~200 ms and the third time window the maximum of the sustained response peaking at ~500 ms in all subjects. The strength and timing of the activation were estimated by measuring within each time window the maximum or minimum amplitude and the time point when the waveform reached this value. To better characterize the sustained N400m responses, we measured the time points at which the waveform reached 50% of the peak amplitude in the ascending and descending slopes; the 50% latency during the descending slope could not be extracted in the right hemisphere of one subject. Moreover, we calculated the mean ampli-tude between the N100m and N400m responses (160- to 250-ms poststimulus) and during the ascending and descending slopes of the N400m response, defined as 200-ms periods before and after the group mean peak latency of the N400m response in the list-initial word. All measures were calculated individually for each subject and separately for each condition.

These measures were analyzed using 2 separate analyses of variance (ANOVAs), one encompassing the responses to the first 3 words in a 2 (hemisphere: left, right) × 5 (list position: 1st word, 2nd word semantic list, 2nd word phonological list, 3rd word semantic list, 3rd word phonological list) design, and the other focusing on the responses to the list-final words in a 2 (hemisphere: left, right) × 2 (word-list type: semantic, phonological) × 2 (congruence: related, unrelated) design. Amplitude comparison between the hemispheres was justified by first verifying that the distance of the ECDs from the center of the head (and the MEG sensors) did not differ between the hemispheres.

For the first 3 words, planned contrasts (first words vs. following words, linear trend, quadratic trend, and semantic vs. phonological list) were used to determine the categories for which the responses differed, separately for the left and right hemisphere. In speech processing, analysis of sound form is generally assumed to precede analysis of meaning (Hickok and Poeppel 2004), that is, it should appear in the activation that leads up to the robust sustained N400m response. In order to test this hypothesis, planned contrasts (paired samples 2-tailed t-tests) were performed on the response amplitudes at 75–160 and 160–250 ms elicited by phonologically unrelated versus related list-final words and by semantically unrelated versus related list-final words.

Difference waveforms were additionally calculated, for visualization, between the source waveforms to the phonologically unrelated and related list-final words and between the source waveforms to the semantically unrelated and related list-final words.

Minimum Current Estimates

MCE (Uutela et al. 1999) is an implementation of the minimum L1-norm estimate (Matsuura and Okabe 1995). The measured signals are accounted for by a distribution of electric current that has the minimum

total amplitude. For each subject, MCE was computed for all 6 averages (all, sem, phon, semrel, semunrel, phonrel, and phonunrel). For group averaging, the individual calculation points were transformed (Schormann et al. 1996; Woods et al. 1998) into a standard brain (Roland and Zilles 1996).

Results

Areal Mean Signals

Figure 2 displays the areal mean signals for the different list types, averaged across all subjects. Both the largest signals and the differences between stimulus categories were concentrated over the temporal areas in all subjects and were particularly pronounced over the left hemisphere. Each word stimulus elicited a bilateral N100m response followed by a sustained response that reached the maximum at about 500 ms after the stimulus onset. Strongest stimulus effects occurred between 200 and 800 ms.

Field Patterns and Source Analysis

Figure 3 shows the sequence of activation elicited by the list-initial word in one subject. Four salient field patterns appeared over the left hemisphere at ~100, ~200, ~400, and ~700 ms. Over the right hemisphere, clear dipolar field patterns were detected at ~100 and at ~500 ms.

The bilateral dipolar fields around ~100 ms (N100m) were accounted for by a source (represented by an ECD) in each hemisphere, with the center of activation in the Sylvian fissure, close to the primary auditory cortex, and current flow oriented perpendicular to the course of the sulcus, similarly in all subjects.

In the left hemisphere after the list-initial word, the N100m was followed by a distinct field pattern at ~200 ms in 7 subjects (cf. Fig. 3). The pattern typically reflected a source in the posterior temporal cortex with the current flow directed anteriorly (in 5 subjects). In 2 subjects, the pattern was accounted for by an inferior frontal source with current flow to anterior–superior direction. In 2 subjects both an anterior

and posterior source could be localized. This field pattern at ~200 ms was only evident for the list-initial word; the mean amplitude at 160–250 ms was significantly stronger after the first word (9 ± 1 nAm) than words 2–4 (second 1 ± 1 nAm, third 1 ± 1 nAm, and fourth 1 ± 1 nAm) ($F_{3,18}$ = 13.3, P < 0.001). The sources detected at ~200 ms were typically active at a later point as well, at ~500–1000 ms after each word, but with the direction of current flow reversed (opposite polarity of source waveforms). These source waveforms did not differentiate between stimulus categories and are, therefore, not included in the further statistical analysis.

Bilateral dipolar field patterns were visible in all subjects during the sustained activity peaking at ~400–500 ms. The
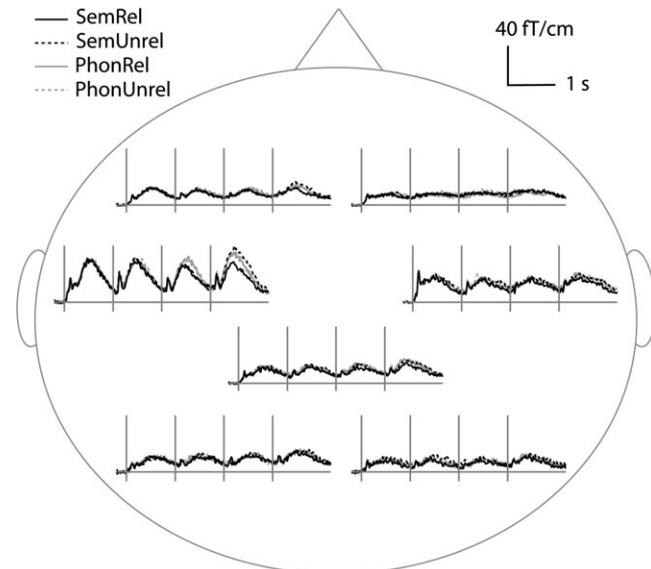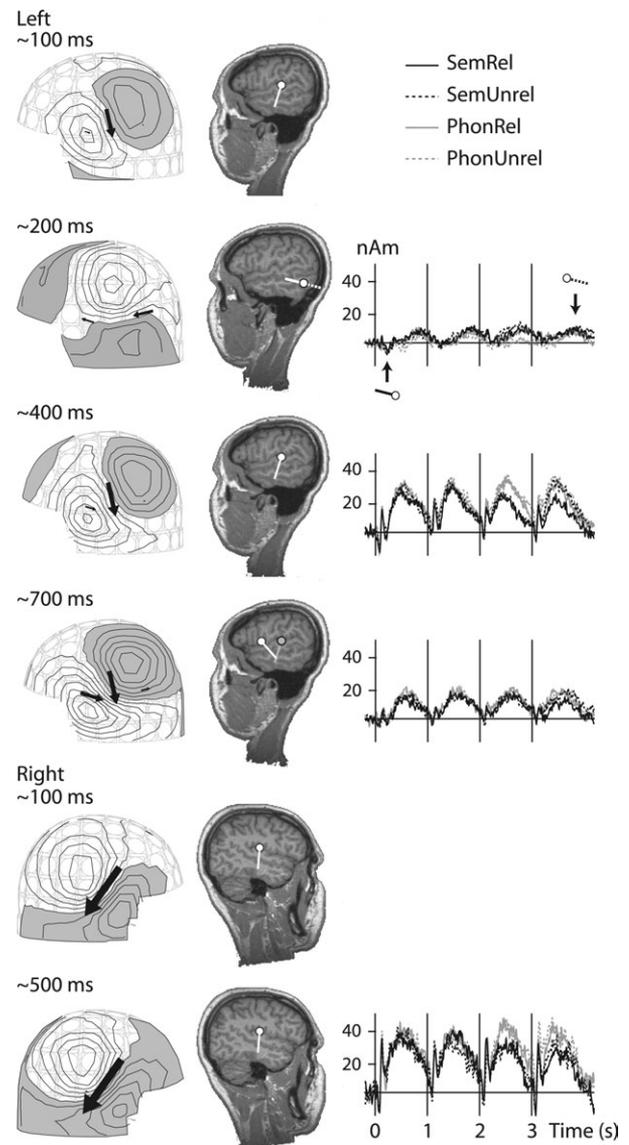

Figure 3. MEG field patterns (left), ECD source localization (middle), and corresponding source waveforms in response to word lists (right) in one subject. The source currents at ~100 and ~400–500 ms were essentially identical in location and orientation, and the source at ~400–500 ms thus accounted for most of the activity also in the earlier time window. In the left hemisphere, a posterior temporal area was active at ~200 ms, with current directed anteriorly (solid line). This same region again showed activation at ~500–1000 ms but with current flow in the opposite direction (dashed line), resulting in the initial negative value and subsequent positive values in the source waveform.

Figure 2. Group averaged areal mean signals over the left and right occipital, temporal, and frontal cortex, and the parietal cortex.

corresponding ECDs were located in the Sylvian fissure, with the center of activation close to the primary auditory cortex and current flow oriented perpendicular to the course of the sulcus (Fig. 4). The location, orientation, and time course of these ECDs suggest that they correspond to the N400m sources described in earlier MEG studies (Helenius et al. 2002; Bonte et al. 2006). Because the location and direction of current flow of these sources were very similar to those of the N100m sources they also accounted for the N100m activity. In order to prevent spurious interactions between these ECDs in the multidipole model, the N100m and N400m activations were thus both represented by the N400m sources (see also Helenius et al. 2002; Bonte et al. 2006). The N100m/N400m source waveforms (Fig. 4) showed a pattern that closely resembled that of the areal mean signals over the left and right temporal lobes (Fig. 2). Henceforth, we refer to the sustained activity peaking at ~400–500 ms as the N400m response.

The results obtained from the distributed MCEs were in accordance with the focal ECD models in each individual subject and at the group level. The areas of strongest activity in the MCE maps corresponded to the N100m/N400m source areas detected with the ECD analysis.

Statistical Tests on the ECD Source Waveforms

The results of the 2 (hemisphere) × 5 (list position) and 2 (hemisphere) × 2 (word-list type) × 2 (congruence) repeated-measures ANOVAs and the planned contrasts are described in the following 2 sections. First, the focus is on the build-up of semantic and phonological expectation over the initial 3 items of the word lists, manifested as reduced activation. Thereafter, we proceed to breakdown of semantic or phonological expectation in the list-final word, and the neural effects of congruency versus incongruency.

Building Semantic or Phonological Expectation

Figure 5a shows the mean time course of activation in the left and right N100m/N400m source area in response to the first 3 words of the semantic and phonological lists. The activation strengths and latencies in the time windows 75–160 (N100m), 160–250, and 250–1000 ms (N400m) are listed in Table 1.
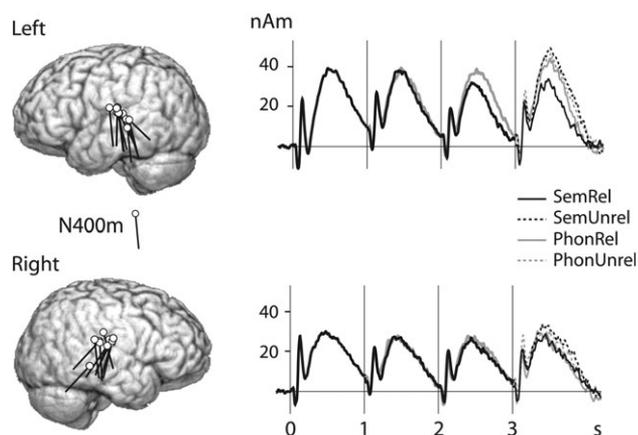
Figure 4. Locations and mean time course of the N400m sources. White dots and black lines indicate the individual source locations and directions of current flow of all 10 subjects in the left and right superior temporal cortex. For the list-initial word, the response is averaged over all word lists and for the second and third words over semantic or phonological lists (see Methods). For the list-final word, all 4 categories are plotted separately.

The N100m peak activation was attenuated in both hemispheres when proceeding in the word list, regardless of the list type. In the left hemisphere, the response was decreased from the second to the third word, following a slight increase from the first to the second word (quadratic trend: $F_{1,9} = 13.4$, $P < 0.01$). In the right hemisphere, the activation was attenuated gradually from the first to the third word (word 1 vs. following words: $F_{1,9} = 8.5$, $P < 0.05$, linear trend: $F_{1,9} = 11.1$, $P < 0.01$). In the left hemisphere, the N100m response was also slightly delayed for the second word in comparison with the first word (delay ~10 ms) and third word (delay ~5 ms) (word 1 vs. following words: $F_{1,9} = 9.6$, $P < 0.05$, linear trend: $F_{1,9} = 5.3$, $P < 0.05$, quadratic trend: $F_{1,9} = 15.6$, $P < 0.005$).
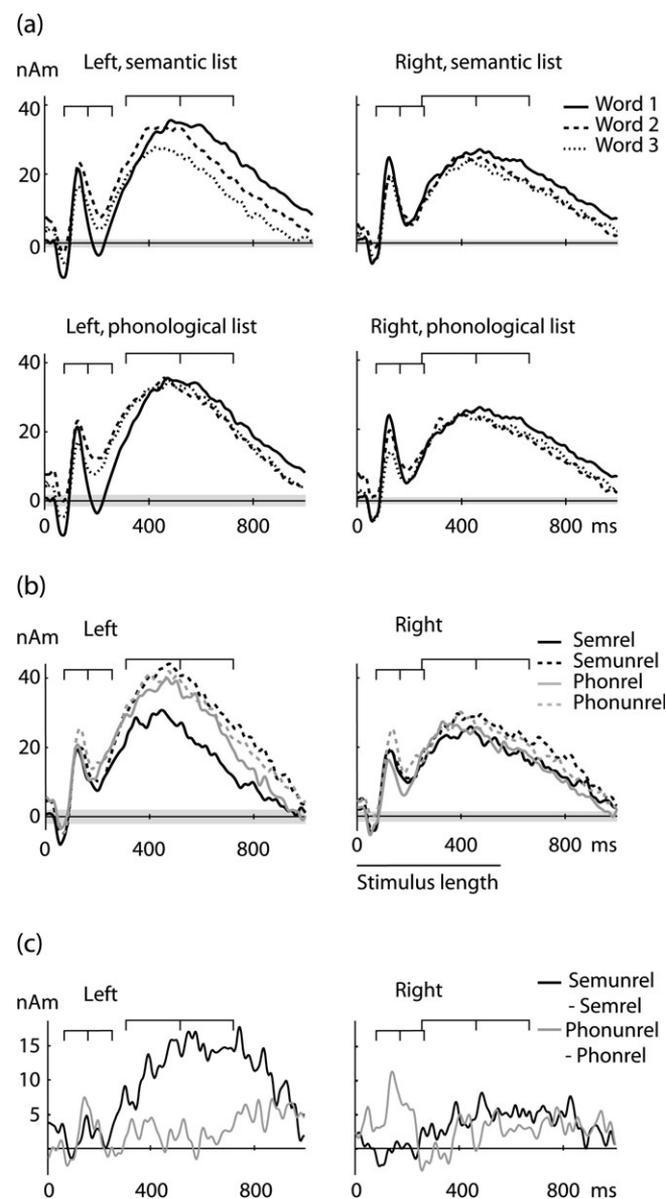
Figure 5. (a) N100m/N400m source waveforms for the semantic (above) and phonological lists (below) in the left and right hemisphere. The responses to the first 3 words are overlaid. (b) N100m/N400m source waveforms for the 4 types of list-final words in the left and right hemisphere. (c) Difference waveforms Semunrel–Semrel and Phonunrel–Phonrel in the left and right N100m/N400m source area, calculated for the list-final word and averaged across subjects. The gray area indicates the noise level (±2 times standard deviation during the 200-ms period immediately preceding the list onset). The markers above the curves indicate the time windows used in the statistical analysis.

Table 1

Strength and timing of activation to the first 3 words in the left- and right-hemisphere N100m/N400m source areas (mean ± SEM), extracted from the individual source waveforms

	Left hemisphere					Right hemisphere				
		Phon		Sem			Phon		Sem	
	Word 1	Word 2	Word 3	Word 2	Word 3	Word 1	Word 2	Word 3	Word 2	Word 3
Time window 75–160 ms										
Peak amplitude (nAm)	23 ± 4	25 ± 4	19 ± 4	25 ± 3	19 ± 4	27 ± 4	21 ± 4	17 ± 4	22 ± 4	20 ± 4
Peak latency (ms)	121 ± 3	133 ± 5	125 ± 3	132 ± 4	129 ± 5	121 ± 3	126 ± 3	128 ± 5	129 ± 3	125 ± 4
Time window 160–250 ms										
Mean amplitude 160–260 ms (nAm)	1 ± 3	14 ± 3	10 ± 3	11 ± 2	7 ± 3	8 ± 3	11 ± 3	8 ± 4	8 ± 2	8 ± 3
Minimum amplitude (nAm)	−4 ± 4	10 ± 3	5 ± 4	6 ± 3	2 ± 4	5 ± 3	7 ± 3	3 ± 4	4 ± 2	4 ± 3
Minimum latency (ms)	197 ± 5	192 ± 9	182 ± 7	211 ± 6	202 ± 7	196 ± 4	192 ± 6	185 ± 7	212 ± 8	192 ± 5
Time window 250–1000 ms										
Mean amplitude ascending slope (nAm)	29 ± 4	31 ± 3	31 ± 4	31 ± 3	25 ± 3	23 ± 4	23 ± 5	22 ± 5	21 ± 4	20 ± 4
Peak amplitude (nAm)	38 ± 5	37 ± 4	37 ± 4	37 ± 4	30 ± 4	31 ± 5	28 ± 5	28 ± 6	29 ± 5	27 ± 5
Mean amplitude descending slope (nAm)	32 ± 4	28 ± 3	29 ± 3	25 ± 2	20 ± 3	25 ± 4	21 ± 5	22 ± 5	22 ± 4	20 ± 5
50% latency ascending slope (ms)	312 ± 13	268 ± 6	271 ± 5	278 ± 7	277 ± 9	281 ± 14	267 ± 6	272 ± 10	279 ± 9	281 ± 14
Peak latency (ms)	505 ± 21	484 ± 19	485 ± 22	460 ± 11	445 ± 15	443 ± 40	427 ± 15	443 ± 54	417 ± 21	464 ± 55
50% latency descending slope (ms)	803 ± 21	776 ± 23	764 ± 21	692 ± 16	676 ± 21	715 ± 58	686 ± 60	657 ± 69	644 ± 61	651 ± 67

Note: The values listed for Word 1 were extracted from the source waveforms averaged across all 4 word-list types (Phonrel, Phonunrel, Semrel, Semunrel), and the values listed for Word 2 and Word 3 were extracted from the source waveforms averaged separately for semantic lists (Semrel, Semunrel) and phonological lists (Phonrel, Phonunrel). The mean amplitude in the ascending/descending slope of the N400m response was computed for the 200-ms interval immediately preceding/succeeding the group mean peak latency in the list-initial word.

In the next time window, 160- to 250-ms poststimulus, the signal strength in the left hemisphere varied with the position of the word in the list, again regardless of the list type; the level of activation was increased sharply from the first to the second word, followed by a small decrease from the second to the third word, as indicated both by the mean amplitude (word 1 vs. following words: $F_{1,9} = 16.0$, $P < 0.005$, linear trend: $F_{1,9} = 7.1$, $P < 0.05$, quadratic trend: $F_{1,9} = 48.1$, $P < 0.001$) and the minimum amplitude in that time window (word 1 vs. following words: $F_{1,9} = 12.7$, $P < 0.01$, linear trend: $F_{1,9} = 5.6$, $P < 0.05$, quadratic trend: $F_{1,9} = 27.4$, $P < 0.005$). Between 160 and 250 ms, the signal reached the minimum 10–20 ms earlier for the phonological than semantic lists in both hemispheres (left: semantic vs. phonological list: $F_{1,9} = 7.7$, $P < 0.05$; right: semantic vs. phonological list: $F_{1,9} = 5.5$, $P < 0.05$).

In the left hemisphere, the subsequent sustained response (N400m) started ~50 ms earlier to the second and third words than to the first word, regardless of list type (50% latency in the ascending slope: word 1 vs. following words: $F_{1,9} = 16.8$, $P < 0.005$, linear trend: $F_{1,9} = 12.7$, $P < 0.01$, quadratic trend: $F_{1,9} = 15.0$, $P < 0.005$). This difference in timing manifested itself also as a larger mean amplitude in the ascending slope for the second than for the first word, regardless of the list type (quadratic trend: $F_{1,9} = 9.6$, $P < 0.05$).

The left-hemisphere N400m response was attenuated during the semantic word lists but not during the phonological lists, as evidenced by the mean amplitude in the ascending slope (attenuation from the second to the third word: quadratic trend: $F_{1,9} = 9.6$, $P < 0.05$, semantic vs. phonological list: $F_{1,9} = 5.2$, $P < 0.05$), the peak amplitude (attenuation from the second to the third word: quadratic trend $F_{1,9} = 6.2$, $P < 0.05$, semantic vs. phonological list: $F_{1,9} = 5.2$, $P < 0.05$), and the mean amplitude in the descending slope (attenuation from the first to the third word: word 1 vs. following words: $F_{1,9} = 6.9$, $P < 0.05$, linear trend: $F_{1,9} = 7.8$, $P < 0.05$, semantic vs. phonological list: $F_{1,9} = 20.2$, $P < 0.005$). The attenuation of the left-hemisphere N400m response specifically to the semantically related words also showed as shortening of the response duration when proceeding from the first to the third word in the semantic lists (50% latency in the descending slope: word 1 vs. following

words: $F_{1,8} = 32.9$, $P < 0.001$, linear trend: $F_{1,8} = 47.3$, $P < 0.001$, semantic vs. phonological list: $F_{1,8} = 28.0$, $P < 0.005$). In addition, the left-hemisphere N400m response reached the maximum ~30 ms earlier for the semantic than phonological lists (semantic vs. phonological list: $F_{1,9} = 6.3$, $P < 0.05$).

Unlike in the left hemisphere, the ascending slope of the right-hemisphere N400m response was not sensitive to the number of preceding words or the word-list type. From the peak latency onwards, however, the N400m activation was attenuated when advancing along the list, for both semantic and phonological lists. This effect was evident in the reduction of the mean amplitude in the descending slope (word 1 vs. following words: $F_{1,9} = 5.4$, $P < 0.05$, linear trend: $F_{1,9} = 5.2$, $P < 0.05$) and shortening of the response duration (50% latency in the descending slope: word 1 vs. following words: $F_{1,8} = 10.9$, $P < 0.05$, linear trend: $F_{1,8} = 15.3$, $P < 0.005$).

In summary, spoken word lists as such had the following effects in the superior temporal cortex: When proceeding in the word list, the N100m activation was attenuated in both hemispheres. In the left hemisphere, from 160 ms onwards, the time course of activation to the first word was clearly distinct from that to the words later in the list: From the first to the second word, the strength of the signal between 160 and 250 ms rose to a markedly higher level, and the N400m response started ~50 ms earlier than for the first word.

Semantic priming played a role from 250 ms onwards. Activation was diminished during the ascending slope of the N400m response, only in the left hemisphere. At the peak and during the descending slope of the N400m response, semantic priming reduced activation in both hemispheres. Phonological priming attenuated activation during the descending slope of the N400m response, reaching significance in the right hemisphere. Phonological and semantic priming thus showed similar effects in the right hemisphere.

Breaking Semantic or Phonological Expectation

Figure 5*b* shows the mean time course of activation in the left and right N100m/N400m source area in response to the 4 types of list-final words. The activation strengths and latencies in the time windows 75–160 (N100m), 160–250, and 250–1000 ms

(N400m) are listed in Table 2. Figure 5c depicts the difference waveforms Semunrel–Semrel and Phonunrel–Phonrel in the left and right N100m/N400m source area, averaged across subjects.

The N100m activation strength showed a salient effect in the phonological but not semantic lists, apparently with strongest activation to phonologically unrelated list-final words (see Fig. 5b,c and Table 2). Indeed, in speech processing, analysis of sound form is generally assumed to precede analysis of meaning (Hickok and Poeppel 2004). Planned contrasts on the N100m peak amplitudes revealed that the difference between phonologically unrelated and related list-final words reached significance in the left hemisphere (phonological lists, word 4, left hemisphere: $t(9) = 2.6$, $P < 0.05$, right hemisphere: $t(9) = 2.0$, $P = 0.08$, n.s.) whereas no difference between semantically unrelated and related final words was found in either hemisphere (semantic lists, word 4, left hemisphere: $t(9) = 0.5$, $P = 0.6$, n.s., right hemisphere: $t(9) = 0.4$, $P = 0.7$, n.s.); in the ANOVA the interactions did not reach significance. In the subsequent time window (160–250 ms), the effect of phonological incongruence on minimum or mean amplitude did not reach significance.

The N100m response was slightly delayed (~10 ms) for the unrelated as compared with related list-final words, in the left hemisphere for both list types (hemisphere × word-list type × congruence $F_{1,9} = 40.9$, $P < 0.001$, left hemisphere: congruence $F_{1,9} = 10.8$, $P < 0.01$) and in the right hemisphere for phonological lists only (right hemisphere: word-list type × congruence $F_{1,9} = 14.4$, $P < 0.005$, right hemisphere, phonological lists: congruence $F_{1,9} = 36.4$, $P < 0.001$). A similar 10-ms delay was detected bilaterally for both list types in the following time window, 160–250 ms (minimum latency: congruence $F_{1,9} = 8.3$, $P < 0.05$).

In the N400m time window (250–1000 ms), the response in the left hemisphere was weakest to the semantically related list-final words as measured by the mean amplitude in the ascending slope (hemisphere × congruence $F_{1,9} = 10.1$, $P < 0.05$, left hemisphere: word-list type × congruence $F_{1,9} = 5.6$, $P < 0.05$, left hemisphere, semantic lists: congruence $F_{1,9} = 12.1$, $P < 0.01$), the mean amplitude in the descending slope (hemisphere × word-list type $F_{1,9} = 7.6$, $P < 0.05$, hemisphere × congruence $F_{1,9} = 10.8$, $P < 0.01$, word-list type × congruence $F_{1,9} = 5.7$, $P < 0.05$, left hemisphere: word-list type × congruence $F_{1,9} = 17.3$, $P < 0.005$, left hemisphere, semantic lists: congruence

$F_{1,9} = 30.2$, $P < 0.001$), and the peak amplitude (hemisphere × congruence $F_{1,9} = 7.3$, $P < 0.05$, word-list type × congruence $F_{1,9} = 6.9$, $P < 0.05$, left hemisphere: word-list type × congruence $F_{1,9} = 13.8$, $P < 0.01$, left hemisphere, semantic lists: congruence $F_{1,9} = 23.4$, $P < 0.005$). The left-hemisphere response was also of shortest duration to the semantically related list-final words (50% latency in the descending slope: hemisphere × word-list type $F_{1,8} = 21.8$, $P < 0.005$, left hemisphere: word-list type × congruence $F_{1,8} = 8.7$, $P < 0.05$, left hemisphere, semantic lists: congruence $F_{1,8} = 19.7$, $P < 0.005$).

In the right hemisphere, significant effects were detected from the peak latency onwards. Activation during the descending slope was stronger after an unrelated than related list-final word, as reflected in the peak amplitude (hemisphere × congruence $F_{1,9} = 7.3$, $P < 0.05$, word-list type × congruence $F_{1,9} = 6.9$, $P < 0.05$, right hemisphere: congruence $F_{1,9} = 9.6$, $P < 0.05$) and mean amplitude (hemisphere × word-list type $F_{1,9} = 7.6$, $P < 0.05$, hemisphere × congruence $F_{1,9} = 10.8$, $P < 0.01$, word-list type × congruence $F_{1,9} = 5.7$, $P < 0.05$, right hemisphere: congruence $F_{1,9} = 9.2$, $P < 0.05$). In addition, the peak and offset latencies were ~60 ms longer for semantic than phonological lists in the right hemisphere (peak latency: hemisphere × word-list type $F_{1,9} = 12.9$, $P < 0.01$, right hemisphere: word-list type $F_{1,9} = 9.6$, $P < 0.05$, 50% latency in the descending slope: hemisphere × word-list type $F_{1,8} = 21.8$, $P < 0.005$, right hemisphere: word-list type $F_{1,8} = 5.9$, $P < 0.05$).

Discussion

Our word-list stimuli revealed both spatiotemporally distinct and overlapping effects of semantic and phonological priming during speech perception. The effects were concentrated to the superior temporal cortex bilaterally, with the center of activation falling close to the primary auditory cortex, in agreement with intraoperative functional lesion studies (see Boatman 2004 for a review). Build-up of expectation over the first 3 words, resulting in reduced activation, and the breakdown of expectation in the fourth word, resulting in enhanced activation, converged on the following sequence, summarized in Figure 6: The N100m activation strength was sensitive to phonological but not semantic mismatch in the left hemisphere, thus indicating processing of sound form (acoustic–phonetic or phonological analysis) at ~100 ms. Starting at ~250 ms, the emphasis

Table 2

Strength and timing of activation to the list-final words in the left- and right-hemisphere N100m/N400m source areas (mean ± SEM), extracted from the individual source waveforms

	Left hemisphere				Right hemisphere			
	Phonrel	Phonunrel	Semrel	Semunrel	Phonrel	Phonunrel	Semrel	Semunrel
Time window 75–160 ms								
Peak amplitude (nAm)	24 ± 3	27 ± 4	21 ± 3	23 ± 3	18 ± 4	26 ± 6	22 ± 4	21 ± 4
Peak latency (ms)	124 ± 3	131 ± 3	123 ± 5	136 ± 5	123 ± 4	137 ± 3	124 ± 5	125 ± 4
Time window 160–250 ms								
Mean amplitude 160–250 ms (nAm)	14 ± 3	16 ± 3	10 ± 2	12 ± 4	10 ± 4	15 ± 4	12 ± 2	13 ± 3
Minimum amplitude (nAm)	8 ± 3	11 ± 3	5 ± 2	7 ± 4	4 ± 4	11 ± 3	7 ± 2	8 ± 3
Minimum latency (ms)	180 ± 5	195 ± 8	187 ± 6	199 ± 7	187 ± 6	201 ± 8	188 ± 6	199 ± 9
Time window 250–1000 ms								
Mean amplitude ascending slope (nAm)	36 ± 4	38 ± 5	28 ± 3	39 ± 5	25 ± 5	26 ± 5	23 ± 4	26 ± 5
Peak amplitude (nAm)	43 ± 5	45 ± 6	34 ± 4	47 ± 6	31 ± 6	34 ± 6	29 ± 5	35 ± 6
Mean amplitude descending slope (nAm)	31 ± 3	33 ± 4	20 ± 3	35 ± 4	21 ± 6	24 ± 5	20 ± 4	26 ± 5
50% latency ascending slope (ms)	268 ± 9	273 ± 7	269 ± 5	275 ± 10	261 ± 6	268 ± 6	272 ± 9	268 ± 10
Peak latency (ms)	466 ± 14	465 ± 20	435 ± 11	459 ± 10	388 ± 20	401 ± 22	446 ± 38	473 ± 38
50% latency descending slope (ms)	711 ± 18	735 ± 26	615 ± 21	757 ± 27	709 ± 57	726 ± 68	774 ± 57	796 ± 62

Note: The mean amplitude in the ascending/descending slope of the N400m response was computed for the 200-ms interval immediately preceding/succeeding the group mean peak latency in the list-initial word.

was on semantic effects, with involvement of the left superior temporal cortex until ~450 ms, after which semantic effects were seen bilaterally. From ~450 ms onwards there was also a subtle effect of phonological priming/mismatch in the right superior temporal cortex. Processing of sound form thus started off left-lateralized at ~100 ms and was then overridden by analysis of meaning. Influence of sound form was present again after ~450 ms, with significant effects in the right superior temporal cortex. Semantic analysis, on the other hand, was initially lateralized to the left hemisphere from ~250 ms onwards but showed bilateral involvement after ~450 ms.

The overall spatiotemporal sequence of cortical activation evoked by spoken words agreed with previous reports (Helenius et al. 2002; Biermann-Ruben et al. 2005; Bonte et al. 2006). A bilateral N100m response was followed by a bilateral N400m response in the superior temporal area. In all 10 subjects, an N100m/N400m source was detected using ECD analysis in both hemispheres, with the center of activation in close vicinity of the primary auditory cortex and current flow perpendicular to the Sylvian fissure. The results of MCE analysis confirmed this source configuration.

The response to the first word of the list was distinct from the responses to words later in the list, but only in the left hemisphere. To the first word, there was a separate posterior and/or anterior activation at ~200 ms (in 7/10 subjects), followed by the N400m response. The sources detected at ~200 ms resemble the ones described recently in response to isolated words (Biermann-Ruben et al. 2005; Bonte et al. 2006). The 200-ms response did not appear for the subsequent words in the list. Instead, activity in the N100m/N400m source area was stronger at ~200 ms and the N400m response started ~50 ms earlier than for the first word. Information thus seems to proceed through a more complex left-hemisphere pathway when a word is heard in isolation than when it is immediately preceded by another word.

The early bilateral suppression of the N100m response when advancing along the word list would seem to agree with the known reduction of the N100m response when (any) auditory
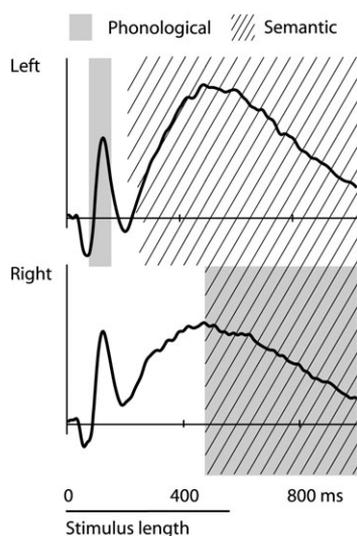

Figure 6. Summary of the main results. Schematic representation of the time windows and hemispheric interplay of phonological and semantic effects, overlaid on the N100m/N400m source waveforms. The gray and striped bars indicate the time windows in which phonological and semantic priming (mismatch) attenuated (increased) the response, respectively.

stimuli are presented successively at a relatively short interstimulus interval (Hari et al. 1982, 1987). However, the exceptionally strong N100m response to the phonologically unrelated list-final words, but no effect to semantically unrelated words, points to acoustic–phonetic and/or phonological analysis within this time window. An increased N100m response to phonological incongruency would certainly be in line with an EEG study using word pairs that shared, or not, the first 2 phonemes (Bonte and Blomert 2004) and MEG studies on speech versus nonspeech analysis that identified phonetic/phonological analysis within this time window in the left hemisphere (Gootjes et al. 1999; Parviainen et al. 2005). Alternatively, the enhanced response at ~100 ms may reflect build-up of an acoustic/phonetic mismatch response (MMF) directly on top of the N100m activation (Vihla et al. 2000) as suggested by the slight delay (~10 ms) of the peak latency as compared with the phonologically related list-final word. The results are in agreement with previously reported sensitivity of the supratemporal auditory cortex to the phonological structure of speech sounds by ~150 ms (Näätänen et al. 1997; Phillips et al. 2000; Vihla et al. 2000).

In the current experiment, in contrast to a number of previous studies (Connolly and Phillips 1994; D'Arcy et al. 2004; Kujala et al. 2004), there was no separate response sensitive to phonological manipulations at 200–350 ms (PMN), that is, during the ascending slope of the N400m response. A possible explanation for this difference is that those studies used sentences or visually primed auditory words as stimuli which might induce different requirements for the word recognition system than the present paradigm. We did identify activation at ~200 ms (in 7/10 subjects) that was separate from the N100/N400m response with anterior and/or posterior temporal locations but this activation was only detected for the list-initial word and showed no sensitivity to phonological manipulation.

The strongest effects were detected in the N400m time window where, in general agreement with earlier reports, we observed both semantic priming effects (e.g., Connolly and Phillips 1994; Helenius et al. 2002) and phonological priming effects that were overall weaker than those for semantic manipulation (e.g., Radeau et al. 1998; Perrin and Garcia-Larrea 2003). The present data suggest that the N400m response consists of 2 functionally separable parts because priming/mismatch affected the ascending and descending slopes in the 2 hemispheres differently. In the earlier time window we found left-lateralized effects to semantic manipulation whereas in the later time window the effects were more bilateral, and phonological manipulation affected the response as well.

The time window of the semantic priming effect detected here (250–1000 ms) agrees with earlier findings (e.g., Radeau et al. 1998; Hagoort and Brown 2000; Helenius et al. 2002; Perrin and Garcia-Larrea 2003). Timing of phonological priming effects, however, seems to be somewhat more variable across studies. Here, an effect of sound form emerged within the first 250 ms and again after 450 ms. Most of the previous experiments on phonological priming have used rhyming word pairs as stimuli (Praamstra and Stegeman 1993; Praamstra et al. 1994; Radeau et al. 1998; Perrin and Garcia-Larrea 2003); the priming effects were observed from 300–400 ms onwards, consistent with the later time window detected here. Different timing of mismatch in the rhyming versus alliterating auditory words might, however, complicate the comparison between these 2 types of studies. When alliterating word pairs were used, the

phonological priming effect began at ~100 ms (Praamstra et al. 1994; Bonte and Blomert 2004), in agreement with the present findings. In those studies there was either no later priming effect (Bonte and Blomert 2004) or it occurred earlier (250–450 ms; Praamstra et al. 1994) than in the present study. One important point to consider is that our MEG study allows analysis at the source level, with easy separation of activation in the left versus right superior temporal cortex. One may, thus, expect improved sensitivity of stimulus effects compared with the earlier EEG studies in which analysis was performed on the scalp electrodes that necessarily provide a somewhat blurred image of the underlying neural activation. It is also important to note that the expectation in our experiment was stronger than when using word pairs. Accordingly, effects of sound form in the early time window (~100 ms) appear to be the most reliable. During the N400 time window, the effects are more variable and clearly weaker than influence of semantic aspects.

Laterality of speech perception has been investigated extensively using hemodynamic neuroimaging methods (for reviews, see e.g., Binder et al. 2000; Hickok and Poeppel 2004; Scott and Wise 2004). In those studies, acoustic–phonetic processing has been consistently associated with bilateral activation of the superior temporal cortex. Processing of phonemes has been suggested to take place bilaterally as well (Binder et al. 2000; Hickok and Poeppel 2004). This agrees with our MEG results that show robust early processing of sound form reaching significance in the left hemisphere, followed by involvement of the right hemisphere at a later time, which would appear as bilateral activation in hemodynamic measurements that are not sensitive to timing. As for lexical–semantic processing, hemodynamic studies mainly propose lateralization to the left hemisphere in accordance with our MEG data that indicate strongest influence of semantic manipulation in the left hemisphere over a long time interval, accompanied by a weaker and later effect in the right hemisphere. Recently, priming paradigms have been modified for functional magnetic resonance imaging use, as well (adaptation paradigms; Kotz et al. 2002; Rissman et al. 2003). In those experiments, semantically unrelated word pairs resulted in stronger activation than related word pairs in areas including superior temporal cortex bilaterally (Kotz et al. 2002) or in the left hemisphere (Rissman et al. 2003), in agreement with the current MEG data.

The influential cohort model (Marslen-Wilson and Welsh 1978) divides the analysis of a spoken word into 3 processing stages: lexical access, lexical selection, and postlexical integration. The lexical access stage denotes an automatic bottom-up process that activates representations of words whose first phoneme(s) match the acoustic input. At the lexical selection stage, the number of activated word representations is reduced to one candidate word that best fits the further input and context created by the preceding words. At the postlexical integration stage, the selected candidate word is merged with this context.

According to behavioral data, spoken words can be identified based on the first 330 ms of input when heard in isolation and based on the first 200 ms within sentence context (isolation point, Grosjean 1980). In the current experiment, the latency at which the word could be identified can be assumed to fall between these 2 estimates. Our early effect of sound form reached the maximum at ~140 ms and was finished by ~200 ms, thus implying that it did not extend to the isolation point but was probably associated with the stage of lexical access (setting up of a cohort of candidate words). The decrease of activation over the course of the first 3 words in the phonological lists and the increased response to a phonologically unrelated final word demonstrated influence of contextual information in this time window. Based on the Cohort model, contextual effects signify lexical selection (exclusion of items from the cohort based on context or further acoustic input). In the present study, however, the words that had the same initial phonemes created a very unusual type of context that was composed solely of (possibly low-level) information about sound form and did not help to exclude words from the cohort, as all words beginning with those initial phonemes were equally probable at any position in the word list. Therefore, the neural effects cannot be readily accounted for by variation in the ease of lexical selection. Instead, they are likely to reflect the effort needed in setting up a cohort of candidate words, that is, lexical access, which is influenced by the context (auditory–phonetic–phonological similarity). Indeed, interactive models such as TRACE (McClelland and Elman 1986) or Logogen (Morton 1969) allow the context to affect processing at any stage, thus promoting the view that lexical access does not occur purely in a bottom-up fashion.

The distinction between bottom-up and top-down processes in speech perception is not clear-cut. In a recent MEG experiment (Bonte et al. 2006), effects of bottom-up and top-down information on speech processing were investigated by comparing responses to syllables that were either cut from sentences, and thus contained acoustic–phonetic bottom-up cues, or were produced in isolation. These syllables were presented in a sequence that contained only (meaningless) syllables or in a context of words and sentences that created an expectation for hearing meaningful speech. Top-down expectation was found to affect the N100m response and both the ascending and descending slopes of the N400m response bilaterally, but bottom-up cues had an effect specifically on the ascending slope of the left N400m response that was interpreted to reflect access of phonetic–phonological processes to lexical–semantic representations.

In the present experiment, the ascending slope of the N400m response (~250–450 ms) was sensitive specifically to semantic context and probably reflects lexical selection (elimination of candidate words from the cohort based on the appropriateness of their meaning in the context set by the preceding words). The effects of both meaning and sound form from 450 ms onwards may be understood in terms of postlexical integration of all available information in the context created by the successive words. Alternatively, the effect of sound form could reflect a postlexical congruency test, suggested by Praamstra et al. (1994) as a possible interpretation of their phonological priming effect in the N400m time window. The subjects may have consciously checked whether the list-final word started with the same phonemes as the previous words. Interestingly, these late effects of sound form started approximately at the time when the word stimulus ended; however, the stimulus length was not systematically varied in the present experiment.

The present detailed spatiotemporal characterization of analysis of sound form and meaning in speech perception thus provides the fine structure that may underlie the activation patterns obtained using temporally or spatially less sensitive neuroimaging methods.

Notes

References

Biermann-Ruben K, Salmelin R, Schnitzler A. 2005. Right rolandic activation during speech perception in stutterers: a MEG study. NeuroImage. 25:793–801.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 10:512–528.

Boatman D. 2004. Cortical bases of speech perception: evidence from functional lesion studies. Cognition. 92:47–65.

Boersma P, Weenink D. 2002. Praat, Institute of Phonetics, University of Amsterdam.

Bonte M, Blomert L. 2004. Developmental changes in ERP correlates of spoken word recognition during early school years: a phonological priming study. Clin Neurophysiol. 115:409–423.

Bonte M, Parviainen T, Hytönen K, Salmelin R. 2006. Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. Cereb Cortex. 16:115–123.

Collins A, Loftus E. 1975. A spreading-activation theory of semantic processing. Psychol Rev. 82:407–428.

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

D'Arcy RC, Connolly JF, Service E, Hawco CS, Houlihan ME. 2004. Separating phonological and semantic processing in auditory sentence processing: a high-resolution event-related brain potential study. Hum Brain Mapp. 22:40–51.

Dehaene-Lambertz G, Dupoux E, Gout A. 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. J Cogn Neurosci. 12:635–647.

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.

Gaskell MG, Marslen-Wilson W. 1997. Integrating form and meaning: a distributed model of speech perception. Lang Cogn Proc. 12:613–656.

Gootjes L, Raij T, Salmelin R, Hari R. 1999. Left-hemisphere dominance for processing of vowels: a whole-scalp neuromagnetic study. NeuroReport. 10:2987–2991.

Grosjean F. 1980. Spoken word recognition processes and the gating paradigm. Percept Psychophys. 28:267–283.

Hagoort P, Brown CM. 2000. ERP effects of listening to speech: semantic ERP effects. Neuropsychologia. 38:1518–1530.

Hämäläinen M, Hari R, Ilmoniemi R, Knuutila J, Lounasmaa O. 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys. 65:413–497.

Hari R. 1990. The neuromagnetic method in the study of human auditory cortex. Adv Audiol. 6:222–282.

Hari R, Kaila K, Katila T, Tuomisto T, Varpula T. 1982. Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. Electroencephalogr Clin Neurophysiol. 54:561–569.

Hari R, Pelizzone M, Mäkelä JP, Hällström J, Leinonen L, Lounasmaa OV. 1987. Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. Audiology. 26:31–43.

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. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 92:67–99.

Hill H, Strube M, Roesch-Ely D, Weisbrod M. 2002. Automatic vs. controlled processes in semantic priming—differentiation by event-related potentials. Int J Psychophysiol. 44:197–218.

Kotz SA, Cappa SF, von Cramon DY, Friederici AD. 2002. Modulation of the lexical-semantic network by auditory semantic priming: an event-related functional MRI study. NeuroImage. 17:1761–1772.

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.

Kutas M, Federmeier KD. 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci. 4:463–470.

Laine M, Virtanen P. 1999. WordMill lexical search program. Turku, Finland: University of Turku, Center for Cognitive Neuroscience.

Lütkenhöner B, Steinsträter O. 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol Neurootol. 3:191–213.

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, Welsh A. 1978. Processing interactions and lexical access during word recognition in continuous speech. Cogn Psychol. 10:29–63.

Matsuura K, Okabe Y. 1995. Selective minimum-norm solution of the biomagnetic inverse problem. IEEE Trans Biomed Eng. 42:608–615.

McClelland J, Elman J. 1986. The TRACE model of speech perception. Cogn Psychol. 18:1–86.

Morton J. 1969. Interaction of information in word recognition. Psychol Rev. 76, 165–178.

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, et al. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 385:432–434.

Parviainen T, Helenius P, Salmelin R. 2005. Cortical differentiation of speech and nonspeech sounds at 100 ms: implications for dyslexia. Cereb Cortex. 15:1054–1063.

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.

Posner M, Snyder C. 1975a. Attention and cognitive control. In: Solso R, editor. Information processing and cognition: the Loyola symposium. Hillsdale (NJ): Lawrence Erlbaum Associates, Inc. p. 55–85.

Posner M, Snyder C. 1975b. Facilitation and inhibition in the processing of signals. In: Rabbitt P, Dornic S, editors. Attention and performance. New York: Academic Press. p. 669–698.

Praamstra P, Meyer A, Levelt W. 1994. Neurophysiological manifestations of phonological processing: latency variation of a negative ERP component timelocked to phonological mismatch. J Cogn Neurosci. 6:204–219.

Praamstra P, Stegeman DF. 1993. Phonological effects on the auditory N400 event-related brain potential. Cogn Brain Res. 1:73–86.

Radeau M, Besson M, Fonteneau E, Castro SL. 1998. Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. Biol Psychol. 48:183–204.

Radeau M, Morais J, Dewier A. 1989. Phonological priming in spoken word recognition: task effects. Mem Cogn. 17:525–535.

Rissman J, Eliassen JC, Blumstein SE. 2003. An event-related fMRI investigation of implicit semantic priming. J Cogn Neurosci. 15:1160–1175.

Roland PE, Zilles K. 1996. The developing European computerized human brain database for all imaging modalities. NeuroImage. 4:39–47.

Sams M, Paavilainen P, Alho K, Näätänen R. 1985. Auditory frequency discrimination and event-related potentials. Electroencephalogr Clin Neurophysiol. 62:437–448.

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, Wise RJ. 2004. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 92:13–45.

Slowiaczek LM, Nusbaum HC, Pisoni DB. 1987. Phonological priming in auditory word recognition. J Exp Psychol Learn Mem Cogn. 13:64–75.

Slowiaczek LM, Pisoni DB. 1986. Effects of phonological similarity on priming in auditory lexical decision. Mem Cogn. 14:230–237.

Tiitinen H, Sivonen P, Alku P, Virtanen J, Näätänen R. 1999. Electromagnetic recordings reveal latency differences in speech and tone processing in humans. Cogn Brain Res. 8:355–363.

Uusitalo MA, Ilmoniemi RJ. 1997. Signal-space projection method for separating MEG or EEG into components. Med Biol Eng Comput. 35:135–140.

Uutela K, Hämäläinen M, Somersalo E. 1999. Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage. 10:173–180.

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.

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.

Vihla M, Salmelin R. 2003. Hemispheric balance in processing attended and non-attended vowels and complex tones. Cogn Brain Res. 16:167–173.

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.