Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 6, 1231–1247

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.6.1231

ERP Evidence for the Time Course of Graphic, Phonological, and Semantic Information in Chinese Meaning and Pronunciation Decisions

Ying Liu, Charles A. Perfetti, and Lesley Hart
University of Pittsburgh

Two words that varied in their relationship were presented sequentially to Chinese readers who made meaning and pronunciation decisions. In the meaning task, they decided whether the words had the same meaning. In the pronunciation task, they decided whether the words had the same pronunciation. In both tasks, the word pairs represented 1 of 4 relationships: graphically similar, homophonic, semantically related, or unrelated. Event related potentials (ERP) recordings made from the onset of the 2nd word suggested a temporal unfolding of graphic, phonological, and semantic effects. Specifically, graphically related pairs produced a smaller P200 in the pronunciation task and a smaller N400 in the meaning task. Homophones produced reduced N400 component with bilateral sources in the meaning task.

Reading a word involves a rapid temporal unfolding of information sources—the appearance of the word, its formal graphic constituents (e.g. letters or other graphic components), its pronunciation (phonological constituents), and its meaning. How these sources of information are used in any given reading event is a different question. In reading for meaning, semantic information must be used, and in reading aloud, phonological information must be used. More generally, skilled readers have knowledge about graphic, phonological, and meaning components that they use in any word-reading task. Theories of word reading are essentially about how these different sources of information become available and provide information for further processing. Theories with very different frameworks share the assumption that all three sources of information can be used when a word is read (e.g., Berent & Perfetti, 1995; Coltheart, Curtis, Atkins, & Haller, 1993; Forster & Davis, 1984; Seidenberg & McClelland, 1989).

Although the above description can apply to reading across all writing systems, most of the research that supports it comes from alphabetic writing systems, especially English. Chinese provides a case of high contrast for alphabetic systems, because its graphic units, characters, do not represent phonemes as alphabetic writing systems do. Instead a character typically represents a syllable-morpheme, usually a whole word, and sometimes a nonword morpheme. Not only does the Chinese system not provide phoneme level units, it allows the reader, in principle, to bypass phonology altogether by associating the character with a meaning. However, because the character also is associated with a spoken word, reading a character may activate both meaning and word-level pronunciation. Indeed, research in recent years has pointed to exactly that conclusion. When Chinese readers read to make meaning judgments, they showed confusion on the basis of pronunciation (Chua, 1999; Perfetti & Zhang, 1995; Xu, Pollatsek, & Potter, 1999; Zhang, Perfetti, & Yang, 1999). For example, Perfetti and Zhang (1995) presented one character followed rapidly by another and asked participants to make either a meaning judgment or a pronunciation judgment. The key result was bidirectional interference. When the pairs of words were homophones (unrelated in meaning), meaning judgments were slower and less accurate. When the words were related in meaning, pronunciation judgments were slower.

Perfetti and Zhang (1995) also examined the time course of these phonological and meaning effects. The interference provided by phonology emerged very rapidly when the stimulus onset asynchrony (SOA) between the first character and the second character was only 90 ms. Semantic interference in pronunciation judgments emerged within 140 ms. Thus, it appears that both the pronunciation and the meaning of a character are very rapidly and automatically available regardless of whether these constituents are required by the task.

In another study of the time course of word constituents, Perfetti and Tan (1998) varied the form and meaning relationships between a prime character and a to-be-named target character. Targets that were graphically similar to their primes showed priming within 43 ms (prime–target SOA). Later effects were observed for phonologically related primes (57 ms) and semantically related primes (85 ms). This temporal ordering of graphic, then phonological, then semantic information may be general across tasks, but it has been established only when pronunciation is required.

Recently, the Chinese meaning and pronunciation judgment tasks have been studied in functional magnetic resonance imaging (fMRI) by Tan et al. (2001). Compared with a fixation baseline, left-middle frontal (BA 9) and left-inferior frontal areas were found to be involved in both meaning and pronunciation tasks. Although the results generally showed a word-reading network for Chinese similar to what has been observed in English, they also showed more right-hemispheric regions (BA 47/45, 7, and 40/39 and the right-visual system) for reading Chinese than have been found in English studies. That these areas were found in both tasks provides evidence for a shared network that is partly independent

---

Ying Liu, Charles A. Perfetti, and Lesley Hart, Learning Research and Development Center, University of Pittsburgh.

This research was supported by National Science Foundation Grant 0113243. We thank Guosheng Ding for assistance in running participants.

Correspondence concerning this article should be addressed to Ying Liu, 3939 O'Hara Street, 645 Learning, Research, and Development Center, Pittsburgh, Pennsylvania 15260. E-mail: liuying@pitt.edu.

of whether the reader targets meaning or pronunciation. Of course, fMRI studies cannot be informative about the time course of activation within the time range of word-identification events. For this we turn to event related potentials (ERP).

ERP records can expose time-course information in the electrical potential shifts associated with processing before the emission of an overt response. Instead of varying the SOA between reading events, which is one way to study time course in behavioral studies, one can observe the temporal dimension more directly with a single word-reading event. An illustration of the potential of this method comes from a study by Bentin, Mouchetant-Rostaing, Giard, Echallier, and Pernier (1999), which examined the time course and scalp distribution of ERP during various target-detection tasks that tapped different levels of information—font-discrimination, rhyme-detection, lexical-decision, and abstractness-decision tasks. In the font task (visual level), words, pseudowords, strings of consonants, strings of alphanumeric symbols, and strings of forms elicited a sharp negative peak at 170 ms (N170) over occipito-temporal sites, larger in left-hemisphere electrodes for orthographic than for nonorthographic stimuli. In the rhyme-decision task (phonological level), the targets were words and pseudowords that rhymed with the French word *vitrail*. The most conspicuous potential was a bilateral temporal lobe negative peak at 320 ms, which was larger over left temporal (T3) than right temporal (T4) sites and larger for pronounceable stimuli than for nonpronounceable ones. In the abstractness task (semantic level), real words and pseudowords elicited an N450 component at frontal sites (F7, F3, FC5), larger for pseudowords. There was no N450 for nonwords, suggesting the component is reflective of orthographic–phonological structures that are typical of words regardless of actual lexicality.

In studies of the time course of lexical constituents, a particularly important issue is the existence of ERP markers for phonology. In addition to the 320-ms component identified by Bentin et al. (1999), research has pointed to 400–450-ms components. Rugg (1984) reported a late (450 ms) negativity component, larger over right hemisphere for nonrhyming pairs in a rhyme-decision task. Also using rhyme tasks, Kramer and Donchin (1987) found smaller N400 components associated with rhymed pairs even when the pairs were not orthographically similar. Barnea and Breznitz (1998) found a similar result in Hebrew, and reported a P200 effect for rhymed pairs. Similarly, in a Chinese pronunciation-decision task, Valdes-Sosa, Gonzalez, Liu, and Zhang (1993) found a reduced N400 when two characters had the same pronunciation. These studies demonstrate that ERP measures are sensitive to graphic and phonological processes that occur when readers make judgments about pronunciation and specifically point to a reduced negativity around 400 ms associated with phonological similarity when the reader's task is to assess phonology.

In all the above studies, participants were required to attend to the phonology. The role of phonology when words are read for meaning is a different question and requires a different experimental task. Although meaning tasks have been investigated with ERP measures, evidence on the role of phonology has not been consistent. Niznikiewicz and Squires (1996), using both word context and sentence context with meaning decision tasks in visual modality, failed to find any N400 difference. Instead, they found an increased N200 (peak at 293 ms) for homophones when compared with orthographic and semantic controls. Ziegler, Benraiess, and Besson (1999) replicated homophone interference effects in categorization (Van Orden, 1987), but failed to find an N400 associated with this effect. Given a semantic category (*food*), their participants produced more errors to homophones (*meet*) than to orthographic controls (*melt*) in the behavior experiment; but, in the ERP experiment, they found no short or midlatency ERP correlates of this effect. Observing some ERP differences in the 800–1600-ms range, they concluded that phonological effects in this task are postlexical.

Our goal in the present experiment was to compare ERP records of two reading tasks, one in which the reader must attend to pronunciation and one in which the reader must attend to meaning. In addition, we examined whether the ERP records can shed light on the time course of graphic, phonological, and semantic constituents that become available during word identification. By using 128 channel recordings, many more than those used in most studies, we were able to do a better source analysis on interesting ERP components. As part of a research program that compares alphabetic and nonalphabetic reading, we focus in this article on Chinese reading, using tasks that we know from behavioral studies provide temporal information on semantic and phonological processes.

For this purpose, native speakers of Chinese made two kinds of decisions when presented with a sequence of two characters. In one task, they decided whether the two characters had the same pronunciation. In another task, they decided whether the two characters had related meanings. To study the effects of form and meaning similarity across tasks, we presented pairs of words that varied in their form and meaning relationships: semantically related characters, same pronunciation characters (homophones), graphically similar characters, and unrelated characters. The tasks and the experimental logic were essentially those introduced by Perfetti and Zhang (1995), with the addition of a condition in which two characters had a high degree of graphical similarity.

This graphically similar condition provides information on visual components in character identification. Will a high degree of graphic similarity of the first and second character affect the reader's decision concerning their pronunciation? In English, one would expect such an effect because of the correlation between letters and sounds. In English, *teem* and *team* not only have the same pronunciation, but also have similar spellings. Because Chinese disassociates visual form and pronunciation, there is less reason to suppose that visual form should affect pronunciation judgments.

In the experiment, most graphically similar characters achieved their similarity by sharing a component radical. Half of them shared right-side phonetic radicals, which can give cues to pronunciation but, often, not valid cues. For example, characters 凉 [pronounced "liang2" (the number following the pronunciation refers to tone), meaning *cool*] and 惊 [pronounced "jing1," meaning *frighten*] both have 京 [pronounced "jing1," meaning *capital*] on the right side as a phonetic radical, but only the second character shares the pronunciation with the phonetic radical. The other half of the graphically similar pairs shared left-side semantic radicals, which can give a cue to meaning, although again it is not always a clear assistance to meaning. For example, characters 伏 [pronounced "fu2," meaning *prone*] and 仗 [pronounced "zhang4," meaning *fight*] both have 亻 [pronounced "ren," meaning *person*] on the left side. Regardless of the phonetic or semantic value of the shared radicals in a given character, the pair of visually similar characters themselves shared no phonological or semantic information.

Homophone pairs were characters that share the same onset, rhyme, and tone in their pronunciation (such as 惊 [pronounced "jing1," meaning *frighten*] and 晶 [pronounced "jing1," meaning *crystal*]). Thus, they were full homophones that share no graphical or semantic information. Semantically related pairs were characters that are related in their meaning (such as 惊 [pronounced "jing1," meaning *frighten*] and 扰 [pronounced "rao3," meaning *disturbing*]) but share no graphical or phonological components.

In addition to the use of visually similar characters, the present experiment had an important procedural departure from the Perfetti and Zhang (1995) experiments. The SOA was set at 640 ms, which included a 500-ms interstimulus interval (ISI). The longest SOA tested by Perfetti and Zhang was 310 ms, at which time they observed both phonological and semantic interference. The longer SOA ensured ample time for processing the first character and reduced ERP carryover effects from the first character.

To reduce the likelihood of strategic effects in the two decision tasks, we introduced another departure from Perfetti and Zhang's (1995) study by adding filler pairs, which were homophonic, semantically related, and visually similar (such as 住 [pronounced "zhu4," meaning *live*] and 驻 [pronounced "zhu4," meaning *stay*]). These fillers require positive decisions, thus eliminating the effectiveness of any strategy to respond negatively to either a homophone pair in the meaning task or a semantically related pair in the pronunciation task. They also work against the strategy of using graphically similar pairs as a cue for negative responses in either task.

In summary, the experiment was designed to expose the temporal unfolding of word constituent information in Chinese single-character processing as a function of whether the reader's task is to retrieve phonology (pronunciation-decision task) or semantics (meaning-decision task). By presenting two characters in succession and taking ERP recordings from the onset of the second character, we expect findings on the use of information sources over time during the processing of the second character. The key to this is the manipulation of form and meaning relationships between the two characters. For example, when the task is to decide whether the two characters have the same meaning, the encoding of the second character and the comparison process might be affected by the prior appearance of a visually similar or phonologically similar character. The negative trials can provide evidence for interference of the kind reported by Perfetti and Zhang (1995) and others (Hung, Tzeng, & Tzeng, 1992; Xu et al., 1999; Zhang & Perfetti, 1993; Zhang et al., 1999) and extend this evidence to cases of graphical similarity. Because the positive trials have a mixed effect of both congruency and making a positive response, we analyze only reaction time data for them. For ERP recording, the positive trials served as fillers.

Method

Participants

Five men and 5 women ranging in age from 22 to 32 years were recruited from the University of Pittsburgh and paid for their participation. All were native Chinese speakers who had been living outside of China for not more than 2 years. All 10 participants had normal or corrected-to-normal vision and were right-handed. All were free of medication for at least 1 week before the experiment and had no history of neurological diseases.

Stimuli

Tasks of meaning judgment and pronunciation judgment used the same 60 core characters, each of which was paired with three analogues: a phonologically identical character (a homophone), a character of related meaning, a character of similar graphic form (but without any shared phonology), and a control character. The materials are shown in the Appendix. In both tasks, a trial consisted of a core character paired with one of the analogues or the control. In the meaning-judgment task, pairs of related meaning characters required a *yes* response, and all other trials required a *no* response. In the pronunciation-judgment task, pairs of homophones required a *yes* response, and all other trials required a *no* response. The trials were organized into two sessions for each task with four blocks in each session. Each block included 4 types of experimental pairs: 15 semantically related pairs, 15 homophone pairs, 15 graphically similar pairs, 15 control pairs, and 20 filler pairs. The fillers added pairs of homophones in the pronunciation task and semantically related pairs in the meaning task, so that the numbers of *yes* and *no* trials were approximately balanced. The order of blocks was counterbalanced for participant and session. Thus, an experimental block consisted of 80 trials, each containing a pair of characters. The second session of each task was a repetition of the first session to increase the number of trials for each experimental condition. The two tasks together had four sessions totaling 1,280 trials presented in 16 blocks.

Procedure

All participants made both meaning judgments and pronunciation judgments. The meaning judgment came first for 5 participants, and the pronunciation judgment came first for the other 5. The length of each task varied from 1 to 1.5 hr across participants. For all participants, the experimental sessions for the two tasks were separated by 2 weeks.

A trial sequence is illustrated in Figure 1. Each trial began with a warning signal (~***~) that stayed on the center of the screen until the participant decided to begin the trial. Because any eye blink or movement resulted in a loss of the trial, participants were told to try hard to not move or blink once a trial had begun. The warning signal allowed participants to pace themselves and to blink between trials. After the participant pressed the space bar, a fixation cross appeared on the screen for 500 ms, followed by the first character, which remained for 140 ms. Following a 500-ms blank interval, the second character appeared at the same location and stayed on the screen until a decision button was pressed. The two characters appeared in the same font and size. In the pronunciation-judgment task,

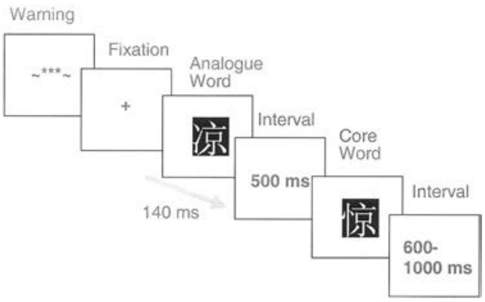

Figure 1. Experiment trial sequence.

the participants decided whether the two sequentially presented characters were homophones. In the meaning-judgment task, they decided whether the two characters were similar in meaning. Core characters always appeared as the second character. Thus, the key measurements were always on the same core character, and what preceded it was the variable. There was a blank interval that lasted between 600 and 1,000 ms randomly before the next warning signal. The experiment took place in a dedicated ERP lab, located in an isolated, quiet room. Each participant viewed the trials on a computer screen in the testing room, while the experimenter monitored the ERP recordings in an adjacent room.

Electroencephalogram (EEG) Recording and Averaging

A 15-in. cathode-ray tube (CRT) monitor working at a 60 Hz refresh rate displayed the instructions and stimuli. The experimental trials were controlled by commercial software, Eprime, which presented the trials and recorded the reaction times. It also sent event information to the EEG recording system. A 128-channel geodesic sensor net (EGI net station, Electrical Geodesics Inc., Eugene, Oregon) was used to collect the EEG data. All impedances were kept below 40KΩ (Ferree, Luu, Russell, & Tucker, 2001). A vertex reference was used in the recording, and the data were recomputed off-line against the average reference (Lehmann & Skrandies, 1980). Six eye channels were recorded to allow rejection of trials with eye movements and blinks. The signals were recorded continuously at 500 Hz by the net station with a 12-bit A/D converter. The hardware filter setting was between 0.1 and 200 Hz. The EGI net station also recorded all event onset times, reaction times, and accuracy for later use in data analysis.

Event-related potentials were averaged off-line over the four conditions in each task after eliminating eye movements and other artifacts. After a baseline correction and a 30 Hz software low-pass filter, grand mean ERPs over participants were calculated.

Data Analyses

Behavioral measures. One participant was rejected because of the high error rate. Reaction time and accuracy data were calculated for 9 participants. Reaction times larger than two standard deviations were excluded from further analysis. Repeated measure analyses of variance (ANOVA) were performed on both reaction time and accuracy data.

ERP measures. Another participant was eliminated because of too many eye blinks. Thus, eight participants provided data for ERP analysis. The average EEG waveform peak was calculated first to show some possible components. Then, a principle-component analysis (PCA) was performed from the onset of the target to 1,000 ms after it. The 500 sample points were down sampled to 250 4-ms time points to reduce the computational burden. The combination of 129 electrodes, 8 participants, and 10 stimuli types totaled 10,320 observations at each time point for the PCA. Correlation matrix and Varimax rotation (Chapman & McCrary, 1995; Picton et al., 2000; van Boxtel, 1998) were used to do the analysis. PCA scores were computed for each component in each condition and for each participant followed by repeated measure multivariate analyses of variance (MANOVA) to test the experiment effects.

Spatial analysis of EEG. Low resolution electromagnetic tomography (LORETA) was used to locate the spatial source of significant ERP components. LORETA, an algorithm that solves the inverse problem of EEG (Pascual-Marqui, Michel, & Lehmann, 1994), is implemented by finding the "smoothest" of all possible solutions consistent with the scalp distribution. A particular advantage of LORETA is that it does not require any assumptions about the number and location of possible sources. The solution space of LORETA consists of 2,394 pixels with 7 mm resolution. Although this is a coarse analysis compared with fMRI, it is informative when used together with the temporal information provided by ERP. The LORETA-KEY software (Pascual-Marqui, Michel, & Lehmann, 2000) was used in the analysis. The version used was registered to the Talairach brain atlas (Talairach & Tournoux, 1988). The weights for computing solutions (transformation matrix) used in the present study were computed based on the position of electrodes on our recording net by using a tool provided with LORETA-KEY.

Results

Behavioral Results

There are three factors in the experiment: task (pronunciation and meaning), repetition (first and second session), and trial type (homophone pairs, semantically related pairs, graphically similar pairs, and control pairs). In a few cases, eye blinks in one of the sessions were too frequent; to keep the ERP data and reaction time data consistent, the reaction time data corresponding to lost ERP data were replaced by the participant treatment mean. Table 1 shows the average reaction times and accuracies of the four conditions in the two tasks. For reaction time, the ANOVA confirmed reliable main effects of trial type, $F(3, 6) = 9.11$, $p < .05$, and repetition $F(1, 8) = 20.03$, $p < .01$. The main effect of task was not reliable, $F(1, 8) = 1.37$, $p = .28$, nor were there any significant interactions. ANOVA on accuracy showed a same test result. Because there were no interactions, we collapsed the data over the first and second session for subsequent analyses.

In the pronunciation task, graphically similar pairs required significantly longer (100 ms) time to reject than control pairs, $F(1, 14) = 22.22$, $p < .01$, and made more errors, $F(1, 14) = 13.56$, $p < .01$. A 24-ms slowdown for semantic pairs was not reliable, $F(1, 14) = 3.67$, $p = .076$. In the meaning task, homophones (23 ms) and graphically similar pairs (32 ms) took longer to reject than controls, but neither difference was reliable, $F(1, 16) = 2.93$, $p = .11$, and $F(1, 16) = 2.98$, $p = .10$, respectively. The accuracy of graphically similar pairs was significantly lower than controls, $F(1, 16) = 9.35$, $p < .01$. The times to reach correct *yes* decisions

Table 1
Mean Reaction Time in Milliseconds and Percentage Accuracy in Two Tasks

	Word type											
	Semantically related			Homophone			Graphically similar			Control		
Task	*M*	*SE*	%	*M*	*SE*	%	*M*	*SE*	%	*M*	*SE*	%
Pronunciation	838	42	95.07	829	68	94.67	914*	54	90.87*	814	42	95.53
Meaning	844*	30	82.65*	808	29	91.88	817	38	89.35*	785	29	92.24

* $p < .05$ relative to control.

were very similar across the two tasks. However, there was a significant 59-ms reaction time difference and a 10% accuracy difference between a *yes* decision and a *no* decision for controls in the meaning task, $F(1, 16) = 7.40$, $p < .05$, and $F(1, 16) = 9.78$, $p < .01$, respectively, which suggests that reaching a positive meaning decision is more difficult relative to controls than reaching a positive pronunciation decision relative to controls.

ERP Results

Outputs from nine international 10–20 system channels (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4) are shown in Figures 2 and 3, in uV over the 1000-ms recording period from the onset of the second character. These nine electrodes were selected because most phonology and meaning-related ERP shifts were within their coverage area when we visually inspected the 129 electrode waveforms. Figure 2 shows the data for the four stimulus types in the pronunciation task, and Figure 3 shows these data in the meaning task.

EEG waveform and peaks. Four peaks alternating between negative and positive are visible at Cz: a negative peak around 100 ms (N100), a positive peak around 200 ms (P200), a negative peak around 400 ms (N400), and a positive peak around 300 ms (P300). The N100 peaks at 130 ms. The P200 peaks at 200 ms and is visible at other frontal and central electrodes. The N400 complex has two negative peaks: The earlier one peaks at 340 ms and the second one peaks at 430 ms. The N400 is also visible at left frontal and central electrodes (F3, Fz, and C3). The P300 shows a positive peak at around 570 ms for positive response conditions (homophones pairs in the pronunciation task and semantically related pairs in the meaning task).

PCA components. As a data-driven method, PCA extracts a small number of independent components from a large number of variables. The data we submitted to PCA had 250 variables, each corresponding to a 4-ms window of ERP. In the PCA solution, nine components with eigenvalues larger than 1 explained 96.8% of the total variance. There is a sharp decrease in the eigenvalue from Components 4 to 5 (32.99 vs. 5.45; see breakpoint in Figure 4). Furthermore, the component loadings after Component 4 did not match any grand ERP waveform amplitude change. Accord-

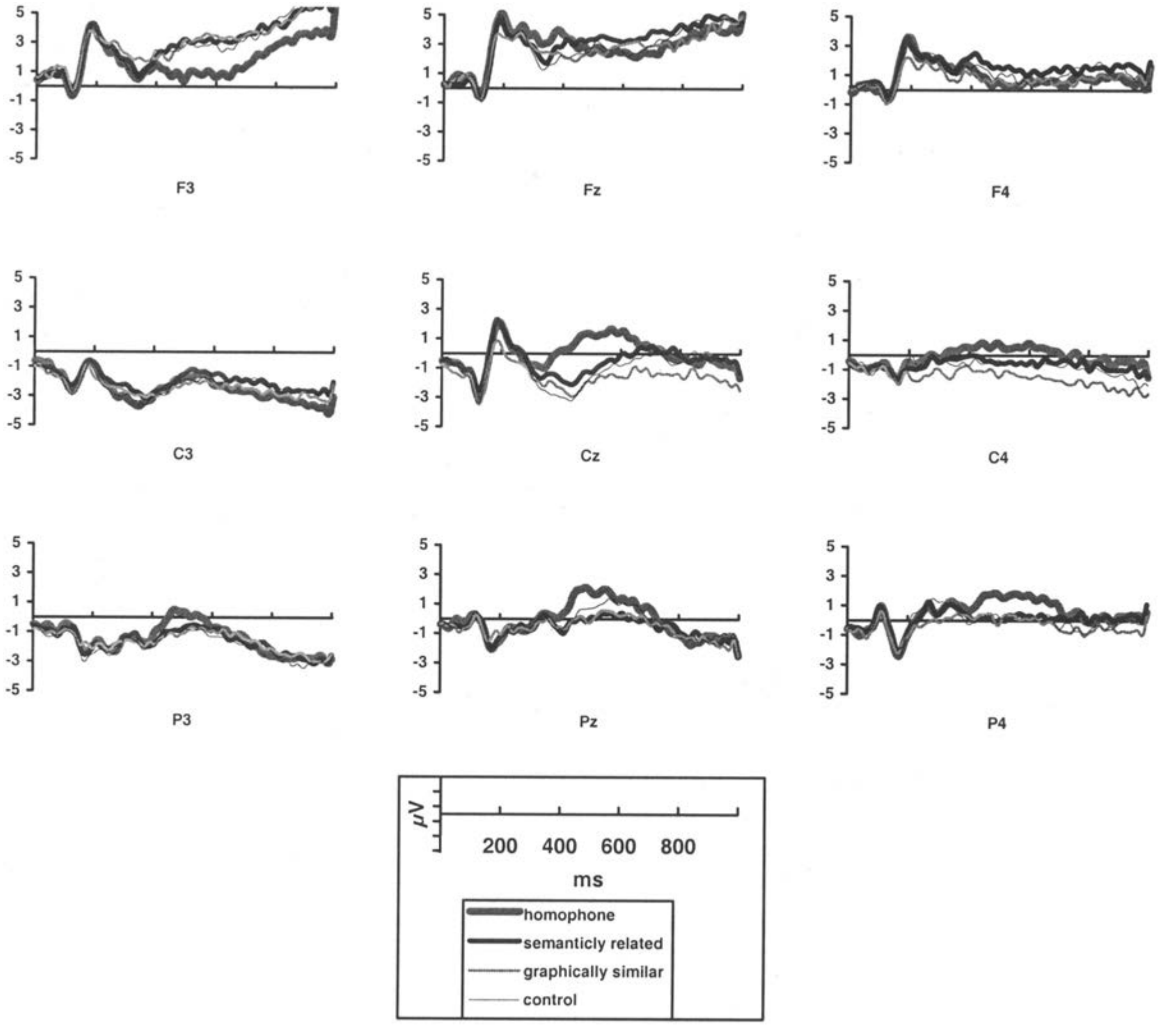

Figure 2. Grand average event related potentials in the pronunciation task.

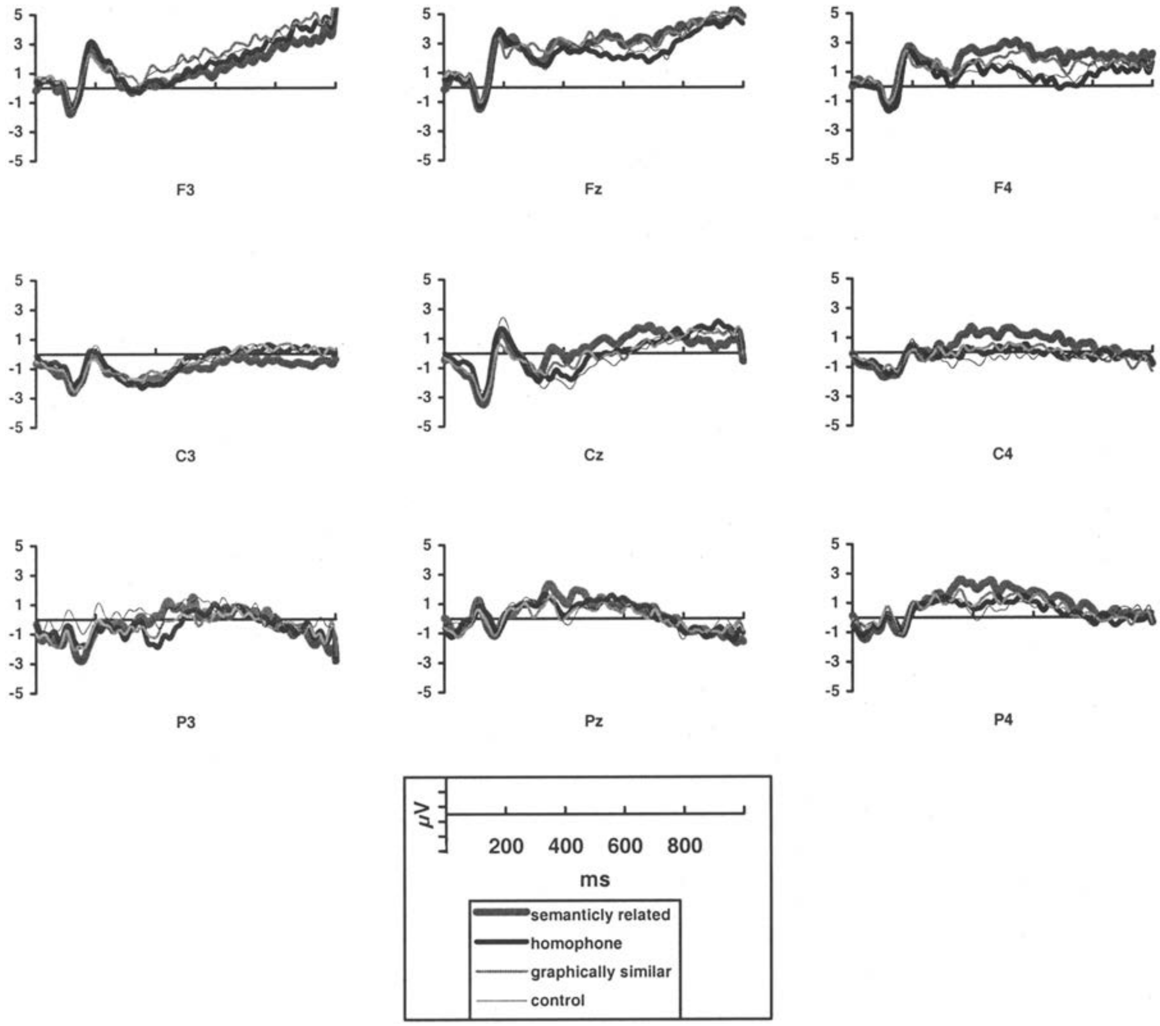

Figure 3. Grand average event related potentials in the meaning task.

ingly, only Components 1 to 4, which explained 90.4% of the total variance, were included in further statistical tests. Figure 5 shows the four major PCA component loadings over time. Figures 6 and 7 show the averaged factor scores of the nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4).

Component 1 (44.03% explained variance) is a slow wave component that has positive component scores. Similar components are widely found in other temporal PCA studies. Although such a component may reflect cognitive activity, more often it is the result of the autocorrelated nature of the data and reflects direct current drift over the trial (Wastell, 1981).

Component 2 (19.35% explained variance) rises from 200 ms and lasts until 1,000 ms, with a peak loading at 376 ms and a negative average component score. This component matches the negative complex in our grand averaged ERP data on Cz (Figures 2 and 3) and reflects the fact that the N400s are larger in the

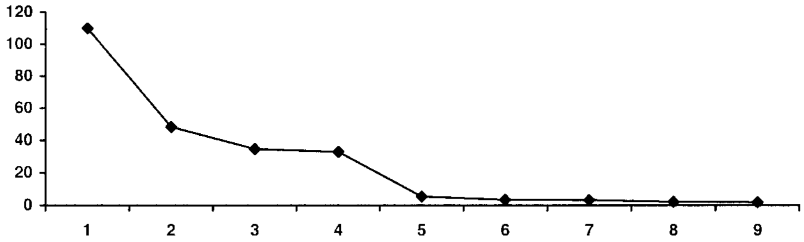

Figure 4. Eigenvalues of nine components.

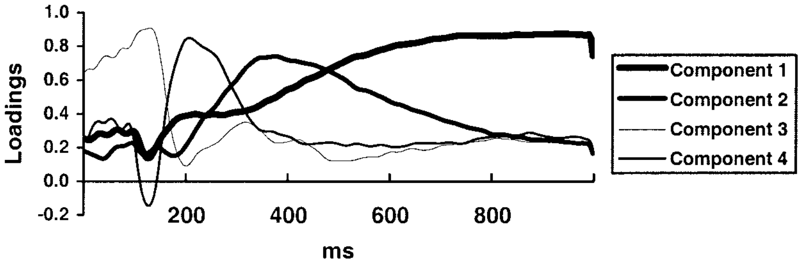

Figure 5. Principle component analysis (PCA) factor loadings of four major components.

incongruent conditions of both tasks. This complex component is most clearly observed at medial-left frontal and central sites. Although there are two visible negative peaks between 300 and 500 ms, the PCA did not separate them because they were functionally correlated across tasks, conditions, and participants in the present study.

Component 3 (13.88% explained variance) shows a large loading from the onset and peaks at 132 ms. The average component score of this component is negative. It matches the N100 peak in the grand averaged ERP at frontal and central sites. This component turns into a positive shift at parietal electrodes.

Component 4 (13.19% explained variance) rises from 160 ms and very sharply reaches a peak at 208 ms. Its loadings drop under 0.5 at 304 ms. The average component score is positive at frontal and central sites. The positive peak of grand averaged ERP at 192 ms at frontal and central sites is consistent with this component. The peak is more positive at frontal than central sites and turns into a negative shift at the parietal sites.

MANOVA and t tests. To determine the effects of the experimental variables on the observed PCA components, we carried out repeated measure MANOVAs on the component scores for the nine electrodes shown in Figure 2 and 3. The full set of electrodes can, however, be used meaningfully in the source analysis, which is reported in next section. For this nine-electrode analysis, a separate MANOVA was carried out for each component and each task with trial type and electrode location (lobe and hemisphere) as

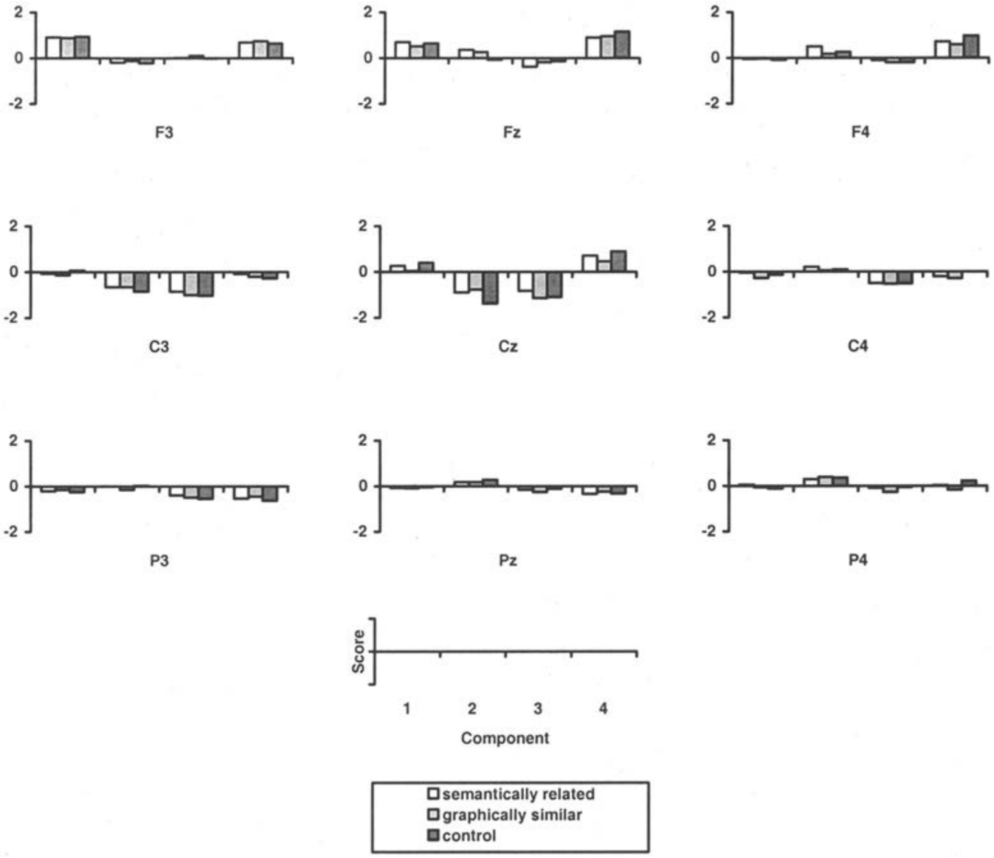

Figure 6. Average component scores of three types of negative trials in the pronunciation task.

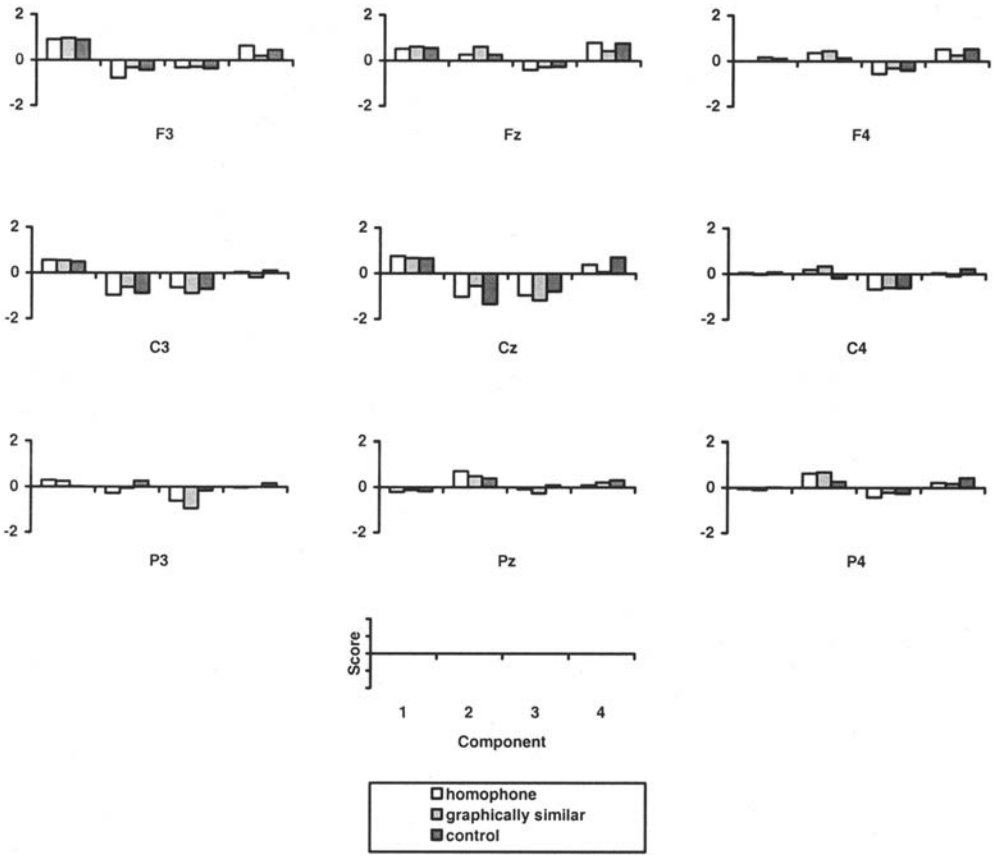

Figure 7. Average component scores of three types of negative trials in the meaning task.

factors. Because the main interest was in the interference effects on negative trials, only negative trials were analyzed. Pronunciation and meaning tasks were analyzed separately because the negative trial types were different in the two tasks. F and p values are reported in Tables 2 and 3.

Component 1 showed significant trial type and hemisphere effects in the pronunciation task. Pairwise t tests between three conditions showed that the component score of graphically similar pairs was more negative than control pairs at C3. In the meaning task, only the lobe effect was significant.

Component 2 showed no significant effects in the pronunciation task. The meaning task produced a trial type effect of borderline reliability, $p = .057$, and a reliable interaction of trial type and lobe. Post hoc paired t tests showed the factor scores of homophone pairs were more positive than control at Cz ($p < .01$), C4 ($p < .05$), and P4 ($p < .05$). The effect was reversed at F3 ($p <$

Table 2
Analysis of Negative Trial Types in Pronunciation Task

MANOVA	*df*	Component 1	 2	 3	 4
Trial type	2, 6	$F = 5.56$, *	$F = 2.39$, $p = .173$	$F = 0.66$, $p = .551$	$F = 4.53$, $p = .063$
Lobe	2, 6	$F = 1.03$, $p = .413$	$F = 4.30$, $p = .069$	$F = 10.24$, *	$F = 4.28$, $p = .070$
Hemisphere	2, 6	$F = 12.09$, **	$F = 4.20$, $p = .072$	$F = 1.82$, $p = .242$	$F = 7.01$, *
Trial Type × Lobe	4, 4	$F = 5.59$, $p = .062$	$F = 5.30$, $p = .068$	$F = 2.94$, $p = .160$	$F = 6.10$, $p = .054$
Trial Type × Hemisphere	4, 4	$F = .60$, $p = .681$	$F = 1.21$, $p = .429$	$F = 2.69$, $p = .180$	$F = 4.96$, $p = .075$

Note. MANOVA = multivariate analysis of variance.
* $p < .05$. ** $p < .01$.

Table 3
Analysis of Negative Trial Types in Meaning Task

MANOVA	*df*	Component 1	Component 2	Component 3	Component 4
Trial type	2, 6	$F = 0.22$, $p = .812$	$F = 4.77$, $p = .057$	$F = 1.02$, $p = .416$	$F = 2.88$, $p = .133$
Lobe	2, 6	$F = 6.11$, *	$F = 9.53$, *	$F = 4.26$, $p = .071$	$F = 3.09$, $p = .120$
Hemisphere	2, 6	$F = 4.58$, $p = .062$	$F = 11.65$, **	$F = 0.78$, $p = .502$	$F = 18.53$, **
Trial Type × Lobe	4, 4	$F = 0.52$, $p = .731$	$F = 6.97$, *	$F = 1.13$, $p = .456$	$F = 3.74$, $p = .115$
Trial Type × Hemisphere	4, 4	$F = 1.54$, $p = .342$	$F = 2.22$, $p = .229$	$F = 1.52$, $p = .347$	$F = 0.28$, $p = .876$

Note. MANOVA = multivariate analysis of variance.
* $p < .05$. ** $p < .01$.

.05), which showed a more negative score for homophone pairs. Graphically similar pairs were significantly more positive than control pairs at Cz ($p < .01$), C4 ($p < .01$), and P4 ($p < .01$).

Component 3 showed a lobe effect in the pronunciation task. Component 4 had a marginally significant trial type and lobe interaction ($p = .054$) in the pronunciation task. Post hoc paired *t* tests showed that the factor scores of graphically similar pairs were significantly more negative than control pairs at F4 ($p < .05$), Cz ($p < .01$), and P4 ($p < .01$). In the meaning task, only the hemisphere effect was significant ($p < .01$).

Source analysis (LORETA). We used LORETA to determine the spatial maps of ERP sources, subjecting differences to statistical parametric mapping (SPM). Several studies have successfully applied parametric statistical tests on LORETA maps with either region of interest (ROI) or pixel-by-pixel tests (Anderer, Pascual Marqui, Semlitsch, & Saletu, 1998; Pizzagalli, Lehmann, Koenig, Regard, & Pascual-Marqui, 2000; Strik, Fallgatter, Brandeis, & Pascual-Marqui, 1998). Because SPM assumes Gaussian distributions of pixels (Friston, Frith, Liddle, & Frackowiak, 1991), we carried out a log transformation of the value of each LORETA pixel, producing a Gaussian distribution (see histogram in Figure 8). The second assumption of SPM, smoothness across neighboring pixels, is satisfied directly by the LORETA output. The *t* test threshold was set to $p < .01$, and cluster size was set to ≥ 3 pixels (1,029 ms^3 volume size) as is common in positron emission tomography (PET) and fMRI studies.

LORETA focused on three comparisons we found in the PCA results: graphical versus control in the pronunciation task (Component 4 at 192 ms), graphical versus control in the meaning task, and homophone versus control in the meaning task (Component 2 at 386 ms). Both time points showed a peak on global field power (Lehmann & Skrandies, 1980) and matched the ERP waveform peaks of Components 4 and 2. The 8 participants were computed separately. For all participants, 129 electrodes were submitted to LORETA. Then, pixel-by-pixel paired *t* tests (within participants) were performed on the three comparisons. Because the scalp electrodes showed significant differences, the LORETA test was used only to identify the source of differences, not as an additional test of whether there were differences.

The source for the graphical effect at 192 ms in the pronunciation task is shown in Figure 9. The figure is a threshold *p* map of the above *t* test result. For 2,394 cortex pixels (noncortex pixels are in gray), each cortex pixel has three possible values: 0 for $p > .01$ (white pixels in figures); 1 for $p \leq .01$ with a positive mean difference (not shown in figures because there are no significant positive pixels in the present study); -1 for $p \leq .01$ with a negative mean difference (black pixels in figures). All significant pixels are negative in this map, showing less activation for graphically similar pairs than control pairs in those areas. The result is consistent with the smaller factor score of graphical pairs. Those areas are right hemisphere Brodmann areas (BA) 6 and 8 (x = 25, y = 17, z = 50, 6 pixels).

Figure 10 shows the test between graphically similar pairs and control pairs in the meaning task at 386 ms. There is less activation for graphically similar pairs than control pairs at right BA 44 and 22 (x = 50, y = 15, z = 8, 3 pixels). Figure 11 shows the

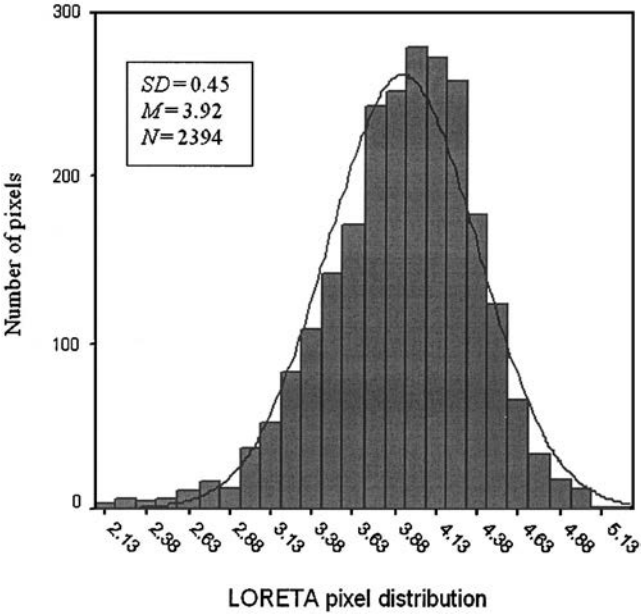

Figure 8. Distribution of mean log transformed low resolution electrical tomography (LORETA) density across 2,394 pixels.

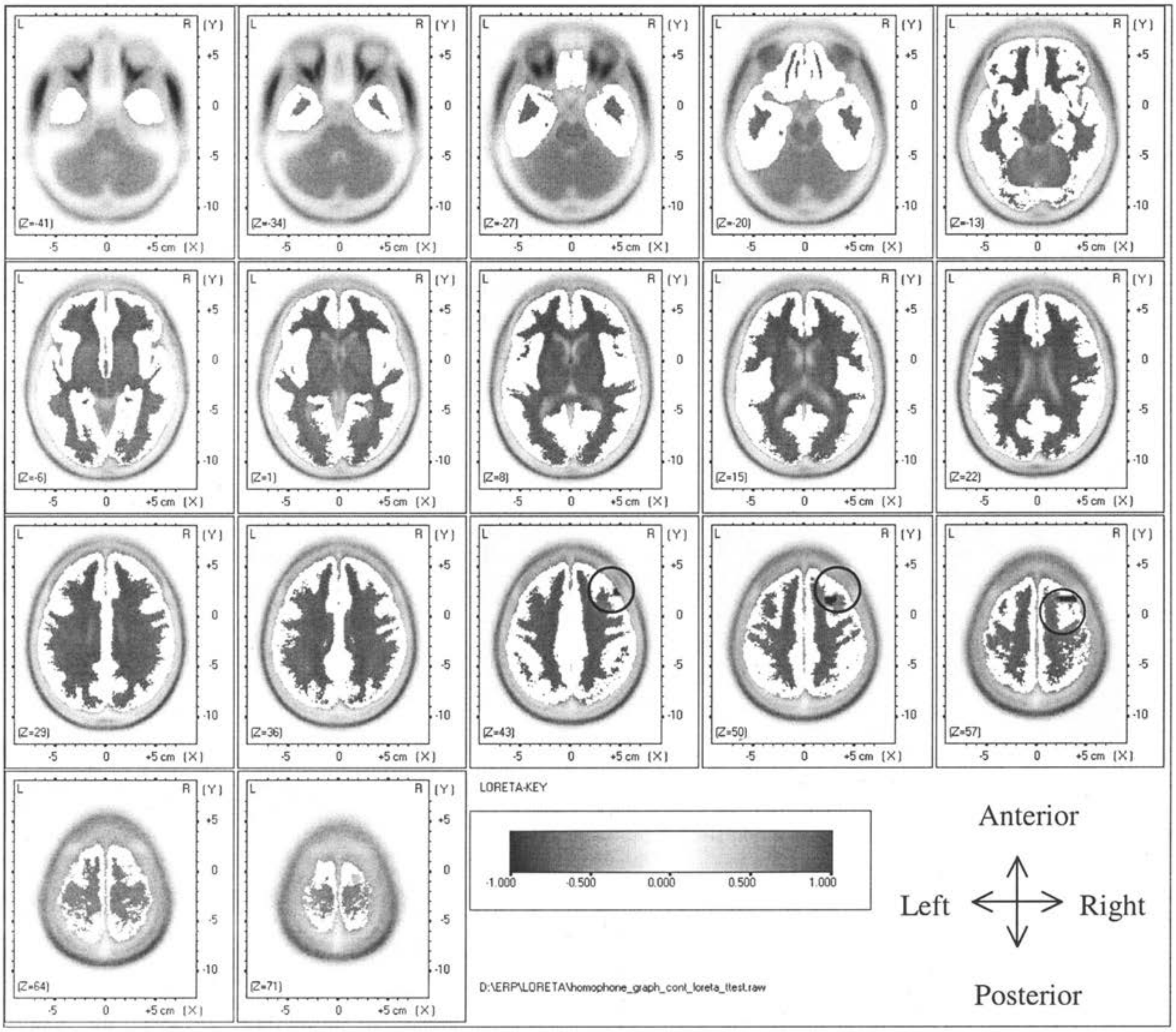

Figure 9. Top view low resolution electromagnetic tomography (LORETA) *p* map of graphical similar versus control pairs in pronunciation task at P200 ($p < .01$). Gray = noncortex; white = $p > .01$ cortex; black = $p \leq .01$ cortex with higher density for control pairs (circled for easy identification).

phonology effect in the meaning task at 386 ms. In the right hemisphere, BA 22 (x = 46, y = 10, z = 1, 2 pixels, superior temporal, auditory association) and inferior BA 44 (x = 50, y = 10, z = 8, 4 pixels, Broca's area) show less activation for homophone pairs than for control pairs. In the left hemisphere, BA 6 (x = −50, y = −2, z = 50, 6 pixels), BA 3, and BA 4 (near central sulcus, 4 pixels) also show less activation for homophone pairs than for control pairs.

Discussion

The present results extend and replicate the pattern of decision times observed by Perfetti and Zhang (1995) and Zhang et al. (1999). The key new result is a graphically based interference effect in both meaning decisions and pronunciation decisions. This result adds to the phonological interference effect observed in meaning decisions in previous studies. The phonological interference effect itself was replicated. The two previous studies found highly significant homophone interference effects of 46–52 ms over several experiments, whereas the present results showed a 23-ms effect with lower statistical reliability in the decision times, coupled with a phonological interference marker in the ERP data. It is likely that the robustness of the decision-time effects across experiments varies with specific materials, such as SOAs, fillers, and the inclusion of graphically similar characters. For example, the SOA of the present study was longer than the longest one used in previous studies (640 ms compared with 350 ms), allowing additional time for the participants to focus on meaning of the first character. Nevertheless, the interference was clearly in evidence in the ERP records, pointing to the value of ERP data for revealing effects that escape easy detection by behavioral measures. In the

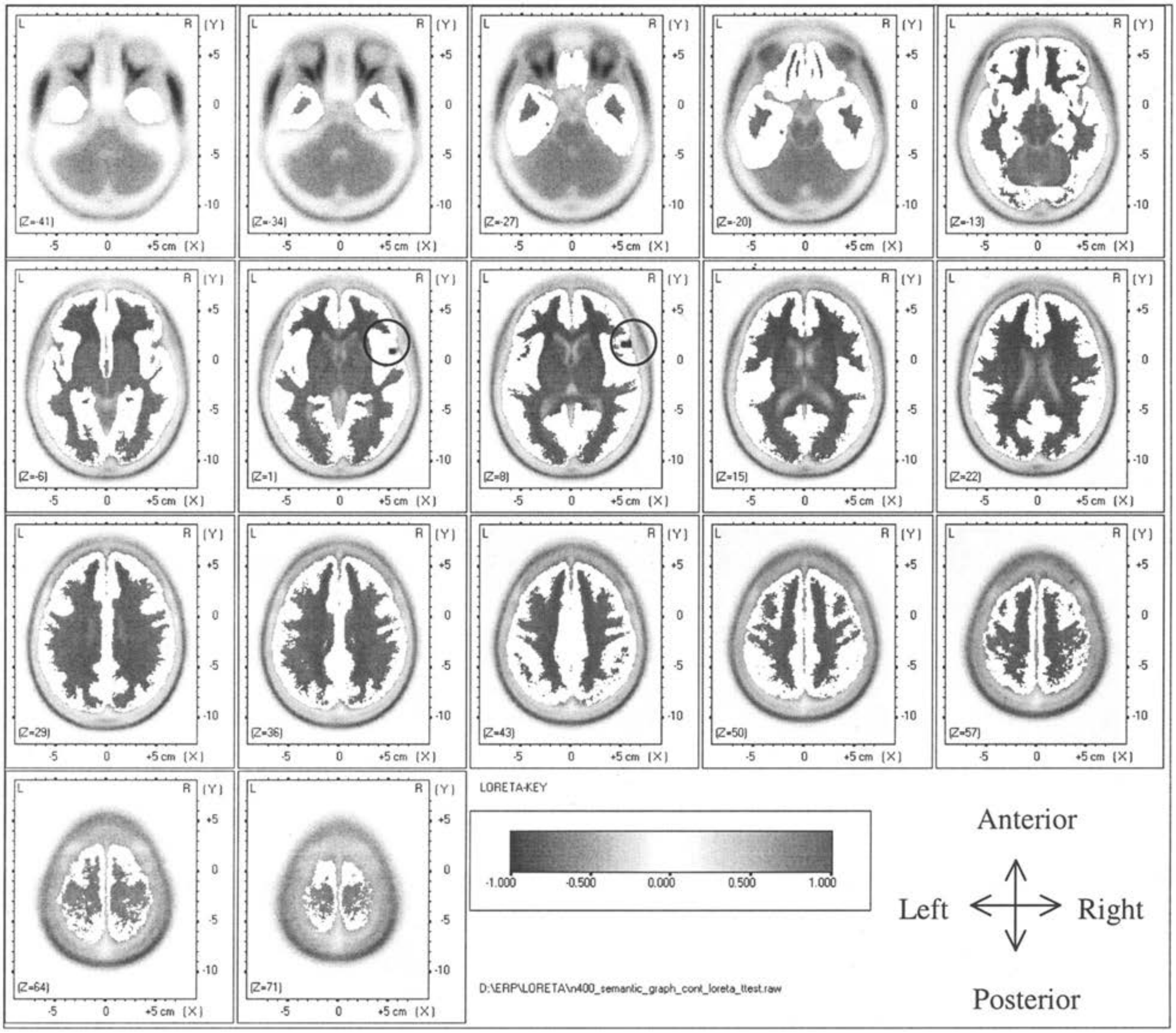

Figure 10. Top view low resolution electromagnetic tomography (LORETA) *p* map of graphical similar versus control pairs in meaning task at N400 ($p < .01$). Gray = noncortex; white = $p > .01$ cortex; black = $p \leq .01$ cortex with higher density for control (circled for easy identification).

present case, ERPs produced evidence for the phonological interference within the first 500 ms of processing. We interpret these interference effects as reflecting phonological activation that occurs automatically during access to the character.

More generally, the ERP evidence shows a temporal unfolding of character information as a function of the reader's task and the properties of a preceding character. It also shows some generalizations for the two tasks. Basically, graphic processes are observed first, then phonological and semantic processes are observed.

Graphic processing is reflected in the P200 component, the earliest component showing experimental effects. Its distribution over the scalp, as shown in Figure 12, is positive at frontal and central electrodes and negative at parietal, temporal, and occipital electrodes. When the second character was preceded by a graphically similar character in the pronunciation task, the P200 component scores were reduced, leading us to characterize this component as reflecting the processing of graphic form. A similar graphic component was found in a study comparing Chinese- and English-word reading of Chinese–English bilinguals (Liu & Perfetti, 2003). Although this component was named N250 in that study, its scalp distribution was the same as the P200 observed in the present study. In that study, the positive shift of the graphic N250 at bilateral frontal electrodes was sensitive to Chinese character frequency and had different peak latencies for Chinese and English materials.

Because P200 shifts have been observed in other research, it is useful to link our interpretation to that of other studies. For example, the P200 has been interpreted as reflecting selective attention (Hackley, Woldorff, & Hillyard, 1990) and visual feature detection processes (Luck & Hillyard, 1994). In a study, like the

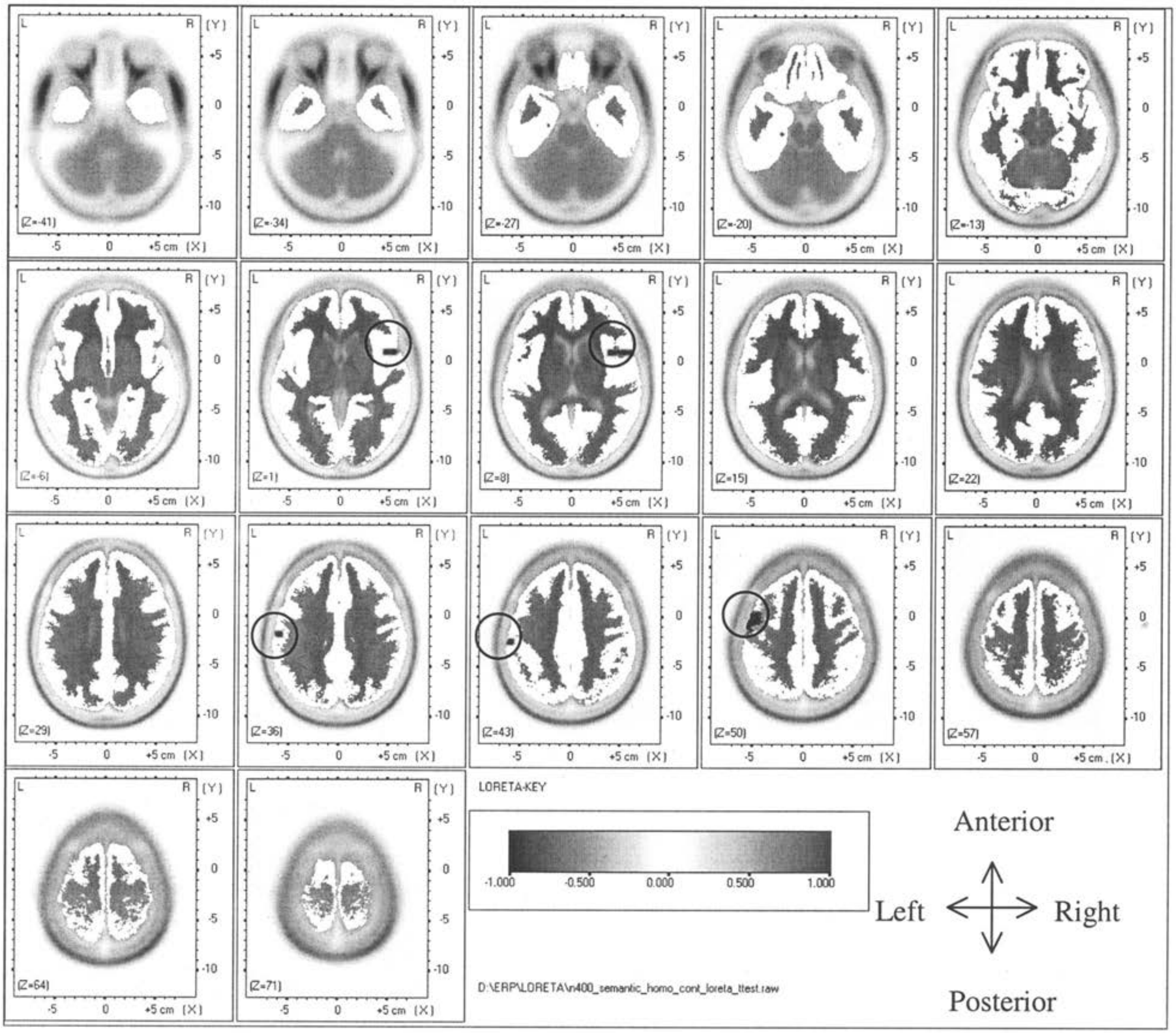

Figure 11. Top view low resolution electromagnetic tomography (LORETA) p map of homophone versus control pairs in meaning task at N400. Gray = noncortex; white = $p > .01$ cortex; black = $p \leq .01$ cortex with higher density for control (circled for easy identification).

present study, that used PCA, Chapman, McCrary, and Chapman (1978) reported a positive component in letter and number comparisons that peaked at 250 ms, which they interpreted as a short-term memory storage component. Thus, our graphic-similarity effect may reflect the coding processes for graphic forms that require attention or temporary memory. More directly related to the present result is the finding by Barnea and Breznitz (1998) that the P200 is associated with both orthographic and phonological processing.

Also consistent with the graphic interpretation is a study using intracranial implanted electrodes (Nobre, Allison, & McCarthy, 1994), which found an N200 specific to orthographic stimuli. Scalp-recorded ERP also found that the N200 is sensitive to the processing of the perceptual features of target stimuli and is larger for orthographic stimuli than nonorthographic stimuli in a visual task (Bentin et al., 1999). These N200s were observed mostly at posterior electrodes. In the present study, a negative 200-ms shift was obtained at posterior electrodes (see Figure 12) and showed a significant effect at right parietal (P4). Thus, the P200 observed here may be related to orthographic N200s reported in other research.

The fact that the present P200 reflects graphic similarity is especially interesting in terms of models that compute orthographic representations for Chinese-word reading. In the present experiment, the major source of visual similarity is a shared radical. Thus, it may not be mere global visual form but rather the activation of orthographic units in the lexicon. This conclusion is in line with a theoretical model first proposed by Perfetti and Tan (1998) and later developed in more detail in Perfetti, Liu, and Tan (2002). The model posits radical units as basic inputs to an

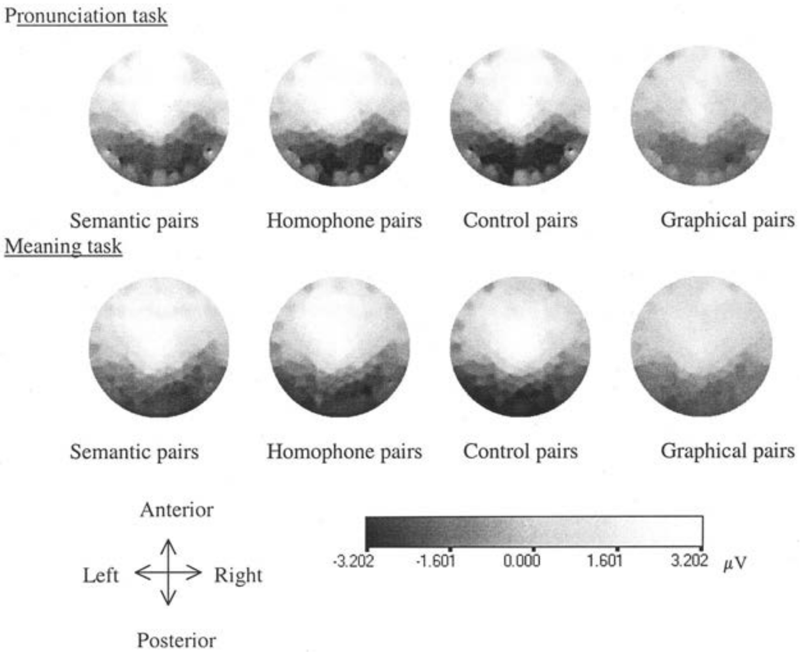

Figure 12. Top view topographic map of P200 amplitude in both tasks at 192 ms.

orthographic-character lexicon and allows the activation of all characters that share the radical. This model would predict the kind of graphic similarity effect observed in the present ERP data, as well as in the behavioral naming data reported by Perfetti and Tan (1998).

The analysis using LORETA maps provides complementary information on the cortical sources of graphic effect. Among areas showing high activation for the P200 component were bilateral occipital and precentral motor areas that were also observed in a LORETA analyses for Chinese reported by Liu and Perfetti (2003). For the pronunciation task, LORETA *t* test results identified the sources of the P200 difference at BA 6, a motor-function area, and BA 8, a large portion of the frontal oculomotor field that is connected with the occipital lobe. BA 6 was also found to be an area activated by Chinese reading in an fMRI study (Tan et al., 2001).

Beyond its role as a marker for graphic processing, a more detailed interpretation of the P200 effect comes from considering the behavioral data in conjunction with the ERP record. The behavioral data are clear in that the effect of this graphic similarity was inhibitory relative to control characters. Whereas the reduction of the P200 reflects that a radical in the second character was present in the first character. Thus it is a graphic-similarity effect that can lead to either facilitation or inhibition, depending on the task and SOA (Perfetti & Tan, 1998). The source of the P200 reduction effect must reside in the radical, demonstrating that the radical functions as an orthographic unit in character identification. The inhibition occurs because the two characters share a radical but differ in pronunciation. This divergence of phonology in the presence of identical orthography is the source of the interference when the participant has to generate a pronunciation, a process required for pronunciation decisions. This interpretation seems to be reinforced by the fact that the graphic effect was more robust for the pronunciation task, which requires a phonological code to be generated for both characters. In addition, this interpretation is consistent with the inhibition effects on naming found by Perfetti and Tan (1998) and with the prediction of the constituency model (Perfetti et al., 2002) that inhibition effects will occur only above identification threshold levels, that is, beyond the point where the orthographic character is recognized. This model also predicts that a brief period of facilitation occurs prior to the threshold being reached; however, this effect would not be observed in the long SOA of this experiment.

We turn now to the phonological effect, reflected in our second component, the complex component containing N400, which is most visible at central and left sites (Figure 13). This component showed also a second graphic-interference effect. For both homophones and graphically similar pairs, related trials were less negative than control trials. In the meaning task, the N400 for graphically similar and homophone pairs was significantly smaller than it was for control pairs. The reaction-time data showed an inhibition tendency for both graphically similar pairs and homophone pairs when compared with controls, although neither of the differences was significant.

Raw LORETA maps showed that there were both frontal and occipital areas involved in the N400 component. The source of the graphic effect at 400 ms was a small area at BA 44 and BA 22 in the right hemisphere, a location close to the 200-ms graphic effect in the pronunciation task, but more inferior. BA 44 is a motor speech area and BA 22 is an auditory-association area. The timing of this graphical effect suggests a later stage of radical activation.

The reduced N400 of homophone pairs in our meaning task was traced by LORETA to areas left-BA 6 and BA 3/4, and right-BA 22 (superior temporal), and BA 44. The right hemisphere areas were the same as those for the graphic effect. However, the left hemisphere results were found for only homophone pairs. This

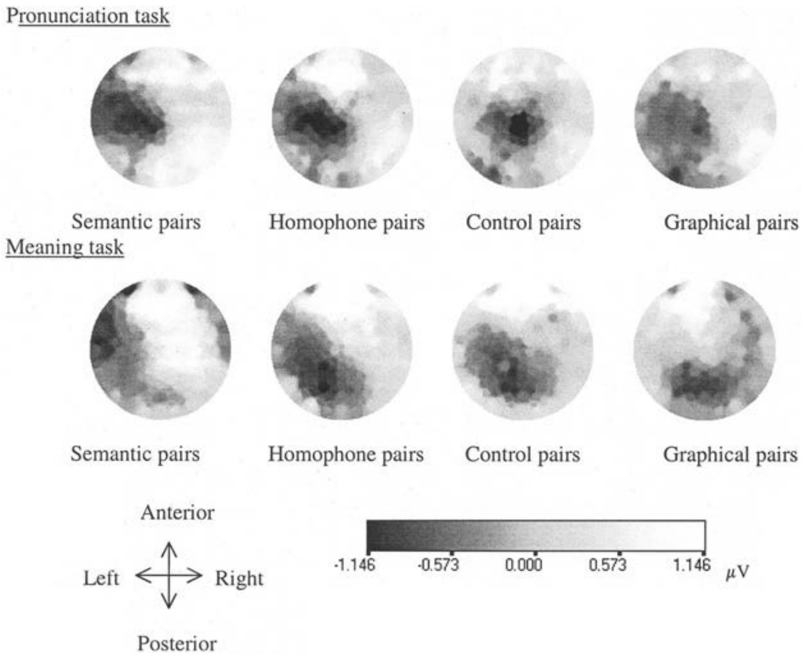

Figure 13. Top view topographic map of N400 amplitude in both tasks at 386 ms.

result might indicate more left-hemisphere language area activation at the character level in the processing of phonology. Especially interesting is that all the brain areas mentioned above were also found to be active in an fMRI study using this task (Tan et al., 2001). Thus, the areas for Chinese character constituent processing were successfully separated by ERP with localization as well as by fMRI.

In a study using magnetoencephalography (MEG), the source of the N400 for Finnish reading was identified as left-superior temporal (Helenius, Salmelin, Service, & Connolly, 1999). However, another MEG study using English materials (Halgren et al., 2002) reported more widely distributed sources, including left temporal, frontal, and a weakly right-anterior temporal activation. Given the convergence of the sourcing of the present N400 with the fMRI results of Tan et al. (2001), a possibility is that the N400 reflects processes that vary in their sources as a function of the writing system, with English, a deep orthography, and Chinese being more similar than Finnish.

Our finding of an N400 reduction for homophones suggests that this ERP component can pick up the automatic activation of phonology in a meaning task. Other ERP research has produced some inconsistency on this point across various tasks that involve meaning (Connolly & Phillips, 1994; Connolly, Phillips, & Forbes, 1995; Niznikiewicz & Squires, 1996; Ziegler et al., 1999). It is difficult to be certain about the source of these different results—the studies used different tasks and varied in the number of electrodes. Our use of high-density recordings, compared with Ziegler et al.'s (1999) study of category decisions, may increase the likelihood of detecting homophone effects. Another interpretation is theoretical. Our results speak to automatic phonology, not specifically to mediation. The comparison of the two words on meaning is very direct in our task, and the N400 homophone effect reflects congruence—the homophones are similar (in pronunciation), just as meaning-related pairs are similar in meaning. By contrast, the categorization task is less direct in exposing phonology. The category name (*food*) and the foil *meet* show their similarity only through the relationship that each has to the unpresented word, *meat.* Furthermore, *meat* is only one candidate among many types of *food.* So, the foil is actually unexpected given the category (*meat* not *meet* is congruent). As a result, the phonology effect of *meet* may be delayed to the postlexical stage (800–1600 ms), as reported by Ziegler et al. On the other hand, it is possible to find an even earlier phonological effect when the final word is highly predictable, as Connolly and his colleagues (Connolly & Phillips, 1994; Connolly et al., 1995) have shown in a sentence context. In their studies, a word having the same initial phoneme as a semantically congruent final word produced a reduced N270 (PMN; e.g., "The gambler had a streak of bad *luggage*"; Connolly & Phillips, 1994, p. 256).

In general terms, the ERP evidence on Chinese provides a picture of temporal unfolding that should be quite general across writing systems, with initial potential shifts reflecting first graphic form and, somewhat later, phonological and semantic information. In the constituency model (Perfetti et al., 2002), the activation of graphic, phonological, and semantic constituents occurs in this same order, with partly overlapping processes. It is interesting to compare Chinese with English in these temporal effects. In an English experiment using the same tasks and general design, Hart, Perfetti, and Liu (2001) found a 200-ms latency component in which homophones were distinguished from graphically similar words (nonhomophones that shared spelling). In the present study, Chinese showed a homophone effect only at 400 ms. Thus, according to ERP evidence, both English and Chinese processing show phonological processing within 400 ms, but English, as an alphabetic system, allows an earlier coupling of graphic with phonological form.

By combining behavioral, temporal, and spatial ERP (LORETA), we find a rich and convergent picture of constituent effects in Chinese-character judgments. When readers judged pronunciation, shared radicals produced a graphic-similarity effect observable as a 200-ms latency component in ERP and an inhibition effect in decision times. This graphic-similarity effect had a later time point (400 ms) and different generators (more inferior) when the reader judged meaning. This contrast implies that the reader's task has an immediate effect on the use of the radical, with a more rapid use of sublexical (radical) information when phonology is required.

Activation of phonology is also seen when readers make meaning judgments, as observed by Perfetti and Zhang (1995). Homophone interference effects were seen in the ERP evidence (reduced N400) and (nonsignificantly) in the decision times. A reverse effect—meaning interference in pronunciation—was not observed. Thus, in addition to illuminating the task-dependent nature of the temporal unfolding of word-constituent information, the results suggest, in agreement with previous results, that automatic activation of phonological information occurs even when the task is not focused on pronunciation.

References

Anderer, P., Pascual Marqui, R. D., Semlitsch, H. V., & Saletu, B. (1998). Differential effects of normal aging on sources of standard N1, target N1 and P300 auditory event-related brain potentials revealed by low resolution electromagnetic tomography (LORETA). *Electroencephalography and Clinical Neurophysiology: Evoked Potentials, 108,* 160–174.

Barnea, A., & Breznitz, Z. (1998). Phonological and orthographic processing of Hebrew words: Electrophysiological aspects. *Journal of Genetic Psychology, 159,* 492–504.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11,* 235–260.

Berent, I., & Perfetti, C. A. (1995). A rose is a REEZ: The two-cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146–184.

Chapman, R. M., & McCrary, J. W. (1995). EP component identification and measurement by principal components analysis. *Brain and Cognition, 27,* 288–310.

Chapman, R. M., McCrary, J. W., & Chapman, J. A. (1978). Short-term memory: The storage component of human brain responses predicts recall. *Science, 202,* 1211–1214.

Chua, F. K. (1999). Phonological recoding in Chinese logograph recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 876–891.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6,* 256–266.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 94,* 276–287.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Journal of Clinical Neurophysiology, 112,* 536–544.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 680–698.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1991). Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism, 11,* 690–699.

Hackley, S. A., Woldorff, M., & Hillyard, S. A. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. *Psychophysiology, 27,* 195–208.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17,* 1101–1116.

Hart, L., Perfetti, C. A., & Liu, Y. (2001). *Event related potentials comparisons of English and Chinese word processes I: English results.* Unpublished manuscript, University of Pittsburgh, Pennsylvania.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1999). Semantic cortical activation in dyslexic readers. *Journal of Cognitive Neuroscience, 11,* 535–550.

Hung, D. L., Tzeng, O. J. L., & Tzeng, A. K. Y. (1992). Automatic activation of linguistic information in Chinese character recognition. In R. Frost (Ed.), *Orthography, phonology, morphology, and meaning* (pp. 119–130). Amsterdam: North-Holland.

Kramer, A. F., & Donchin, E. (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 76–86.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multi-channels potential fields. *Electroencephalography and Clinical Neurophysiology, 48,* 609–621.

Liu, Y., & Perfetti, C. A. (2003). The time course of brain activity in reading English and Chinese: An ERP study of Chinese bilinguals. *Human Brain Mapping, 18,* 167–175.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31,* 291–308.

Niznikiewicz, M., & Squires, N. K. (1996). Phonological processing and the role of strategy in silent reading: Behavioral and electrophysiological evidence. *Brain and Language, 52,* 342–364.

Nobre, A. C., Allison, T., & McCarthy, G. (1994, November). Word recognition in the human inferior temporal lobe. *Nature, 372,* 260–263.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology, 18,* 49–65.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (2000). LORETA-KEY software. [Computer Software]. Retrieved from http://www.unizh.ch/keyinst/loreta.htm.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2002). How the mind meets the brain in reading: A Comparative writing systems approach. In H. S. R. Kao, C. K. Leong, & D. G. Gao (Eds.), *Cognitive neuroscience studies of the Chinese language* (pp. 35–59). Hong Kong: Hong Kong University Press.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 101–118.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 24–33.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37,* 127–152.

Pizzagalli, D., Lehmann, D., Koenig, T., Regard, M., & Pascual Marqui, R. D. (2000). Face-elicited ERPs and affective attitude: Brain electric microstate and tomography analyses. *Clinical Neurophysiology, 111,* 521–531.

Psychological Software Tools, Inc. (2000). Eprime (Version 1.0b) [Computer software]. Pittsburgh, PA: Author.

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language, 23,* 225–240.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Strik, W. K., Fallgatter, A. J., Brandeis, D., & Pascual Marqui, R. D. (1998). Three-dimensional tomography of event-related potentials during response inhibition: Evidence for phasic frontal lobe activation. *Electroencephalography and Clinical Neurophysiology: Evoked Potentials, 108,* 406–413.

Talairach, J., & Tournoux, P. (1988). *Co-planar sereotaxc atlas of the human brain* (M. Rayport, Trans.). New York: Thieme Medical Publishers.

Tan, L. H., Liu, H.-L., Perfetti, C. A., Spinks, J. A., Fox, P. T., & Gao, J.-H. (2001). The neural system underlying Chinese logograph reading. *NeuroImage, 13,* 836–846.

Valdes-Sosa, M., Gonzalez, A., Liu, X., & Zhang, X. (1993). Brain potentials in a phonological matching task using Chinese characters. *Neuropsychologia, 31,* 853–864.

van Boxtel, G. J. M. (1998). Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments and Computers, 30,* 87–102.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition, 15,* 181–198.

Wastell, D. G. (1981). On the correlated nature of evoked brain activity: Biophysical and statistical considerations. *Biological Psychology, 13,* 51–69.

Xu, Y., Pollatsek, A., & Potter, M. C. (1999). The activation of phonology during silent Chinese word reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 838–857.

Zhang, S., & Perfetti, C. A. (1993). The tongue-twister effect in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1082–1093.

Zhang, S., Perfetti, C. A., & Yang, H. (1999). Whole word, frequency-general phonology in semantic processing of Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 858–875.

Ziegler, J. C., Benraiess, A., & Besson, M. (1999). From print to meaning: An electrophysiological investigation of the role of phonology in accessing word meaning. *Psychophysiology, 36,* 775–785.

Appendix

Experiment Materials

	Core	Semantic	Graphical	Homophone	Control		Core	Semantic	Graphical	Homophone	Control
1	惊	扰	凉	晶	寂	31	察	观	祭	茶	输
2	功	斥	玫	公	狗	32	视	听	砚	事	然
3	旺	衰	蛀	忘	途	33	偿	补	傥	肠	稿
4	扬	举	肠	洋	律	34	晨	晚	震	沉	呼
5	零	整	雪	凌	蚀	35	呈	递	号	成	发
6	仗	战	伏	帐	析	36	词	字	伺	磁	滑
7	订	改	计	锭	砖	37	戴	穿	截	带	科
8	裂	破	袭	猎	鸦	38	登	爬	豆	灯	乱
9	拱	环	哄	贡	脾	39	第	序	策	帝	汽
10	焦	灼	集	郊	咽	40	断	分	斯	段	朋
11	竿	棍	第	甘	趁	41	峰	顶	峥	封	哭
12	错	杂	银	挫	绅	42	割	切	辖	歌	简
13	挥	舞	浑	灰	险	43	顾	管	顿	故	竟
14	恋	怀	峦	炼	斧	44	画	描	函	话	侯
15	筹	谋	等	愁	控	45	汇	集	江	绘	居
16	阅	看	问	跃	烦	46	寄	邮	奇	记	利
17	挽	卷	换	碗	刑	47	佳	丽	住	家	用
18	池	塘	他	迟	胆	48	据	靠	握	具	路
19	稍	轻	税	烧	幸	49	酷	极	醋	库	漠
20	逃	奔	过	淘	雇	50	矿	场	扩	况	派
21	镇	屯	钥	振	堆	51	窥	望	窍	亏	驾
22	挨	靠	抬	哀	耐	52	类	族	娄	泪	掉
23	伴	友	胖	办	拿	53	帘	遮	窜	联	赶
24	爸	妈	斧	霸	扇	54	临	将	监	林	考
25	搬	移	投	班	讨	55	蜜	甜	虫	觅	奈
26	豹	虎	勺	报	油	56	男	雄	另	南	容
27	辈	世	非	备	院	57	盼	等	扮	判	守
28	币	款	巾	必	远	58	疲	累	疾	脾	罗
29	博	富	傅	勃	砸	59	凭	依	赁	平	难
30	财	钱	贝	裁	柔	60	述	讲	达	数	群

Received September 19, 2001
Revision received April 22, 2003
Accepted April 30, 2003 ■