Brain Responses to Segmentally and Tonally Induced Semantic Violations in Cantonese

Annett Schirmer[1], Siu-Lam Tang[2], Trevor B. Penney[2], Thomas C. Gunter[1], and Hsuan-Chih Chen[2]

Abstract

■ The present event-related potential (ERP) study examined the role of tone and segmental information in Cantonese word processing. To this end, participants listened to sentences that were either semantically correct or contained a semantically incorrect word. Semantically incorrect words differed from the most expected sentence completion at the tone level, at the segmental level, or at both levels. All semantically incorrect words elicited an increased frontal negativity that was maximal 300 msec following word onset and an increased centro-parietal positivity that was maximal 650 msec following word onset. There were differences between completely incongruous words and the other two violation conditions with respect to the latency and amplitude of the ERP effects. These differences may be due to differences in the onset of acoustic deviation of the presented from the expected word and different mechanisms involved in the processing of complete as compared to partial acoustic deviations. Most importantly, however, tonally and segmentally induced semantic violations were comparable. This suggests that listeners access tone and segmental information at a similar point in time and that both types of information play comparable roles during word processing in Cantonese. ■

INTRODUCTION

Speech comprises segmental and suprasegmental information. The term segmental information refers to the individual speech segments or phonemes that can be combined to form words. For example, the speech segments /f//ɔ://r//b//e//r/ produced in this particular order form the word ‘‘forebear’’ (example taken from Cutler, 1986). The term suprasegmental information refers to acoustic cues that can extend over several speech segments and may thus characterize larger entities (e.g., syllables, words). Pitch, duration, and amplitude represent such acoustic cues. In the above example, these cues differentiate between the two possible meanings of ‘‘forebear.’’ Moreover, by modulating pitch, duration, and amplitude, speakers stress the first syllable when referring to the meaning ‘‘ancestor’’ and the second syllable when referring to the meaning ‘‘endure.’’

For a long time, researchers exclusively considered segmental information when modeling word recognition (Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson, Moss, & van Halen, 1996; Norris, 1994; Zwitserlood, 1989; Marslen-Wilson, 1987; McClelland & Elman, 1986). In this context, it has been proposed that upon hearing a string of speech segments, listeners activate word entries in the mental lexicon that partially or completely map the acoustic input. These activated entries are thought to compete until a word is unambiguously identified. In contrast to segmental information, suprasegmental information was considered insignificant for lexical activation and competition. Rather, its role has been restricted to speech segmentation, that is, the identification of individual words in continuous speech (McQueen, Norris, & Cutler, 1994).

These assumptions have been formulated on the basis of research conducted in Indo-European languages such as English and Dutch. These languages are characterized by a high degree of covariation between segmental and suprasegmental information. As a consequence, words such as ‘‘forebear,’’ where suprasegmental information alone is important for meaning identification, are rare. In general, segmental information is sufficient for listeners to unambiguously identify the meaning of a word. Thus, it is not too surprising that studies investigating word processing find that segmental information constrains the activation of entries in the mental lexicon more efficiently than suprasegmental information (Friedrich, Kotz, Friederici, & Alter, 2004; Cutler & Van Donselaar, 2001; Cutler, 1986). However, the degree of covariation between segmental and suprasegmental information differs between languages. Therefore, it is possible that the role of suprasegmental information for word processing differs between languages. Work by Soto-Faraco, Sebastián-Gallés, and Cutler (2001) revealed evidence for this assumption. The authors investigated word recognition in Castilian Spanish, which

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [2]Chinese University of Hong Kong

comprises a larger number of words that differ only in suprasegmental information (e.g., ''saBAna'' vs. ''SAbana''; upper case denotes stressed and lower case denotes unstressed syllables; translation: ''savannah'' vs. ''sheet''). The authors presented these words in a cross-modal fragment priming experiment. Participants saw visual targets that differed from the auditory prime in either suprasegmental or segmental information. Reaction times revealed comparable inhibition for suprasegmentally and segmentally incongruous targets indicating that word processing may be similarly constrained by suprasegmental and segmental information.

An even lower degree of covariation between segmental and suprasegmental information can be found in tone languages such as Chinese. Cantonese, a Chinese dialect spoken in the southern Chinese provinces of Guangdong and Guangxi, as well as in Hong Kong and Macau, distinguishes 6 lexical tones. For example, whether the word ''yau'' means ''rest,'' ''paint,'' ''thin,'' ''oil,'' ''friend,'' or ''again'' depends on whether it is spoken with a high level, high rising, mid level, low falling, low rising, or low level tone, respectively. Given the importance of tone, one might assume that, as in Castilian Spanish, suprasegmental (i.e., tone) and segmental information play comparable roles during word processing.

Cutler and Chen (1997) conducted a series of behavioral experiments to test this assumption. Participants heard real words and pseudowords that differed from real words in either tone or segmental information. Surprisingly, listeners accepted pseudowords more readily as words when tone, as compared to segmental information, differentiated the pseudowords from real words. In addition, a same–different task in which participants heard word pairs that differed in either tone or segmental information indicated that Cantonese listeners relied more strongly on segmental than suprasegmental information. Cantonese listeners were faster and more accurate in deciding that two words were different when the difference was segmental as compared to suprasegmental.

Based on these findings, the authors proposed time-course differences in the processing of segmental and suprasegmental information. Tone processing was thought to require more time and to be compromised relative to segmental processing in a speeded judgment task. Furthermore, although the authors pointed out that ''tone does participate fully in the lexical access code,'' they suggested that the usability of tone is limited relative to segmental information. This conclusion is consistent with findings from Ye and Connine (1999) who reported faster responses to vowel mismatches as compared to tone mismatches in a syllable monitoring task. However, Ye and Connine additionally suggested that the role of tone during word processing depends on the presence of contextual information. They demonstrated that strong semantic context, as is available in idioms, helps to preactivate tone information and thus disrupts the advantage of segmental information found for words presented in neutral sentence context or in isolation.

Although behavioral measures are sensitive to aspects of word processing, they also reflect processes that are associated with the response. Thus, it is possible that the results of Ye and Connine (1999) and Cutler and Chen (1997) reflect different response strategies associated with the detection of tone and segmental violations. Moreover, differences may occur because tone information is relative, whereas segmental information is absolute. Tones differ in height and contour depending on factors such as speaker, emotional state, and acoustic sentence context (e.g., preceding tones). This suggests that listeners have to be more ''flexible'' in interpreting tones as compared to segmental information. Thus, the response to target words in lexical decision or monitoring tasks, may be influenced by this greater flexibility or willingness to tolerate variability in tone. Moreover, listeners might more readily accept tone as compared to segmental violations even though they detect these violations just as reliably.

To better understand the significance of tonal and segmental information, it is therefore helpful to examine how stimulus processing unfolds in time. The event-related potential (ERP) is a tool that allows measurement of speech processing on-line. Several ERP components—characterized by polarity, latency, and scalp distribution—have been implicated in specific aspects of speech processing. For example, processing the semantic meaning of a word has been associated with the N400, a negativity that peaks approximately 400 msec following word onset over centro-parietal scalp regions. Compared to incongruous words, words that are congruous with the preceding sentence context elicit smaller N400 amplitudes (Hahne & Friederici, 2002; Hagoort & Brown, 2000; Van Petten, Coulson, Rubin, Plante, & Parks, 1999; Friederici, Pfeifer, & Hahne, 1993; Kutas, Van Petten, & Besson, 1988; Kutas & Hillyard, 1984). The N400 effect is independent of phonological similarity: Incongruous words that share phonemes with the expected word (e.g., ''The bill was seventeen *scholars*.'') elicit comparable N400 amplitudes as incongruous words that differ completely from the expected word (e.g., ''The bill was seventeen *teachers*''; Van den Brink, Brown, & Hagoort, 2001; Van Petten et al., 1999). However, the onset of the N400 effect is modulated by when in time the incongruous word deviates from the expected congruous word. Later acoustic deviation (e.g., ''The bill was seventeen *dolphins*.'') elicits a later onset of the N400 effect (Van Petten et al., 1999). Note that some authors propose that the latency of acoustic deviation modulates a component preceding the N400 rather than the N400 itself (Van den Brink et al., 2001; Connolly & Phillips, 1994). This proposal will be presented in more detail in the Discussion.

Compared to segmental information, ERP correlates for the processing of suprasegmental information have received less attention in the literature. Bocker, Bastiaansen, Vroomen, Brunia, and de Gelder (1999) investigated how listeners process words starting with either a stressed or an unstressed syllable. They found that initially unstressed words elicited a larger negativity 325 msec following word onset as compared to initially stressed words. They proposed that the so-called N325 reflects the processing of a word's stress pattern. In an attempt to replicate these findings, Friedrich, Alter, and Kotz (2001) determined that the N325 is modulated by vowel quality rather than suprasegmental information. In Dutch and German, unstressed syllables frequently contain reduced vowels, whereas stressed syllables contain full vowels. When controlling for vowel quality, Friedrich and colleagues failed to observe an N325 effect for word initial stress. Instead, they found larger P200 amplitudes for initially unstressed as compared to stressed words indicating that suprasegmental information is accessed very early during word processing.

Given the importance of suprasegmental information in tone languages such as Cantonese, it is possible that in tone languages suprasegmental information is accessed equally early. However, unlike Indo-European languages such as English and Dutch, the contribution of suprasegmental information to word processing may go beyond speech segmentation and might be comparable to the contribution of segmental information. To investigate this question, we presented sentences auditorily that where either semantically correct or contained words that did not fit into the sentence context. The latter type of stimuli was constructed by modifying the segmental structure (e.g., b***ou***6; "step"), the tone (e.g., beng***2***; "bisquit"), or both segmental structure and tone (e.g., ***gwai3***; "season") so that a previously semantically congruous word (e.g., beng6; "illness") became semantically incongruous. Although these are all semantic and not purely acoustic violations, for the purpose of this article we will refer to them as segmental, tonal, and complete violations, respectively. Note that we tried to match the onset of the segmental violation with the onset of the tonal violation: Both involved the word's rime, whereas the complete violation started from word onset. Participants were asked to listen to each sentence and to indicate by pressing a button whether or not the sentence was semantically congruous. Participants were uninformed about the different types of semantic violations to ensure that any attention they directed to tone would be natural rather than task-enforced. Furthermore, we reduced the influence of the response on the ERP by placing the critical word in the middle of the sentence and asking participants to give their judgment when seeing a question mark that appeared 300 msec following sentence offset.

Based on evidence from Indo-European languages, we hypothesized that critical words in the segmental violation condition would elicit a larger N400 than critical words in the congruous condition (Van Petten et al., 1999; Kutas, Van Petten, et al., 1988; Kutas & Hillyard, 1980, 1984). A similar N400 effect was predicted for the complete violation. However, as this condition differed from the congruous condition in the onset consonant, whereas the segmental violation differed only in the rime, the N400 effect for the complete violation was expected to start somewhat earlier (Van Petten et al., 1999). Our predictions for the tone violations were based on models of tonal processing described in the literature. First, it has been proposed that tone participates in the lexical access code (Cutler & Chen, 1997). Therefore, we predicted that tone violations would elicit an N400 effect. Second, behavioral work suggests that tonal processing lags behind segmental processing (Cutler & Chen, 1997). If true, participants should detect tonal violations later than segmental violations. Moreover, as an earlier detection of acoustic differences between the presented and the expected word results in an earlier N400 effect (Van Petten et al., 1999), we would expect the N400 effect to start earlier for segmental as compared to tone violations. However, as there is ERP evidence that suprasegmental information is available 200 msec following word onset (Friedrich, Alter, et al., 2001), it is possible that tone contributes to word processing as early as segmental information. If true, the latency of the N400 effect should be comparable for tone and segmental violations.

RESULTS

Behavioral Results

The behavioral results for each condition are illustrated in Figure 1. ANOVAs treating Word (congruous, CN; tone violation, TV; segmental violation, SV; complete violation, CV) as a repeated-measures factor were conducted for reaction times and accuracy separately. However, interpretation of the results of this analysis is constrained by the use of a delayed response task. Thus, reaction times and perhaps also accuracy may not have been sensitive to on-line processing differences between the conditions. Moreover, as the detection of a semantic incongruity had to be held in memory until the response probe appeared on the screen, both reaction times and accuracy may reflect working memory processes in addition to other processes associated with the response (see Introduction).

Analysis of reaction times revealed a significant effect of Word [$F(3,69) = 11.16$, $p < .0001$]. Single comparisons between each possible pair were conducted in order to investigate the Word main effect. The congruous condition elicited slower responses than the tone violation [$F(1,23) = 9.55$, $p < .0052$], the segmental violation [$F(1,23) = 13.21$, $p < .0014$], and the complete violation [$F(1,23) = 32.6$, $p < .0001$]. There were no

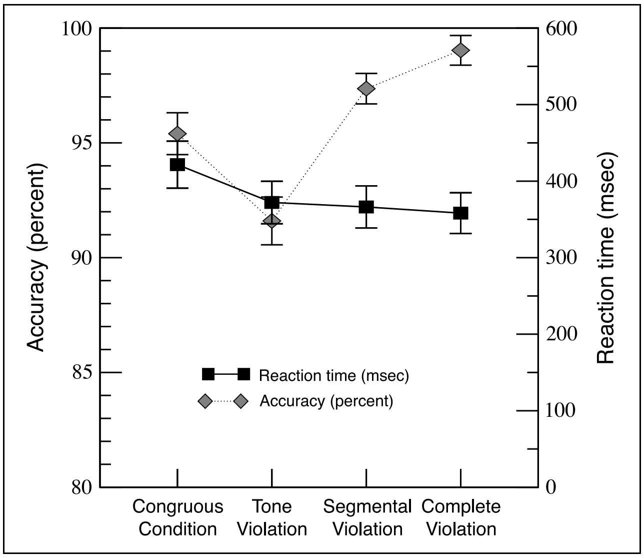

Figure 1. Reaction times and accuracy for the congruous condition, the tone violation, the segmental violation, and the complete violation. Error bars indicate the *SEM*.

significant differences between the violation conditions (all ps > .1).

There was a significant main effect of Word [$F(3,69) = 32.60$, $p < .0001$] on accuracy. Participants made most errors in the tone violation condition. A Bonferroni adjusted p value for post hoc comparisons ($p = .0083$) revealed a marginal difference between this condition and the semantically congruous condition [$F(1,23) = 7.23$, $p = .0131$], and a significant difference between this condition and the segmental [$F(1,23) = 29.31$, $p < .0001$] and complete violation conditions [$F(1,23) = 40.81$, $p < .0001$]. There was a tendency for semantically congruous sentences to elicit more errors than segmental violations [$F(1,23) = 4.96$, $p = .0359$]. Semantically congruous sentences elicited significantly more errors than complete violations [$F(1,23) = 10.48$, $p < .0036$]. Accuracy tended to be lower for the segmental as compared to the complete violation condition [$F(1,23) = 4.6$, $p = .0428$].

ERP Results

Relative to the semantically congruous condition, all violation conditions elicited an N400-like negativity between 200 and 450 msec and a late positive component (LPC) between 450 and 1000 msec (see Figure 2). Fifty-millisecond time window analyses with Word, Hemisphere (left, right), and Site (anterior, central posterior) as repeated-measures factors were conducted to explore the time course of differences between the experimental conditions. Mean amplitudes of significant consecutive time windows were then averaged across the significant time interval and entered into a second ANOVA. Exploratory analyses revealed a significant Word main effect between 200 and 350 msec and between 500 and 1000 msec. The Word main effect interacted with the factor Site between 300 and 1000 msec and with the factor Hemisphere between 600 and 700 msec and between 750 and 1000 msec. An overview of these results is provided in Table 1.

Based on the results of the exploratory analysis, the N400-like negativity was investigated by using a 200 to 300 msec time window appropriate for the Word main effect and a 300 to 450 msec time window appropriate for the Word × Site interaction. As indicated by the exploratory analysis, the Word main effect was significant between 200 and 300 msec following word onset [$F(3,69) = 5.81$, $p < .01$]. Post hoc comparisons revealed a significant difference between the congruous condition and the complete violation condition [$F(1,23) = 19.44$, $p < .001$]. The comparison between the two partial violations and the control condition approached significance [CN vs. TV: $F(1,23) = 6.32$, $p = .0194$; CN vs. RV: $F(1,23) = 4.61$, $p = .0425$] and the comparison between the three violation conditions was nonsignificant [CV vs. TV: $p > .1$; CV vs. RV: $F(1,23) = 4.66$, $p = .0416$; TV vs. RV: $p > .1$]. Visual examination of the ERP effects elicited in this early time window suggests that the complete violation differed from the congruous condition across electrode sites, whereas the effect of the partial violations was restricted to anterior sites. Thus, scalp distribution may have contributed to the finding that only the complete violation differed significantly from the congruous condition between 200 and 300 msec following word onset. However, as the Word × Site interaction failed to reach significance [$F(6,138) = 2.41$, $p = .08$], it seems that this contribution was less powerful than differences between complete and partial violations in the onset of the N400-like negativity.

Between 300 and 450 msec following word onset, Word interacted with Site [$F(3,69) = 9.24$, $p < .0001$]. This interaction can be explained by the frontal distribution of the Word effect [anterior: $F(3,69) = 6.07$, $p < .01$; central: $p > .1$; posterior: $p > .1$]. Post hoc comparisons of this effect indicated that all violation conditions differed significantly from the congruous condition [CN vs. CV: $F(1,23) = 15.50$, $p < .001$; CN vs. TV: $F(1,23) = 10.32$, $p < .0083$; CN vs. RV: $F(1,23) = 9.28$, $p < .0083$]. Comparisons between the three violation conditions were nonsignificant (all ps > .1).

The late positivity was examined within the 450 to 1000 msec time range with respect to the Word × Site interaction and within the 600 to 1000 msec time range with respect to the Word × Hemisphere interaction. As indicated by the exploratory analysis, the Word × Site interaction was significant between 450 and 1000 msec [$F(3,69) = 37.42$, $p < .0001$]. Post hoc comparisons revealed a significant Word effect at central [$F(3,69) = 22.43$, $p < .0001$] and posterior sites [$F(3,69) = 34.66$,

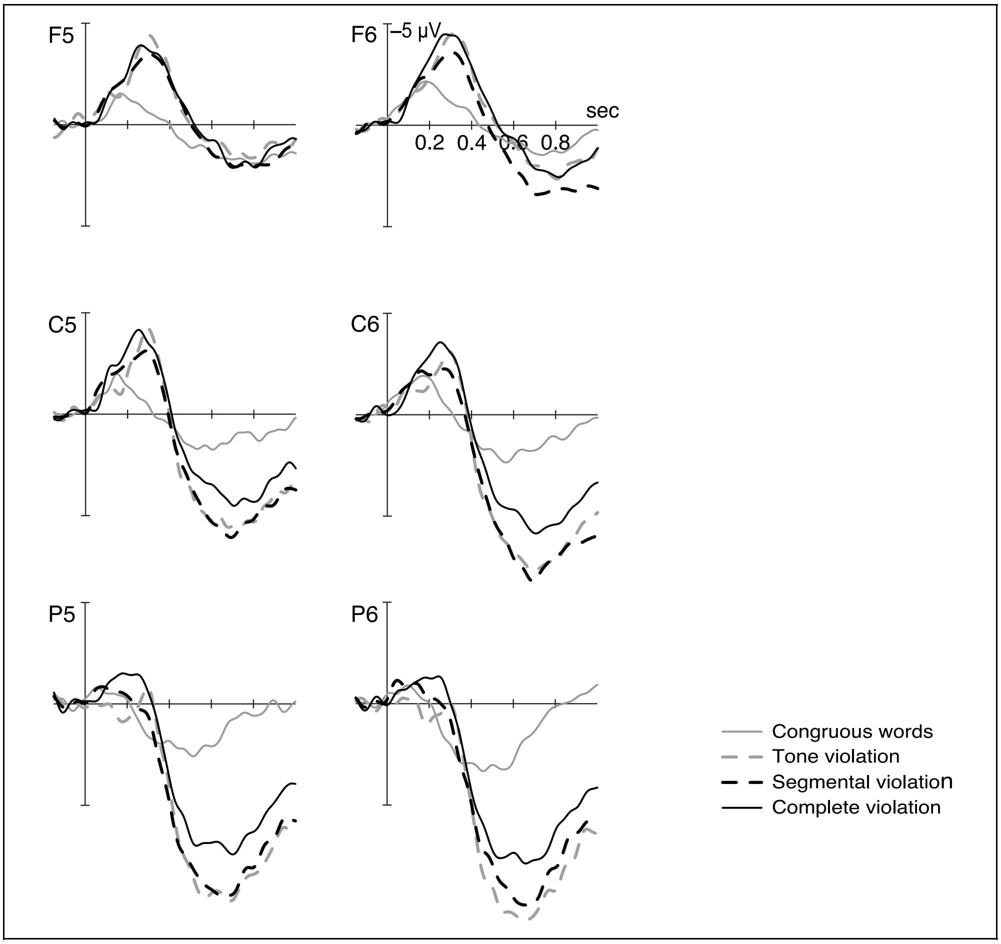

Figure 2. ERP results. Semantic violations elicited an increased N400-like negativity and an increased late positive component. There was no difference between tonally and segmentally induced semantic violations.

$p < .0001$], whereas the Word effect at anterior sites was nonsignificant ($p > .1$) (Figure 3). At both central and posterior sites, violations elicited a larger positivity than the control condition [central, CN vs. CV: $F(1,23) = 15.72$, $p < .001$; CN vs. TV: $F(1,23) = 42.70$, $p < .0001$; CN vs. RV: $F(1,23) = 40.51$, $p < .0001$; posterior, CN vs. CV: $F(1,23) = 40.52$, $p < .0001$; CN vs. TV: $F(1,23) = 64.77$, $p < .0001$; CN vs. RV: $F(1,23) = 54.85$, $p < .0001$]. Compared to the complete violation, partial violations elicited a significantly larger positivity at central cites [TV vs. CV: $F(1,23) = 8.84$, $p < .0083$; RV vs. CV: $F(1,23) = 8.57$, $p < .0083$] and a tendentially larger positivity at posterior sites [TV vs. CV: $F(1,23) = 8.22$, $p = .0087$; RV vs. CV: $F(1,23) = 4.93$, $p = .0365$]. The tone and the segmental violation were comparable both at central and posterior sites (both ps $> .1$). Between 600 and 1000 msec, the Word × Hemisphere interaction [$F(3,69) = 5.88$, $p < .01$] indicated that the Word effect was more pronounced over right [$F(3,69) = 24.45$, $p < .0001$] than left hemisphere regions [$F(3,69) = 14.95$, $p < .0001$]. Nevertheless, the control condition differed significantly from the violation conditions over both cortical hemispheres [right, CN vs. CV: $F(3,69) = 20.03$, $p < .001$; CN vs. TV: $F(3,69) = 59.13$, $p < .0001$; CN vs. RV: $F(3,69) = 46.37$, $p < .0001$; left, CN vs. CV: $F(3,69) = 11.98$, $p < .0083$; CN vs. TV: $F(3,69) = 24.95$, $p < .0001$; CN vs. RV: $F(3,69) = 31.76$, $p < .0001$]. Furthermore, the two partial violations were comparable (both ps $> .1$) and the comparison between the partial violation and the complete violation approached significance [right, TV vs. CV: $F(3,69) = 6.93$, $p = .0149$; RV vs. CV: $F(3,69) = 5.94$, $p = .0229$; left, TV vs. CV: $F(3,69) = 3.28$, $p = .0834$; RV vs. CV: $F(3,69) = 4.67$, $p = .0414$].

To get a better sense of the reliability of the observed ERP effects, we examined the data at the level of individual participants. For the N400-like negativity, we computed the mean amplitudes between 200 and 450 msec following word onset as recorded at anterior sites. Subtracting the obtained values for the congruous condition from the values for the complete violation revealed a negative amplitude in 21 out of 24 participants. Thus, 21 out of 24 participants showed a larger negativity for the complete violation as compared to the congruous condition. This number was somewhat smaller for the partial violations. Eighteen participants showed a larger negativity for the tone and 16 for the segmental violation as compared to the congruous condition. A comparison between tone and segmental violations revealed comparable mean amplitudes in five participants. The amplitude difference for these

Table 1. *F* Values of the Word × Site × Hemisphere ANOVA

Time Window (msec)	*Word (df = 3,69)*	*Word*Site (df = 6,138)*	*Word*LR (df = 3,69)*	*Word*Site*LR (df = 6,138)*
0–50	0.82	0.4	1.58	1.51
50–100	0.57	2.18*	1.57#	1.85
100–150	0.92	0.90	1.00	1.98
150–200	2.69#	1.53	0.48	1.21
200–250	4.30**	2.33*	1.53	1.00
250–300	6.91***	1.89#	1.89	0.79
300–350	6.12***	4.08***	0.76	0.70
350–400	1.10	8.53****	0.42	1.66
400–450	1.93	13.08****	1.22	2.34
450–500	5.58**	22.63****	1.21	2.92*
500–550	10.26****	23.91****	1.66	1.38
550–600	11.45****	25.58****	0.47	1.82#
600–650	18.69****	31.79****	4.66**	1.41
650–700	21.43****	35.84****	4.11**	1.93#
700–750	21.59****	46.36****	2.48#	1.22
750–800	20.95****	35.22****	3.91*	1.03
800–850	16.03****	31.90****	4.54**	0.85
850–900	15.50****	22.75****	4.20**	1.32
900–950	11.99****	25.12****	6.15***	1.02
950–1000	11.78****	19.73****	6.81***	1.47

*$p < .05$.

**$p < .01$.

***$p < .001$.

****$p < .0001$.

#$p < .1$.

participants fell between −1 and +1 μV. Larger amplitude differences indicative of a more negative potential for tone as compared to segmental violations were observed in 11 participants. Eight participants showed the reversed effect, that is, larger amplitudes for the segmental as compared to the tone violation. Individual analyses of the LPC were conducted for central and posterior sites within a time range of 450 to 1000 msec following word onset. These analyses revealed that the complete violation elicited a larger LPC than the congruous condition in 21 out of 24 participants. Similar results were obtained for the partial violations. Tone and segmental violations elicited larger LPCs as compared to the congruous condition in 22 and 21 participants, respectively. Seven participants showed a comparable LPC for the tone and the segmental violation, whereas 10 showed larger amplitudes for the tone violation and 7 showed larger amplitudes for the segmental violation. Thus, similar to the N400-like negativity, individual differences are high with respect to the comparison of tone and segmental violations and may account for the statistical null result. Effect size calculations (early negativity: Cohen's $d = 0.124$; LPC: Cohen's $d = 0.092$) indicate that a sample size of more than 156 participants is necessary to find a statistically significant difference between tone and rime violations.

To summarize, all violation conditions differed from the control condition with respect to an early, frontally distributed negativity that peaked approximately 300 msec following word onset. This effect started 100 msec earlier for the complete as compared to the partial violations. However, that partial violations failed to differ significantly from complete violations suggests that these differences in onset latency are subject to some amount of variability. This was confirmed by an analysis of individual participants, which indicated that the N400-like effect was more reliable for complete as compared to partial violations. The early negativity was

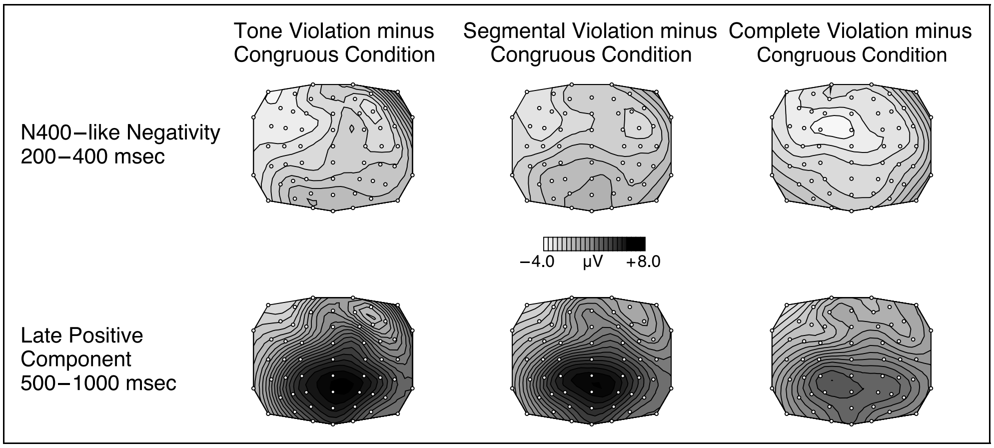

Figure 3. Scalp distribution maps for ERP effects as elicited in the tone, segmental, and complete violation conditions.

followed by an LPC. This component had a centro-parietal distribution and reached its maximum 650 msec following word onset. All violation conditions elicited a significantly larger LPC than the control condition. Additionally, partial violations elicited a larger LPC than complete violations over central regions. Moreover, the violation effects were larger over right than left hemisphere regions. Compared to the N400-like effect, the LPC effect showed more reliably across participants for all three violation conditions.

DISCUSSION

The present study aimed at specifying the role of tone and segmental information for word processing in Cantonese. Previous behavioral evidence suggested that although tone might participate in the lexical code, its role for word processing is restricted relative to segmental information because of temporal constraints (Cutler & Chen, 1997). However, given that Cantonese comprises so many words that differ only in tone, fundamental differences in the role of tone and segmental information for word processing would be counterintuitive. If listeners were not able to use tone as efficiently as segmental information, speech comprehension would be very much context dependent and prone to errors. Thus, the question arises whether previous behavioral studies were really sensitive to on-line tone and segmental processing or whether the reported differences reflect mechanisms that follow initial stimulus processing (e.g., response strategies).

On the Role of Tone and Segmental Information for Word Processing in Cantonese

To answer this question, we compared ERP correlates for the processing of words that differ from a semantically expected word in tone, segmental information, or both tone and segmental information. All three violation conditions elicited an N400-like effect relative to a semantically congruous condition. Statistical analyses indicated that the ERP effect for the complete violation occurred 100 msec earlier than the ERP effects for tone and segmental violations. This difference in onset latency of ERP effects correlates with the onset of acoustic deviation of the presented word from the expected word and is in accordance with previous reports from Indo-European languages (Van den Brink et al., 2001; Van Petten et al., 1999; Conolly & Phillips 1994). However, given that the two partial violations failed to differ significantly from the complete violation and analysis of individual participants suggested that the N400-like effect showed more reliably for complete as compared to partial violations, these latency differences must be interpreted cautiously. More robust latency effects may require acoustic manipulations on a larger temporal scale (e.g., acoustic deviation from word onset as compared to the second syllable of a word).

Most importantly for the purpose of the present study, time course and amplitude of the N400-like negativity were comparable between tone and segmental violations. This suggests that the more accurate detection of deviations in segmental as compared to tone information as observed in the present and previous studies reflect processes that follow initial word recognition (e.g., response strategies, working memory). Furthermore, the present results provide evidence that tone does indeed participate in the lexical code and that its temporal processing constraints are comparable to that of segmental information. This latter conclusion contradicts previous assumptions about temporal differences in the availability of tone and segmental information (Ye & Connine, 1999; Cutler & Chen, 1997). As these assumptions have been derived from studies that presented words in isolation, one might argue that our findings reflect semantic bias produced by sentence context and that when words are presented in isolation word processing does indeed profit more strongly from segmental as compared to tonal information. Ye and Connine (1999) conducted research that compared reaction times for words presented in isolation and in sentence context. They asked participants to monitor syllable sequences for a specific target vowel or tone. Participants responded faster in the vowel-monitoring task as compared to the tone-monitoring task when target syllables were presented in isolation. When embedded in idioms, this effect disappeared. Based on these findings, Ye and Connine concluded that listeners use semantic context to preactivate tone and segmental information of upcoming sentence constituents. One could hypothesize that this preactivation facilitates the processing of tone in sentence context as compared to isolation. However, whether this hypothesis is true remains open as the participants of Ye and Connine were not forced to process tone or segmental information of the target syllable in order to perform the task. Rather, participants could respond as soon as they recognized the idiom and retrieved the corresponding tone and segmental information from memory.

The present study employed sentences with a very high cloze probability similar to the idioms used by Ye and Connine (1999). However, in contrast to their study, the semantic context established here helped participants to process target words only if they were congruous. Moreover, if they were incongruous, tonal and segmental information was unpredictable and participants had to process this information in order to perform semantic acceptability judgments. This, together with the evidence that semantic bias (i.e., cloze probability) modulates N400 amplitude but not latency (Gunter, Friederici, & Schriefers, 2000; Kutas & Hillyard, 1984), suggests that semantic bias does not explain the present finding of a comparable time course of segmental and tonal processing. Moreover, we assume that a similar time course of segmental and tonal processing can be observed when words are presented in isolation. Further research has to confirm this assumption.

Comparing the Present ERP Effects to Findings from Indo-European Languages

Following early work by Kutas and Hillyard (1980), many researchers have attempted to specify the neurophysiological underpinnings of language processing using ERPs. This research, however, has focused on Indo-European languages such as English and Dutch, whereas other languages received less attention. As tone languages differ in many respects from Indo-European languages, a comparison across these languages is necessary to develop an understanding of the processes that are shared by different languages and innate to the human language faculty.

By investigating word processing in Cantonese, an Asian tone language, the present study makes a contribution toward this goal. Similar to findings obtained for Indo-European languages, words that were incongruous to preceding sentence context elicited a larger negativity in the ERP than congruous words. Surprisingly, this negativity differed in some respects from the N400 elicited for semantically congruous and incongruous words in Indo-European languages. Whereas Indo-European language studies usually found this negativity to be largest over central or centro-parietal regions with a maximum peak at around 400 to 500 msec following word onset (Hahne & Friederici, 2002; Hagoort & Brown, 2000; Friederici, Pfeifer, et al., 1993; Holcomb & Anderson, 1993; Holcomb & Neville, 1990), the present negativity was maximal at frontal sites and at 300 msec following word onset. The scalp distribution and latency of this negativity have been replicated in a recently conducted spoken language processing study in Cantonese (Schön, Schirmer, Ng, Penney, & Besson, submitted).

Besides the N400, some Indo-European language studies report a phonological mismatch negativity (PMN) in response to incongruence between a word and a preceding sentence context (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, Brake, & 1992). As the PMN is similar in latency and scalp distribution to the negativity observed in the present study, this component seems of relevance for the current discussion. Connolly and Phillips (1994) investigated the PMN by presenting words that were (1) the most expected sentence completion (e.g., "The piano was out of *tune*."), (2) semantically anomalous but matching the most expected completion in word initial phonemes (e.g., "The gambler had a streak of bad *luggage*."), (3) semantically appropriate, but differing from the most expected completion in word initial phonemes (e.g., "Don caught the ball with his *glove*."), (4) semantically anomalous and differing from the most expected completion in word initial phonemes (e.g., "The dog chased our cat up the *queen*."). All violation conditions elicited a larger negativity relative to the control condition. However, the negativity in violation condition (2) occurred later than the negativity in violation condition (3). Moreover, violation condition (4) elicited both an early and a late negativity. The authors argued that the PMN reflects an early interaction between context and word processing that facilitates the selection of the word's meaning from the mental lexicon and precedes the contextual integration associated with the N400. As the PMN is similar to the negativity observed in the present study, one could argue that both reflect similar processes. However, if our negativity is indeed a PMN, where is the N400? The complete violation employed here corresponds to condition (4) in the Connolly and Phillips study and therefore should elicit both components. That combined semantic and word–initial phoneme violations elicit only one rather than two negativities has been reported previously. Van Petten et al. (1999) found that words deviating from the expected word in initial or later phonemes elicit an earlier or later onset of the N400 effect, respectively. Thus, it seems that comparing word initial and word medial phoneme manipulations does not reliably reveal a PMN. Given that in Indo-European languages semantic violations reliably elicit an N400 effect and semantic violations in the present study elicited only one negativity, we assume that this negativity is more likely to reflect processing associated with the N400. Moreover, differences in latency and scalp distribution between the present negativity and the N400 found in Indo-European languages may reflect language-specific characteristics.

Although there are many ways in which Cantonese differs from nontone languages such as English, here we focus on two aspects. First, Cantonese makes stronger use of suprasegmental information on a lexical level. As a consequence, brain structures that process suprasegmental information might be more strongly engaged in tone as compared to nontone language processing leading to differences in the scalp topography of ERPs elicited to spoken words. Note that even though we assume that acoustic stimulus characteristics drive the scalp topography of the N400-like negativity, we argue that semantic processes underlie the N400-like effect—that is, the amplitude difference between semantically congruous and incongruous words. A second aspect that distinguishes Cantonese, as well as other Chinese dialects, from Indo-European languages is greater context dependency (see Chen, Cheung, Tang, & Wong, 2000; Chen, 1992). Even though Chinese words might be defined both segmentally and suprasegmentally, in many instances, they are still ambiguous because of the large number of monosyllabic words and the high degree of homophony (Chen & Shu, 2001; Chen & Juola, 1982). Moreover, although on average 7.6 Chinese characters share the same syllable, 2.95 characters additionally share the same tone in Cantonese (Zhang & Zhang, 1987). Thus, these words require context in order to be correctly interpreted and it is plausible that this necessity modulates language processing. Accordingly, we propose that differences in the latency of context effects found between the present study and studies in Indo-European languages can be explained by differences in context dependency. A stronger context dependency may result in an earlier effect of semantic context on word processing.

The N400-like negativity observed in the present study was followed by an LPC. The LPC was larger for violations as compared to the control condition and for partial as compared to complete violations. Its maximum was located at right parietal sites and had a latency of approximately 650 msec. Stimulus-locked LPCs have been reported in a number of tasks. For example, in memory retrieval tasks, previously studied items elicit a larger positivity than new items. This old/new effect has been interpreted as a reflection of controlled recollection of the item and the study episode (Mecklinger, 2000; Allan, Wilding, & Rugg, 1998). In language studies, an LPC is typically elicited for syntactic violations and has been associated with processes of repair and reanalysis. However, there are also studies that revealed an LPC in response to semantic violations (Hahne & Friederici, 2002; Van den Brink et al., 2001; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998; Kutas & Hillyard, 1980). The functional implications of this effect have not excited much interest so far and are, as a consequence, unknown. We assume that this effect is not linked to task- or response-related processes as it has also been observed when participants were passively listening (Van den Brink et al., 2001; Kutas & Hillyard, 1980). Therefore, we would like to propose that a similar function underlies this effect as has been associated with the LPC elicited to syntactic violations. Listeners, after detecting a semantic error, probably engage repair mechanisms. If this is true, the present findings suggest differences in the repair of partial and complete violations. In the earlier case, acoustic overlap between the presented and the expected word may lead to an activation of both the presented and the expected word (Connine et al., 1997; Zwitserlood, 1989). After context renders the presented candidate semantically incorrect, listeners may replace it with the semantically correct candidate. However, if there is no acoustic overlap between the presented and the expected word, there are no acoustic cues to activate the expected word. Rather, listeners have to rely entirely on semantic context or world knowledge to restore the meaning of the sentence. Given the similarities in latency and scalp distribution of the present LPC effect and the LPC effect observed in Indo-European languages, it seems that, in contrast to the detection of semantic incongruities, the processes associated with the LPC are comparable across languages.

Conclusions

In tone languages such as Cantonese, tone and segmental information specify the meaning of a word. In the present study, we investigated the influence of both information types on semantic processing. A comparison between the ERPs elicited by semantically congruous words and by semantic violations that were induced at the segmental level, at the tone level, or at both levels suggests that tone and segmental information have a similar impact on semantic processing and that the processing of both information types has a comparable time course. Both elicited an N400-like negativity and an LPC relative to semantically congruous words. In contrast to the LPC, the N400-like negativity differed in latency and scalp topography from findings in Indo-European languages. The more frontal distribution and the shorter latency observed in the present study suggest that language-specific characteristics such as the degree of homophony or the use of tone modulate the latency of context effects and the neuronal structures involved in word processing. Thus, the present findings highlight the significance of cross-language comparisons for modeling both the processes that all languages share and the processes that are language specific.

METHODS

Participants

Thirty Cantonese native speakers were invited to participate in the experiment. Because of excessive eye movements, 6 participants had to be excluded from the data analysis. The remaining 24 participants had a mean age of 21.1 (*SD* 1.5) and were right-handed except for one participant who was left-handed. Fifteen of the 24 participants were women.

Stimuli

The stimulus material consisted of 60 three-clause sentences (e.g., Ah Ming doesn't feel well; therefore he sees the doctor to treat his __ and then applies for sick leave to recover.). The target word was positioned at the end of the second clause. Each sentence was combined with four different one-syllable target words: a semantically congruous word (e.g., /beng6/; "illness"), a word that differed from the semantically congruous word only in tone (e.g., /beng2/; "biscuit"), a word that differed from the semantically congruous word only in segmental information (e.g., /bou6/; "step"), and a word that differed from the semantically congruous word in both tone and segmental information (e.g., /gwai3/; "season"). There was no semantic relationship between target words in the four experimental conditions. The complete violation condition deviated from the semantically congruous condition at word onset, whereas the pure segmental violation was restricted to the rime of the syllable. By violating only the rime of the word we intended to approximate the violation onset in the segmental and the tone condition. In 46 out of 60 cases, the tone violation affected only the rime of the syllable. In the remaining 14 cases, the tone violation included the syllable onset because target words started with voiced consonants (i.e., /j/, /l/, /m/, /n/). However, due to co-articulation, syllable onsets also contained some of the segmental information of the rime. Thus, although the onset of tone and segmental violations may not be identical, they closely overlap in time. Moreover, they both start later than the complete violation, which allowed us to look at the influence of early versus late acoustic deviations from the semantically expected word.

We asked a group of 20 participants that did not serve in the ERP experiment to complete a cloze probability questionnaire, in which they were asked to write down a word that would fit in the target position of each sentence. Sentences and corresponding congruous words that were selected for the study elicited a mean cloze probability of 94% (*SD* 0.84). Forty-five of the 60 sentences had a cloze probability above 90%. The cloze probability for incongruous sentence completions used in the three violation conditions was 0.

As there is no complete database for word frequency in Cantonese, we conducted a rating study with 24 participants in which we presented target words in a questionnaire. We asked participants to use a 5-point scale ranging from 1 to 5 when rating the familiarity of a given word. Most target words were rated as being familiar; only 4 words received a mean familiarity lower than 3 (2.8, 2.4 in the tone violation; 2.9, 2.8 in the segmental violation). The mean familiarity was 4.46 (*SD* 0.25), 4.09 (*SD* 0.49), 4.19 (*SD* 0.37), and 4.16 (*SD* 0.31) for the congruous condition, the tone violation, the segmental violation, and the complete violation, respectively.

A female Cantonese native speaker produced all 240 sentences. Sentences were taped with a DAT recorder and digitized at a 16-bit/44-kHz sampling rate. The duration of target words was 350 msec (*SD* 52), 368 msec (*SD* 36), 354 msec (*SD* 70), and 331 msec (*SD* 57) for the congruous condition, the tone violation, the segmental violation, and the complete violation, respectively. Furthermore, the delay between sentence onset and the occurrence of the target word was 3191 msec (*SD* 306), 3118 msec (*SD* 293), 3108 msec (*SD* 329), and 3078 msec (*SD* 319) for the congruous condition, the tone violation, the segmental violation, and the complete violation, respectively. Each participant in the present study listened to a set of 120 sentences with 30 sentences per condition. Thus, the carrier sentences were presented twice to each listener. Across participants and conditions, the order of sentence repetitions was counterbalanced and the sentence order within sessions pseudorandomized. Furthermore, none of the participants heard a particular target word more than once.

Procedure

Participants were seated in a comfortable chair in a sound-attenuating room facing a computer monitor. Auditory sentences were presented over headphones. A fixation cross preceded every sentence by 500 msec and remained on the computer screen for the entire sentence duration. We chose a delayed response task to prevent influences of the response on the critical period in the ERP. The response window was indicated by a question mark that followed the sentence offset by 300 msec. Participants indicated whether the preceding sentence was semantically congruous or incongruous by pressing one of two buttons on a response box. Half the participants were asked to use their left hand to indicate "congruous" and their right hand to indicate "incongruous." The remaining participants had the reversed hand-response assignment. With response execution the question mark turned off and the next trial began following an interval of 2500 msec.

Electroencephalogram Recording

The electroencephalogram (EEG) was recorded from 36 electrodes that were mounted in an elastic cap according to the modified expanded 10–20 system. The reference electrode was placed on the nose tip. Horizontal eye movements were recorded using electrodes placed on the outer canthus of each eye. Electrodes placed above and below the right eye were used to record vertical eye movements. The electrode impedance was kept below 5 kΩ. The electrical signal was low pass filtered at 70 Hz and recorded at a 256 Hz sampling rate. A digital 20 Hz low pass filter was applied off-line.

Data Analysis

ERPs were computed for the critical words in each condition using a 150-msec prestimulus baseline and an averaging time window of 1000 msec starting from word onset. Only correctly responded trials were included in this process. In order to explore the time course of differences between the experimental conditions, we averaged amplitude modulations for successive 50-msec time windows starting from stimulus onset. Thus, twenty 50-msec time window analyses were performed with Word (congruous, tone violation, segmental violation, complete violation), Hemisphere (left, right), and Site (anterior, central, posterior) as repeated-measures factors. For the factors Hemisphere and Site, electrodes were grouped as follows: anterior–left, AF3, F7, F5, F3, FC5; anterior–right, AF4, F8, F6, F4, FC6; central–left, T7, C5, C3, FC1, CP1; central–right, T8, C6, C4, FC2, CP2; posterior–left, CP5, P3, P5, P7, O1; posterior–right, CP6, P4, P6, P8, O2. Because of the increased likelihood of Type I errors associated with the large number of analyses, only effects that reached significance in two or more consecutive time windows were considered significant. For significant effects, a second ANOVA was performed summarizing significant consecutive time windows into one larger time window. p values of post hoc single comparisons were corrected using a modified Bonferroni procedure (see Keppel, 1991). Consequently, post hoc analyses of Word main effects had a significance threshold $p < .0083$. The Greenhouse–Geisser procedure was applied to correct for violations of sphericity.

Acknowledgments

This research was supported by a Direct Grant for Research from the Social Science and Education Panel of the Chinese University of Hong Kong to Hsuan-Chih Chen, a travel grant under the Germany/Hong Kong Joint Research Scheme from the Deutscher Akademischer Austauschdienst and the Research Grants Council of Hong Kong to Hsuan-Chih Chen and Thomas C. Gunter (D/01/220442), and grants from the Research Grants Council of Hong Kong Special Administrative Region, China (Project Numbers CUHK4322/01H) and (CUHK4142/04H) awarded to Trevor B. Penney and Hsuan-Chih Chen, respectively.

Reprint requests should be sent to Annett Schirmer, Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA, or via e-mail: schirmer@cbs.mpg.de, or to Hsuan-Chih Chen, Department of Psychology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China, or via e-mail: hcchen@psy.cuhk.edu.hk.

REFERENCES

Allan, K., Wilding, E. L., & Rugg, M. D. (1998). Electrophysiological evidence for dissociable processes contributing to recollection. *Acta Psychologica, 98,* 231–252.

Bocker, K. B., Bastiaansen, M. C., Vroomen, J., Brunia, C. H., & de Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology, 36,* 706–720.

Chen, H.-C. (1992). Reading comprehension in Chinese: Implications from character reading times. In H.-C. Chen & O. Tzeng (Eds.). *Language processing in Chinese* (pp. 175–205). Amsterdam: North-Holland.

Chen, H.-C., Cheung, H., Tang, S. L., & Wong, Y. T. (2000). Effects of antecedent order and semantic context on Chinese pronoun resolution. *Memory & Cognition, 28,* 427–438.

Chen, H.-C., & Juola, J. F. (1982). Dimensions of lexical coding in Chinese and English. *Memory & Cognition, 10,* 216–224.

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin and Review, 8,* 511–518.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. G. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37,* 463–480.

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., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43,* 1–18.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201–220.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics, 59,* 165–179.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44,* 171–195.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1,* 183–192.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport, 12,* 3189–3191.

Friedrich, C. K., Kotz, S. A., Friederici, A. F., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cognitive Brain Research, 20,* 300–308.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12,* 556–568.

Hagoort, P., & Brown, C. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia, 38,* 1518–1530.

Hahne, A., & Friederici, A. D. (2002). Differential task effects

on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13,* 339–356.
Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8,* 379–411.
Holcomb, P. J., & Neville, H. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes, 5,* 281–312.
Keppel, G. (1991). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice Hall.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.
Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69,* 218–233.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71–102.
Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–1392.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.
McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 621–638.
Mecklinger, A. (2000). Interfacing mind and brain: A neurocognitive model of recognition memory. *Psychophysiology, 37,* 565–582.
Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for the specificity of the syntactic positive shift. *Neuropsychologia, 36,* 217–226.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.
Schön, D., Schirmer, A., Ng, E., Penney, T., & Besson, M. (submitted). *Effects of phonemic and tone changes on semantic processing in Cantonese: A behavioral and ERP study.*
Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412–432.
Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–985.
Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394–417.
Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14,* 609–630.
Zhang, L. Y., & Zhang, S. Y. (1987). *Common Chinese characters pronounced according to Mandarin and Cantonese.* Hong Kong: Zhong Hua.
Zwitserlood, P. (1989). The locus of the effects of sentential–semantic context in spoken word processing. *Cognition, 32,* 25–64.