V . 33, . 2, . 2004 (2004)

W D L, M H ?
A E -R B P A T
P M C

W [1,2] [1]

(, 1973). (, 1998; V & , 1973). 25 (-) - 400 . A 400 150 1000 . 5.1(400) 2130.15(

that is attached to the rhyme in a syllable (see also Chen, 1999; Leben, 1978; Wan & Jaeger's [1998] description of an autosegmental framework for a different perspective). "Lexical tone" is different from the type of intonation used in English to give emphasis and mark interrogatives (Repp & Lin, 1990; Wang, 1973). In Chinese, lexical tones have phonemic status, meaning that they provide semantic and syntactic information of a word (Ho & Bryant, 1997; Wang, 1973). Errors in the pronunciation of these lexical tones can produce semantic and syntactic errors. For example, in the sentence "Wo ai wo <u>mā</u>" (I love my <u>mother</u>), changing the tone of "mā" could produce the words [illegible], [illegible], or the verb [illegible] 6973).Milner words differing only in the initial consonant. The third set was a series of five hums, produced at the pitches of the five different Thai tones. Results of a word identification task revealed a left hemisphere (right ear) advantage for both the tone and consonant tasks. There was no difference in accuracy between the two ears for the hum task, even though the hums contained tones. These results suggest that when tone is linguistically processed, it is primarily processed in the left hemisphere, as with other linguistic stimuli. In contrast to the Thai speakers, English speakers showed no significant ear advantage for either the tone or hum conditions, but did show a right ear (left hemisphere) advantage for the consonant condition. However, it is important to point out that not all types of linguistic information are left lateralized; right hemisphere dominance has been reported in the processing of speech intonation (Shipley-Brown *et al.*, 1998), and damage to the right hemisphere disturbs the production of prosody (see Edmondson *et al.*, 1987).

Lexical tones in Mandarin Chinese have been argued to be processed in a similar fashion to lexical stress in English and unlike segmentals (Chen, 1999). While we are unaware of any studies using the event-related brain potential (ERP) methodology to study lexical stress, there have been a few papers on the electrophysiological correlates of prosodic processing. These studies have primarily focused on comparisons between music and language and have found prosodic errors (mismatches between syntax and prosody) to elicit ERP components (e.g., P600) similar to those elicited by syntactic errors signaled by nonprosodic information (Besson, 1998; Steinhauer *et al.*, 1999).

A cornerstone of research in the lateralization of language debate comes from research with patients suffering from aphasia. Aphasia is a type of language impairment generally linked to left hemisphere brain damage (Parkin, 1996). If tone production and/or comprehension by native speakers of tone languages were impaired more by right than left hemisphere damage, this would support the hypothesis that the right hemisphere is dominant for lexical tones. However, research with left-hemisphere-damaged aphasics indicates that there is no difference in language impairment between speakers of tonal and nontonal languages (Gandour, 1998; Packard, 1986). After reviewing a large number of neuropsychological findings, Gandour (1998) found no evidence of right hemisphere specialization for lexical tone function in speakers of tone languages. For the purposes of the current study, these findings imply that online processing of lexical tones will likely resemble processing of other kinds of linguistic information, rather than alinguistic or pure tonal processing.

Additional support for the idea that the physical nature of the linguistic information (tonal or otherwise) has little impact on the processing of that information comes from research with deaf aphasics who communicate in sign languages. The domain-specific viewpoint would hold that because sign languages are expressed in the visual and spatial realms, damage to the right hemisphere would cause serious language impairments. As it turns out, damage to the right hemisphere of signers impairs nonlinguistic visuo-spatial abilities whereas only damage to the left hemisphere of signers causes serious language impairment (Corina, 1998; see also Langdon & Warrington, 2000). This is the same pattern found in patients of spoken languages (Parkin, 1996; Zurif, 1995). In the cases of both sign and tone languages, it appears that it is not the physical characteristics of the stimuli, but rather their linguistic nature that determines the specialization of the left hemisphere (but see Bavelier *et al.* [1997, 1998] for evidence of right hemisphere activation during ASL comprehension in neurologically intact deaf individuals).

A remaining question concerns how lexical tone information is integrated with other types of linguistic information. The neuropsychological evidence mentioned above suggests that tonal information is processed no differently than other types of linguistic information. In addition, developmental

Yet, it is not known how tonal information is used during online sentence processing. One way of approaching this question is by thinking about when, during the pronunciation of a word, tonal information can determine word meaning. In Chinese, lexical tone is heard on the vowel. The basic unit of speech in Chinese is the morpheme. Generally, each morpheme or syllable has two parts: the initial and final segments. The initial segment is composed of the beginning consonant(s). The final segment is made up of the ending vowel(s) plus the ending consonant(s). This final segment can also be called the rhyme, and it is during the rhyme that lexical tone appears (Ho & Bryant, 1997). For example, the word gǒu (dog) has an initial consonant "g" and rhyme "ǒu," on which the tone is realized. Because words are detected incrementally, with each bit of sound information listeners are making guesses as to what the word might be (Van Petten [illegible]. 1999). As a result, during spoken word recognition, tone information is heard slightly later than word-initial-segmental information and likely contributes to word recognition slightly later. In a series of experiments and explicit models of word recognition processes, Allopenna [illegible]. (1998) demonstrated that as a word unfolds in time, other words with overlapping sound information become activated in time. Hearing the word "beaker" activates not only "beaker" and cohorts like "beetle," but also activates rhymes like "speaker." Note, however, that the activation of "speaker" is slightly delayed relative to that of "beetle," likely due to the fact that the overlapping sound information between "beaker" and "speaker" occurs milliseconds later than that between "beetle" and "beaker." Given a continuous model of spoken word recognition, we expect that the lexical tones in Mandarin words will influence potential lexical candidates in real-time during the process of word recognition, and that this kind of information may be available slightly later than initial consonant information.

Another piece of information suggesting that syllabic and tonal information are treated differently comes from a series of implicit priming production studies by [illegible]. (2002). In their task, Mandarin speakers responded to previously memorized compound word pairs by uttering the second word upon seeing the first word as a cue on a screen. In the learning phase of the three experiments described below, participants memorized four word pairs. Following this, they were tested by flashing a cue word, to which they were required to respond with the corresponding target (second) word. In Experiment 1b, the initial syllable + tone (bolded) of the four target words were the same (see examples below).[3] In Experiment 2a, the first syllable of the four target words (bolded) shared the same syllable but not the tone.

[3] Examples adapted from Appendix 1 of Chen [illegible]. (2002).

[illegible] 1	[illegible]	[illegible]
pair1	chi1-bing1	[illegible] **1**-liang2
pair2	zi4-sha1	[illegible] **1**-sheng1
pair3	kun1-chong2	[illegible] **1**-ting2
pair4	shu1-cai4	[illegible] **1**-jiao1
[illegible] 2[illegible]	[illegible]	[illegible]
pair1	hang2-kong1	[illegible] **1**-ji1
pair2	su4-shen1	[illegible] **2**-pang4
pair3	zhu1-bao3	[illegible] **3**-cui4
pair4	chou1-yan1	[illegible] **4**-yan2
[illegible] 2	[illegible]	[illegible]
pair1	hang2-kong1	[illegible] **1**-ji1
pair2	dian4-nao3	

Finally, in Experiment 2b, the first syllables of the four target words (bolded) shared tone but were different syllables.

Speech onset times (SOTs) for target words containing overlapping sound information were compared to SOTs for word pairs that did not contain any overlapping sound information. They found priming (faster SOTs as compared to the baseline unrelated word pairs) for both syllable + tone and syllable alone conditions and no (or sometimes slightly negative) priming for the tone condition. These timing differences suggest that at least in production, the syllable is easier to prepare ahead of time than the tone, and that tone and syllable may be differently processed. Implications for the current study will be discussed in a later section.

[illegible]

In recent years, the recording of event-related brain potentials has been a profitable method for research in language processing. The recording of scalp electrical potentials provides a detailed time course of the electrical activity in the brain associated with specific events. One of the best-described brain electrical potentials in the ERP language community is the N400. The term “N400” is used because this component is negative relative to a baseline and peaks at approximately 400 ms following stimulus presentation, although N400 amplitude differences can be seen as early as 50 ms after stimulus presentation (Holcomb & Neville, 1991; cf. Kutas, 1993). The N400 reflects a word’s expectancy and degree of association with its context. Presentation of any word will generally elicit an N400, but less-associated and less-expected words tend to elicit N400s with larger amplitudes (Brown & Hagoort, 1993; Kutas, 1993; Kutas & Hillyard, 1984; Kutas ., 1987; Kutas ., 1988).

The distribution of the N400 is typically centro-parietal and often seen with larger amplitudes in the right hemisphere (Kutas & Hillyard, 1980; Kutas & Van Petten, 1994). The N400 and its underlying cognitive processes are thought to be modality independent (Kutas & Kluender, 1994; Kutas ., 1987; Nigam ., 1992). However, modality-modulated effects have been reported; for example, auditorially elicited ERPs are reported to show left (rather than right) hemisphere prominence (Holcomb & Neville, 1990; Kutas & Van Petten, 1994) and tend to onset earlier and last longer than those in the visual modality (Holcomb & Neville, 1991, 1990).

Within the sentence processing community using the ERP technique, the presentation of anomalous sentences is a standard technique used to examine the details of human sentence processing. An “anomaly” or anomalous word is generally defined as a word that deviates from an expected word or set of words in one or more dimensions. For example, the word “run” in the sentence context “Jackie Joiner run fast” deviates from the expected “runs” in its marking for third person. Comparisons of “run” versus “runs” in this sentence context arguably reveals something about the processing of the third person particle. Numerous 6216.auby.uender, 1994; Tc2316.

(1a) and (1b)[4] have high and low contextual constraints and Cloze probabilities, respectively:

(1) a. He mailed the letter without a [illegible].
    b. He was soothed by the gentle [illegible].

Kutas and Hillyard (1984) presented subjects with sentences like (1a) and (1b) and found that N400 amplitude to sentence-ending words decreased as Cloze probability increased. This finding indicates that semantic incongruity is not a necessary condition for the elicitation of an N400; word expectancy is an additional important factor in determining N400 characteristics.

In later research, Kutas (1993) describes the N400 component as sensitive to both contextual constraint and semantic association. Additionally, meaningful sentences that are terminated unexpectedly will also generate N400s (Kutas *et al.*, 1987). Finally, larger N400s are elicited by content words when compared to function words (Kutas *et al.*, 1987; Kutas *et al.*, 1988).

Some authors have proposed that N400s can be elicited by factors other than semantic expectation. For example, presentation of the second word of a non-rhyming pair (in contrast with a rhyming word) elicits a late negative component termed N450 that is largest over temporal sites (Rugg, 1984a). Rugg argues that the N450 belongs to the same family as the N400 but suggests that the N450 reflects processing at the phonological level. Radeau *et al.* (1998) also found the N400 component to be sensitive to phonological processing. In an auditory priming task, N400 amplitudes were attenuated for words preceded by semantic primes; a smaller but significant attenuation was found for phonologically related primes. These priming effects first became significant in the 300–800 ms window following stimulus presentation, with larger priming effects at posterior electrode sites. The findings by Rugg and by Radeau *et al.* are particularly relevant to the current study. Specifically, we wondered whether lexical tone deviations would elicit the typical auditory N400 effect or if we would observe data patterns more similar to those in these phonological processing studies. One difficulty in comparing the current study with the Rugg and the Radeau *et al.* findings is that these latter two studies investigated the processing of isolated words whereas the current study uses words embedded in sentence contexts.

Kutas (1993) argues that the N400 is representative of a default response to potentially meaningful items such as words, pseudowords, lexical strings, line drawings, real pictures (Nigam *et al.*, 1992), and ASL handsigns. Importantly, deviations in familiar melodies and physical or grammatical anomalies fail to produce the N400 effect (Kutas *et al.*, 1987).

Kutas *et al.* (1987) argue that if N400s are linked to phonological processes, then we ought not to find them in deaf signers. They collected ERP data from

[4] Sentences from Kutas & Hillyard (1984).

three groups of people: hearing adults in a visual presentation, hearing adults in an auditory presentation, and deaf signers with visual presentation of signs. All three groups showed the N400 effect, supporting their claim that the N400 reflects language processing in general and is not necessarily linked to phonological processing. Given these findings, it appears that the N400 and N450 are two different components: the first reflective of linguistic processing and the latter reflective of phonological processing.

The current study was aimed at examining the online processing of lexical tones in Mandarin Chinese. Given previous offline and neuropsychological findings (Packard, 1986; Shipley-Brown et al., 1988; Van Lanker & Fromkin, 1973), we expected to find that lexical tones are processed like other types of linguistic information. Because the N400 reflects semantic expectancy violations during linguistic processing, we would expect tone violations to elicit N400s only if they are being linguistically processed. An alternate possibility, given the findings of Rugg (1984a; 1984b) and Radeau et al. (1998) is that we may find an N450 for the phonological processing of lexical tones that could appear distinct from the processing of non-tonal-based anomalies. A final motivation for the current study was to see if the different types of deviations (changes in lexical tone, segmental information) between anomalous and expected ending words would cause different ERP effects either in latency or in amplitude.

Of the numerous, mutually unintelligible Chinese dialects, Mandarin was used in the current study because it is the major dialect that varies least from region to region and is considered the national language (Wang, 1973). Additionally, of the various dialects, Mandarin Chinese has the least number of tones (Li & Thompson, 1979). Cantonese, for example, has eight tones whereas Mandarin has four (Ladefoged & Buchholz, 1993). Many Chinese speakers who speak another dialect are also fluent in Mandarin. The fact that Mandarin has fewer tones than the other dialects suggests that processing Mandarin tones may be potentially easier for a speaker of a dialect with more tones than vice versa. Evidence from tone perception research indicates that while speakers will be best at perceiving tones of their own dialect, Cantonese speakers may be more versatile in their ability to perceive tones than Mandarin speakers (Lee et al., 1996).

To investigate the online processing of lexical tones in Chinese, we conducted an event-related potential study in Mandarin Chinese. A second experiment investigated the plausibility of the sentences presented in Experiment 1. In Experiment 1, native Chinese speakers were auditorially presented with sentences that had congruent or anomalous endings. Three different types of semantic anomalies were created by manipulating the tone, syllable, or both tone and syllable (double-anomaly) of sentence final words. We hypothesized that an N400 effect would be elicited by all three types of semantic anomalies, independently of the specific source of the semantic information (syllable or

(2) [illegible]: Sentences ended with an expected, plausible syllable but an implausible tone. “When I go to the movies, I always eat some popcorn and [illegible].” (“government office” = táng, second tone).

(3) [illegible]: Sentences ended with an implausible, unexpected syllable with a tone from the expected ending word. “When I go to the movies, I always eat some popcorn and [illegible].” (“book” = shū, first tone)

(4) [illegible]: Sentences ended with both an implausible, unexpected syllable and an unexpected tone. “When I go to the movies, I always eat some popcorn and [illegible].” (“political party” = dǒng, third tone).

[illegible]

Twenty-five participants (11 men and 14 women) were recruited from the greater Portland, Oregon, area for the primary study. Mean participant age was 31.6 years (SD = 13.82, range = 18–64 years). All participants had normal or corrected-to-normal vision, no known neurological impairments, and were right-handed according to self-report. Additionally, all participants were native Chinese speakers who spoke Mandarin Chinese fluently. Most participants also spoke English. For the two participants who did not, a Chinese–English bilingual was present at the time of the study in order to translate if necessary. In addition, one of the researchers spoke Mandarin when necessary.

Although bilingualism was not a requirement for participation in this study, participants in both experiments completed a language questionnaire that yielded information about languages spoken, age of language acquisition, degree of bilingualism, and estimations regarding the amount of use of English and Chinese. On two different scales of 1–10, participants rated degree of bilingualism and use of Chinese and English on a daily basis. On the bilingualism scale, a rating of 1 indicated fluency in Chinese only, a rating of 5 indicated equal or comparable fluency in both Chinese and English, and a rating of 10 indicated fluency in English only. Mean rating on bilingualism scale was 4.23 (SD = 1.78, range = 1–7). On the daily use scale, a rating of 1 indicated primary use of Chinese, and a rating of 10 indicated primary use of English. Mean rating on the daily use scale was 5.2 (SD = 2.69, range = 1–9). Mean age of first exposure to Chinese was 0.5 years (SD = 1.24, range = 0–5 years). Mean age of first exposure to English was 11.22 years (SD = 9.06, range = 0–41 years). All participants in the study were fluent, native speakers of Chinese. All spoke Mandarin fluently, but not all spoke Mandarin as their primary dialect.

Participants were paid $12.00 for participation in the experiment. Additionally, bus fare was made available to all participants.

[illegible]

Auditory and visual stimuli were presented using a Macintosh Quadra 840 AV and a program written in PsyScope, Version 1.02 (Cohen [illegible]., 1993). Participants were seated in a moderately lit, electrically shielded, sound-attenuated chamber facing a computer monitor. Sentences were presented using the computer's internal speaker, which was to the left of the participants. An intercom speaker to the right was used to communicate with the experimenter during the study. A button box with three different colored response buttons was placed on a table in front of participants. A mesh-covered window on the right-hand side of the chamber allowed the experimenter to monitor the behavior of participants and the progress of the experiment.

[illegible]

An Electro-Cap International brand electrode cap was used to record brain activity from six specific sites on the scalps of participants; a seventh electrode on the cap served as ground. Conductive paste was used to lower impedance between the scalp and the electrodes. Two active electrode sites along the midline, Cz and Pz, were located in the electrode cap according to the International 10–20 System (Jasper, 1958). Four additional nonstandard electrode sites were used: a frontal electrode was located approximately over Broca's area (B3), and a symmetrical electrode was located in the homologous site of the right hemisphere, called Broca's right (B4). These electrodes were sewn into the electrode cap midway between F7/8 and T3/4, respectively. A temporoparietal electrode was located approximately over Wernicke's area (WL), and a symmetrical electrode was placed over the homologous site in the right hemisphere, referred to as Wernicke's right (WR). These electrodes were placed in the electrode cap 13% posterior to Cz and 30% lateral of the interaural distance, on each side of the head. These four nonstandard electrode sites are commonly used in ERP language research (Kluender & Kutas, 1993; Pratarelli, 1994). Each of the six electroencephalogram (EEG) electrodes were referenced to a pair of linked mastoid electrodes (Lehtonen & Koivikko, 1971; Stephenson & Gibbs, 1951).[5] In addition to the six scalp sites, eye movements were monitored with a bipolar electro-oculogram (EOG) recording by placing one electrode above the right eyebrow and the

[5] The linked electrodes used as the mastoid reference were created by Robert Ormond, Electronics Support at Reed College. A 12 kΩ resistor was inserted into each of the electrode wires before they joined to form one single pin.

other on the cheekbone below the right eye. An electrode sewn into the electrode cap just posterior to the center of the forehead was used as ground.

P511 Grass amplifiers with filter band-pass settings of 0.01 to 100 Hz were used to amplify the EEG and EOG signals. A 60-Hz notch filter was used throughout the recording. After being amplified, the EEG and EOG signals were sent to an analog-to-digital (A/D) converter; the sampling rate was 200 Hz. A Macintosh Centris 650 was equipped with an NI NB-MIO-16X board, enabling it to digitize and store the EEG data using LabVIEW software. Following EEG recording, trials with artifacts, excessive noise, or amplifier blocking were rejected by hand. High-frequency noise was removed with a 30-Hz filter. Data analysis was performed both before and after the 30-Hz filter. As the general pattern of results did not change by adding in this filter, we will be presenting only the filtered data.

- 0 -31.0.1 4 31. - - -

The experimental sentence frames were the same across the four lists, but ending words for each sentence frame were rotated using a modified Latin squares design. Each list had an equal distribution of the four critical (ending) word types for the experimental sentences. In addition, a pseudorandomization procedure limited the number of times the same condition could appear in a row.

[illegible]

For each sentence frame, critical word, filler sentence, and probe word, a sound file was created using SoundEdit 16. A female native speaker of Mandarin Chinese from Beijing, who spoke clearly and had good articulation, was recruited to record the sound files. All sentences were recorded at a natural speed into the computer using the congruent ending words.[6] Next, the final word of each target sentence was spliced off using SoundEdit (these were not heard by participants). All ending words (both plausible and implausible) were recorded in isolation and stored in separate files. Sentence frames and ending words were presented consecutively with, on average, a 3-ms break between sentence frame and ending words. Because all of the target sound files were created in the same way, there were not differences in the “choppyness” or “naturalness” of the sound files among experimental conditions. Filler sentences were recorded in the same manner as target sentences, but ending words were spliced off the original sentence recording and presented in a separate file[7]. A native Chinese speaker checked over the sound files and found that the splicing of ending words was not disruptive and that the sentences were easy to understand. Probe words were spliced from target and filler sentences using SoundEdit and stored in a separate file. All 240 sentence frames, ending words, and probe words were auditorially presented.

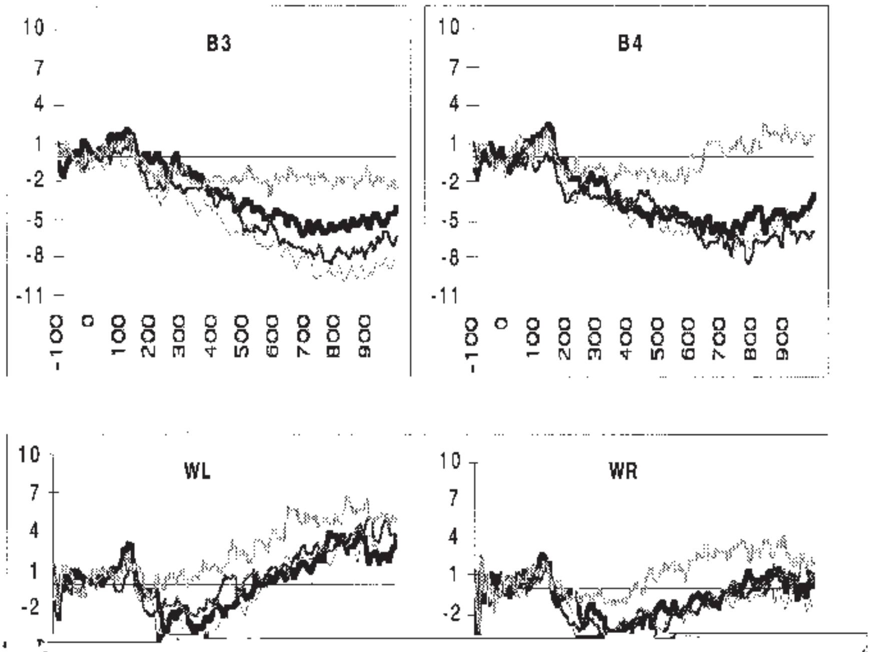
[illegible]

Total experimental session duration was approximately 2 h. Participants were given both oral and written (in both Chinese and English) instructions regarding the study. Participants were instructed to listen to a series of auditorially presented sentences in Mandarin Chinese for comprehension. They were also told that after each sentence, they would hear a single word. Participants

[6] The tone with which some words in Chinese are pronounced changes depending on the tones of neighboring words. This phenomenon is called “tone sandhi.” Stimulus materials were designed with these tone changes in mind. Words immediately preceding ending words were always presented as if the canonical ending word was next, regardless of actual ending word.

[7] We have no reason to believe that participants noted any difference between the filler and experimental sentences in this regard. In addition, we did not analyze the electrical activity associated with filler sentences.

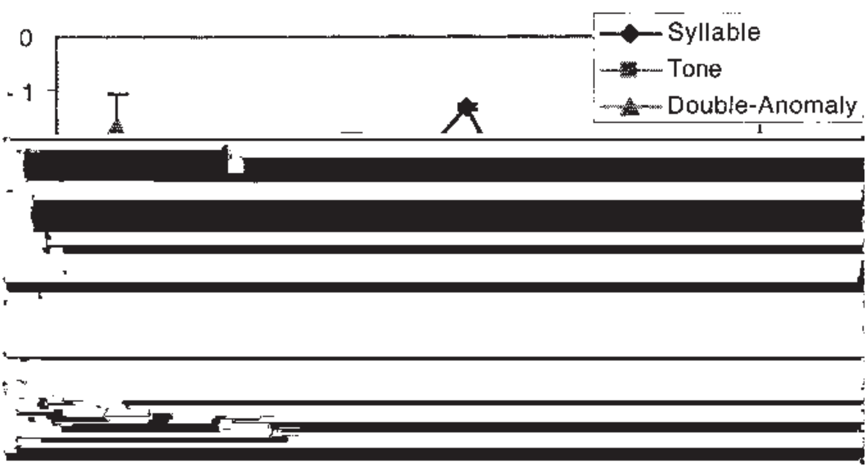
tone = 1502.26 ms (SD = 351.64); and double-anomaly = 1561.01 ms (SD = 351.30). An ANOVA revealed a main effect of condition, with significance levels adjusted using the Greenhouse–Geisser adjustment for sphericity of variance: G–G e = 0.809, $F(3,19) = 6.68$; $p < 0.01$.[8] Post-hoc analyses were used to explore this main effect of condition. Scheffé tests for condition only revealed significant differences between tone and syllable conditions, $p < 0.01$ (mean difference = −135.66 ms), and between syllable and control conditions, $p < 0.01$ (mean difTj/F1 1 Tf1.0728 0v.13q1]ts.D08e


B3
B4
WL
WR
10
7
4
1
-2
-5
-8
-11
-100
0
100
200
300
400
500
600
700
800
900

analysis, in an attempt to analyze the pattern of effects as they unfolded over time, we separately analyzed condition effects at consecutive 50-ms windows from 150 to 550 ms poststimulus (see Van Petten . [1999] for a similar analysis). Finally, a late window between 800–1000 ms poststimulus was used to evaluate the offset of the predicted condition effects. In summary, ERP individual averages were initially subjected to three sets of analyses: (1) a large-window ANOVA to determine the basic effects, (2) a set of ANOVAs that were used to ascertain the onset of the condition effects, and (3) an ANOVA to examine the offset of the condition effects.

A

800 . In general, in this large window, anomalous conditions elicited ERPs with larger negativities when compared to the control condition. Mean voltage differences were calculated by subtracting the mean voltage amplitude for each anomalous condition from the control condition. In Fig. 2, we depict the mean voltage differences for each area and for each condition. The zero line on this graph indicates the amplitude of the control condition. Upon inspection of Fig. 2, the largest differences between the three anomalous conditions appear to be at the B3, WR, and Cz electrodes. To analyze condition and area effects, mean voltages from 200 to 800 ms poststimulus were calculated relative to the 100-ms baseline and submitted to a 4 × 6 repeated measures ANOVA with condition (levels: control, tone, syllable, double-anomaly) and area (levels: B3, B4, WL, WR, Cz, Pz) as within subject factors. This time window was chosen based on previous



. 2. Amplitude differences between control (μV) and each experimental condition for each area in the 200–800-ms poststimulus window. Error bars denote one standard error.

auditory ERP studies and on visual inspection. The ANOVA confirmed the apparent main effect of condition (G–G e = 0.743, F[3,19] = 11.92, $p < 0.001$). A main effect of area was also observed (G–G e = 0.510, F[5,19] = 12.09, $p < 0.001$).

Post-hoc analyses were used to explore the main effects of condition and area. A Scheffé test for condition effects revealed significant differences between control and all three anomalous conditions: tone versus control, $p < 0.001$ (mean difference = −3.44); syllable versus control, $p < 0.0001$ (mean difference = −3.84); and double-anomaly versus control, $p < 0.001$ (mean difference = −3.16). There were no differences between the anomalous conditions.

Significant area effects were largely due to larger negativities seen at anterior electrode sites. Scheffé tests (all $p < 0.01$) indicated that B3 and B4 were significantly more negative than all other areas (except Cz). Finally, the ANOVA indicated only a trend for a condition by area interaction (G–G e = 0.331, F[15,19] = 2.00, $p = 0.09$).

As predicted, the anomalous conditions elicited larger negativities than the control condition. The lack of a difference between anomalous conditions points to the fact that they all elicited an N400 effect of a similar magnitude. It is only in onset latency (see the early window analyses below) where some differences among the anomalous conditions were observed. Finally, the lack of a significant condition by area interaction indicates that the effect of condition was widely distributed.

In order to verify the differences in condition effects across the time windows, we submitted the first eight 50-ms time windows (beginning at 150 ms poststimulus) to a three-way ANOVA including condition, area, and time window as factors. Replicating the findings of the large-window analysis, we found main effects of both condition (G–G e = 0.795, F[3,19] = 8.92, $p < 0.0001$) and area (G–G e = 0.530, F[5,19] = 4.18, $p < 0.05$). Additionally, we observed a significant main effect of time window (G–G e = 0.322, F[7,19] = 3.54, $p <$

= 5 1 . 0 . 7 4

. Results of Analyses of Variance with Condition (Experimental vs. Control) and Area as Factors: Main Effect of Condition.

Double-anomaly

trol at each of the eight consecutive 50-ms time windows (based on analyses in Van Petten, ., [1999]). The results of these analyses are shown in Table I.[10]

In general, across the time range analyzed, condition effects became stronger the later the time window (as evidenced by larger F values). The tone and syllable conditions onset first, differing significantly from control in the 200–250-ms window. However, the effect for the syllable condition was much larger than that of the tone condition. The double-anomaly condition onset the latest; while it first differed from control at the 250–300-ms window, it was not until the 350–400-ms window that the double-anomaly condition began to consistently differ from control.

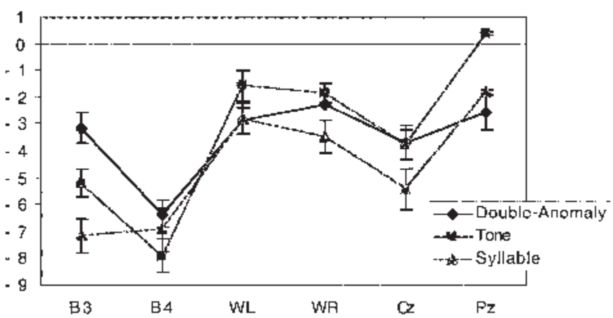
In summary, this pattern of data suggests that while condition effects onset early (200 ms poststimulus), these effects were consistent only for the syllable and tone conditions. It was not until a later window (350–400 ms) that all three conditions were consistently and simultaneously different from control. The lack of a condition by area interaction in the large-window analysis suggests that the main effect of condition was widespread; however, visual inspection indicates that the condition effects were largest at Cz and the frontal electrodes. Additionally, the lack of a difference between the anomalous conditions in the large-window analysis suggests a uniform effect across experimental conditions. The differences between them appeared to be primarily one of timing, with the syllable and tone conditions eliciting the earliest significant negativities.

1000

Analysis at this late window was primarily used to observe when effects seen in earlier windows tapered off. Mean voltages from 800 to 1000 ms post-stimulus were submitted to a similar ANOVA as above. The ANOVA indicated a main effect of condition (G–G e = 0.834, F[3,19] = 5.63, $p < 0.01$) and a main effect of area (G–G e = 0.431, F[5,19] = 21.38, $p < 0.001$]. Although smaller than in earlier windows, the anomalous conditions were still eliciting negativities in comparison to control. Figure 3 shows the mean voltage differences by area and condition.

The main effect of condition was explored in post-hoc analyses. A Scheffé test for condition revealed significant differences between syllable versus control, $p < 0.01$ (mean difference = −4.59) and double-anomaly versus control conditions, $p < 0.05$ (mean difference = −3.46). The difference between tone and control was borderline, $p = 0.06$ (mean difference = −3.30). There were no differences between the anomalous conditions.

In addition, the ANOVA revealed a significant condition by area interaction (G–G e = 0.388, F[15,19] = 4.70, $p < 0.001$). Separate ANOVAs were conducted for each electrode site, revealing significant condition effects at B3 (F = 11.30, $p < 0.001$), B4 (F = 9.10, $p < 0.001$), WR (F = 3.09, $p < 0.05$), and Cz (F = 4.04, $p < 0.05$). Scheffé tests (all effects reported here reached at least < 0.05 of significance) revealed that in this late window, the only condition consistently different from control across these four areas was the syllable condition. The tone condition differed from the control only in anterior areas. Finally, the double anomaly was different from the control only in B4. In summary, while some of the condition effects taper off at this later window, many of the N400 effects are long lasting. The anomaly effect elicited by the syllable condition not only showed the strongest and earliest effects, but it was also the



3. Amplitude differences between control (μV) and each experimental condition for each area in the 800–1000-ms poststimulus window. Error bars denote one standard error.

longest lasting and the most consistent across areas. The other condition effects remained evident only in anterior areas.

2

This experiment served as a follow-up to confirm the quality of the stimulus materials used in Experiment 1. Participants in Experiment 2 were asked to rate the plausibility of the 160 experimental sentence frames used in Experiment 1. As in Experiment 1, each of the 160 experimental sentence frames could end with (1) a semantically plausible word, (2) a semantically implausible word due to a tone error, (3) a semantically implausible word due to a syllable error, or (4) a semantically implausible word due to both a tone and syllable error.

Twenty native and fluent speakers of Mandarin Chinese who were able to read and write in Chinese participated in this experiment. Participants were recruited from the same pool of participants who participated in Experiment 1. Thirteen of the 20 participants in Experiment 2 also participated in Experiment 1.

Participants were given a questionnaire with written versions (simplified Chinese characters) of the 160 experimental sentences from Experiment 1. Participants were asked to rate each sentence on a scale of −4 to 4 for how much the sentence "made sense" by circling one of the numbers. A rating of −4 meant the sentence did not make sense whereas a rating of 4 meant the sentence made perfect sense. Each sentence frame appeared once on each questionnaire. Four different questionnaires were designed so that each sentence frame appeared only once with each of its four possible endings. Questionnaires were delivered and returned through the mail. Participants were paid $5.00 muchstionnaires w27 .078frame apTf0c simpapTfosstrolnumbers. A rat.9623

were conducted based on similar work in the auditory modality by Holcomb and Neville (1991), and Kutas *et al.* (1987). Concurrent with the findings in both of these studies, we found large and long-lasting negativities for anomalous conditions. The late plausibility effects (at 800–1000 ms) were strongest at B3, B4, WR, and Cz.

An additional note is that the effects we observed were in general strongest at Cz and frontal areas. The N400 effect is typically described as having a centro-parietal distribution. To be sure, we did find significant condition effects in all areas, and the lack of an area by condition interaction in the large-window analysis suggests our effects were widespread. Additionally, the timing of the effect is consistent with other N400 studies using auditorially presented stimuli (Holcomb & Neville, 1990, 1991). A possible explanation for this atypical distribution is that we are observing the superposition of the N400 with another effect with a more frontal distribution. One suggestion is that we are observing the effects of a left anterior negativity (LAN), which is typically thought to be associated with language tasks that tax working memory (see Rōsler *et al.*, 1998). However, a working memory account of our findings is not clear, and the broad distribution of our effects (with a bilateral anterior maximum) do not clearly support a LAN account. We wish to conclude tentatively that our results are consistent with an N400 interpretation, but we cannot rule out other interpretations. What is important to draw from our data is the fact that all three anomalous conditions elicited negativities that were significantly different from control, and that we observed some timing differences between anomalous conditions.

With respect to the long-lasting nature of the negativities we observed, Kutas *et al.* (1987) suggest that variations in auditory stimulus length may create a "jitter" in auditory ERP effects, resulting in longer lasting effects in auditory, as compared to visual, conditions. This type of jitter may as well have been responsible for the long-lasting negativities we found. The grand average plots reported in their paper show a similar pattern to ours, with condition effects lasting longest primarily in anterior areas.

The fact that our effects onset so early suggests that we are not simply seeing effects of phonological processing in the form of an N450 as reported by Rugg (1984a, 1984b). If we were looking at N450 effects, we would expect the data to show larger amplitude differences between anomalous conditions, as they are quite different from one another on a phonological level. Additionally, our N400s differed from those seen in Rugg (1984a), and the phonological word priming study by Radeau *et al.* (1998) both in their timing and distribution. While our effects were large at all electrode sites, we observed the largest negativities at anterior sites and Cz. In contrast, the N450 described in Rugg (1984a) was largest over temporal sites. A cautionary note is that the Rugg study did not have lateral frontal electrodes; the only frontal electrode was Fz (which we did not use). Given the differences in electrode placement, comparisons between the

different from the other anomalous conditions). Therefore, the delayed N400 effect for the double-anomaly condition is unlikely to be due to plausibility differences.

We can think of another, unlikely yet possible explanation for the slight delay in the onset of the effect for the double-anomaly condition. We need to present the argument in two parts. First, there are a large number of homophones in Chinese. If you look in a Chinese dictionary at a particular syllable + tone combination, you are likely to see listed at least two different potential meanings for that sound. In our experimental situation, we talk about ,., "expected word," but individual subjects may have considered a different or multiple other possible words as the ending word for the sentence frame. In addition, the large number of homophones in Chinese increases the likelihood that a participant could briefly expect a different ending word with a similar onset to the presented ending word.[11] While experimental sentences were designed to be highly predictiTfT*0.finentng wosu5re that shared features (or were homophones with) the presented, erroneous ending word but were consistent with the sentence context. The essence of the argument is that it may take longer to identify something as erroneous when it is completely unrelated as compared to situations when the erroneous word is similar enough to the target word to be identified as a deviation from it.

In conclusion, the data gathered in the current study are consistent with findings in previous research using the N400 in the auditory modality. We replicated previous findings that auditorially elicited N400s onset earlier and last longer than those elicited in the visual modality. We demonstrated a robust N400 effect for all three types of anomalies in Mandarin Chinese. It is now clear that tonal information is used in online sentence processing and is quickly used to identify word meaning (within 200 ms of word onset) and to identify inconsistencies in sentential semantic structure (revealed by an early N400 response to tonal anomalies). Given the range of ideas about how tonal information might be processed (as linguistic information or as pure tonal information), these results clearly suggest that lexical tones are processed as linguistic information.

While lexical tones are similar to pure tonal information in form, the former clearly influence semantic processing. Though there may be some similarities between English lexical stress and Mandarin's lexical tones, there are important differences between the two (see Chen [1999] for some of their similarities). It is clear that the stress patterns in English influence segmentation processes (Cutler & Norris, 1988) and that contrastive stress can influence the earliest moments of reference resolution (Sedivy [illegible]., 1999). However, even these dramatic effects pale in comparison to the effect of lexical tone in Mandarin. Unlike stress in English, in Chinese many words have minimal pairs that can be distinguished only by lexical tone. Additionally, changes in lexical tone can cause a change in meaning (e.g., [illegible] for [illegible]) but also syntactic category information (e.g., [illegible] for the verb [illegible]). In future work, a comparison of the processing of lexical stress and lexical tone might yield interesting insights about how these two cues are related and could potentially lend support to Chen's theory about the similarities between lexical stress and lexical tone.

The current study thus supports the wide range of offline and neuropsychological evidence in favor of the idea that lexical tones are processed as linguistic information. Even more exciting than the replication of the N400 effect of when and how different features of words are used to guide our predictions about the meaning and form of sentences.

Differences in the identification of anomalous words as such, or so-called jittery detection of errors, is likely directly related to the use of auditorially presented stimuli. Auditory stimuli may be more likely than visual stimuli to reveal effects of the incremental processing of words (see Allopenna *et al.*, 1998; Van Petten *et al.*, 1999). The delayed effect for the double-anomaly condition we observed is likely due to a jittery or delayed detection of the anomaly in this

Edmondson, J. A., Chan, J-L., Seibert, G. B., & Ross, E. D. (1987). The effect of right-brain damage on

Lehtonen, J. B., & Koivikko, M. J. (1971). The use of a non-cephalic reference electrode in recording cerebral evoked potentials in man. [illegible], *31*, 154–156.

Li, C. N., & Thompson, S. A. (1977). A mechanism for the development of copula morphemes. In C. Li (Ed.), [illegible]. Austin: University of Texas Press.

Li, C. N., & Thompson, S. A. (1979). Chinese: Dialect variations and language reform. In T. Shopen (Ed.), [illegible] (pp. 295–335). Cambridge, MA: Winthrop Publishers, Inc.

Lou, L-G., Fan, S-L., & Kuang, P-Z. (1989). Event-related brain potentials (ERPs) reflect mismatch between Chinese character and its mental template. *A*[illegible], *21*(3), 321–327.

McCallum, W. C., Farmer, S. F., & Pocock, P. K. (1984). The effects of physical and semantic incongruities on auditory event related potentials. [illegible], *59*, 477–488.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. [illegible], *4*(1), 15–22.

Packard, J. L. (1986). Tone production deficits in nonfluent Aphasic Chinese speech. [illegible], *29*, 212–223.

Parkin, A. J. (1996). [illegible]. Cambridge, MA: Blackwell.

Pratarelli, M. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. [illegible], *24*, 137–157.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A-S., & McNamara, J., Eds. (1997). [illegible]. Sunderland, MA: Sinauer Associates.

Radeau, M., Besson, M., Fonteneau, E., & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: Behavioral and electrophysiological evidence. [illegible], *48*, 183–204.

Rasmussen, T., & Milner, B. (1977). Clinical and surgical studies of the cerebral speech areas in man.

Van Lanker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. [illegible], *1*, 101–109.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. [illegible], *8*(4), 485–531.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. [illegible], *25*(2), 394–417.

Wan, I-P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. [illegible], *15*, 417–461.

Wang, W. S-Y. (1973). The Chinese language. [illegible], *228*, 50–60.

Zurif, E. B. (1995). Brain regions of relevance to syntactic processing. In [illegible] *1*. L. R. Gleitman, M. Liberman, & D. N. Osherson (Eds.), Cambridge, MA: MIT Press.