LANGUAGE AND SPEECH, 1989, 32(4), 337–354 

# PROSODIC STRUCTURE IN LANGUAGE UNDERSTANDING: EVIDENCE FROM TONE SANDHI IN MANDARIN*

SHARI R. SPEER
*Yale University and Haskins Laboratories*
*New Haven, Connecticut*
CHI-LIN SHIH
*Rutgers University, Newark, New Jersey*
and
MARIA L. SLOWIACZEK
*Bell Communications Research*
*Morristown, New Jersey*

Two experiments show that prosodic information plays a crucial role in the processing of sentences of Standard Mandarin Chinese, where local lexical ambiguities may occur due to the operation of a tone sandhi rule. In Chinese, each word is associated with a tone; in this paper, the term "Mandarin tone sandhi" refers to a phonological rule that changes the first of two consecutive low tones (Tone 3) to a rising tone (Tone 2). As a result, a two-syllable sequence with a rising tone followed by a low tone is ambiguous. In Experiment 1, listeners identified lexical tones for ambiguous, unambiguous, and nonsense words in phrasal contexts where the tone sandhi rule might have applied. Comparable results in the lexical versus nonsense conditions indicate that judgments did not rely simply on lexically stored tonal information, but also made reference to the tonal context of the phrase. In Experiment 2, subjects chose the most likely written English translation for auditory sentences of Mandarin. Global prosodic information was manipulated to create different levels of "prosodic closeness" between two critical items in a tone sandhi environment, while the syntactic relation between these items was held constant. Results show that listeners relied on the prosodic structure of the phrases to determine whether or not the tone sandhi rule had applied, and consequently to identify individual lexical items. The evidence is taken to support the notion that prosodic structure influences auditory language comprehension processes.

*Key words:* Mandarin Chinese, tone sandhi, prosody, sentence processing

---

The research reported here was supported by NICHD Grant HD01994 to Haskins Laboratories, by NIMH Grant T3216745 to the Psychology Department of the University of Massachusetts, and by NICHD Grant T3207035 to the first author. We are grateful to Chuck Clifton for allowing us to conduct these experiments in his laboratory, to Chen Zhe for introducing us to the Mandarin speakers who participated in the experiments, to Cheryl Wilson for help with constructing materials, and to Alice Faber and Haskins Laboratories for assistance with the acoustic analyses. We also thank Chuck Clifton, Robert Crowder, Bruno Repp, Jaqueline Vaissière, and Marjorie Chan for helpful comments on the manuscript.

Correspondence may be sent to Shari R. Speer, Department of Psychology, P.O. Box 11A Yale Station, New Haven, CT 06520.

## INTRODUCTION

Researchers in psycholinguistics have increasingly recognized that prosodic structure plays a more fundamental role in sentence production and comprehension than just that of adding expressive emphasis and connotative meaning to speech. Prosodic structure refers to the pattern of pitch, timing, and intensity differences that correspond to the qualities of intonation, stress, pausing, and rhythm in spoken sentences. Although there is a relationship between prosodic and syntactic structures, the two are not necessarily isomorphic, since prosodic structure reflects not only the influence of syntax, but that of phonology and pragmatics as well.[1] Prosody is a potential source of disambiguating information when the syntax of a sentence is ambiguous (Beckman and Edwards, in press; Cooper and Sorensen, 1977; Klatt and Cooper, 1975; Sorensen and Cooper, 1980). In sentence comprehension, listeners can resolve syntactic ambiguities by making use of prosodic cues such as pauses and phrase-final lengthening (Lehiste, Olive, and Streeter, 1976; Slowiaczek, 1981; Carroll and Slowiaczek, 1987; Streeter, 1978; Wales and Toner, 1979).

This paper investigates whether or not listeners can use prosodic information to disambiguate phrases and sentences in Standard Mandarin Chinese. Mandarin is a tone language; that is, it makes lexical distinctions using variation in tone. Nearly every morpheme in Mandarin is a monosyllable with an affiliated lexical tone. A change in tone changes the meaning of a word; for example, the word 'leng' when spoken with a rising tone means 'corner', but when spoken with a low tone it means 'cold' There are four tones in Mandarin: a high, level tone [H] (Tone 1); a rising tone [LH] (Tone 2); a low tone [L] (Tone 3); and a falling tone [HL] (Tone 4).[2] The low tone (Tone 3) optionally surfaces as falling-rising in pre-pausal positions. Figure 1 shows schematic versions of the fundamental frequency contours for each of the four Mandarin tones, and shows the

---

[1] A sentence with a single syntactic structure can be spoken with a variety of well-formed prosodic structures. Here are some two-phrase examples from Mandarin Chinese and English: (multiple-phrase readings would also be possible; "/" indicates a prosodic break between phrases).

Mandarin:

An-Dong-ni / xi-wang Tang-mu qu Nan Jia-Zhou kai yi-ge hui-yi.
An-Dong-ni xi-wang Tang-mu / qu Nan Jia-Zhou kai yi-ge hui-yi.
An-Dong-ni xi-wang Tang-mu qu Nan Jia-Zhou / kai yi-ge hui-yi.

[Anthony wish Tom go Southern California participate one meeting]

English:

Anthony wished / that Tom was attending the meeting in Southern California.
Anthony wished that Tom / was attending the meeting in Southern California.
Anthony wished that Tom was attending the meeting / in Southern California.

[2] Tone 3 occurs with a variety of shapes; here we refer to the shape of Tone 3 in non-phrase-final position. There is also a 'neutral' or 'atonic' tone in Mandarin, referring to the lack of tone on an unstressed syllable. Its pitch value is dependent on its tonal environment.

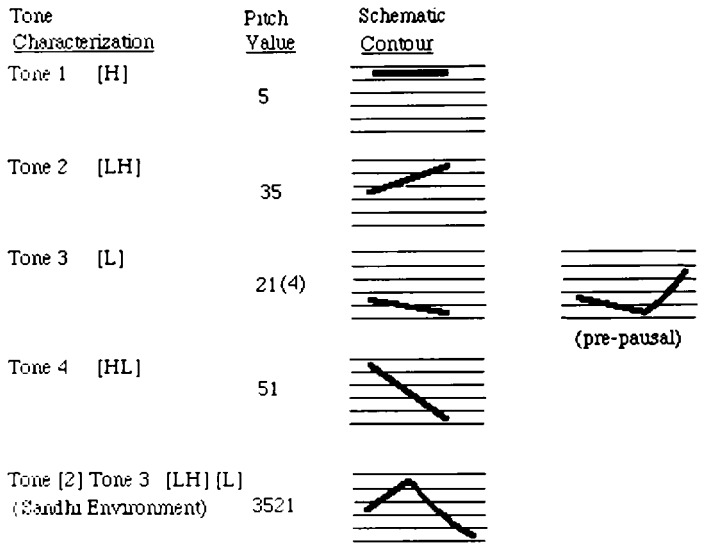


Fig. 1. Representations of the four Mandarin tones and the tone sandhi environment. [2] indicates a surface Tone 2 with underlying Tone 3 (after Chao, 1968).

"pitch values" (on a scale of 1–5, with 5 as the highest pitch) that are commonly associated with tone description in Asian languages (cf. Chao, 1968).

When a Tone 3 syllable is followed by another Tone 3 syllable, the application of a phonological rule changes the first Tone 3 to a Tone 2. The rule can be formally stated [L] → [LH] /_ [L]. Rules of this class, involving the local interactions of specific phonological features, are often called sandhi rules.[3] A schematic version of the fundamental frequency contour associated with the application of the tone sandhi rule is shown in Figure 1. Tone sandhi in Mandarin can lead to lexical ambiguities. For example, consider the following phrase:

du2 = to study
du3 = to gamble

du2 jiu3–tian1
to study / gamble nine days

The word 'du' means 'to study' if it is said with a rising tone (du2). However, 'du' said with a low tone (du3) means 'to gamble'. In the above sentence, a Tone 3 word such as

[3] The phonological rule of 3/2 tone sandhi discussed here is not the only such rule in the family of Mandarin languages. There are other, generally optional, rules of tone sandhi that apply in some dialects and/or in very rapid speech. Our experimental materials did not contain instances of the application of such rules.

'du3' would be changed to 'du2' because it is followed by a Tone 3 word (jiu3). This would create a lexical ambiguity in this sentence between 'study' and 'gamble'

Perceptual experiments have shown that listeners are unable to distinguish between these two kinds of Tone 2–Tone 3 surface sequences. In these experiments, speakers were tape-recorded while reading Chinese characters aloud. Although the written characters were unambiguous, the spoken words were ambiguous because of the application of the tone sandhi rule. When listeners were given a written list of Chinese characters to identify with the spoken words, they were unable to distinguish which member of the sandhi pairs had been intended by the speaker. This was true even when the speakers listened to recordings of themselves (Wang and Li, 1967; see also Zue, 1976).

Although the application of the tone sandhi rule is common in Standard Mandarin, it does not occur in all contexts. There is general agreement among researchers that the rule requires a certain degree of "closeness" between the two adjacent words with underlying Tone 3; without this, the application of the rule is blocked. Some researchers have argued that the operation of certain phonological rules, such as palatalization, is restricted to the domain of a syntactic constituent (Cooper, Egido, and Paccia, 1978). The domain of application for the tone sandhi rule in Mandarin has also been hypothesized to be determined by the strength of syntactic and morpho-syntactic boundaries between the words with the tones involved (Cheng, 1970; Kaisse, 1985; Selkirk, 1984). From this perspective, the "closeness" that allows the operation of the phonological rules is based on the syntactic grouping of the words. An example phrase that eludes the syntactic juncture-based analyses is shown below:

| | | | | | |
|---|---|---|---|---|---|
| Lao[2] | Li[2] | mai3 | xiao[2] | bi3 | |
| Old | Li | buys | small | writing brush. | |
| L | L | L | L | L | underlying tones |
| LH | LH | L | LH | L | surface tones |

The example shows the most common pronunciation for a string consisting entirely of words with underlying Tone 3 (L).[4] The problem for syntactic juncture-based accounts (acknowledged in Cheng, 1968; Kaisse, 1985; Selkirk, 1984) is that the tone sandhi rule has applied across the strongest syntactic boundary (the subject–predicate boundary between Li[2] and mai3), but has not applied across the lesser syntactic boundary within the verb phrase (the verb–object boundary between mai3 and xiao[2]). Thus, the example contradicts the notion that the strength of syntactic "closeness" determines the domain of application for the tone sandhi rule.

An alternative perspective suggests that the domain of tone sandhi application is determined by prosodic factors (Zhang, 1988; Hung, 1987; Shih, 1986). That is, the "closeness" that determines the operation of the rule is based on the prosodic grouping of the words. In general, prosodic structure consists of a finite number of hierarchically

---

[4] The cited sentence is one of many possible surface variants. The pronunciation Lao[2] Li3 mai3 xiao[2] bi3, which fits the juncture-based analysis, is also acceptable at some speech rates.

organized levels, each associated with certain phonological or phonetic properties (cf. Nespor and Vogel, 1986; Pierrehumbert and Beckman, 1988; Selkirk, 1978, 1980). At the lowest level of structure, the syllables of an utterance are grouped into small units, which are in turn grouped into larger, higher-level units. As units increase in size and in level within the hierarchy, they incorporate more spoken material, and they are increasingly likely to be followed by a pause. The existence of an intentional pause (but not a hesitation pause) is likely to block the application of the tone sandhi rule (Cheng, 1970; 1973).

According to the theory proposed by Shih (1986), the units of prosodic structure in Standard Mandarin are disyllabic feet (composed of a disyllabic word or two adjacent monosyllabic words), super feet (composed of disyllabic feet and adjacent, as yet unadjoined syllables), and phrases (composed of feet and super feet). Shih showed that although the tone sandhi rule may be blocked across many types of syntactic juncture, it is never blocked within a foot (see Hung, 1987, and Zhang, 1988, for discussion of the constraints on the formation of Mandarin prosodic feet). When two low tones are grouped in the same foot, the tone-bearing syllables are maximally prosodically close, and the tone sandhi rule must apply. When two low tones are grouped into separate prosodic units, a prosodic boundary (or boundaries) occurs between the tone-bearing syllables, and the rule is not required to apply. As the two adjacent Tone 3 words become more distant in the prosodic structure, the application of the tone sandhi rule is increasingly likely to be blocked. Thus, prosodic "closeness" determines the domain in which the phonological rule applies.

For example, in Figure 2(a), application of the tone sandhi rule is mandatory for the syllable gou3, because gou3 'dog' and yao3 'bites' are part of the same prosodic unit at the lowest level of structure, the disyllabic foot. Also in 2(a), application of the tone sandhi rule is optional for the syllable yao3, because yao3 'bites' and hao3 'good' occur in separate feet, but are part of the same prosodic unit at a higher level of structure, the phrase. In Figure 2(b), application of the tone sandhi rule is very likely to be blocked for the syllable ying3 because lu4–ying3 'video' and chang3 'space' occur in separate prosodic phrases.[5] Although prosodic structures and syntactic structures may sometimes

[5] Speech rate is a factor in determining the domain of tone sandhi for all the analyses discussed here (Cheng, 1970; Kaisse, 1985; Selkirk, 1984; Shih, 1986). According to Shih's theory, tone sandhi applies cyclically in a prosodically-determined, hierarchically-structured domain. The initial level for the application of the tone sandhi rule is tempo-determined: The size of the prosodic domain for the mandatory application of tone sandhi in the first cycle is larger the faster the tempo. That is, the domain of application for tone sandhi in rapid speech (the super-foot or phrase) is larger than the domain for slow speech (the disyllabic foot). This is consistent with theories (e.g., Selkirk, 1984) that posit 'ideal' time as the initial determinant of whether a rule will apply. Shih (1986) derives the problematic 'Lao[2] Li[2] mai3 xiao[2] bi3' as follows: A foot formation rule assigns the prosodic structure ((lao li)f (mai (xiao bi)f')P. In the first cycle, tone sandhi is mandatory at the level of the disyllabic foot, resulting in tonal change for 'lao' (old) and 'xiao' (small). A second cycle allows tone sandhi to optionally apply at the level of the phrase, resulting in a tonal change for 'li'. There is no change in tone for 'mai' in the second cycle because the first-cycle application of the tone sandhi rule changed the context tone 3 (on 'xiao'), destroying the sandhi environment for 'mai'.

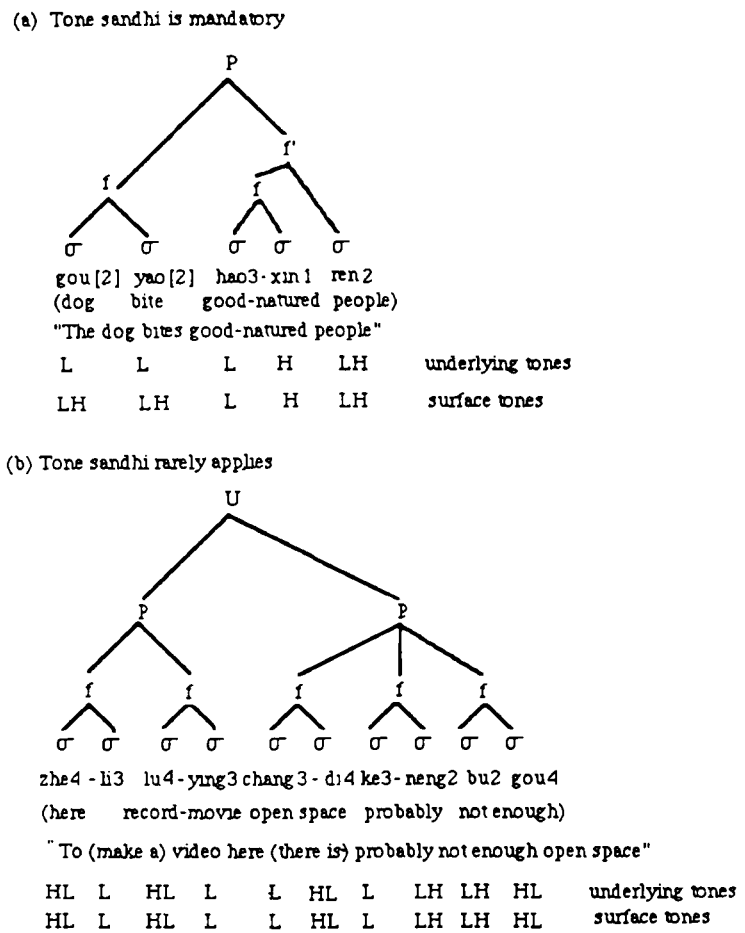


Fig. 2. Prosodically determined domains for the application of tone sandhi. (Symbols: σ syllable, f prosodic foot, f' prosodic super foot, P prosodic phrase, and U utterance.)

share boundaries, (a) shows that this is not always the case: Although a high level syntactic boundary (between subject and predicate) occurs between gou3 'dog' and yao3 'bite', the two words form a single prosodic unit. Because the domain of application for tone sandhi is prosodically rather than syntactically determined, the rule applies even across the strong syntactic boundary. Thus, Shih's theory explicitly predicts that prosody, and not syntax, controls the application of the tone sandhi rule. The experiments presented here were constructed to demonstrate the influence of prosodically determined domains when syntactic structure is held constant. If the Mandarin tone

sandhi rule is sensitive to prosodic information, we expect the rule to be less likely to apply across the boundary between two prosodic constituents as the hierarchical level of the constituents increases. On the other hand, if the controlling factor for Mandarin tone sandhi is syntax, we should find no difference in likelihood of rule application across the constant syntactic juncture.

We conducted two experiments to test how well listeners can use information from prosodic structure to help determine whether or not the tone sandhi rule has applied. In the first experiment, we asked listeners to identify the lexical tones of ambiguous, unambiguous, and nonsense words in phrasal contexts. Some of the items contained a surface Tone 2–Tone 3 sequence, and we were interested to see whether the listeners would report Tone 2 (the acoustic surface tone) or Tone 3 (the tone associated with the word before the tone sandhi rule applied, i.e., the underlying tone). In the nonsense conditions, where materials contained no clear lexical, syntactic, or semantic information, listeners were forced to make their identifications based on prosodic information.

In the second experiment, we asked listeners to choose a written phrase of English that was most closely related in meaning to an auditory phrase of Mandarin. Listeners heard a sentence or phrase that contained a potential environment for the application of the tone sandhi rule. Lexical ambiguities were created by placing a word that was spoken with Tone 2 before a Tone 3 word. The Tone 2 words were chosen to have two possible underlying sources, one a Tone 2 word and the other a Tone 3 word. Since the tone sandhi rule would change Tone 3 to Tone 2 in the environment of a second Tone 3, the Tone 2 words were ambiguous. The sentences were constructed so that the phrase-level prosodic structure would influence the likelihood of the application of the sandhi rule, and thus influence the ease of disambiguation. Four conditions were created to vary how close the ambiguous word and the Tone 3 context word were in the prosodic structure of the sentence. As described by Shih (1986), the closer the two words are prosodically, the more likely it is that the tone sandhi rule will apply. If listeners use this information from prosodic structure in the process of determining lexical tone, there should be less ambiguity for the sentences where the two words are prosodically more distant. In the prosodically distant case, the tone sandhi rule is less likely to apply, so that a word that appears as a Tone 2 word probably originated as a Tone 2 word. Consequently, we predicted that as the prosodic distance increased across the four conditions, subjects would be increasingly likely to report that the ambiguous word was a Tone 2 word.

## EXPERIMENT 1

### *Method*

*Subjects.* Twenty native speakers of Mandarin Chinese from the University of Massachusetts community served as subjects. Ten speakers were from the PRC and ten were from the ROC. All spoke English as well as Chinese; however, their skill in English speaking varied widely. The subjects were paid for their participation in the experiment.

*Materials and design.* Forty phrases in Mandarin Chinese were constructed, with 10

phrases in each of four conditions: In condition 1 (Unambiguous Sandhi), the phrases contained a Tone 3 word which preceded another Tone 3 word, so that the tone sandhi rule applied. There was no Tone 2 word in the language with the same segmental structure as the Tone 3 word; that is, condition 1 phrases were unambiguous. In condition 2 (Disambiguated Sandhi), the phrases contained a Tone 3 word that was ambiguous because there was a corresponding Tone 2 word, but the word could be disambiguated from the syntactic and semantic context of the sentence. In condition 3 (Nonsense), strings of ambiguous morphemes were presented as if they were sentences, and in half of these trials a Tone 2 syllable was followed by Tone 3 syllable, creating an environment in which the tone sandhi rule could have operated. These nonsense strings were composed of homophonous Mandarin monosyllables, chosen so that each monosyllable occurs in Mandarin with more than one tone, and so that even with its specified tone, the monosyllable remains ambiguous (e.g., 'ji' can occur with all four tones: ji1 means "chicken", or "engine"; ji2, "residence", or "to attack"; ji3, "spear", or "oneself"; and ji4, "medicine", or "to mail"). Nonsense strings had no discernable syntactic or semantic structure. The nonsense condition served to investigate how the subjects would characterize the tone of these words in the absence of all but phonological clues to the associated lexical item(s) (i.e., subjects had to respond on the basis of only the acoustic tones and their order in the string). Finally, in condition 4 (Filler), ten unambiguous filler phrases contained no words to which sandhi rules could apply. Examples of these materials are presented in Table 1.

The forty sentences were spoken by the second author, a female native speaker of Mandarin.[6] The sentences were recorded and presented to subjects in one of four random orders. An equal number of subjects heard the sentences in each of the four orders.

*Procedure.* Subjects were seated in a quiet room, separate from the experimenter, and listened to the sentences over headphones. Subjects were instructed in English, "Listen to each sentence of Mandarin on the tape. At the same time, please read the sentence, transcribed in Roman characters, on the answer sheet in front of you. As you listen, respond by marking each syllable on the answer sheet with the symbol for the tone you hear, corresponding to the four tones found in Mandarin Chinese" Four tone markings are customarily used by Mandarin speakers; all subjects used the same scheme: '–' Tone 1 (high level); '/' Tone 2 (rising); '\/' Tone 3 (low, falling-rising), and '\' Tone 4 (falling). Subjects were allowed 4 seconds after each phrase to finish marking the tones for that phrase. At the beginning of each tape, four practice trials were included so that subjects could ask questions and become accustomed to the transcription and to the tone marking task. The sentences were transcribed according to the Pinyin system. Since there are several systems for transcribing Chinese words in Roman characters, subjects were told that the spelling of the transcribed words was not important, and that their main task was to listen to and identify the tones on the tape. Subjects were

---

[6] Standard Mandarin, the national language of China (both the PRC and the ROC), is spoken in schools and public places over a huge geographic area. For this reason, different accents of Standard Mandarin are common. Our speaker grew up in Taiwan, and has a southern accent. The sandhi rule discussed here seems to be a deep-rooted process, in that it is shared across all variants of Standard Mandarin and applies under substantially identical grammatical conditic[illegible]

TABLE 1

Experiment 1: Example Mandarin phrases with English glosses.
(Numbers indicate tone; [2] indicates the tone sandhi rule could have applied;
* indicates ungrammaticality.)

| | | | | |
|---|---|---|---|---|
| *Condition 1* | Unambiguous Sandhi | kao3<br>*kao2 | 'bake' | kao[2] bing3-gan1<br>'bake biscuit' |
| *Condition 2* | Disambi-guated Sandhi | zhu3<br>zhu2 | 'cook'<br>'bamboo' | zhu[2] hao3 fan4<br>'the meal is cooked' /<br>*'the meal is bamboo' |
| *Condition 3* | Nonsense potential Sandhi | jiang4<br>ji2<br>ji3<br>wu3<br>zhuan4 | 'stubborn/general'<br>'residence/to attack'<br>'spear/oneself'<br>'five/cover'<br>'to turn/profit' | jiang4 ji[2] wu3 zhuan4 |
| | non-Sandhi | zhen1<br>ming2<br>xin1<br>dou4 | 'needle/truth'<br>'name/bright'<br>'heart/new'<br>'fight/beans' | zhen1 ming2 xin1 dou4 |
| *Condition 4* | Filler | | | dian4-nao3 yin1-yue4<br>'computer music' |

encouraged to mark the tones as quickly as possible, and to always record their first impression of the tone they heard. Finally, subjects were warned that some of the phrases would be nonsense strings.

*Results and discussion*

The results of the experiment are presented in Table 2. Neither order of presentation nor nation of subject origin influenced subjects' judgments of tone; accordingly, the data were collapsed across the four orders and two nationalities of subjects. As Table 2 shows, subjects were most likely to report the underlying tone when the tone sandhi rule had applied. In conditions 1 and 2, the tone sandhi conditions, subjects consistently reported Tone 3 (the underlying tone) rather than Tone 2 (the surface tone) when surface Tone 2 was followed by Tone 3. In condition 1, the unambiguous sandhi condition, subjects reported Tone 3 82% of the time. This is comparable to the 84% overall accuracy for all tones in condition 1 phrases.[7] In condition 2, the disambiguated sandhi condition,

[7] In both nonsense strings and phrases of lexical items, subjects were scored as having correctly reported the tone when they marked the tone that the speaker had intended to assign to the syllable.

TABLE 2

Experiment 1: Percentage of subjects reporting Tone 2 vs. Tone 3 when surface Tone 2 was presented, and percent correct for non-sandhi tones

| | | *% Tone 3* | *% Tone 2* | *% Other* | *% Correct non-sandhi tones* |
|---|---|---|---|---|---|
| *Condition 1* | Unambiguous sandhi | 82 | 13 | 5 | 84 |
| *Condition 2* | Disambiguated sandhi | 79 | 18 | 3 | 82 |
| *Condition 3* | Nonsense | | | | |
| | potential sandhi | 51 | 38 | 11 | 79 |
| | non-sandhi | 10 | 63 | 27 | 79 |
| *Condition 4* | Filler | 9 | 70 | 21 | 81 |

there were syntactic and semantic cues to indicate whether sandhi had applied, and subjects reported Tone 3 79% of the time, a rate again compatible with their overall rate of 82% accuracy for all tones in that condition. In the fillers, subjects reported tone correctly 81% of the time.

By comparison, when there was a potential tone sandhi environment in nonsense strings, subjects reported Tone 3 only 51% of the time, while their overall accuracy for tones in the nonsense phrases was 79%.[8] Comparison of the nonsense condition is particularly important: Even without the input from lexical words, the prosodic information from the tone sandhi environment resulted in a tone sandhi judgment half the time. Table 2 shows that subjects clearly distinguished between surface Tone 2 in the tone sandhi environment and surface Tone 2 in other contexts: Outside the sandhi environment, surface Tone 2 words were unlikely to be reported as Tone 3, whether they appeared in lexical items (9%) or in nonsense (10%).

Significantly more underlying tones than surface tones were reported for all items in sandhi conditions. In the nonsense condition, no one item in a tone sandhi environment was clearly more likely to elicit a Tone 2 or Tone 3 judgment than any other; this finding suggests that neither special acoustic cues to underlying tone nor extraneous variables such as relative frequency of ambiguous word meanings can account for the tendency for subjects to report Tone 3 when they heard a surface Tone 2 in a tone sandhi environ-

[8] We speculate that subjects' overall accuracy for this task was substantially below 100% because most adult Mandarin speakers have not engaged in a tone-marking task since elementary school, and thus our subjects were 'out of practice'. Also, the task had to be performed as quickly as possible, and time pressure may have decreased accuracy.

ment. Unfortunately, it was not possible to conduct an acoustic analysis comparing the duration and fundamental frequency of the words with underlying Tone 2 to that for words with underlying Tone 3. Because duration and fundamental frequency are influenced by segmental information (e.g., vowel height), such an analysis would require that words with identical segmental information occur in all conditions. This would, of course, be impossible because it would require a word to be simultaneously unambiguous and ambiguous. (For example, the word 'kao[2]' ('bake') in the unambiguous sandhi condition would ideally be compared to a segmentally identical 'kao2' (not a word) in the disambiguated sandhi condition. However, in order for a version of 'kao' to occur in condition 2, it would need to be an ambiguous word.)

## EXPERIMENT 2

### *Method*

*Subjects.* The same as in Experiment 1.

*Materials and design.* Twenty-four sets of sentences were constructed, with four versions of each sentence. Each sentence contained a tone sandhi environment, a Tone 3 word preceded by a word which was spoken with a Tone 2 but could have originated as either a Tone 2 or a Tone 3 word. All of the phrases were ambiguous, since they were constructed so that either the Tone 2 or Tone 3 word could make a grammatical and semantically sensible phrase. According to Shih's theory, the prosodic distance between the Tone 2 and Tone 3 words determines whether or not tone sandhi will apply. Therefore, the four versions of each sentence were constructed to present four levels of prosodic distance between the ambiguous word and the Tone 3 word which created the Sandhi environment. The four conditions are as follows: 1) Mandatory tone sandhi, 2) tone sandhi often applies, 3) tone sandhi applies sometimes, 4) tone sandhi rarely applies.

Materials for this experiment contained a variety of syntactic and prosodic structures. Each sentence had only one possible syntactic interpretation, but, of course, multiple prosodic readings were possible. The reader pronounced each sentence grammatically, producing boundaries appropriate to the four experimental conditions. Boundary information included segmental lengthening, pausing, declination, and/or catathesis (see the acoustic analysis below). An example of four versions of one phrase, with prosodic structures indicated, is presented in Table 3. For the example phrase in condition 1, the Tone 2 word and the Tone 3 word occurred within the same trisyllabic super foot; in condition 2, the Tone 2 word occurred in a disyllabic foot, while the Tone 3 word occurred in an adjacent super foot within the same phrase; in condition 3, the Tone 2 word and the Tone 3 word occurred in adjacent phrases, and in condition 4, the Tone 2 word and the Tone 3 word occurred in adjacent phrases that were lengthened to increase boundary strength. The syntactic relationship between the Tone 2 and Tone 3 words was identical for all four versions of each sentence: In the example phrases, the complex verb phrase 'fan1–tu[2] zhao3', "turn the map/ground to look for" appears in all four conditions.

TABLE 3

Experiment 2: Example Mandarin phrases in Roman transcription and English translation. (Numbers indicate tone; [2] indicates the ambiguous word to which the tone sandhi rule could apply; 0 indicates the neutral tone)

*Condition 1. Mandatory tone sandhi*

Surface: fan1-tu[2] zhao3
(( )f )f'
turn the map/ground look for

Underlying Tone 2: 'check the map to search'
Underlying Tone 3: 'dig and search'

*Condition 2. Tone sandhi often applies*

Surface: hai3-dao4 fan1-tu[2] zhao3 zhu1-bao3
(( )f ( )f ( ( )f)f'P
pirate turn the map/ground look for treasure

Underlying Tone 2: 'The pirate checks the map looking for treasure'
Underlying Tone 3: 'The pirate digs the ground looking for treasure'

*Condition 3. Tone sandhi sometimes applies*

Surface: hai3-dao4 fan1-tu[2] zhao3 cong2-gian2 de0 zhu1-bao3
(( )f ( )f)P (( ( )f )f' ( )f)P
pirate turn the map / ground look for previous treasures

Underlying Tone 2: 'The pirate checks the map looking for old treasures'
Underlying Tone 3: 'The pirate digs the ground looking for old treasures'

*Condition 4. Tone sandhi rarely applies*

Surface:
hai3-dao4 ri4-ye4 fan1-tu[2] zhao3 cong2-gian2 liu2-xia4 de0 zhu1-bao3
(( )f ( )f ( )f)P (( ( )f)f' (( )f )f' ( )f)P
pirate day-and-night turn the map/ground look for previous left treasures

Underlying Tone 2: 'The pirate checks the map day and night looking for old treasures'
Underlying Tone 3: 'The pirate digs the ground day and night looking for old treasures'

Each subject listened to 36 sentences, with six sentences in each of the four experimental conditions, and twelve unambiguous filler sentences. Materials were ordered across the four tapes via a Latin square such that each subject heard only one version of

TABLE 4

Experiment 2: Percent of Sandhi vs. No Sandhi responses by metrical closeness condition

| *CONDITION* | *Sandhi* | *No Sandhi* | *Errors/missing* |
|---|---|---|---|
| Mandatory | 80 | 18 | 2 |
| Often | 58 | 41 | 1 |
| Sometimes | 44 | 55 | 1 |
| Rare | 20 | 78 | 2 |

each sentence, but across subjects all 24 experimental sentences were presented in all four conditions.

*Procedure.* The sentences were spoken by the same female native speaker of Mandarin Chinese and recorded. The subjects, again seated in a quiet room away from the experimenter, listened to the sentences over headphones. A warning tone was presented 300 msec before each sentence. Subjects were given an answer sheet that had, for each Mandarin sentence, an English gloss of the two meanings of the ambiguous Mandarin word. Subjects were allowed 4 seconds after each sentence to circle the word or phrase that most closely corresponded to the meaning of a word or phrase that they had just heard. Subjects were told that if none of the words in the auditory sentence corresponded to the written English words for that trial, they should simply skip that item and wait for the next auditory sentence.

*Results and discussion*

Table 4 shows the percentage of responses in each of the four prosodic closeness conditions. As predicted, subjects were most likely to choose the English item that corresponded to an underlying Tone 3 word in condition 1 (Mandatory). As shown in the table, subjects perceived the lexical item which had been transformed by the tone sandhi rule 80% of the time when the Tone 2–Tone 3 words were closely related in the prosodic tree. Subjects were least likely to choose the English word that corresponded to an underlying Tone 3 word in condition 4 (Rare), where the theory of prosodic distance predicts that the tone sandhi rule would not apply. (They did so only 20% of the time.) Conditions 2 and 3 fell appropriately in between: In condition 2 (Often), subjects were slightly more likely to perceive that the tone sandhi rule had applied than they were in condition 3 (Sometimes) (58% vs. 44%).

Analyses of variance with subjects as the random variable showed no effect of order of presentation ($F(3, 16) = 2.15$, $p < 0.13$), and no effect of subjects' nationality

TABLE 5

Experiment 2: Duration (msec) and fundamental frequency ($F_0$) values (Hz) for ambiguous words by condition

| | *Mandatory* | *Often* | *Sometimes* | *Rare* |
|---|---|---|---|---|
| Duration (msec) | 267 | 249 | 241 | 245 |
| Lowest $F_0$ (Hz) | 210 | 200 | 194 | 195 |
| Ending $F_0$ (Hz) | 245 | 228 | 217 | 218 |
| Change in $F_0$ (Hz) | 35 | 28 | 23 | 23 |

($F(1, 18) < 1$), but a significant effect of the prosodic closeness variable ($F(3, 48) = 68.29$, $p < 0.0001$). An analysis of variance with items as the random variable also showed a significant effect of metrical closeness condition ($F(3, 69) = 26.17$, $p < 0.0001$). Planned orthogonal comparisons showed all differences among the means to be statistically significant ($p < 0.01$).

For the ambiguous morphemes, the mean log frequency of occurrence of the associated Tone 2 character was 1.91, while the mean log frequency of occurrence of the associated Tone 3 character was 2.03 (Suen, 1979). The slightly higher frequency of occurrence of the characters associated with the Tone 3 meaning of the ambiguous words may explain why some subjects continued to choose the English item associated with the Tone 3 word even in condition 4 (Rare), where the application of the tone sandhi rule is likely to be blocked. The relative frequency of the meanings of the ambiguous words did not, of course, differentially influence the results of Experiment 2, since the same ambiguous items occurred across all four conditions.

*Acoustic analysis.* Table 5 shows summary data from an acoustic analysis of the 24 Tone [2] items in each of the four prosodically-determined sandhi conditions: Mandatory, Often, Sometimes, and Rare. Four parameters, generally taken as the acoustic indices of tone (cf. Dow, 1972; Ho, 1976; Howie, 1976) were evaluated for each ambiguous word. These were: 1) the shape of the $F_0$ contour across the word; 2) duration (word duration and the duration of any pause following the critical word were measured separately); 3) lowest fundamental frequency ($F_0$); and 4) ending $F_0$. Visual inspection of the $F_0$ coutour for each Tone [2] word showed that the pitch shapes were clearly those of a rising tone for all items in all conditions; in fact the smallest mean increase in $F_0$ (for the Rare condition) was 23 Hz. The greatest rise in $F_0$ appeared in the Mandatory condition (35 Hz), that is, the Tone 2 shape was the least like the low, falling Tone 3 shape in the condition where subjects most frequently reported hearing a Tone 3 word. As Table 5 shows, the Tone [2] words were longest in the Mandatory

conditions (267 msec), and word duration decreased as the prosodic closeness between the Tone 2 and Tone 3 words decreased.[9] In duration, the Tone [2] words in the Mandatory condition were also the least like a Tone 3 word: When Tone 3 appears in non-pre-pausal positions, it is the shortest of the tones (Shih, 1988). Like duration, lowest $F_0$, ending $F_0$, and change in $F_0$ decreased as the prosodic closeness between the Tone 2 and Tone 3 words decreased. For 13 of the 24 items, in each of the four conditions, there was no pause between the Tone [2] word and the following Tone 3 word. The remaining 11 items did contain a pause between the Tone [2] word and the Tone 3 word. However, there were no differences in pause duration across the four conditions ($F < 1$). In these items, the Tone 3 word began with a stop consonant, and it is likely that the measurable between-word pause occurred due to the stop closure. A multivariate analysis of variance showed a significant effect of the prosodic closeness variable; univariate protected F tests with alpha levels adjusted according to the Bonferroni procedure indicated that prosodic closeness significantly influenced the duration ($F(3, 69) = 7.58$, $p < 0.001$), lowest $F_0$ ($F(3, 69) = 7.68$, $p < 0.001$), and ending $F_0$ ($F(3, 69) = 12.37$, $p < 0.001$) of the ambiguous Tone [2] words. *Posthoc* comparisons of the means using a Scheffé-Bonferroni decision rule (Bird, 1975) indicated that for all three measures, significant differences ($\alpha = 0.05/3$) occurred between the Mandatory condition and the combined Sometimes and Rare conditions. These indications of decreasing individual word length, a decline in absolute $F_0$, and a narrowing in the range of $F_0$ values are consistent with those to be expected due to declination and catathesis ("down-stepping") as the position of the critical word moves farther from the beginning of the phrase.

The pattern of results found in this acoustic analysis indicates that, although there were significant physical differences between the Tone [2] words across the four conditions, these differences should be attributed to the natural declination and catathesis commonly associated with the production of Mandarin phrases of increasing length (cf. Connell, Hogan, and Rozsypal, 1983; Gårding, 1987; Shih, 1988). The differences in duration and $F_0$ for the ambiguous words did *not* render the surface Tone [2] in the Mandatory condition somehow more similar to a surface Tone 3, while leaving the surface Tone [2] in the corresponding Rare condition more similar to a surface Tone 2. On the contrary, the acoustic analysis shows that the critical words most

[9] The progressively decreasing length of individual words across the four conditions may indicate a corresponding increase in the spoken tempo of the phrases. However, this is not problematic for our interpretation. According to most theories (e.g., Cheng, 1970; Shih, 1986) the domain of application of the tone sandhi rule increases in length (spans more and more words) as rate of speech increases, and the relative length of individual words correspondingly decreases (see also footnote 2). An explanation that relied on speech rate alone, then, would predict that the tone sandhi rule would be *least* likely to apply in the Mandatory condition, where the longest words appear. Our findings, of course, are opposite this prediction: Our results show that tone sandhi was *most* likely to apply in the Mandatory condition, where word length suggests a slower speaking rate, which would correspond to a decreased span of application for the tone sandhi rule.

likely to be judged by subjects to have an underlying low tone (Tone 3) were those that had the longest, steepest, most obviously rising tone (Tone 2) on the surface. Furthermore, the critical words least likely to be judged by subjects to have an underlying low Tone 3 were those that had the shorter, flatter (more Tone 3-like) shape on the surface.

## General Discussion

The results from Experiment 1 indicate that subjects overwhelmingly report the linguistically determined underlying tone rather than the acoustically determined surface tone. This result is consistent with previous perception studies (Wang and Li, 1967; Zue, 1976). The tone reports did not simply rely on lexically stored tonal information, because subjects reported Tone 3 when the sandhi environment occurred in nonsense strings, where the tonal context provided the only clue to the potential operation of the tone sandhi rule. This pattern of results suggests that listeners use a representation of speech that incorporates rule-based phonological structure.

Experiment 2 further determined the nature of the phonological representation used in phrase and sentence processing. Subjects relied on the prosodic structure of the phrases to determine whether or not the tone sandhi rule had applied and consequently to identify individual lexical items. Note that although the global prosodic structure was manipulated to create different degrees of prosodic closeness between the two critical items in the tone sandhi environment, the phrase level syntactic relations between the two items did not vary. Thus, when syntax is held constant, prosodic structure determines the domain of application of the tone sandhi rule for comprehenders of Standard Mandarin Chinese. In future experiments, factorial manipulation of syntactic and prosodic structures may allow further clarification of the respective contributions of syntactic and prosodic structures to the ambiguity resolution process.

The acoustic analyses showed that differences in the fundamental frequency and duration of the word to which the tone sandhi rule had applied were those due to intonational phrasing. There was no indication of any word-level 'cueing' of rule application, such as similarity of the surface and underlying tones. In sum, the evidence is highly suggestive and encouraging that phrase-level prosodic structure plays a critical role in auditory language comprehension processes.

*(Received April 27, 1989; accepted March 7, 1990)*

## References

BECKMAN, M., and EDWARDS, J. (in press). Lengthenings and shortenings and the nature of prosodic constituency. In M. Beckman and J. Kingston (eds.), *Papers in Laboratory Phonology*. Cambridge, U.K.: Cambridge University Press.

BIRD, K.D. (1975). Simultaneous contrast testing procedures for multivariate experiments. *Multivariate Behavioral Research*, **10**, 343–351.

CARROLL, P.J., and SLOWIACZEK, M. (1987). Modes and modules: Multiple pathways to the language processor. In J. Garfield (ed.), *Modularity in Knowledge Representation and Natural Language Understanding* (pp. 221–247). New York: Academic Press.

CHAO, Y.R. (1968). *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

CHENG, C.C. (1968). *Mandarin Phonology.* Doctoral dissertation, University of Illinois at Urbana.

CHENG, C.C. (1970). Domains of phonological rule application. In J.M. Sadock and A.L. Vanek (eds.), *Studies Presented to Robert B. Lees by His Students* (pp. 39–59). Edmonton: Linguistic Research.

CHENG, C.C. (1973). *A Synchronic Phonology of Mandarin Chinese.* The Hague: Mouton.

COOPER, W., EGIDO, C., and PACCIA, J.M. (1978). Grammatical control of a phonological rule: Palatalization. *Journal of Experimental Psychology: Human Perception and Performance,* **4,** 264–272.

COOPER, W., and SORENSEN, J. (1977). Fundamental frequency contours and syntactic boundaries. *Journal of the Acoustical Society of America,* **62,** 683–692.

CONNELL, B.A., HOGAN, J.T., and ROZSYPAL, A.J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics,* **11,** 337–351.

DOW, F. (1972). *An Outline of Mandarin Phonetics.* Canberra: Australian National University Press.

GÅRDING, E. (1987). Speech act and tonal pattern in standard Chinese: Constancy and variation. *Phonetica,* **44,** 13–29.

HO, A.T. (1976). The acoustic variation of Mandarin tones. *Phonetica,* **33,** 353–367.

HOWIE, J.M. (1976). *Acoustical Studies of Mandarin Vowels and Tones.* Cambridge, U.K.: Cambridge University Press.

HUNG, T. (1987). *Syntactic and Semantic Aspects of Chinese Tone Sandhi.* Doctoral dissertation, University of California at San Diego.

KAISSE, E.M. (1985). *Connected Speech: The Interaction of Syntax and Phonology.* New York: Academic Press.

KLATT, D., and COOPER, W.E. (1975). Perception of segment duration in sentence contexts. In A. Cohen and S.G. Nooteboom (eds.), *Structure and Process in Speech Perception* (pp. 69–89). Heidelberg: Springer-Verlag.

LEHISTE, I., OLIVE, J., and STREETER, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America,* **60,** 1199–1202.

NESPOR, M., and VOGEL, I. (1986). *Prosodic Phonology.* Dordrecht, Holland: Foris Publications.

PIERREHUMBERT, J., and BECKMAN, M. (1988). *Japanese Tone Structure.* Cambridge, MA: MIT Press.

SELKIRK, E.O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic Prosody II* (pp. 47–68). Trondheim, Norway: TAPIR.

SELKIRK, E.O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and M.-L. Kean (eds.), *Juncture* (pp. 24–42). Saratoga, CA: Arma Libri.

SELKIRK, E.O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure.* Cambridge, MA: MIT Press.

SHIH, C. (1986). *The Prosodic Domain of Tone Sandhi in Mandarin Chinese.* Doctoral dissertation, University of California at San Diego.

SHIH, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory,* **3,** 83–109 (Ithaca, NY).

SLOWIACZEK, M. (1981). *Prosodic Units as Processing Units.* Doctoral dissertation, University of Massachusetts at Amherst.

SORENSEN, J., and COOPER, W. (1980). Syntactic coding of fundamental frequency in speech production. In R.A. Cole (ed.), *Perception and Production of Fluent Speech* (pp. 399–440). Hillsdale, NJ: Lawrence Erlbaum.

STREETER, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America,* **64,** 1582–1592.

SUEN, C.Y. (1979). *Computational Analysis of Mandarin.* Boston, MA: Birkauser Verlag.

WALES, R., and TONER, H. (1979). Intonation and ambiguity. In W.E. Cooper and E.C.T. Walker (eds.), *Sentence processing: Psycholinguistic Studies Presented to Merrill Garrett* (pp. 135–158). Hillsdale, NJ: Lawrence Erlbaum.

WANG, W., and LI, K. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, **10**, 629–636.

ZHANG, Z.S. (1988). *Tone and Tone Sandhi in Chinese.* Doctoral dissertation, The Ohio State University.

ZUE, V.W. (1976). Some perceptual experiments on the Mandarin tones. Paper presented at the 92nd Meeting of the Acoustical Society of America, San Diego, California.