# THE RELATIONSHIP BETWEEN RHYTHM AND THE TONE SANDHI DOMAIN IN MANDARIN

*J. Wang and R.H. Mannell*
*Speech, Hearing and Language Research Centre*
*Macquarie University, Sydney, Australia*

## ABSTRACT

In this paper, we develop the notion that the relationship between prosody and the tone sandhi domain in Mandarin is rhythm-based, determined by the strong rhythmic tendency to group words into disyllabic and trisyllabic units. Three experiments were conducted to test the relationship between rhythmic grouping and tone sandhi, as well as the effect of speech rate and focus position.

## INTRODUCTION

Tone sandhi (TS) of the 3rd tone in Mandarin has long been of phonological and phonetic interest particularly because of its relationship to other prosodic features, and to syntactic structures. With the aim of developing rules to predict the TS domain, phonological studies have focused on syntactic, metrical and prosodic approaches and there is a general consensus that TS domain is determined by prosodic structures [1] [4] [7]. Acoustic phonetic studies have also been carried out on tones and tone sandhi in different tonal and prosodic contexts [3].

Prosodic approaches to TS, have tended to suggest that the TS domain is based on the prosodic grouping of words, with the units of grouping being disyllabic feet, super feet, and phrases [4], but such approaches (eg. *Foot Formation Rules*), have limited predictive power as little is known about how syllables and words are grouped in different prosodic contexts.

In this study we examine the following hypotheses: (1) the prosodic conditioning of TS is rhythm-based and is strongly related to the tendency of speakers to group words into disyllabic and trisyllabic units; (2) in most situations, word-level rhythmic grouping is relatively stable, although changing speech rate and focus position could, in certain circumstances, affect rhythmic structures, leading to variations in the TS domain.

## RHYTHMIC GROUPING AND THE TONE SANDHI DOMAIN

In connected Mandarin speech, syllables and words are grouped into rhythmic units. The predominant rhythmic tendency is to use disyllabic constituents, either lexical words or the combination of two monosyllabic words. This tendency implies that monosyllabic words tend not to stand as independent prosodic (rhythmic) units. As a result, monosyllabic lexical words are normally grouped with neighbouring words to form basic rhythmic units. In fluent, natural speech, priority is given to grouping monosyllabic words with neighbouring monosyllabic words, then disyllabic words and, finally, polysyllabic words with more than two syllables — unless there is good reason not to, such as pausing. The constituents resulting from this type of grouping are basically disyllabic or trisyllabic independent prosodic units. We prefer to call them prosodic words (PW), rather than feet, since the term "foot" is traditionally related to the alternation of stressed and unstressed syllables. Lexical words (rather than syllables) are the basic elements of this type of rhythmic grouping, and there are three main factors determining the formation of prosodic words: a syllable count of each lexical word; a syllable count of each of its neighbours; and the speaker's desire to form a balanced rhythmic structure. A syllable count of each word determines if it is obligatory to group the word with neighbouring words, while syllable counts of neighbouring words influence the grouping priority (the way words are grouped to form balanced rhythmic structures). In fluent speech with a normal speech rate, for example, the rhythmic structures 2/2, 3/2, 3/3 are likely in the following sentences:

*a.* laoli / **mai** gou.
*b.* laoli **mai** / xiaogou.
*c.* laoli **mai** / xiaomugou.

Here, the monosyllable *mai* is grouped with either the ensuing word *gou* or the preceding word *laoli*, because of the speaker's desire to form balanced rhythmic structures, rather than being determined by syntactic branch directions or semantically related elements. In practice, the monosyllabic verb *mai* can be grouped with either *xiaogou* or *laoli* in example *b.* without affecting the rhythmic balance of the sentence, although grouping it with *laoli* is usually preferred. However, for sentence c, it is virtually mandatory in normal speech to group *mai* with *laoli*, as 3/3 is a more balanced structure than 2/4.

Tone sandhi for the 3rd tone in Mandarin is motivated by dissimilating adjacent low tones and tends to operate within units of speech planning [1]. Our research indicates that speech planning units are created from rhythmic groupings rather than *syntactically or semantically related elements.* Although Foot Formation Rules work well in most situations, and were originally designed to *modify syntactic structure and produce rhythmically balanced structures for TS to operate on* [4], limitations arise because they refer to conditions for syntactic branch direction, rather than to rhythmic groupings.

If the TS domain is identical to the rhythmic grouping of words, as we have previously argued, then factors which influence rhythmic groupings must also affect the TS domain (such as speech rate, focus stress and pauses), since the surface TS patterns should reflect the underlying groupings. Therefore, close examination of the effects of these factors on the TS domain should further clarify how words are rhythmically grouped in Mandarin speech. There has been a lot phonological discussion on the basic TS domain and variations in the TS domain, due to speech rate and contrastive stress [4] [7], but there has been little acoustic investigation carried out to date [6]. In this study, three related experiments were carried out to examine, both aurally and acoustically, the relationship between rhythmic groupings and the TS domain, as well as the effect of speech rates and focus position on rhythmic groupings and TS domain.

## MATERIALS AND PROCEDURE

12 sentences, comprising of only 3rd tone syllables, were read by four female Mandarin speakers from Beijing (XY, LP, LL, WJ). The subjects were asked to read the sentences naturally with three different speech rates, two or three times initially, and then according to specific contexts that required changing contrastive stress positions across syllables and words. All the sentences were recorded in studio conditions using DATs and the speech samples were then digitised onto a Sun computer at a sample rate of 20 kHz, and manually segmented and labelled using the Waves program. Relevant duration data and pitch contours were extracted and calculated using the mu+ system [2].

a1 wo xiang mai gou.
a2 wo xiang mai xiaogou.
a3 wo xiang mai xiaomugou.
a4 wo mai gou.
a5 wo mai xiaogou.
a6 wo mai xiaomugou.
b1 laoli xiang mai gou.
b2 laoli xiang mai xiaogou.
b3 laoli xiang mai xiaomugou.
b4 laoli mai gou.
b5 laoli mai xiaogou.
b6 laoli mai xiaomugou.

## RESULT 1: BASIC GROUPING

In this experiment, 12 sentences were read, fluently and naturally, three times by each of the subjects. The following table

shows the the TS domain. The results are relatively consistent, with identical PWs used by each of the four subjects. An analysis was carried out on the TS patterns with each syllable's tone being identified aurally. Rhythmic structures (prosodic word boundaries) were determined according to the prosodic rules outlined earlier. In the following table, 'R' refers to TS occurring with the syllable's tone being transformed into a rising tone, while 'L' refers to the syllable keeping its underlying tone. The slash '/' indicates PW boundaries.

| | |
|---|---|
| a1 RL/RL | b1 RL/RRL |
| a2 RRL/RL | b2 RL/RL/RL |
| a3 RRL/RRL | b3 RL/RL/RRL |
| a4 RRL | b4 RL/RL |
| a5 RL/RL | b5 RRL/RL |
| a6 RL/RRL | b6 RRL/RRL |

Each sentence was read three times by each of the subjects and some idiosyncratic variations of TS were observed. These variations can be categorised into three types: Type I — using alternate TS patterns, such as LRL instead of RRL in trisyllabic PWs with 1+2 syntactic structures, like *(xiao(mu(gou)))* in b3; Type II — joining two PWs together, so that RRRL occurs instead of RLRL, as in b4; Type III — grouping monosyllabic words in different ways in certain contexts, such as in b1 and b5 where the rhythm could be either 2/3 or 3/2. Our results indicate that there are conventional rhythmic structures for each of the sentences but there is also a certain degree of individual freedom.

## RESULT 2: SPEECH RATE

We also asked the subjects to read each of the sample sentences at different speech rates for purposes of testing the effect of speech rates on the TS domain and PW grouping. The speech rates (syllables per second) are, as one would expect, slightly different across the four subjects.

The following table gives the average speech rates for each of the four speakers at slow, medium and fast speech rates.

| | XY | LP | LL | WJ |
|---|---|---|---|---|
| medium | 4.5 | 4.1 | 4.4 | 4.8 |
| fast | 5.4 | 6.5 | 5.2 | 6.2 |
| slow | 3.2 | 3.1 | 3.3 | 3.7 |

We found that TS domains in the majority of the sentences were reasonably consistent for the various speech rates. As was the case in the previous section, three types of variations were observed. Type I occurred more frequently in slow speech, while type II was more common in fast speech. Type III had no definite correlation with speech rate. While it appears that more tonal transformations occur in fast speech than in slow speech, either through using alternate TS patterns or by grouping two PWs together, the basic rhythmic groupings remain relatively stable, regardless of speech rate.

## RESULT 3: FOCUS POSITION

The position of the accent has been reported to influence the TS domain as TS always starts from accented syllables [7]. In this experiment, the location of contrastive accents is designed to change across syllables and words according to different contexts, and subjects were asked to read a dialogue in which they could stress appropriate syllables or words.

Q: Ni xiang mai xiaogou ma?
(Do you want to sell that small dog?)
A: Bu, wo xiang **mai** xiaogou.
(No, I want to buy a small dog.)

There are 54 contexts for the 12 test sentences in which the location of contrastive accent varies. Our results showed that changes in TS domain do occur according to the position of accented words or syllables in certain contexts, but that it is extremely difficult to determine whether these changes are accent-conditioned. First, TS domains in most of the sentences are quite stable, regardless of the location of accented syllables. For example, TS domain in b4, b5, b6 is always 2/2, 3/2, 3/3, regardless of whether *mai* is accented or not.

We then looked at prosodic contexts which caused rhythmic regrouping and found there were only two examples that consistently occurred across each of the four subjects. The normal TS domain of b2 is RL/RRL, but RRL/RL when *mai* is accented; while in a1 it is normally RL/RL, but L/RRL when *xiang* is accented. If TS domain changes are caused by different planning groupings, it seems to be associated with not only the position of accented syllables but also PW structures. It was found that rhythmic regroupings only occurred in sentences with more than two successive monosyllabic words and accented-related regrouping never split polysyllabic lexical words into two separate PWs. For example, *xiang* can be grouped with *laoli* in b4, if it is necessary to accent *mai* using a full rising tone, but this is not the case in b5 when *xiao* is accented. There is a basic distinction between PWs consisting of only one lexical word and those with one or more words. The latter have a certain flexibility to be grouped in different ways, which is not only illustrated in the TS domain, but also in the acoustic data. For example, the syllable duration of *gou* (see the table below) not only varies according to whether it is accented or unaccented but also according to whether it is a monosyllabic word or the last syllable of a polysyllabic word. (In this table, 1 is *gou* in a monosyllabic, 2 in a disyllabic, and 3 in a trisyllabic word. "a1" etc. are sentence numbers. "238" etc. are durations in ms).

| | unaccented | accented |
|---|---|---|
| 1 (a1, a4, b1, b4) | 238 | 354 |
| 2 (a2, a5, b2, b5) | 220 | 297 |
| 3 (a3, a6, b3, b6) | 220 | 295 |

When *gou* is accented as a monosyllabic word, the degree of lengthening is significantly increased. This implies that monosyllabic words, as meaningful units, have a relatively loose relationship with neighbouring words within PWs. This may explain why accented words can affect rhythmic structures and TS domains.



## CONCLUSION

This study has attempted to combine phonological and phonetic approaches to the TS domain within a rhythmic framework. We have found that basic rhythmic structures are relatively stable, but that certain spontaneous factors can influence grouping. Relatively stable rhythmic structures would appear, therefore, to be the logical starting point for developing phonological rules for predicting the TS domain, but the types of spontaneous factors which can occur make the perfection of these rules extremely difficult. This rhythmic framework has been tentatively applied in the hierarchical labelling of rhythmic units in a Mandarin speech database aimed at establishing a prosodic model for Mandarin speech synthesis [5]. However, at this stage, further acoustic investigation of rhythmic grouping in Mandarin speech is still required.

## REFERENCES

[1] Chen, M. (1990), "What must phonology know about syntax", in *The Phonology-Syntax Connection*, S. Inkelao and D.Zek (eds.), U.Chicago Press, 19-46.
[2] Harrington, J., Cassidy, S., Fletcher, J., McVeigh, A. (1993), "The mu+ system for corpus based speech research", *Computer Speech and Language*, 7, 305-331.
[3] Speer, S.R., Shih, C.L., Slowiaczek, M.L. (1989), "Prosodic structure in language understanding: evidence from tone sandhi in Mandarin", *Language and Speech*, 32(4), 337-354.
[4] Shih, C. L. (1986), *The prosodic domain of tone sandhi in Chinese*, PhD dissertation, Univ.of California, San Diego.
[5] Wang, J. (1994), "Syllable duration in Mandarin", *Proc. 5th Australian Int. Conf. Speech Science and Technology*, 322-327.
[6] Zee, E. (1980), "A spectrographic investigation of Mandarin tone sandhi", *UCLA Working Papers Phonetics*, 98-116.
[7] Zhang, Z. S. (1988), *Tone and tone sandhi in Chinese*, PhD dis., Ohio State U.