Acoust. Sci. & Tech. **24**, 3 (2003)

PAPER

Prosodic variations in disyllabic meaningful words focused with different stress patterns in Mandarin Chinese

Zhenglai Gu[1,*], Hiroki Mori[2,†] and Hideki Kasuya[2,‡]

[1] *Graduate School of Engineering, Utsunomiya University, 7–1–2, Yoto, Utsunomiya, 321–8585 Japan*
[2] *Faculty of Engineering, Utsunomiya University, 7–1–2, Yoto, Utsunomiya, 321–8585 Japan*

(*Received 8 March 2002, Accepted for publication 14 November 2002*)

Abstract: Focus plays a vital role in spoken communication. Unlike some languages in which word-level accent is assigned lexically, most Chinese words have no lexical constraint on accentuation at the word level. Thus, if a disyllabic word is to be focused on, Chinese speakers must decide which syllable, or indeed if both syllables, should be stressed. Moreover, the stress assignment of the focal words is affected by the semantic structure of the words, which are composed of meaningful characters. In this paper, the manner in which disyllabic focused words affect variations in the prosodic correlates is investigated, in terms of the semantic structures, stress patterns and tonal combinations of the words in Mandarin Chinese. The findings are summarized as follows. (1) The effects of semantic structures of the focused words on the prosodic correlates are completely mediated by the stress patterns. (2) The prosodic correlates of the second syllable of the focused word are more variable than those of the first syllable. Specifically, there is a significant asymmetry of vowel duration as well as fundamental frequency (F_0) range between the pre-stressed and post-stressed syllables. Moreover, the tonal combination significantly affects the variations of both the vowel duration and F_0 range.

Keywords: Focus, Stress pattern, Semantic structure, F_0 range, Duration

PACS number: 43.70.Fq, 43.72.Ja

1. INTRODUCTION

In Chinese, a monosyllabic word is composed of one character that conveys semantic information. The meaning of polysyllabic words is generally derived from the combination of these monosyllabic characters, as in the word "gang1 cai2 (steel material)," which consists of the characters "gang1 (steel)" and "cai2 (material)." Moreover, unlike other languages in which word-level accent is assigned lexically, as in the Japanese examples of "a'me" (rain) and "ame" (candy), most Chinese words have no lexical constraint on accentuation at the word level. Thus, if a disyllabic word is to be focused on, Chinese speakers must decide which syllable, or indeed if both syllables, should be stressed. This stress assignment can be particularly confusing to second language learners when speaking Chinese, for they tend simply to place stress accent on all syllables of focused words. Although it is acceptable to apply this simple rule in certain circumstances, such as in the reading of a government report in which very prominent stress is required, and in reading aloud by young primary school students, it is not appropriate to communicate in this manner in daily life. The stress patterns of disyllabic Chinese words are defined conventionally as follows [1]: stress placement on the first syllable (Sp10), the second syllable (Sp01) and on both syllables (Sp11). Hereafter, Sp10 and Sp01 are referred to as "single stress" and Sp11 as "double stress" patterns.

It is widely held that the focally stressed syllable is accompanied by the changes of fundamental frequency (F_0), duration, intensity, formant and spectral tilt [2–8]. F_0 movements, moreover, are generally seen as the reliable prosodic cues to focus. As a tonal language, the F_0 movements in Mandarin Chinese are generally described on the basis of a top- and bottom-line intonation model [9,10]. A study showed further that the F_0 movements of maxima are significant and relatively free, but those of minima are limited by the bottom-line, which results in the expansion or compression of the local F_0 range.

*e-mail: gzl@klab.jp
†e-mail: hiroki@klab.jp
‡e-mail: kasuya@klab.jp

There have been several reports on the nature of simpler double-stressed focal words and their surroundings in very short sentences, including: (1) a radical asymmetry of fundamental frequency (F_0) range occurs around the focused word [7,8], i.e., the F_0 range of on-focus words is significantly increased and the F_0 range of post-focus words is decreased, while the F_0 range of the pre-focus words is little changed; and (2) a remarkable increase in duration of on-focus words is observed, which is associated with small changes in the duration of both the pre- and post-focus syllables [8], a phenomenon we call symmetry of duration. Despite these accumulated findings, however, little attention has been paid to the nature of single-stressed focal words, which are fundamental components of fluent Chinese speech. Meanwhile, it has been shown that the assignment of stress patterns is strongly affected by the semantic structures of the focal words [11]. We therefore conjecture that the stress patterns function as a mediator between semantic structures and prosodic correlates of the focal words.

The present study attempts to answer the following questions: (1) what is the relationship between semantic structures, stress patterns, and prosodic correlates of the focal words in Mandarin Chinese, and specifically, are the effects of semantic structures on the prosodic correlates completely mediated by the stress patterns? (2) How do the three aforementioned stress patterns and the tone combinations affect the prosodic correlates?

We deal with disyllabic words in this paper, because not only are they most frequently used, but also they are the fundamental constituents of other polysyllabic words of three syllables or more.

2. SEMANTIC STRUCTURES AND CLASSIFICATION OF WORDS

We classify disyllabic words of Chinese into the following four categories according to different semantic structures:

Category 1 (C1): words in which the contribution of the first syllable is more important than that of the second syllable to the meaning of the word.

Category 2 (C2): words in which the contribution of the second syllable is more important than that of the first syllable to the meaning of the word.

Category 3 (C3): words in which both the first and second syllables contribute equally to the meaning of the word.

Category 4 (C4): words in which the relative semantic weight is subject to context.

In a previous study [11], we found that Chinese speakers tend to retain stress positions on the sense-contributing syllables of the words (Table 1). The statistical results of CCTV news speech have shown that (1) focused words of C1 and C2 are conventionally accompanied by stress patterns Sp10 and Sp01, respectively; (2) C3 words are often focused with stress pattern Sp10, but other stress patterns are preferred depending on the prosodic contextual factors, such as phrase finality; (3) the assignment of the stress patterns of C4 words is somewhat complicated, depending on whether the focus is narrow or broad [2]. In the case of narrow focus, the stress is assigned to the narrow focused syllable. In the case of broad focus, the stress pattern Sp11 is most frequently used, while the other stress patterns are used depending on the prosodic contextual factors. By and large, it is obvious that the semantic structures of a focused word strongly influence the assignment of its stress pattern. Thus, the issue of the relationship between semantic structures, stress patterns and prosodic correlates of the focused words depends on whether or not the effects of semantic structures of the focused words on the prosodic correlates are completely mediated by the stress patterns.

Table 1 Stress patterns for each category of semantic structure.

Cat.	Example	English translation	Stress patterns*		
			Sp10	Sp01	Sp11
C1	Zhuan1Yi1	Concentrated	Yes		
C2	Yi2Ding4	Must		Yes	
C3	Gang1Cai2	Just	Yes	Yes	Yes
C4	Gang1Cai2	Steel material	Yes	Yes	Yes

*Note that no entry refers to unnatural stress placement on the specific syllable(s).

3. EXPERIMENT 1

3.1. Materials and Analytical Procedure

Four pairs of disyllabic words were selected from different semantic categories as target words (Table 2). Each pair of words had the same tone combination and the same or similar phonemic structure, but a different semantic structure. To control the stress patterns of the focal word, we used the carrier sentence, "Zhe4 shi4 <u>TW</u> er2 bu2 shi4 <u>CW</u>. (This is <u>TW</u> but not <u>CW</u>.)," into which

Table 2 Target words for Experiment 1.

Word pairs (A, B)		English translation	Cat.
Gang1gou3	A	Congo	C3
Gan1gou3	B	Dried fruit	C4
Yi2ding4	A	Must	C2
Yi2ding4	B	Tablet	C4
Gang1cai2	A	Just	C3
Gang1cai2	B	Steel material	C4
Zhuan1yi1	A	Concentrated	C1
Zhong1yi1	B	Traditional doctor	C4

target and contrastive words were inserted for all four pairs of disyllabic words.

For example, given the word "gang1 cai2" ("steel material"), we have the following four derivatives.
"Zhe4 shi4 gang1 cai2 er2 bu2 shi4 mu4 cai2." ("This is steel material but not wood.")
"Zhe4 shi4 gang1 cai2 er2 bu2 shi4 gang1 chan3 ping3." ("This is steel material but not a steel product.")
"Zhe4 shi4 gang1 cai2 er2 bu2 shi4 mu4 chan3." ("This is steel material but not a wood product.")
"Zhe4 shi4 gang1 cai2 er2 bu2 shi4 qi2 ta1." ("This is steel material but not anything else.")

In the above sentences, the stressed syllables are underlined. We also made the speakers aware that sentences with target words of C1 and C2 are used in a special context as "metalinguistic correction" [2], because words of C1 and C2 are conventionally focused with stress patterns of Sp10 and Sp01, respectively. Five native Chinese speakers born in Beijing, two males and three females, read one sentence twice. All of the utterances were recorded on DAT in a soundproof room before being digitized at a sampling rate of 11.025 kHz. The syllable and vowel segmentation was carried out manually by observing the waveform and sound spectrogram of an utterance. F_0 contours in semitone (ref. 1 Hz) were automatically analyzed at 5-ms intervals and erroneous F_0 values, if any, were corrected manually. The largest and smallest F_0 values of the F_0 contour of the vowel segment of a syllable were extracted as the F_0 maximum and F_0 minimum values of the syllable, respectively. The F_0 range was defined as the difference between the two extreme values.

3.2. Results

Table 3 shows the mean ratios of the vowel duration of the disyllabic words in the focused utterance to that in the neutral utterance, and the results of the one-way ANOVA. Table 4 is the same as Table 3 except that the mean differences are taken for the F_0 range. In addition, we performed a two-way ANOVA, with "Word A" and "Word B" as a between-item factor, and the stress patterns as a within-item factor.

The one-way ANOVA found that neither vowel duration ($F < 1$) nor F_0 range ($F < 1$) was affected by the different semantic structures of words A and B focused with the same stress pattern. The two-way ANOVA revealed that: (1) stress patterns significantly affected changes in both vowel duration ($p < 0.05$) and F_0 range ($p < 0.05$), but (2) there were no interaction effects between stress patterns and the pair of words ($F < 1$).

3.3. Discussion

The results show that the prosodic correlates of the words focused with the same stress pattern are insignificantly affected by their semantic structures. Stress patterns do influence prosodic correlates. Thus, by associating with the previous results [11] that the semantic structures strongly affect the assignment of stress patterns as mentioned in section 2, it can be concluded that the effects of semantic structures of the focal words on the prosodic correlates are completely mediated by the stress patterns (Fig. 1).

These findings answer the first question posed in the introduction, and form the foundation for the answer to the

Table 3 Mean ratio of vowel duration of focus utterance to neutral utterance.

Subjects	Stress pattern	Syllable 1 W_A	Syllable 1 W_B	Syllable 1 F	Syllable 1 p	Syllable 2 W_A	Syllable 2 W_B	Syllable 2 F	Syllable 2 p
MZJ	Sp11	1.66	1.67	**0.003**	0.958	1.25	1.29	**0.417**	0.536
	Sp10	1.79	1.76	**0.023**	0.882	0.94	0.95	**0.032**	0.862
	Sp01	1.11	1.19	**0.146**	0.711	1.32	1.41	**0.295**	0.601
MXB	Sp11	1.16	1.15	**0.100**	0.762	1.23	1.49	**0.531**	0.622
	Sp10	1.20	1.19	**0.012**	0.916	0.84	0.85	**0.103**	0.759
	Sp01	1.08	1.06	**0.222**	0.654	1.64	1.47	**0.413**	0.544
FHJ	Sp11	1.55	1.51	**0.244**	0.638	1.59	1.51	**0.486**	0.512
	Sp10	1.64	1.59	**0.151**	0.711	0.87	0.84	**0.381**	0.560
	Sp01	1.16	1.14	**0.118**	0.743	1.82	1.69	**0.376**	0.562
FSY	Sp11	1.11	1.09	**0.618**	0.461	1.11	1.03	**0.832**	0.397
	Sp10	1.16	1.13	**0.708**	0.432	0.91	0.90	**0.020**	0.891
	Sp01	1.09	1.04	**0.344**	0.576	1.21	1.13	**0.264**	0.626
FZM	Sp11	1.29	1.37	**0.372**	0.564	1.31	1.47	**0.207**	0.664
	Sp10	1.38	1.36	**0.014**	0.910	0.74	0.81	**0.124**	0.746
	Sp01	1.16	1.13	**0.321**	0.591	1.53	1.59	**0.101**	0.760

Table 4 Mean difference of F_0 range between focus utterance and neutral utterance.

Subjects	Stress pattern	Syllable 1 W_A	Syllable 1 W_B	Syllable 1 F	Syllable 1 p	Syllable 2 W_A	Syllable 2 W_B	Syllable 2 F	Syllable 2 p
MZJ	Sp11	2.27	1.97	**0.148**	0.709	4.24	3.91	**0.025**	0.877
	Sp10	2.21	2.53	**0.049**	0.831	1.58	1.81	**0.036**	0.854
	Sp01	1.03	1.05	**0.001**	0.974	5.42	4.79	**0.094**	0.766
MXB	Sp11	1.14	1.30	**0.045**	0.838	1.50	1.10	**0.845**	0.393
	Sp10	0.98	1.03	**0.005**	0.949	−0.93	−0.59	**0.106**	0.755
	Sp01	0.54	0.49	**0.037**	0.855	3.35	2.86	**0.137**	0.724
FHJ	Sp11	0.75	0.56	**0.244**	0.638	1.55	1.56	**0.001**	0.991
	Sp10	0.98	0.69	**0.079**	0.783	−0.37	−0.23	**0.019**	0.893
	Sp01	0.48	0.24	**0.918**	0.355	2.86	2.36	**0.153**	0.709
FSY	Sp11	−0.02	0.21	**0.589**	0.472	3.27	2.53	**0.075**	0.793
	Sp10	0.21	0.33	**0.851**	0.391	−0.81	−1.39	**0.291**	0.608
	Sp01	−0.06	−0.03	**0.021**	0.896	3.12	3.10	**0.001**	0.991
FZM	Sp11	0.79	0.49	**0.110**	0.751	1.33	1.67	**0.032**	0.863
	Sp10	0.50	0.87	**0.128**	0.732	−0.02	−0.23	**0.061**	0.813
	Sp01	0.36	0.11	**0.474**	0.517	3.68	3.19	**0.027**	0.876

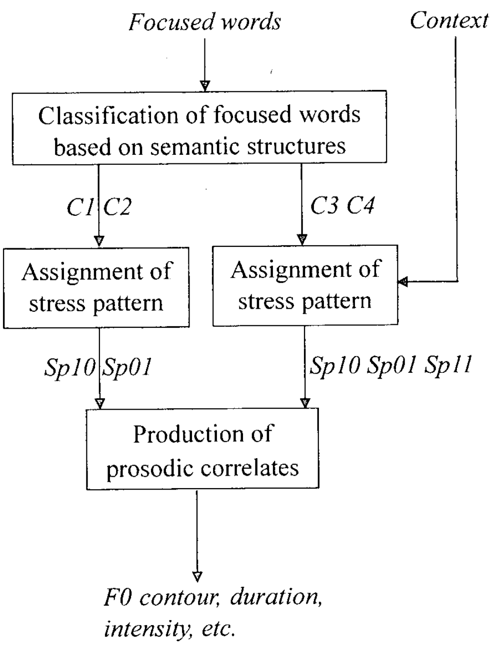


Fig. 1 Relationship between semantic structures of words and prosodic correlates.

second question, because it is reliable to investigate the prosodic correlates of the focal words regardless of what kind of semantic structure they have.

4. EXPERIMENT 2

4.1. Materials and Analytical Procedure

The experimental procedure was the same as that for Experiment 1, except that the target words reflected all 15 tonal combinations (Chinese 4 tones: H, R, L, F), excluding tone L-L (tone L will be changed into tone R when it precedes another tone L). Two words were selected for each tone combination (see the Appendix for a full list of all 30 target words). Thus, 1,200 (= 30 words × 3 stress patterns and 1 neutral × 2 repetitions × 5 speakers) utterances were recorded and prosodic variations were examined through the analysis of vowel duration, F_0 maximum, F_0 minimum, and F_0 range.

The values of the prosodic correlates of each tone combination were averaged across the two target words of the same tone and across both repetitions of the target words in the same sentence uttered with the same stress pattern. Thus, each stress pattern has a total of 75 (= 15 tone combinations × 5 speakers) mean values for the representation of the variation of vowel duration and F_0 range.

4.2. Results

4.2.1. Effects of stress patterns

Vowel Duration

Figure 2 illustrates the distribution of the vowel duration ratios (α_1, α_2), where α_1 and α_2 are the ratios of the vowel duration between focused and neutral utterances for syllable 1 and syllable 2, respectively. The 68% density ellipses for the stress patterns Sp10 and Sp01 are shown in the figure.

As can be seen in Fig. 2, there is a significant asymmetry of vowel duration ratios, because firstly, 97% of Sp10 lie in the left region of the ordinate, which

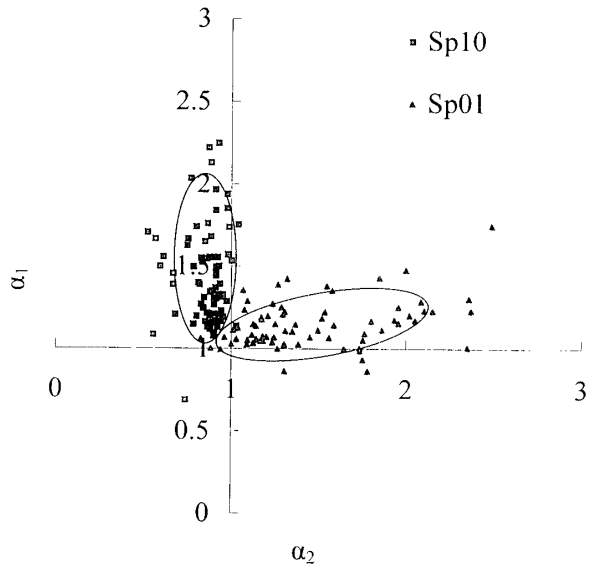


Fig. 2 Distribution of the vowel duration ratio (α_1, α_2), where α_1 and α_2 are the vowel duration ratios between focused and neutral utterances for syllable 1 and syllable 2, respectively.

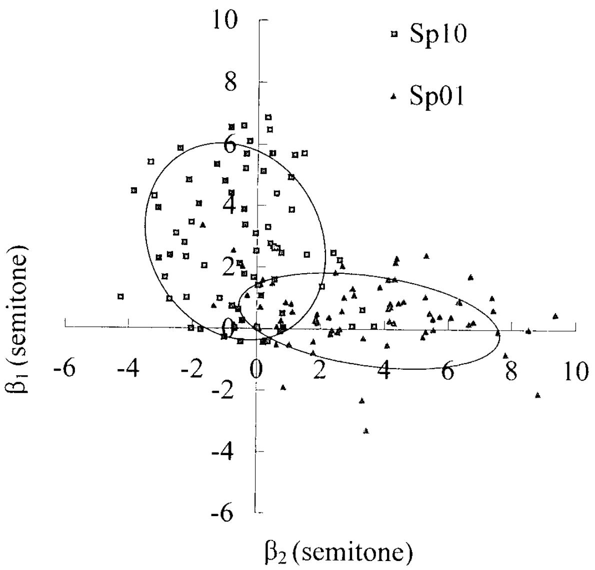


Fig. 3 Distribution of the F_0 range difference (β_1, β_2), where β_1 and β_2 represent the F_0 range differences between focused and neutral utterances for syllable 1 and syllable 2, respectively.

Table 5 Mean values $\overline{\alpha}_1$ and $\overline{\alpha}_2$ of vowel duration ratios in syllables 1 and 2 (α_1, α_2), respectively.

Stress pattern	$\overline{\alpha}_1$	$\overline{\alpha}_2$
Sp11	1.39	1.32
Sp10	1.43	0.85
Sp01	1.15	1.49

Table 6 Mean values $\overline{\beta}_1$ and $\overline{\beta}_2$ of F_0 range differences in syllables 1 and 2 (β_1, β_2), respectively.

Stress pattern	$\overline{\beta}_1$ [semitone]	$\overline{\beta}_2$ [semitone]
Sp11	1.86	2.18
Sp10	2.65	−0.45
Sp01	0.51	3.31

indicates that the vowel duration of syllable 2 is shortened when there is single stress on syllable 1, and secondly, only 8% of Sp01 lie below the abscissa, which indicates that the vowel duration of syllable 1 is rarely shortened when single stress is placed on syllable 2. Table 5 shows the mean values $\overline{\alpha}_1$ and $\overline{\alpha}_2$ of the vowel duration ratios in syllables 1 and 2 (α_1, α_2), respectively. As can be seen in Table 5, $\overline{\alpha}_2$ is decreased significantly to an average of 85% for Sp10 and increased remarkably to an average of 149% for Sp01. This is in contrast to the increase in $\overline{\alpha}_1$ to an average of 115% and 143% for Sp01 and Sp10, respectively. In addition, the difference between the mean vowel duration ratio of pre-stressed syllable ($\overline{\alpha}_1 = 115\%$) and mean vowel duration ratio of post-stressed syllable ($\overline{\alpha}_2 = 85\%$) is significant ($p < 0.05$), indicating that single stress on the syllable of disyllabic words not only lengthens the stressed syllable but also systematically affects the length of the unstressed syllable.

F_0 Range

Figure 3 shows the distribution of the F_0 range differences (β_1, β_2), where β_1 and β_2 represent the differences in F_0 range between focused and neutral utterances for syllables 1 and 2, respectively. The 68% probability density ellipses for the single stress patterns Sp10 and Sp01 are also shown in the figure. Firstly, 57% of Sp10 lie in the left region of the ordinate, indicating that the F_0 range of syllable 2 is more likely to be decreased in the case of single stress on syllable 1; secondly, 27% of Sp01 lie below the abscissa, indicating that the F_0 range of pre-stressed syllable 1 is less likely to be decreased when syllable 2 is stressed. Table 6 shows the mean values $\overline{\beta}_1$ and $\overline{\beta}_2$ of F_0 range differences in syllables 1 and 2 (β_1, β_2), respectively. As can be seen, the mean F_0 range differences increase more for the single stress patterns ($\overline{\beta}_1 = 2.65$, $\overline{\beta}_2 = 3.31$) than those for the double stress pattern ($\overline{\beta}_1 = 1.86$, $\overline{\beta}_2 = 2.18$), with significant differences at the level of $p < 0.05$. Another substantial difference is that the mean F_0 range difference of the pre-stressed syllable 1 ($\overline{\beta}_1 = 0.51$) is larger than that of the post-stressed syllable 2 ($\overline{\beta}_2 = -0.45$) ($p < 0.05$).

4.2.2. Effects of tone combinations

We performed an extended CHAID (Chi-square Automatic Interaction Detection) analysis [12,13], with the standardized values of α_1, α_2, β_1 and β_2 as dependent variables and the identities of the tones and speakers as independent variables (Table 7). The standardization was made on the basis of their means and standard deviations,

Table 7 Independent and dependent variables used in the CHAID analysis.

Independent variables (I_V)	
T_I	Tone identities of syllable 1 (H, R, L, F)
T_F	Tone identities of syllable 2 (H, R, L, F)
Spk	Subjects who uttered the target word (F-1, F-2, M-1, M-2, M-3)
Dependent variables (D_V)	
$\hat{\alpha}_1$	Standardized value of α_1
$\hat{\alpha}_2$	Standardized value of α_2
$\hat{\beta}_1$	Standardized value of β_1
$\hat{\beta}_2$	Standardized value of β_2

where the standardized values were denoted as $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively. The CHAID algorithm recursively splits the dependent variable space into two or more subspaces. An exhaustive search is made to find the best split resulting in the most statistically significant difference among all possible splits with regard to the independent variables. The splitting process is repeated until insignificant difference is found after splitting. The one-way ANOVA is used to indicate the degree of the difference. Thus the larger the F-value, the greater the difference becomes. We also calculated the reduction ratio of variance (RR) after the split. Tables 8 and 9 show the results for

Table 8 Major results of CHAID analysis for vowel duration. * indicates that the subspace represented by the independent variables was further split into a subspace.

Stress pattern Sp01 Depth of split	Independent variables		Dependent variables $\hat{\alpha}_1$	F	RR[%]
1st split	Spk:	F-1, M-1, M-2, M-3* F-2	0.07 −0.56	28	28
2nd split	T_F:	H, R, F L	−0.13 0.89	11	15
			$\hat{\alpha}_2$	F	RR[%]
1st split	T_F:	H, R, F L	0.01 2.17	97	57
Stress pattern Sp10 Depth of split	Independent variables		Dependent variables $\hat{\alpha}_1$	F	RR[%]
1st split	T_I:	H, R L, F	0.67 1.07	8	10

Table 9 Major results of CHAID analysis for F_0 range. * indicates that the subspace represented by the independent variables was further split into a subspace.

Stress pattern Sp01 Depth of split	Independent variables		Dependent variables $\hat{\beta}_1$	F	RR[%]
1st split	T_I	H, F R, L	−0.69 0.42	8	10
			$\hat{\beta}_2$	F	RR[%]
1st split	T_F:	H R, L, F	−0.61 0.99	62	46
Stress pattern Sp10 Depth of split	Independent variables		Dependent variables $\hat{\beta}_1$	F	RR[%]
1st split	T_I:	H L R, F	−0.36 0.68 1.1	45	55
			Dependent variable $\hat{\beta}_2$	F	RR[%]
1st split	T_I:	H, R L, F*	−0.37 −0.92	15	18
2nd split	T_F:	H R, L, F	−0.45 −1.11	10	10

vowel duration and F_0 range, respectively. As can be seen, most of the CHAID-based splitting process underwent at least one split, except for the process for the dependent variable $\hat{\alpha}_2$ under stress pattern Sp10. Moreover, the identities of the tones were mostly selected as the major independent variable for the dependent variables $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_1$ and $\hat{\beta}_2$, indicating that the tone combinations significantly affect the variations of both durations and F_0 range.

Vowel Duration

The tonal categories of the second syllables (T_F) under stress pattern Sp01 strongly affect not only the value of $\hat{\alpha}_2$ but also that of $\hat{\alpha}_1$. As can be seen in Table 8, the standardized vowel duration ratio of the stressed tone L ($\hat{\alpha}_2 = 2.17$) was much larger than that of the other tones ($\hat{\alpha}_2 = 0.01$), with a significant difference ($F = 97$, $RR = 57\%$), which even anticipatively affects the standardized vowel duration ratios of the pre-stressed syllables ($F = 11$, $RR = 15\%$). The results for the stress pattern Sp10 less clearly show the tonal effects with a difference ($F = 8$, $RR = 10\%$).

F_0 Range

As can be seen in Table 9, the tonal categories strongly affect the variations of the F_0 range of both stressed syllables 1 and 2 with significant difference ($F = 45$, $RR = 55\%$ for syllable 1; $F = 62$, $RR = 46\%$ for syllable 2). In general, the F_0 ranges of the static tones H and L were increased less than those of the dynamic tones R and F, except that the F_0 range of the word-final tone L was increased to the same extent as that of the dynamic tones. Moreover, the high-offset of the stressed tones H and R increased the onset F_0 of the post-stressed syllables, while the low-offset of the stressed tones L and F decreased the onset F_0 of the post-stressed syllables. As a result, there is a significant difference between the effects of the two tonal categories ($F = 15$, $RR = 18\%$).

4.3. Discussion

The results for the effects of the stress patterns of the focused word illustrate that the stressed syllables were consistently produced with significantly increased vowel duration and F_0 range, but with regard to the pre- and post-stressed syllable in the words, there was an overall asymmetric variation in vowel duration and F_0 range, i.e., a slight increase in pre-stressed syllable and a considerable decrease in post-stressed syllable. However, as can be seen in Figures 3 and 4, the asymmetrical property of F_0 range is less obvious than that of vowel duration, implying that tonal interaction probably affects the variation of the F_0 range. In addition, neither the distribution of vowel duration ratios nor the differences of F_0 range is concentrated within the domain of each stress pattern, indicating that there may still be other factors which might systematically affect the variations of vowel

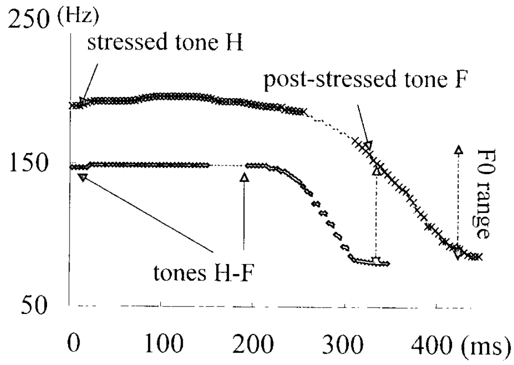


Fig. 4 F_0 contours of a tone combination H-F produced by speaker MZJ with a stress pattern Sp10 and a neutral pattern.

duration and F_0 range. Thus it is worth considering the detailed effects of tone combinations.

The results for the effects of the tone combinations illustrate that: (1) the word-final tone L of a stressed syllable shows a unique lengthening effect on the variation of vowel duration, which may be due, at least in part, to the phenomenon that a rising F_0 contour is appended to the word-final tone L, while the word-initial tone L keeps falling only; (2) the F_0 range of the post-stressed syllable is strongly affected by the high-or-low offset of the tone of the preceding stressed syllable. Contrary to the expectation that the F_0 range of the post-stressed syllable would be decreased according to the asymmetric property of the F_0 range, it is increased when preceded by the high-offset tones. The tone combination H-F is a typical example in our data. As can be seen in Fig. 4, the F_0 contour of the first stressed syllable is raised significantly, which is followed by a considerable increase of the F_0 maximum in the second post-stressed syllable by 22 Hz and a slight increase of the F_0 minimum by 5 Hz. As a result, the F_0 range is increased in the post-stressed syllable. Xu [14] examined F_0 contours of bi-tonal sequences and found a carry-over tonal effect on F_0 contours: the starting F_0 of a tone is assimilated to the offset of a previous tone. It is considered that the carry-over effect resulted in the increase of the F_0 range of the post-stressed syllable.

5. SUMMARY

The results of our two experiments reveal that: (1) the effects of semantic structures of the focused words on the prosodic correlates are completely mediated by the stress patterns, (2) there is a significant asymmetry of vowel duration as well as F_0 range between the pre-stressed and post-stressed syllables, implying that different strategies are employed in the task of focusing disyllabic words with the single stress patterns Sp10 and Sp01, i.e., emphasizing the first syllable as well as weakening the second syllable for the former, but emphasizing the second syllable only for the latter. Moreover, the tonal combinations significantly affect the variations of both the vowel duration and F_0 range.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Zongge Li and his colleagues for their assistance in speech recording carried out at Fudan University, Shanghai, China.

REFERENCES

[1] R. Iwata, "Tone and accent across the Chinese Dialects," *J. Phonet. Soc. Jpn.*, **5**, 18–27 (2001).

[2] D. R. Ladd, *Intonational Phonology* (CUP, Cambridge, 1996), pp. 153, 180.

[3] N. Campbell and M. E. Bechman, "Stress, prominence, and spectral tilt," *ESCA Tutorial/Research Workshop on Intonation: Theory, Models, Applications, Athens, Greece*, pp. 67–70 (1997).

[4] G. Fant, A. Kruckenberg and J. Liljencrants, "Acoustic-phonetic analysis of prominence in Swedish," in *Intonation: Analysis, Modelling and Technology*, A. Botinis, Ed. (Kluwer Academic Publisher, Dordrecht, 2000), pp. 55–86.

[5] K. Maekawa, "Effects of focus on duration and vowel formant frequency in Japanese," *Computing Prosody: Computational Models for Processing Spontaneous Speech*, Y. Sagisaka, N. Campbell and N. Higuchi, Eds. (Springer Verlag, New York, 1996), pp. 129–153.

[6] M. Heldner and E. Strangert, "Temporal effects of focus in Swedish," *J. Phonet.*, **29**, 329–361 (2001).

[7] E. Garding, "Speech act and tonal pattern in standard Chinese constancy and variation," *Phonetica*, **44**, 13–29 (1987).

[8] Y. Xu, "Effects of tone and focus on the formation and alignment of f_0 contours," *J. Phonet.*, **27**, 55–105 (1999).

[9] Y. R. Chao, *A Grammar of Spoken Chinese* (CUP, Cambridge, 1968), p. 35.

[10] J. L. Zhang, "Acoustic parameters and phonological rules of a text-to-speech system for Chinese," *IEEE ICASSP 86*, pp. 2023–2026 (1986).

[11] Z. Gu, H. Mori and H. Kasuya, "Stress accent placement of focused words in Mandarin Chinese," *Tech. Rep. IEICE*, SP 2001-18, pp. 23–27 (2001).

[12] G. Kass, "An exploratory technique for investigating large quantities of categorical data," *J. Appl. Stat.*, **29**, 119–127 (1980).

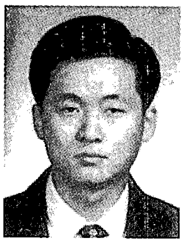
[13] D. Biggs, B. de Ville and E. Suen, "A method of choosing multiway partitions for classification and decision trees," *J. Appl. Stat.*, **18**, 49–62 (1991).

[14] Y. Xu, "Contextual tonal variations in Mandarin," *J. Phonet.*, **25**, 61–83 (1997).

APPENDIX

Table A.1 Target words used in Experiment 2.

Target word	English translation	Tones
Zhong1Xing1	Center	H-H
Zhong1Xing1	Cordial	H-H
Zhang1Jie2	Chapters and sections	H-R
Zheng1Jie2	Chastity	H-R
Zhong1Gu3	Bell tower	H-L
Zhong1Gu3	Medieval times	H-L
Zhong1Ri4	China and Japan	H-F
Zhong1Ri4	All day	H-F
Nan2Fang1	South area	R-H
Nan2Fang1	Husband's side	R-H
Chang2Chang2	Often	R-R
Chang2Qing2	Reason	R-R
Tou2Nao3	Brains	R-L
Hou2Nao3	Monkey head	R-L
Hua2Gui4	Luxurious	R-F
Hua2Gui4	Slide cabinet	R-F
Shi3Zhong1	All along	L-H
Shi3Guan1	Official historian	L-H
Huo3Shi2	Mess	L-R
Huo3Shi2	Flint	L-R
Ma3Ke4	Mark	L-F
Ma3Ku4	Riding breeches	L-F
She4Shi1	Missing a shot	F-H
She4Shi1	Facilities	F-H
Dong4Liang2	Ridgepole and beam	F-R
Dong4Ling2	Icicle	F-R
Kan4Fa3	View	F-L
Kan4Hao3	Take a fancy to	F-L
Hou4Bei4	Posterity	F-F
Hou4Bei4	Reserve	F-F



Zhenglai Gu received the B.E., degree in Computer Science from Fudan University, Shanghai, China, in 1995, and the M.S. degree in Information Engineering from Utsunomiya University, Utsunomiya, Japan, in 1999. He is now a Ph.D. candidate in Production and Information Sciences at Utsunomiya University. His current research interests include speech synthesis, prosody and digital signal processing.

Hiroki Mori received the B.E., M.E. and Ph.D. degrees from Tohoku University, in 1993, 1995 and 1998, respectively. He was with the Graduate School of Engineering, Tohoku University in 1998. He is now a Research Associate of Utsunomiya University. His research interests include speech recognition, speech synthesis, spoken dialogue systems, and natural language processing. He is a member of the Acoustical Society of Japan, the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan and the Japan Society of Logopedics and Phoniatrics.

Hideki Kasuya received the B.S., M.S., and Ph.D. degrees in Electrical Communication Engineering all from Tohoku University, Sendai, Japan, in 1963, 1965, and 1970, respectively. In 1968, he joined the Research Institute of Electrical Communication as a research associate at Tohoku University, where he was primarily engaged in speech analysis, perception and recognition. From 1974 to 1977 he was a visiting researcher at the Speech Communications Research Laboratory, Inc., California, U.S.A., working on speech recognition. Since 1978 he has been with the Faculty of Engineering, Utsunomiya University, where he is now a professor in the Department of Electrical and Electronic Engineering. His research interests include various areas of speech science and technology, digital signal processing, and image processing.