Tone and Intonation in Mandarin*

Chi-lin Shih

This paper discusses the basic facts about Mandarin intonation. I start with a description of tone shapes and tonal targets in monosyllabic and disyllabic sequences, and then proceed to discuss other factors such as catathesis, prominence, and pitch raising and lowering in discourse structure. Although it is possible that we will never be able to give an exhaustive list of the various factors that may influence the realization of F0, a step-by-step study will bring us closer to our goal.

1. MANDARIN TONES IN ISOLATION

Mandarin Chinese distinguishes four lexical tones: high level, rising, low falling, and falling, which are traditionally referred to as tone 1, tone 2, tone 3, and tone 4 respectively. A numeral following the segmental transcription refers to the tone. Figure (10) illustrates a set of minimal pairs with the syllable *ma*: *ma1* 'mother', *ma2* 'hemp', *ma3* 'horse', and *ma4* 'to scold'. The figure is a display of the time function of F0 values, with the *y* axis representing F0 in Hz, and the *x* axis representing time. In addition to four lexical tonal contrasts, a limited number of lexical items, mostly suffixes and sentential particles, do not have an underlying tone, and their F0 values are predictable from the surrounding tones. The absence of a tone on a syllable is traditionally referred to as neutral tone or tone 0.

A high level tone, or tone 1, starts in a speaker's high pitch range and remains high. As the name implies, there is no drastic pitch movement except a slight dip in the middle of the vowel, and a slight rise toward the end of the syllable. A rising tone, or tone 2, starts at the speaker's mid pitch range, remains level or drops slightly during the first half of the vowel, and rises up to high at the end. A low tone, or tone 3, is phonetically a low falling tone. It starts at the speaker's mid range and falls to the low range. It is often accompanied by laryngealization over the second half of the syllable. A falling tone, or tone 4, usually peaks around the vowel onset, then falls to the low pitch range at the end. In syllables with initial voiceless consonants, the small rising slope is often invisible

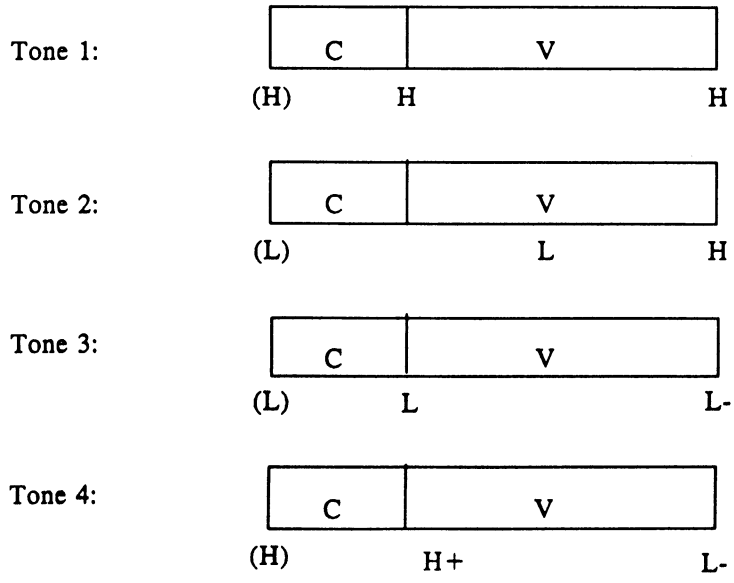
* Most of the experimental work reported on in this paper was done while I was doing postdoctoral research at AT&T Bell Labs. I am grateful for the support that I received from Mark Liberman. This paper has greatly benefited from comments and suggestions from the following people: Nick Clements, Mark Liberman, Janet Pierrehumbert, Kim Silverman, and Richard Sproat. I extend my appreciation to them.

and the pitch contour is a straight falling line. In syllables with an on-glide, the pitch is rising through the glide and gives the impression of a delayed peak.

The shape of tone 3 varies the most among all tones due to phonological processes. The low-falling pattern shown in Figure (1) has the highest frequency in speech. It is the pattern that occurs in non-final position. In isolation and in sentence final position, tone 3 may have a rising tail, and that is known as the falling-rising tone. Southern speakers often keep the low-falling pattern even in the final position in casual speech, and use the falling-rising pattern only in deliberate, emphatic speech, or in yes-no question. Northern speakers often use the falling-rising pattern sentence finally in all speech acts. The falling-rising pattern is considered the base form in much of the tonal literature, as in Woo (1969), Yip (1980), and Tseng (1981), because it is the citation form. Another complication at non-final position comes from a tone sandhi process which converts the first of two low tones into a rising tone, see Cheng (1973), Shih (1986). The falling-rising version of tone 3 has the longest duration among all tones, whereas the low-falling version has the shortest duration.

It is interesting to note that the beginning and end points of all tones fall on three distinct levels rather than scattering across a continuum. Tone 1 and tone 4 both start high, very close to where tone 1 and tone 2 end. Tone 2 and tone 3 both begin in the middle range, while tone 3 and tone 4 both fall to the low pitch range. More than 20 repetitions of each tone with various syllable structures confirm our observations so far. The following table summarizes our understanding of the relative values and the placement of tonal targets for each tone.

TABLE (1)



The table is obviously more complicated than a phonological representation consisting of only H and L. However, the additional variation is minimal. H+ represent a pitch level slightly higher then H, and that corresponds to the peak of tone 4. L- is lower than L, that is the end point of tone 3 or tone 4. The values of H, H+, L or L- have to be adjusted for individual speakers and for style of speech.

I included tonal targets in parentheses at the beginning of the consonantal region in order to account for the tone shapes of the consonant region in isolation and in sentence initial position. In non-initial position, the consonantal region is where the tonal transition occurs; see Figure (2). The F0 values there are derivable by interpolating surrounding tonal targets, and there is little evidence for the existence of a real target.

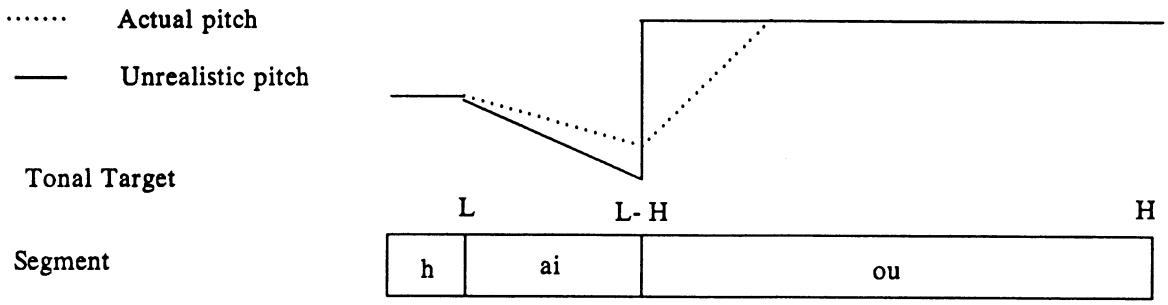
Contra Garding (1987), who proposes that the "turning points" of pitch contours are located at C/V boundaries, I find that a more flexible placement of targets as suggested in the above table gives us the best fit for tones in isolation as well as in connected speech. According to my data, tone 1 and tone 3 do have targets at the C/V boundaries, while tone 4 has slightly delayed target, furthermore, tone 2 doesn't have a target until the middle of the vowel. The delayed target of tone 2 ensures a relatively late rising contour, and predicts correctly that the first half of tone 2 varies according to the tonal transition in connected speech.

Table (1) enables us to generate stylized tonal contours that match tones in isolation. From there, we can proceed to study the more complicated interactions of tones, and tone and intonation. I should make it clear that I am not assuming that tones in isolation are the underlying tonal values, for monosyllables are subject to intonation effects just as longer sequences are. But by comparing tones in isolation to tones in longer sequence, and by controlling our test sentences, we have a good chance of isolating individual effects, and being able to predict pitch contours of uncontrolled sentences. To illustrate where we are and how far we need to go, Figure (3) compares tones generated by Table (1) to the natural pitch contour of the sentence *Mu4-diao1 ti2-cai2 you3 bu4-shao[2] qu3 zi4 min2-jian1 gu4-shi0*, 'The themes of wood carving are often taken from folklores'. I assign 250, 300, 200 and 150 to H, H+, L and L- respectively. The generated pitch countour requires improvements in many aspects. Firstly, some tonal targets need adjustment due to tonal co-articulation; secondly, the scaling of F0 values need more control. I turn to tonal co-articulation in the following section, and discuss two of the pitch scaling factors, catathesis and paragraph structure, in later sections.

2. TONE SEQUENCES

Firstly, we need to investigate tonal co-articulation. Pitch movement in natural speech is gradual. It takes time to reach a L target from a high point,

and possibly takes more time to rise to a H. Tonal targets within a syllable, say, H and L of a falling tone, are not adjacent to each other on the time scale: H is placed toward the beginning of the vowel and L at the end, and there will be sufficient time for the pitch to fall. The physical constraint limits the number of tonal targets in a syllable, and explains why a zig-zag tone is unheard of in natural languages. When different tones are strung together in words and sentences, the situation often arises that adjacent tonal targets have opposite values. That is where tonal co-articulation is expected. Some problems are resolved by not positing a tonal target at the beginning of the consonant. Doing that frees up the consonantal region for tonal transition. Figure (2), *mai3 mao1*, 'to buy a cat', illustrates this point. When there is no consonant on the second syllable, the tonal transition takes place at the beginning of the vowel, as if part of the vowel is interpreted as functioning as a consonant. The tone and segment template of *hai3 ou1*, 'seagull' is shown below. The solid line represents the unrealistic pitch by taking the face value of each tonal target. The dotted line gives the adjusted pitch contour that is closer to natural speech. The actual pitch track is shown in Figure (4).



Theoretically, there are several other possibilities to resolve the transition problem: a target may move backward; the value of H and L may be neutralized; or some targets may simply be deleted. In Mandarin, tonal targets rarely move back. Peak delay and adjustment of pitch level are quite common. In the template above and in Figure (4), the end of the syllable *hai3* is not as low as it would be in isolation.

The following set of data allow us to look into tonal co-articulation of all disyllabic tonal combinations. They consist of 16 minimal pairs that have the same segments "fu-ji" but contrast in tones. The whole set was recorded three times in random order. *Fu3-ji3* changes to *Fu[2]-ji3* as a result of tone sandhi, reducing 16 tonal combinations to 15 on the surface. Words/phrases in the *fu-ji* set differ in syntactic/morphological structures but are similar in prosodic structure: all are in a foot and with primary stress on the final syllable.

fu1-ji1	V N	'to hatch a chicken'
fu1-ji2	N's N	'husband's residence'
fu1-ji3	N N	'a hatchet and a spear'
fu1-ji4	Mod N	'apply-medicine, ointment'
fu2-ji1	$(V\ N)_{V/N}$	'planchette'
fu2-ji2	Mod V	'ambush'
fu2-ji3	V N	'to hold onto a spear'
fu2-ji4	Mod N	'the medicine that should be taken internally'
fu3-ji1	Mod N	'a spare engine'
fu3-ji2	Mod V	'dive to attack'
fu3-ji3	V N	'to assist oneself'
fu3-ji4	$(Mod\ V)_{N}$	'the assistant in a sacrifice ceremony'
fu4-ji1	Mod N	'abdominal muscles'
fu4-ji2	V N	'to carry the book case, to study at far away place'
fu4-ji3	V N	'to carry the spear, to fight'
fu4-ji4	Mod V	'attached-mail, to mail with an attachment'

Tables (2) and (3) below list the mean values of the tonal targets of *fu* and *ji* respectively. Both tables arrange the tones in columns and the tonal contexts in rows. Tones in isolation are given in the first row for comparisons. Pitch trackings of some sample sets are given in Figures (5)-(8).

Table 2 Tones in the Initial Position

	fu1	fu2	fu3	fu4
isolation	266-266	211-253	214-154	287-159
before tone 1	258-270	219-**245**	213-176	299-**218**
before tone 2	274-291	216-257	223-168	300-**227**
before tone 3	273-290	223-281	_	310-**238**
before tone 4	268-286	216-**243**	225-178	300-**227**

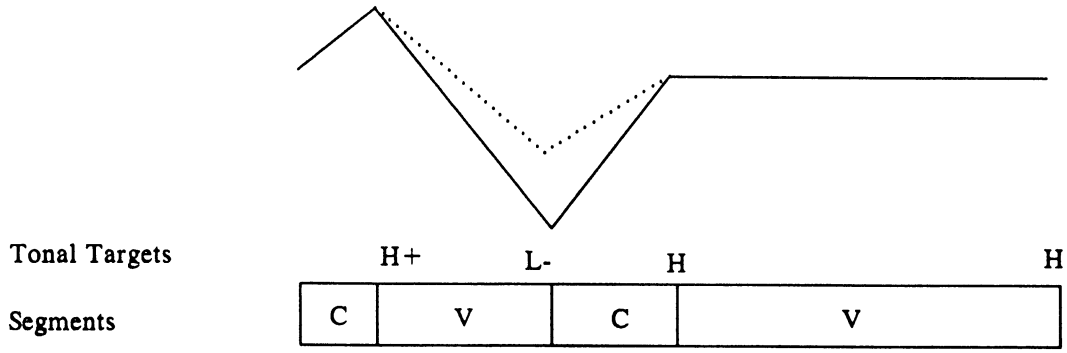
Table 3 Tones in the final Position

	ji1	ji2	ji3	ji4
isolation	259-258	209-262	211-153	291-162
after tone 1	269-276	209-235	219-144	296-176
after tone 2	266-273	217-249	**247**-144	281-177
after tone 3	263-272	**188**-258	—	284-175
after tone 4	262-266	207-252	201-141	278-174

There may be a reason behind every variation we see in the tables above. Some of the variations are not consistent across all syllable types as discussed in Shih (1987). They are possibly affected by consonant perturbation (Silverman 1987 and works cited therein) and intrinsic vowel height (Steele 1985 and works cited therein), two topics that I won't address in this paper. Details aside, I highlight the most consistent variations in boldface and discuss them briefly. The main issues could be accounted for by three most general tonal co-articulation rules below.

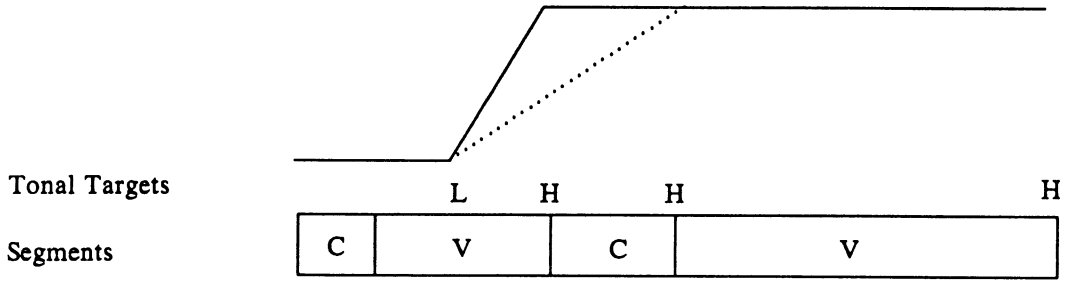
(1) Non-final tone 4 ends in L, rather than L-.
(2) The final H of tone 2 is deleted if the following tone starts with H.
(3) The beginning L of tone 3 tends to assimilate to the previous H.

In the initial position, as shown in Table 2, a tone 4 never reaches the L- target. Instead, the final value is comparable to the L value at the beginning of a tone 2 or tone 3 in isolation. Moreover, a following tone 2 or tone 3 starts exactly where tone 4 ends. I schematize this situation below. The dotted line shows the changes in pitch contour caused by coarticulation,

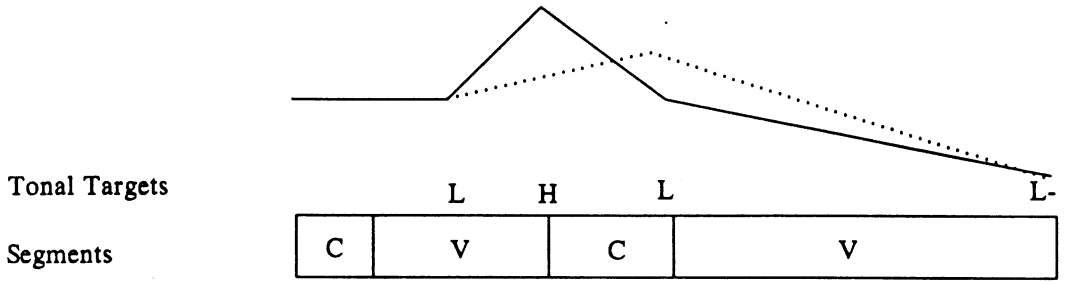


A tone 2 ends lower (245, 243) when the following target is H (tone 1 and tone 4); but ends higher (257, 281) when the following target is L (tone 2 and tone 3). The cause of this situation is not dissimilation. Rather, when a H target follows, the rising slope ends not at the end of the tone 2 syllable, but ends where the following H target is. This phenomenon suggests that the final H of a tone 2 is deleted in the presence of the next H target. Tone 2 seems to end lower in this context only because the value at the syllable boundary does not represent the H

target, but is a point on the rising slope. When the following target is L, the position of the final H is unaffected and the pitch value at the end of the syllable reflects the H value. The right-shift of the H target explains the surface dissimilation effect. The dotted line in the following picture shows the modification.



Everything being equal, tone 2 and tone 3 usually start with the same pitch height in isolation. We account for this by giving them the same beginning target L. However, when the preceding target is high, the beginning point of tone 3 is very often higher than that of tone 2. This situation is unexpected because the pitch value at the vowel onset of tone 2 should be somewhere between H and L, due to a delayed initial target, and be higher than a tone 3 with an earlier L target. The unexpected result arises from rule (3) above. The initial L of tone 3 is sometimes pulled up in the direction of the preceding H. This happens more frequently in fast speech across sonorant consonants. And the effect is more apparent when tone 2 precedes. It seems to me that the falling pitch (with absolutely no initial rise as in tone 4) and the very low ending is the characteristic of tone 3. The height of the initial target is less rigid. The co-articulation change is shown below.



Some other general points shown in the above tables include the following. An utterance initial H target has a much wider range than a L target. Initial H of tone 1 falls between 250-280 Hz, and H+ of tone 4 280-335 Hz. There is a difference of 30-50 Hz. The difference among utterance-initial L target is less than 20 Hz. The picture is reversed for the second syllable, where H target falls between 245-275 in tone 1, 260-300 in tone 4, but L target spreads from 200 to 255 in tone 3 and 175 to 225 in tone 2. The deviation of L targets is largely due to tonal co-articulation, for L target tends to be deleted or be assimilated to the

preceding target.

In the final position, a tone 2 starts lower when it follows a tone 3 (188 Hz). This situation is explained by the absence of a target at the vowel onset. The L- value of a preceding tone 3 will affect the beginning of tone 2, but not so much for other tones. Tone 4 does not have the same lowering effect because non-final tone 4 ends in L, the same as the initial tone 2 target, so the pitch at the first part of tone 2 would be level.

Figure (9) shows the improvement after implementing tonal co-articulation rules. What needs to be done at this stage is to study the factors that control F0 scaling. We discuss two of these factors below, catathesis and paragraph structure.

3. CATATHESIS

Catathesis (or down-step) refers to F0 lowering due to specific tonal combination. The most common case is in a sequence H L H, the second H is considerably lower than the first because of the intervening L (Liberman and Pierrehumbert, 1984). The following experiment was designed to test the difference in the catathesis effect of tones with a L target, namely, tone 2, tone 3, and tone 4. All test sentences are five syllables long and have tone 1 in the first, third, and final position. The intervening syllables have one of the four possible tones. The catathesis effect, if any, can be measured by comparing the F0 value of the tone 1 syllables.

The data

1. jil shi1 xiu1 tuo1 che1.	'The mechanic fixes the cart.'
2. gong1 ren2 shou1 fang2 zu1.	'The worker collects the rent.'
3. Jing1 li3 he1 guo3 zi1.	'The manager drinks juice'
4. shang1 dian4 chu1 jiu4 shu1.	'The shop publishes old books.'

The data set was recorded in random order, 4 times in natural speech, and 4 times in reiterant speech (Liberman and Streeter, 1978), in which all syllables in the target sentence are replaced by *da*, while maintaining the original prosodic pattern. The purpose of reiterant speech is to avoid interference of segmental effects from various syllables in the target sentence. Pitch values of the initial, medial, and final tone 1 were measured and averaged. The measurement was taken from the center area of tone 1, avoiding the consonantal effect at the beginning, and optional final raising at the end. Table 4 and Table 5 list the mean pitch value from reiterant and natural speech respectively. The values are arranged by sentential-position in columns, and by sentence types in rows. The sentences are mnemonically numbered after the conditioning tones in the second and fourth position. Figures (10)-(13) give a sample of each sentence.

Table 4 Reiterant Speech

	Initial	Medial	Final
1	271	260	249
2	278	246	227
3	287	247	217
4	283	247	229

Table 5 Natural Speech

	Initial	Medial	Final
1	284	265	252
2	286	255	239
3	299	260	236
4	280	259	241

In general, the final tone is the lowest and the initial tone is the highest. This situation is found in sentence 1 as well, where initial, medial, and final tone 1 scale down at a rate of 2% per syllable in reiterant speech, and slightly more in natural speech. Since there is no L target in sentence 1, the lowering effect could not have come from catathesis. I attribute it to declination.

In both reiterant and natural speech, the sentences with intervening tone 2, 3, or 4 have lower medial and final tone than the sentence that has only tone 1, suggesting that all tones with a L target have some catathesis effect. There is not much difference on tone 2 and tone 4, while tone 3 exhibits more lowering effect on the following tone.

Roughly speaking, the catathesis effect of tone 2 and tone 4 is 12% on the medial syllable, and 7% on the final syllable. The catathesis effect of tone 3 is 14% on the medial syllable, 12% on the final syllable.

There are two reasons why the medial syllable is lowered more than the final one in sentence 2, 3, and 4. 1. The medial syllable is surrounded by L tones, which might pull down the pitch level. 2. The medial syllable is prosodically weak,[1] and a weak syllable is more susceptible to the surrounding environment.

1. All medial syllables in the test sentences are monosyllabic verbs, which are usually weaker than nouns in Mandarin.

The effects of catathesis seem to be related to the actual pitch level of the preceding L. Tone 3 has the lowest pitch level, L-, in the non-final position, so it has the strongest catathesis effect. Tone 2 starts at the L level, and tone 4 only falls to L in the non-final position, therefore both have less catathesis effect. When tone 4 is followed by a neutral tone, in which pitch falls to a even lower point than tone 3, the catathesis effect is indeed much stronger than all our test sentences here (Shih 1987).

Figure (14) shows the result after implementing tonal co-articulation rules, catathesis effects, some other factors discussed in Shih (1987), and smoothing. Although the two pitch contours do not match perfectly, the re-synthesized speech using rule-generated pitch contour represented by the dotted line already sounds very natural.

4. PROMINENCE

This section investigates the interaction of tones and prominence. There are three questions I am trying to answer. 1. How is prominence realized on tones? 2. Would different combinations of tone sequences affect the realization of prominence? 3. What happens to the post-prominence constituents?

A set of time words *jin1-tian1* 'today', *ming2-tian1* 'tomorrow', *mei3-tian1* 'everyday' and *hou4-tian1* 'the day after tomorrow' is embedded in the context *Lao3 Wang2* ____ *yao4 mai3 yu2* 'Lao Wang wants to buy fish ____.' Each sentence is repeated three times in three ways: 1. plain statement; 2. statement with emphasis on the subject 'Lao-Wang'; and 3. statement with emphasis on the time words. The data are listed below.

Lao3-Wang2 jin1-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish today.'
Lao3-Wang2 ming2-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish tomorrow.'
Lao3-Wang2 mei3-tian1 yao4 mai3 yu2. 'Lao Wang wants to buy fish everyday.'
Lao3-Wang2 hou4-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish the day after tomorrow.'

There are three additional sentences to test the interaction of prominence and low tone; and the effect on post-prominence level tones.

Lao[2]-Wang3 jin1-tian1 yao4 mai3 yu2.'Lao Wang3 wants to buy fish today.'
Lao[2]-Wang3 ming2-tian1 yao4 mai3 yu2.'Lao Wang3 wants to buy fish tomorrow.'
Lao3-Wang2 jin1 tian1 gang1 he1 dong1 gua1 tang1. 'Lao Wang just drank winter melon soup.'

In the subject *Lao3-Wang2*, *Lao3* is a prefix, thus weak prosodically. When emphasized, the strong syllable *Wang2* receives more prominence effect. The reverse is found in time words. *Tian1* 'day' is a generic noun, which is a prosodically weak member in Mandarin compounds. When emphasized, the strong syllables *jin1*, *ming2*, *mei3* and *hou4* are the loci of the prominence.[2]

Figures (15) to (21) present samples from test sentences. When multiple pitch contours are presented in the same figure, the solid line shows the normal reading, the dotted line shows the sentence with the highest prominence on *Lao-Wang*, and dashed line has the highest prominence on time words. In the following discussion, I use upper case letters to represent a prominent syllable or word, and use lower case for words with less or no prominence.

It is apparent from the figures that prominence is reflected by expanding pitch range: high targets become much higher, while low targets remain at the same level or are slightly lower. Aside from the increased pitch range, more prominent forms also have longer duration and higher intensity.

While the above generalization is true, a closer look of the data reveals a number of surprising details. While longer duration and higher intensity always fall on the most prominent syllable or word, high pitch is sometimes realized on an adjacent but less prominent word. For example, in Figure (15), the most prominent *WANG2* ends at 300 Hz, while the following *jin1-tian1* is 325 Hz high. In Figure (18), we see no difference in the dotted and dashed pitch tracks of *hou4*, when the dashed *HOU4* should have the highest prominence, and the dotted *hou4* is post-prominence, thus a weak syllable.

The mirror image of this situation is also true. Figure (20) shows that the influence of the final target L of *WANG3* extends into the beginning of the next syllable *ming2*, causing a post-prominence tone 2 to begin at a much lower pitch than where *WANG3* ends.

This situation is a reflection of the tonal co-articulation rule (2), where I discussed the deletion of the final H target of tone 2 in the presence of a following H target. In those situations, a tone 2 takes the later H to be its target, and extends the rising slope beyond the syllable boundary. What we see in Figure (15), (18), and (21) is a more dramatic display of the same thing.

Taking the shifted H or L to be the real target, the pattern of prominence structures begins to emerge. Figure (22) compares sentences with focus at different loci. The high peaks reach the same level, even though the peaks may not coincide with the focused syllable. While it is widely accepted that prominence will raise a high tone, whether low tone will be lowered is more of a debate. It is not entirely clear from our data that the L target of tone 2 would be lowered under prominence. We have some cases that do and some others that don't. The clearer evidence of low tone lowering comes from the final target of tone 3, which is the lowest target among all tones. Figure (20) shows clearly the

---

2. Noun phrases consist of a modifier and a generic noun tend to have initial stress. Other examples include professions with *ren2* 'person', as in *gong1-ren2* 'work-man, worker', or country names with *guo2, country*, as in 'fa4-guo2, France-country'.

lowering of L target on *MEI3*. The effect is evident on the following syllable *ming2*. These figures combined provide support to the claim that prominence causes pitch expansion, rather than just pitch raising.

Following a prominent H or L, the pitch level goes back to normal. The first L tone after the prominent H, and the first H after the prominent L, would bring the pitch level back to normal immediately; see Figures (15), (16). However, a string of like tones changes gradually[3]. Figure (21) shows the gradual fall of tone 1 after a prominent *WANG2*. If we assume that pitch range expansion is done by scaling tonal targets away from the reference line, an abstract line corresponds to the mid pitch range, we would be able to explain automatically why post-prominence targets remember what the normal values are. When pitch range changes, the reference line remains unaffected. And after prominence, pitch range is scaled back in relation to the reference line, and in turn determines the normal values of H and L targets.

5. PITCH HEIGHT AND DISCOURSE STRUCTURE

In this section I will discuss briefly the difference between prominence effect and what is referred to as *initial raising* and *final lowering* of discourse structure.

Hirschberg and Pierrehumbert (1986) assume that discourse exhibits a hierarchical structure, and that hierarchical structure is reflected in intonation by varied pitch ranges: an increase in the pitch range signals the beginning of a discourse segment, while final lowering signals the end of it. The magnitude of increase corresponds to the hierarchical level of discourse structure. While adopting the main concept of Hirschberg and Pierrehumbert, I see strong evidence from Mandarin that initial raising and final lowering actually affects the level of reference line, but does not affect the pitch range.

To study the pitch scaling effect in discourse, I recorded a story without a text, repeated it until I could finish the whole story fluently without undesired pauses and interjections. The story was about thirty sentences long, and had four paragraphs. The story is pitch tracked and H and L targets of all tone 2 are measured for comparison. Tone 2 is chosen because previous study on Mandarin tonal co-articulation suggests that tone 2 is the only tone that will give us a reliable reading of both H and L target, provided that we take the measurement of the final H target in the following tone 1. It is difficult to obtain the value of a H target when a tone 4 follows, so all those samples were discarded.

---

3. The only case in Mandarin where a string of like tones occurs is in a sequence of tone 1, which has only H targets. There is no tone that is just plain L, the so-called low tone, or tone 3, is actually low-falling. Moreover, strings of tone 3 is subject to a tone sandhi rule that change some to tone 2.

Figure (23) plots all the *reliable* tone 2 from the story. The y axis represents the H target of a tone 2, while the x axis represents the L target. The plot shows a linear dependency between the value of H and L targets: the higher the H target, the higher the L target, and vice versa. The higher tone 2 occurs at the beginning of the story and at the beginning of paragraphs. The lowest tone 2 occurs at the end of the story. Unfortunately, the paragraph structure of the story is not represented in the plot, so the correlation between pitch height and discourse structure wouldn't be obvious. However, it disproves the claim that initial high pitch level is achieved by increased pitch range. If that is the case, we should see the samples gather around a vertical line that represents a more or less invariant pitch level for L targets.

This study sheds some light on the confusion between initial raising and prominence effects. While both have higher H targets, The Mandarin data suggests that the distinction between the two lies in the realization of L targets. Initial raising involves register shift, when both H and L targets are realized with higher pitch value. Prominence effects involve pitch range expansion, which causes H target to be higher and L target to be lower.

6. *CONCLUSION*

Several issues discussed in this paper differ from other intonational studies, I will address them briefly in the conclusion.

Han and Kim (1974) reports on the disyllabic tonal sequence of Vietnamese. They found very little tonal co-articulation. Basically, tone shapes, slopes, and the placement of tonal targets in their study are the same in monosyllabic and disyllabic forms. However, pitch height may be influenced by preceding or following tones. The difference of Vietnamese and Mandarin suggests that tonal co-articulation rules as described in this paper may be language specific.

The tonally triggered catathesis poses a problem for Garding's (1987) grid model for Chinese. The grid model assumes a separation of the speech act related intonation and the lexically defined tones. For statements like the test sentences in section 3, a falling grid is drawn first, and then tonal targets (Garding's turning points) will be placed on grid lines. Ideally, sentences of the same speech act/intonation should share the same grid. But a uniform grid fails to capture the variations caused by the nature of tonal targets. Sentences with only tone 1 take an almost level grid, while sentences with tone 3 require a more steeply falling grid. The sentences here are relatively simple. An uncontrolled sentence with freely combined tones would require even more individual adjustment of grid. At that stage, it is difficult to justify the complete separation of intonation and tones, at least for what a 'statement grid' represents. The catathesis experiment shows that F0 value of each target is not solely determined by speech act, but is largely determined by the tonal composition.

Garding (1983, 1984, 1987) proposes that prominence effect could be captured by setting a *jumping* grid, in which the grid line is broken and expanded at the location of a prominent constituent, and compressed afterwards for the de-accented post-prominence constituents. I agree in principal that prominence is related to the expansion and compression of pitch range, however, my data shows a great divergency in what a potential prominence grid should look like. Sequences of post prominence tone 1 suggest that grid line should compress gradually, while a L target would require a sudden compression of grid just where the L target is located. Even more problematic is the higher, but unfocused H target after tone 2. Apparently, pitch range can continue to expand when the grid line would suggest a return. In my view, expansion of pitch range is caused by prominence, but the actual implementation could be mechanical. Extension of rising slope into the following H tone is part of a tonal co-articulation rule, and it applies with no reference to meaning or intention. As a result, expanded pitch range can not be automatically translated back to prominence. Garding's model tries to relate speech acts and their functions directly to surface realization of intonation, represented by grids. But since the surface F0 values are affected by many factors simultaneously, it will be quite difficult to find a grid that is meaningful.

Several studies related prominence effect to high pitch. Eady and Cooper (1986) suggest that, among other acoustic correlates, higher F0 topline is a major manifestation of non sentence-final focus. Inkelas, Leben and Cobler (1986) suggest that prominence is represented phonologically by H register. The fact that a prominent L target either lowers or remains at the same height suggests that pitch range expansion is a more appropriate representation.

The confusion of what high pitch could represent may be a reason why Wells (1986) fails to find a correlation between *the highest peak* and prominence. Although there is a strong correlation between high pitch and prominence, this paper discusses two other possible interpretations: the highest peak in a sentence could be at the sentence/paragraph initial position, signalling the beginning of a discourse structure; it could also be a post-prominence peak, carrying no prominence in itself.

References

Cheng, C. C. (1970) Domains of phonological rule application. In *Studies presented to Robert B. Lees by his students.* ed by J. M. Sadock and A. L Vanek. Edmonton: Linguistic Research. p 39-59.

Eady S. J. and W. E. Cooper (1986) Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, 80(2), p 402-415.

Han, M. S. and K.-O. Kim (1974) Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2, p 223-232.

Hirschberg, J. and J. Pierrehumbert (1986) The intonational structuring of discourse. in *Proceedings of ACL, New York*, p 136-144.

Garding, E. (1983) A generative model of intonation, in *Prosody: models and measurements*, ed by A. Cutler and D. R. Ladd. Springer-Verlag.

Garding, E. (1984) Chinese and Swedish in a generative model of intonation. *Nordic Prosody* V 3, p79-91. University of Umea.

Garding, E. (1987) Speech act and tonal pattern in standard Chinese: constancy and variation. *Phonetica* 44: 13-29.

Inkelas, S. Leben, W. and M. Cobler (1986) The phonology of intonation in Hausa, unpublished ms., Stanford University.

Liberman, M. Y. and L. A. Streeter (1978) Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America*, 63, p 231-233.

Liberman, M. Y. and J. B. Pierrehumbert (1984) Intonational invariance under changes in pitch range and length. in *Language Sound Structure*, the MIT Press.

Pierrehumbert, J. B. (1980) *The phonology and phonetics of English intonation.* Doctoral dissertation, MIT.

Shih, C.-L. (1986) *The prosodic domain of tone sandhi in Chinese.* Doctoral dissertation, UC-San Diego.

Shih, C.-L. (1987) The phonetics of the Chinese tonal system. Technical memo, AT&T Bell Labs.

Silverman, K. (1987) *The structure and processing of fundamental frequency contours.* Doctoral dissertation, University of Cambridge.

Steele, S. A. (1985) *Vowel intrinsic fundamental frequency in prosodic context*, doctoral dissertation, the University of Texas at Dallas.

Tseng, C. Y. (1981) *An acoustic phonetic study on tones in Mandarin Chinese*, doctoral dissertation, Brown Univeristy.

Wells, W. H. G. (1986) An experimental approach to the interpretation of focus in spoken English. in *Intonation in discourse*, ed by C. Johns-Lewis. Croom Helm, London & Sydney.

Woo, N. (1972) *Prosody and phonology*, Doctoral dissertation, MIT, reproduced by the Indiana University Linguistics Club.

Yip, M. (1980) *The tonal phonology of Chinese.* Doctoral dissertation, MIT.

Figure (1)

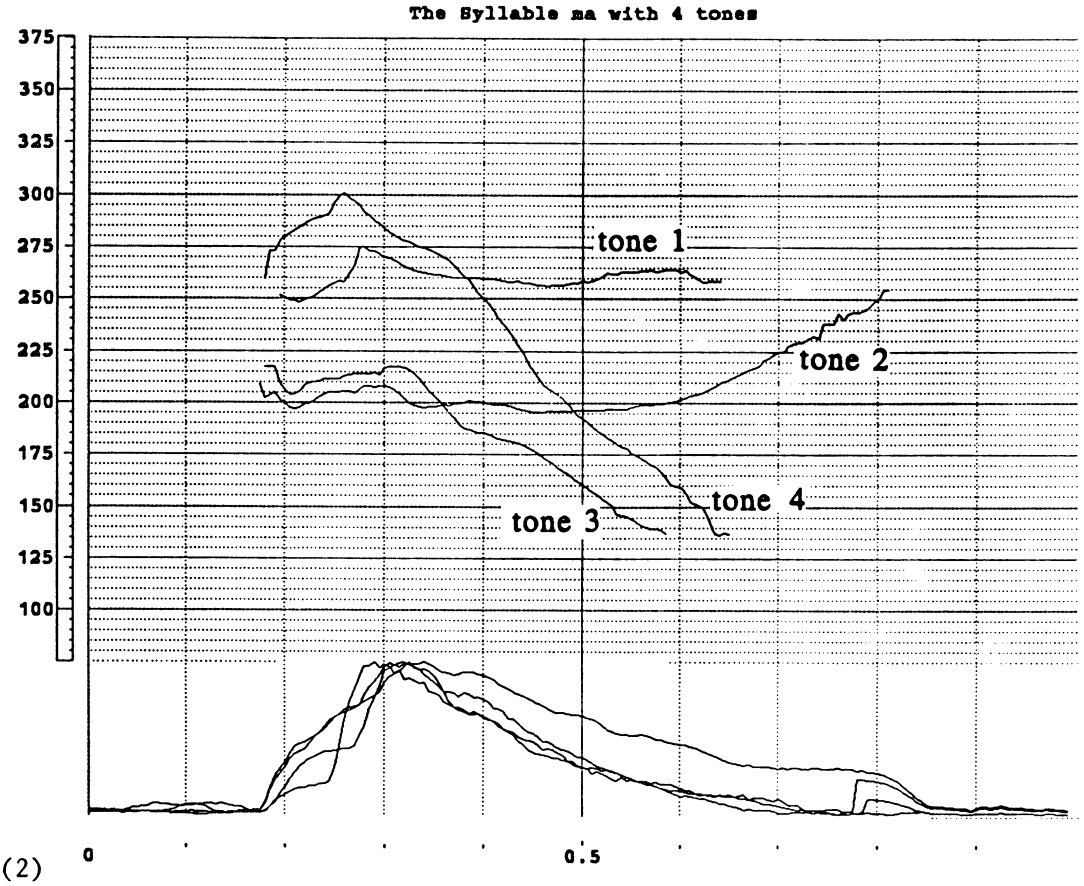


Figure (2)

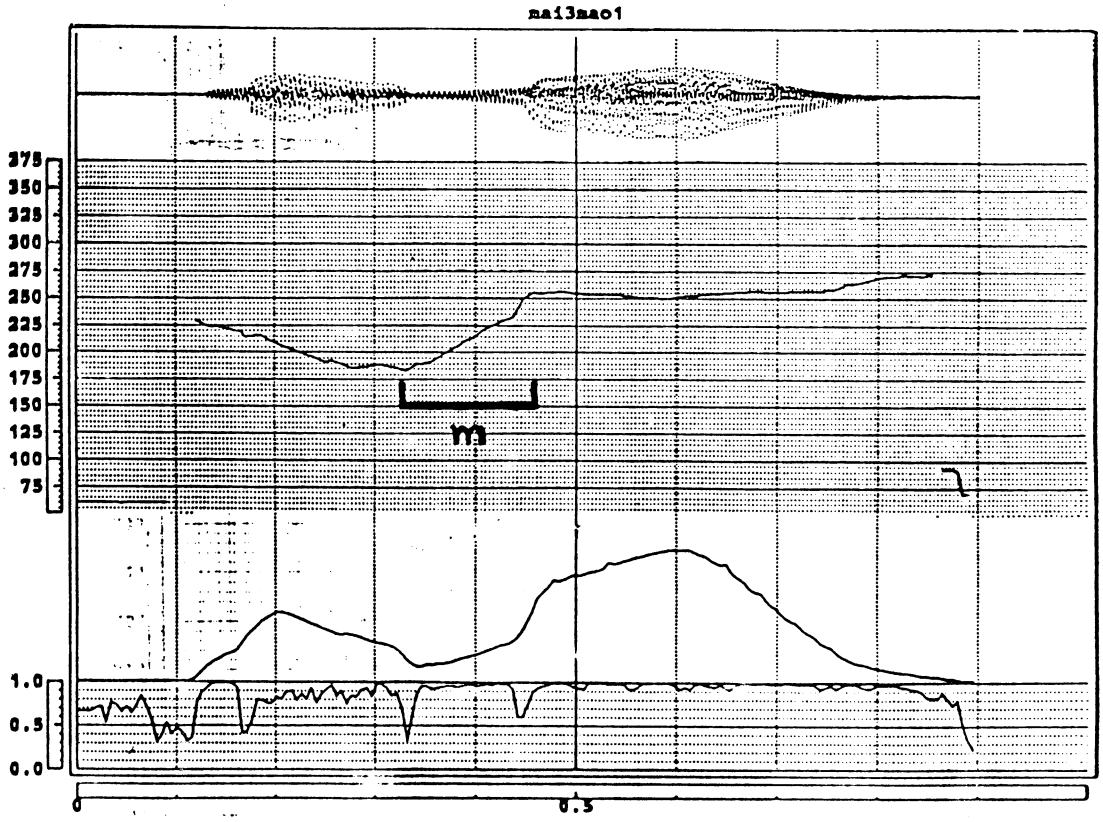


Figure (3)

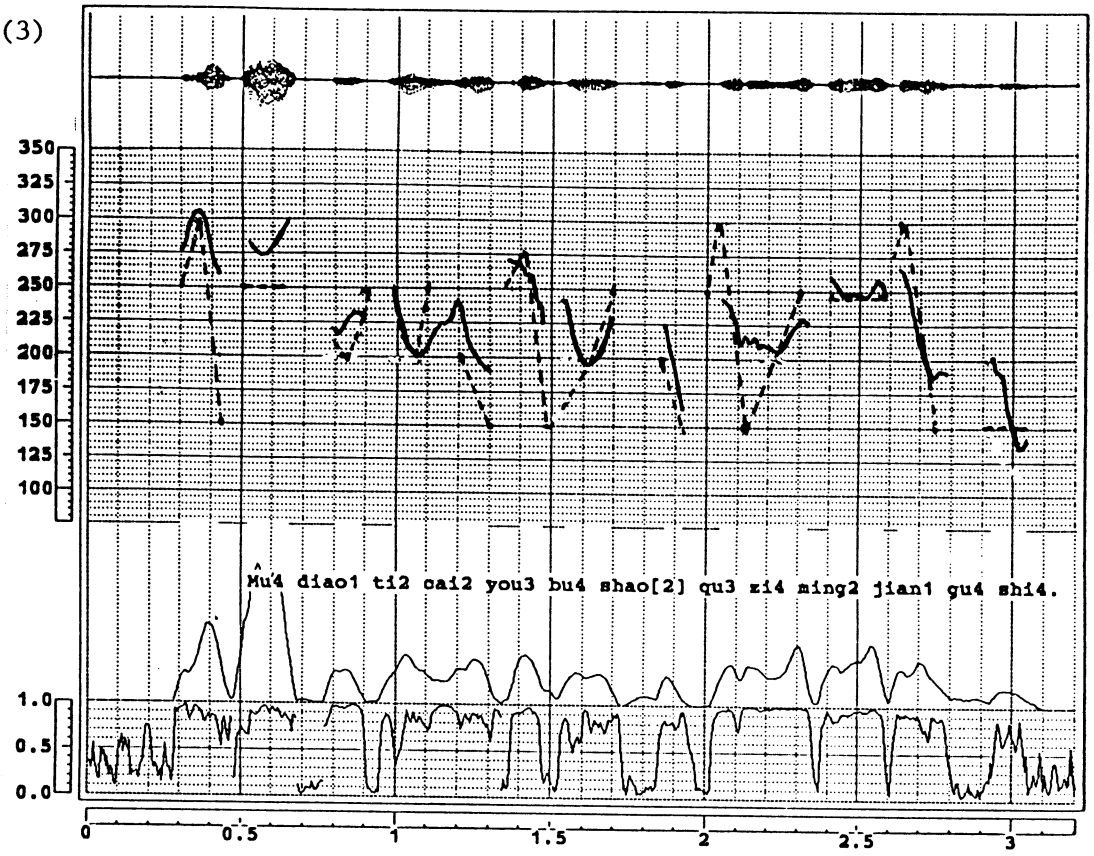


Figure (4)

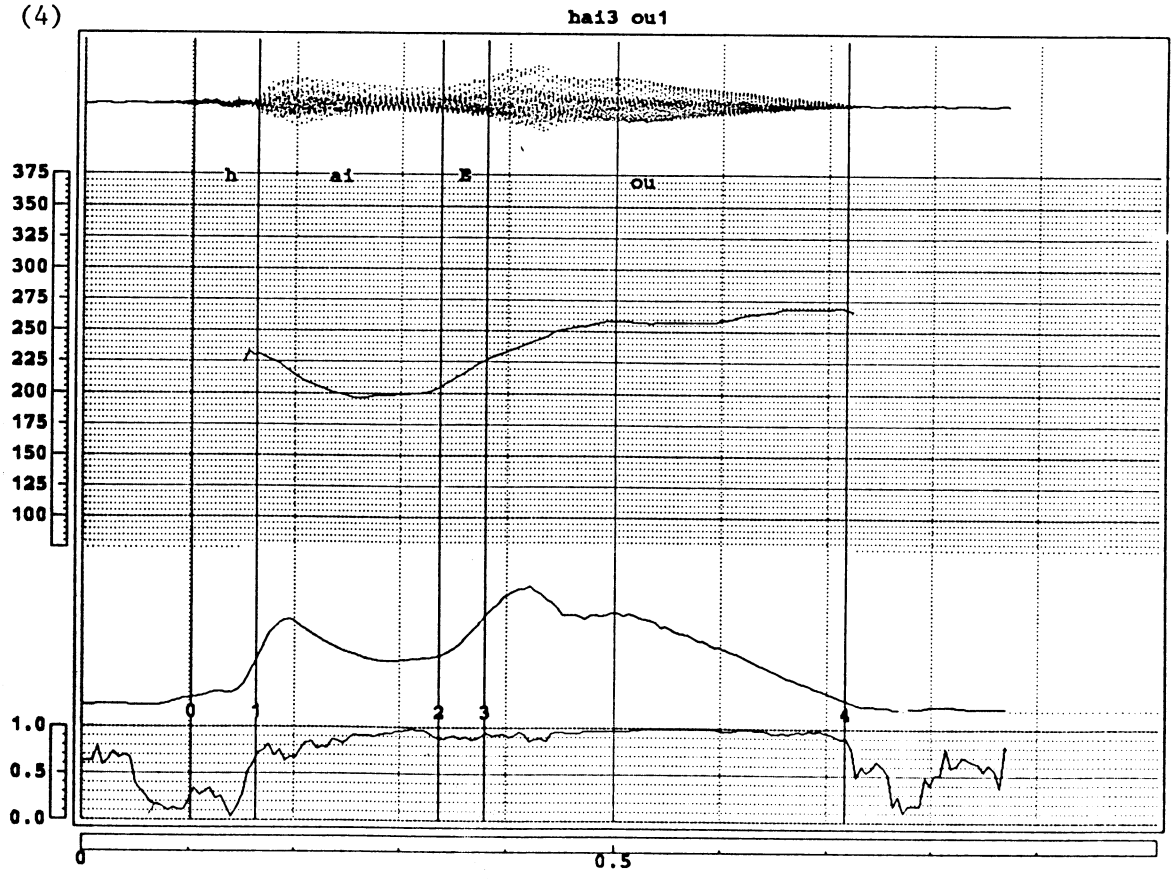


Figure (5)

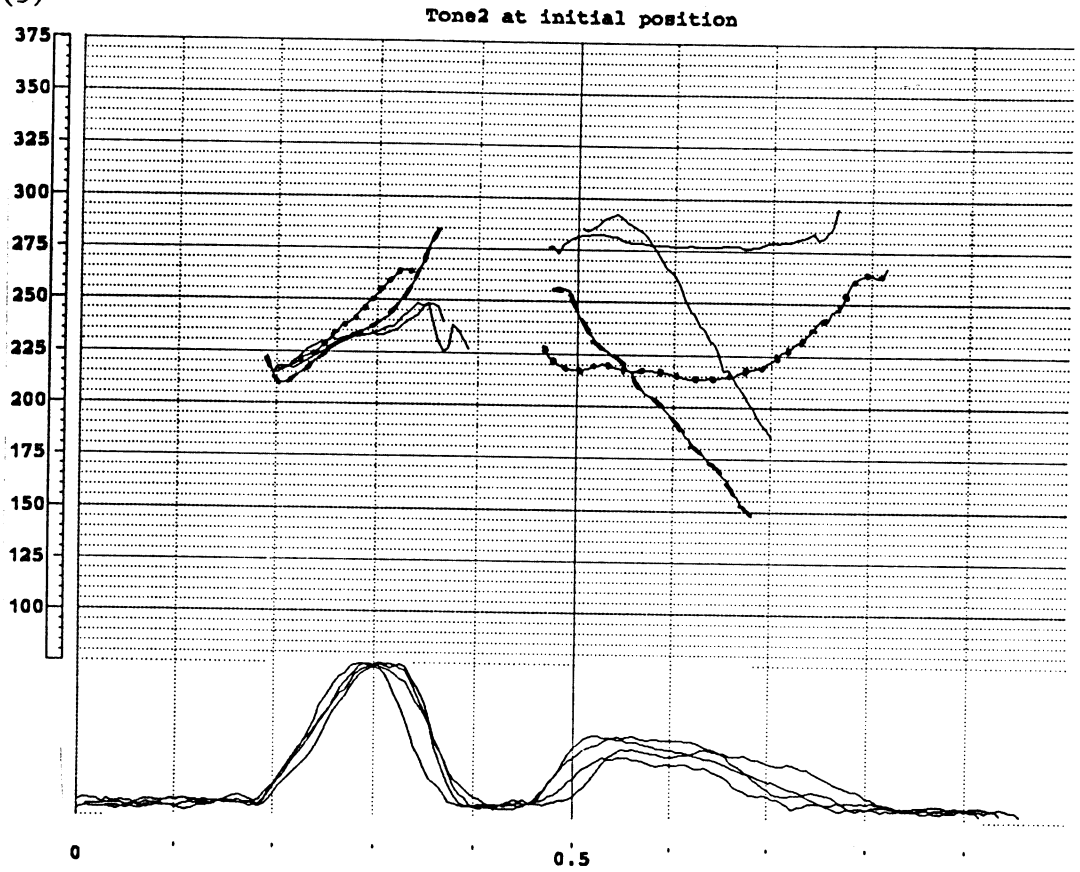


Figure (6)

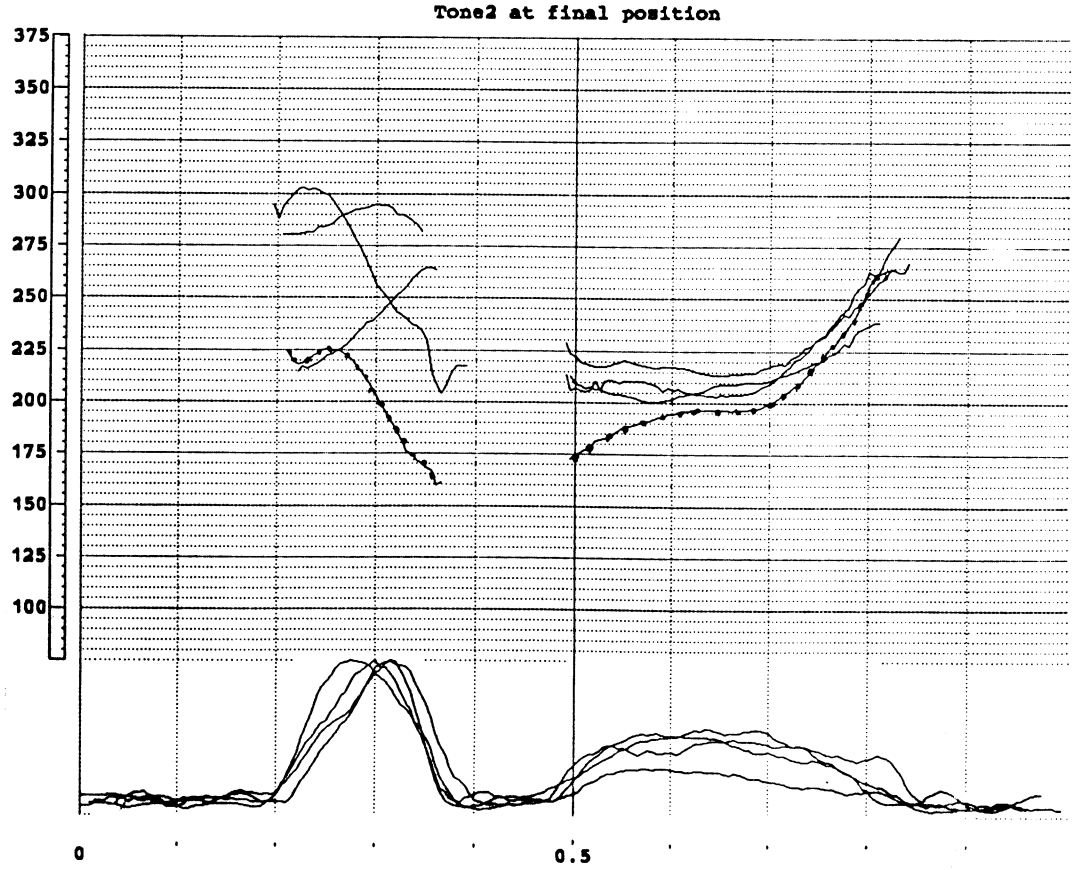


Figure (7)

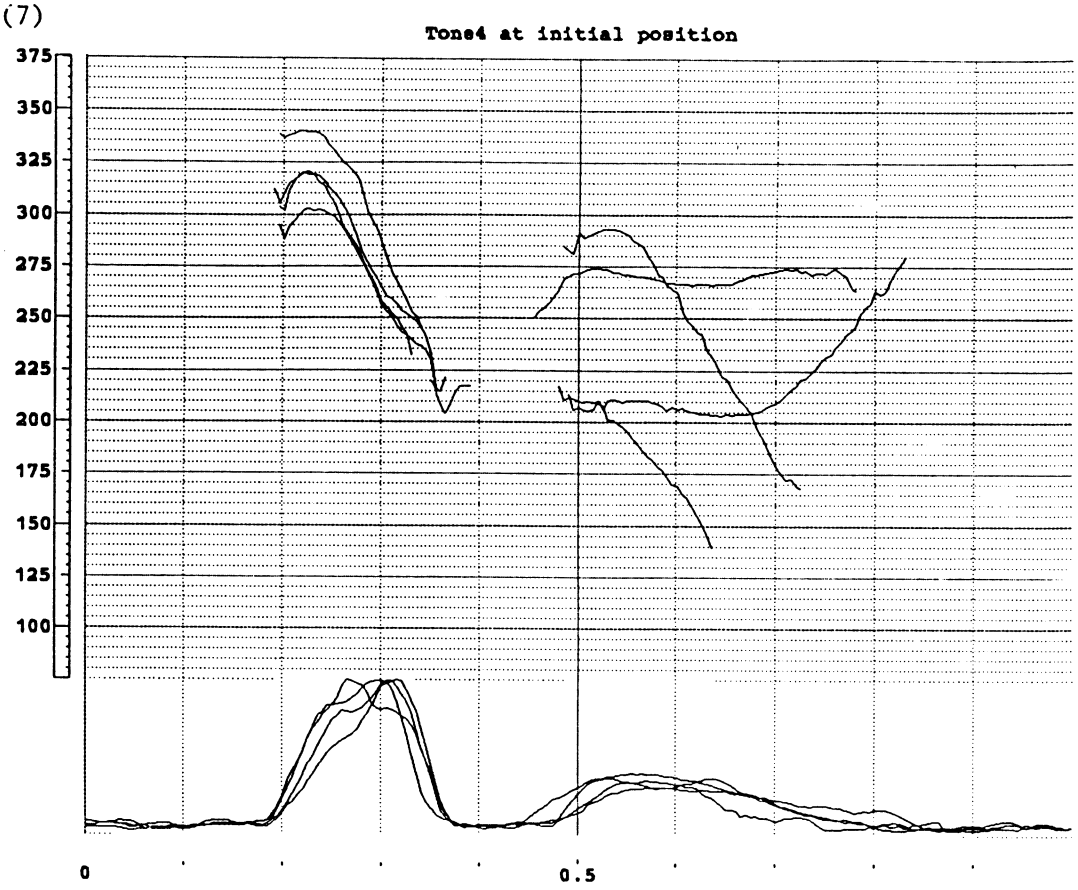


Figure (8)

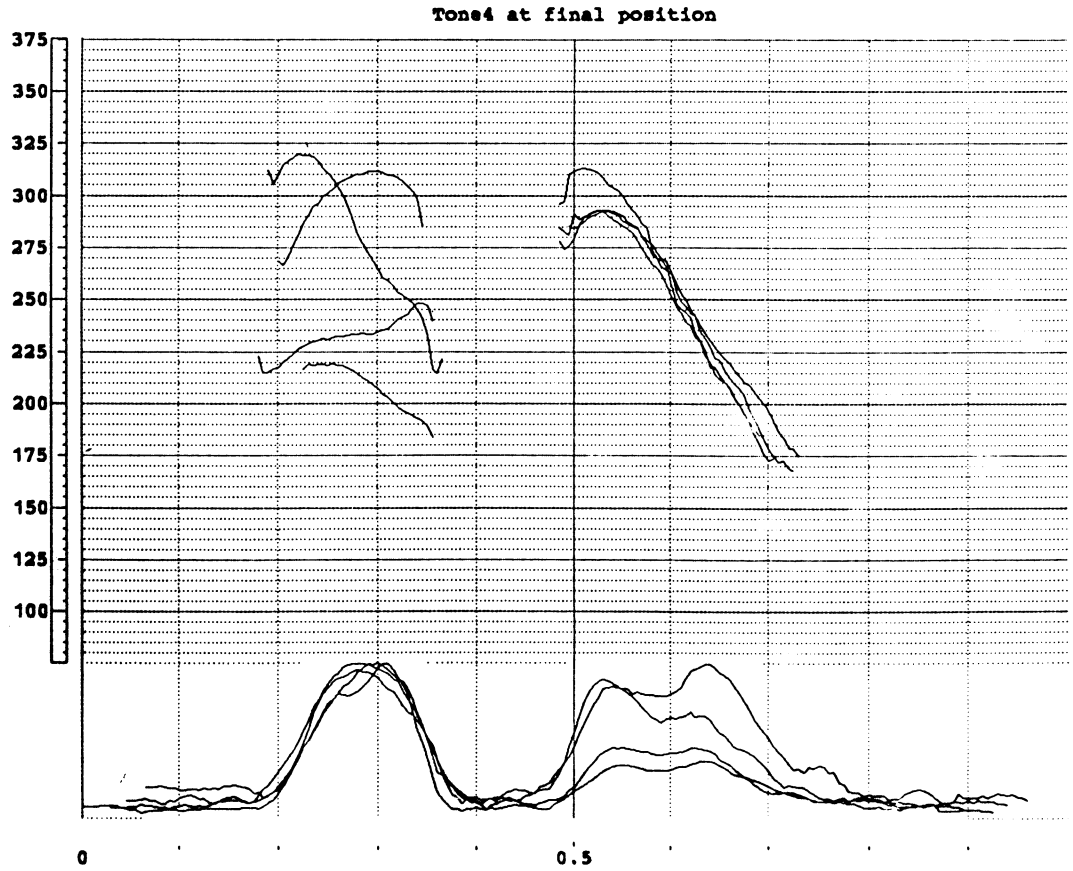


Figure (9)

Figure (10)

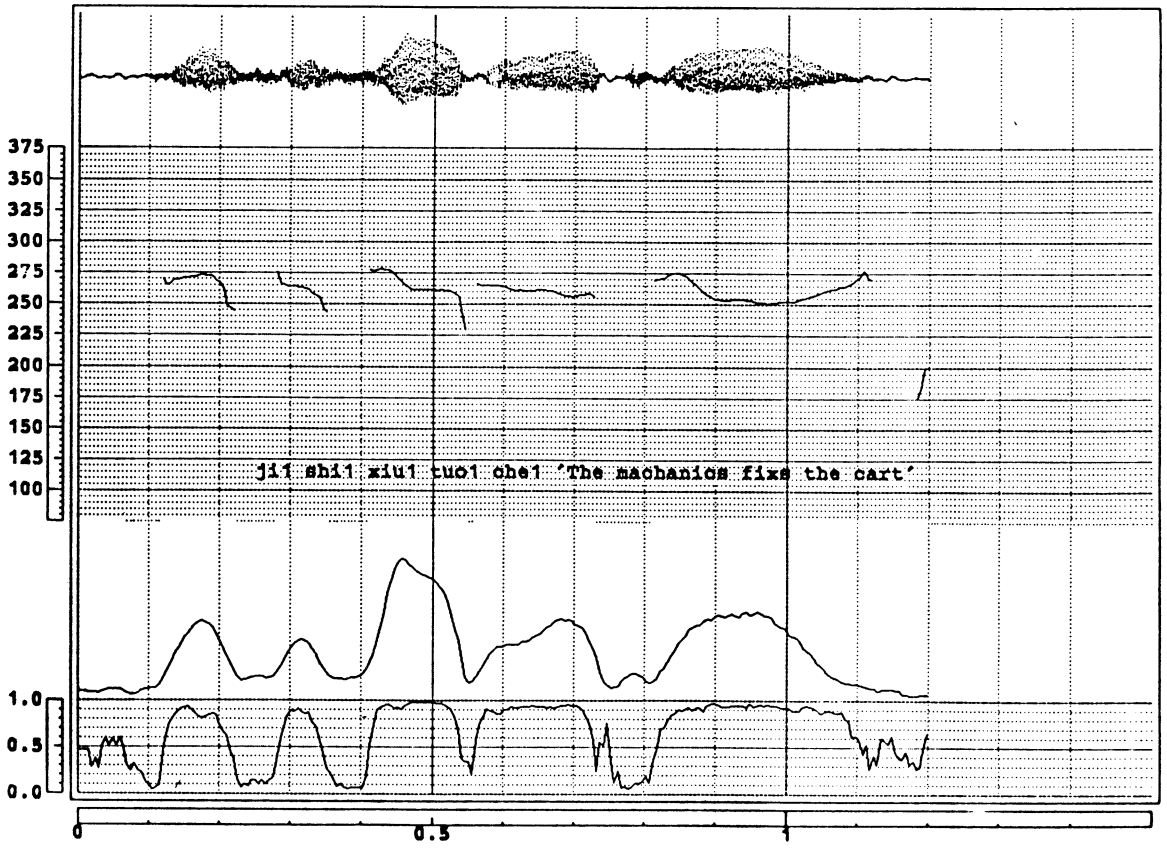


Figure (11)

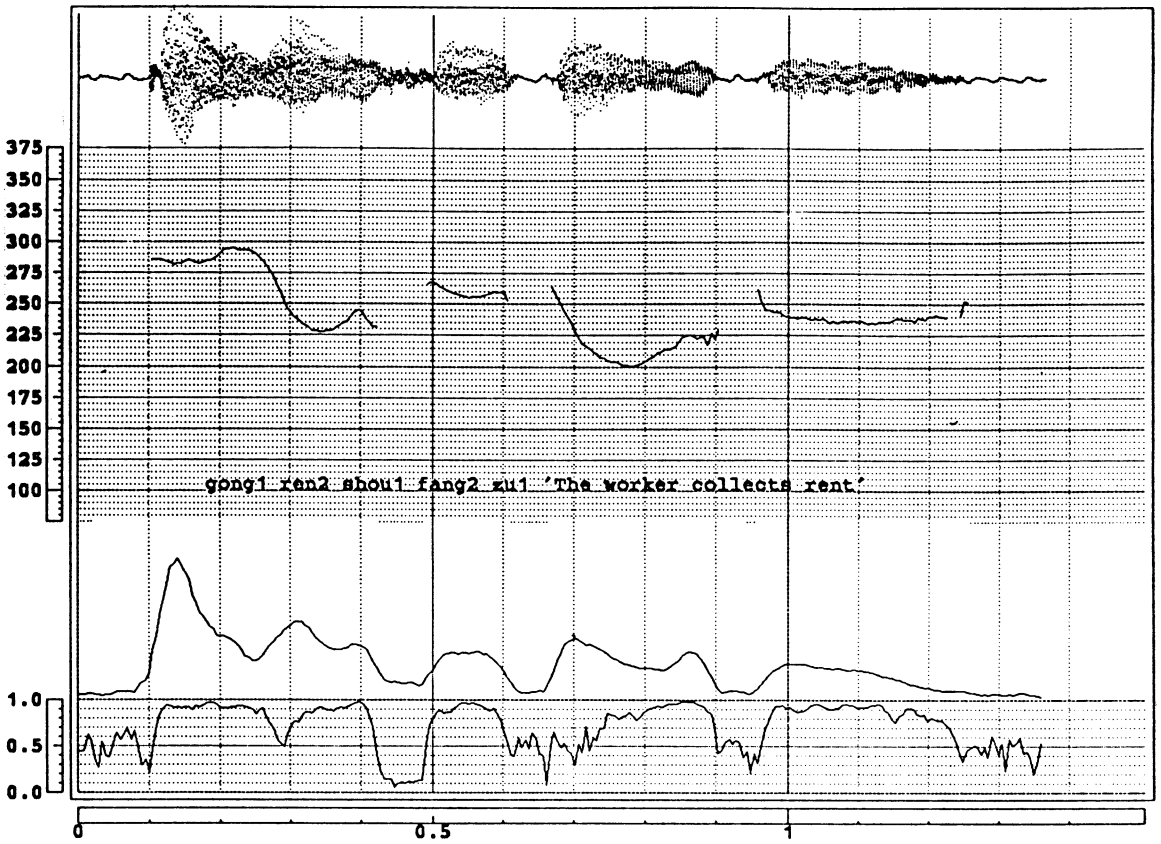


Figure (12)

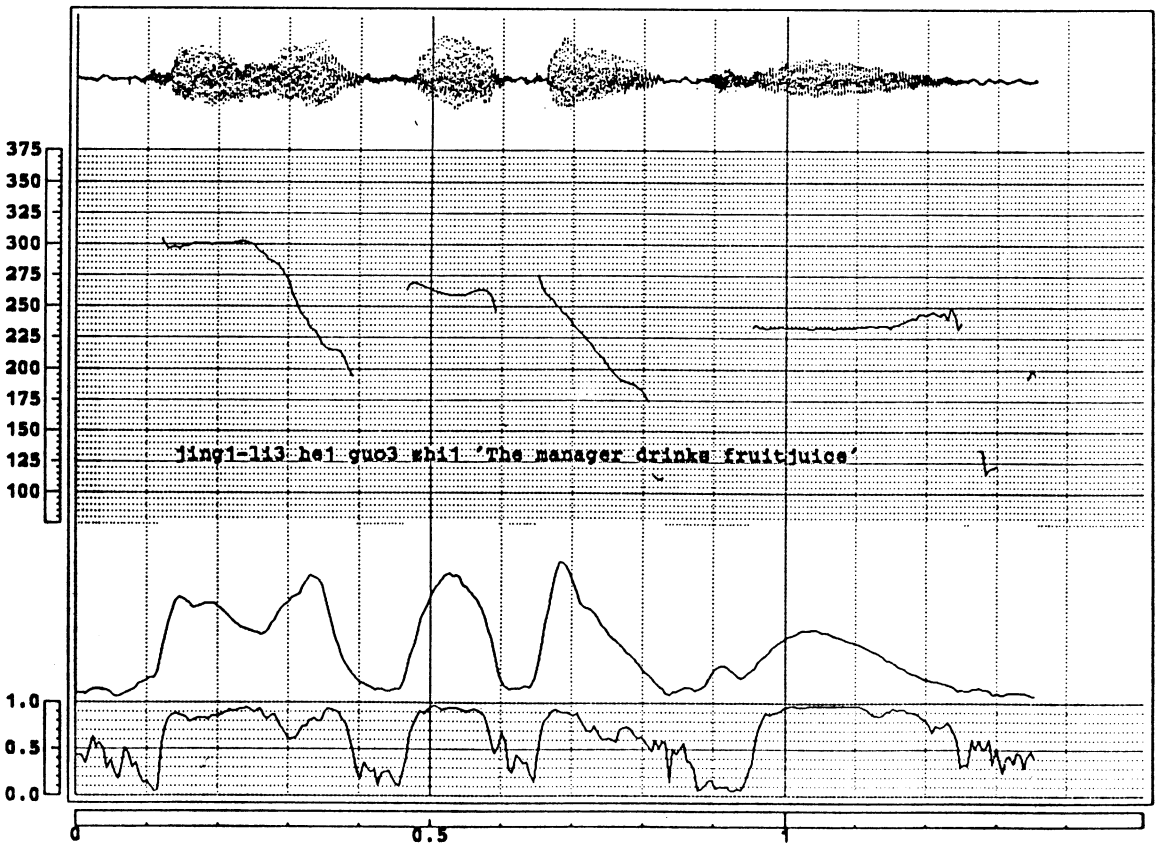


Figure (13)

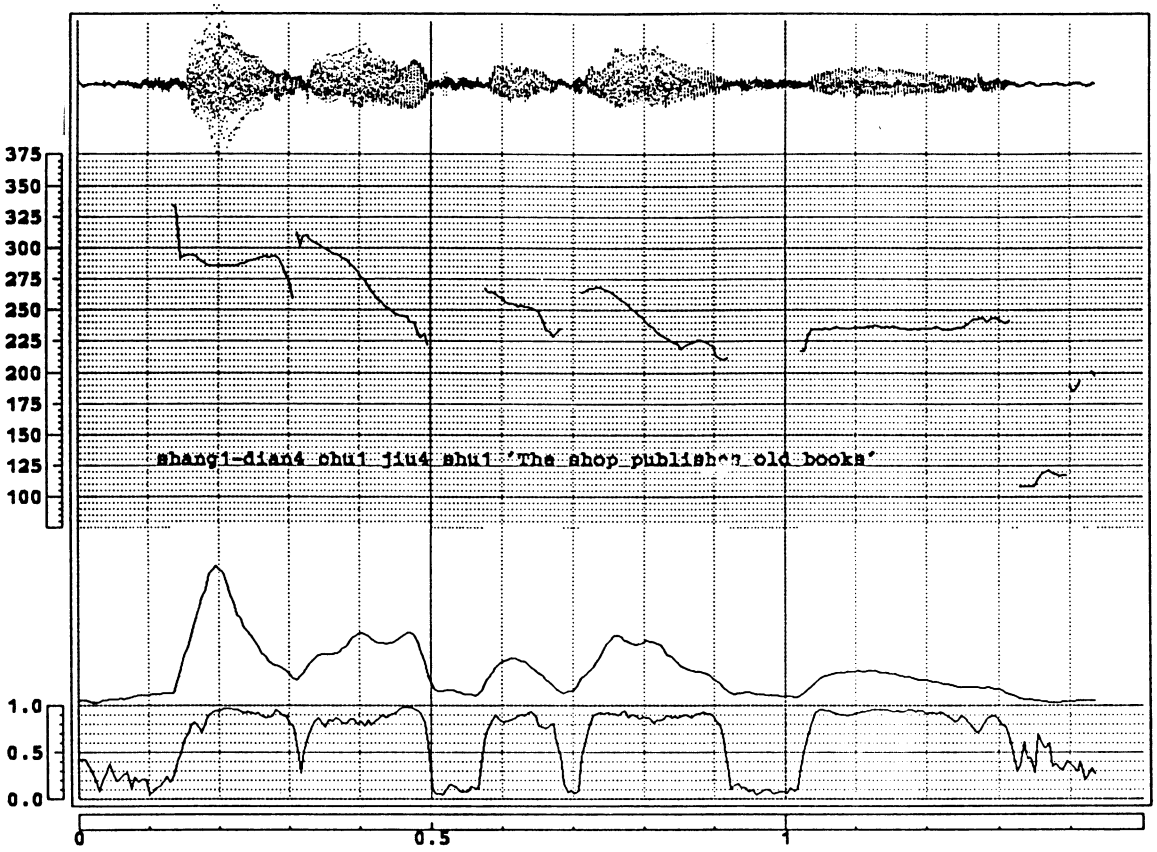


Figure (14)

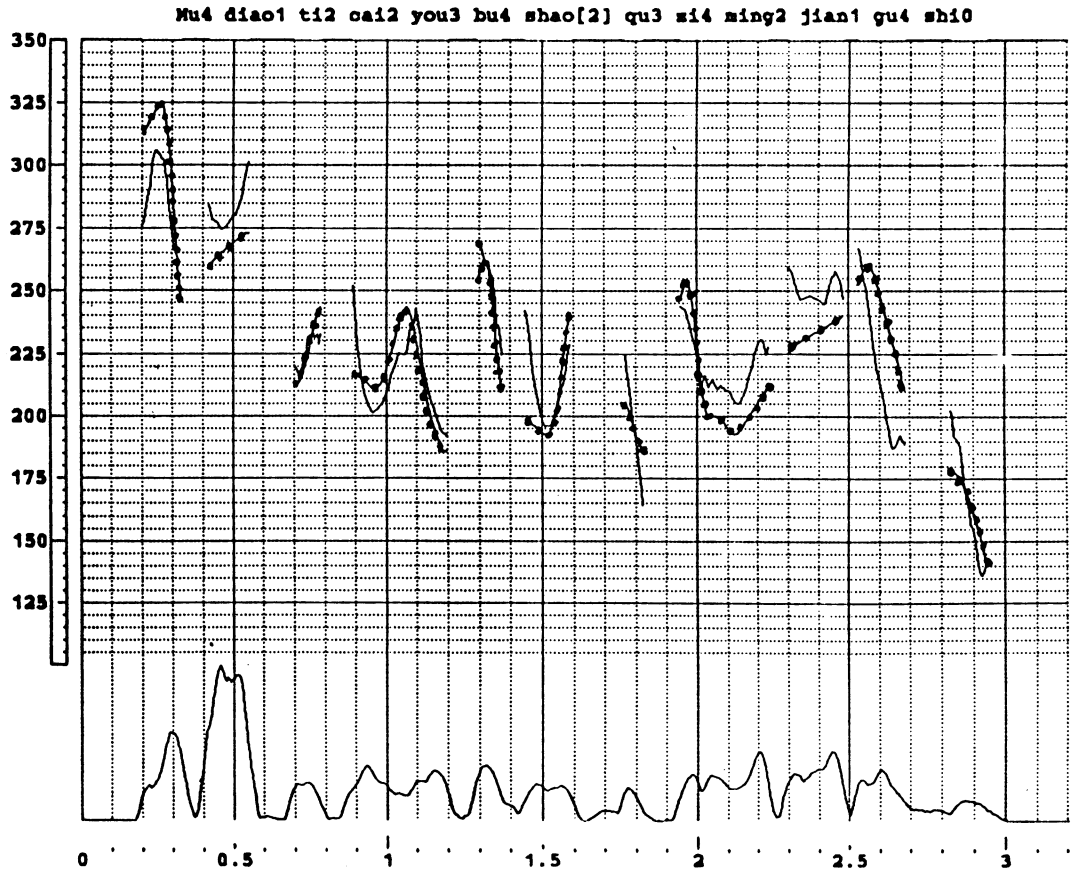


Figure (15)

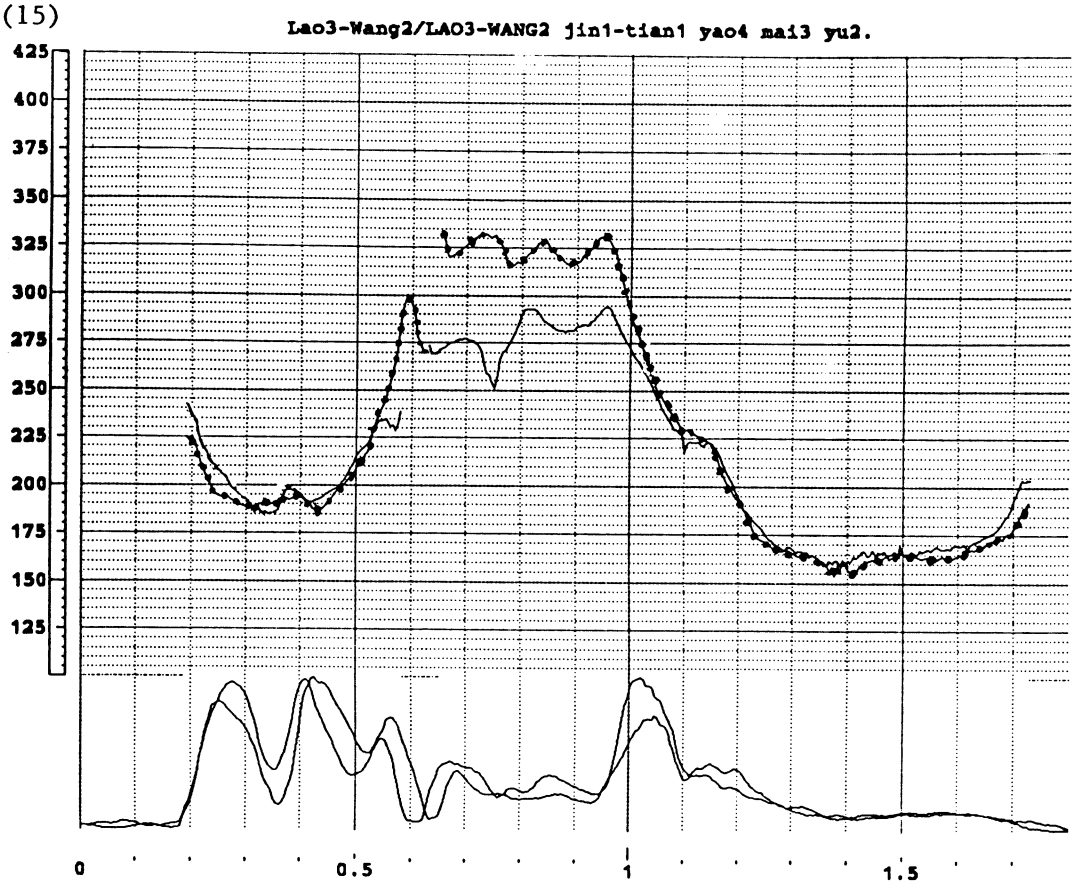


Figure (16)

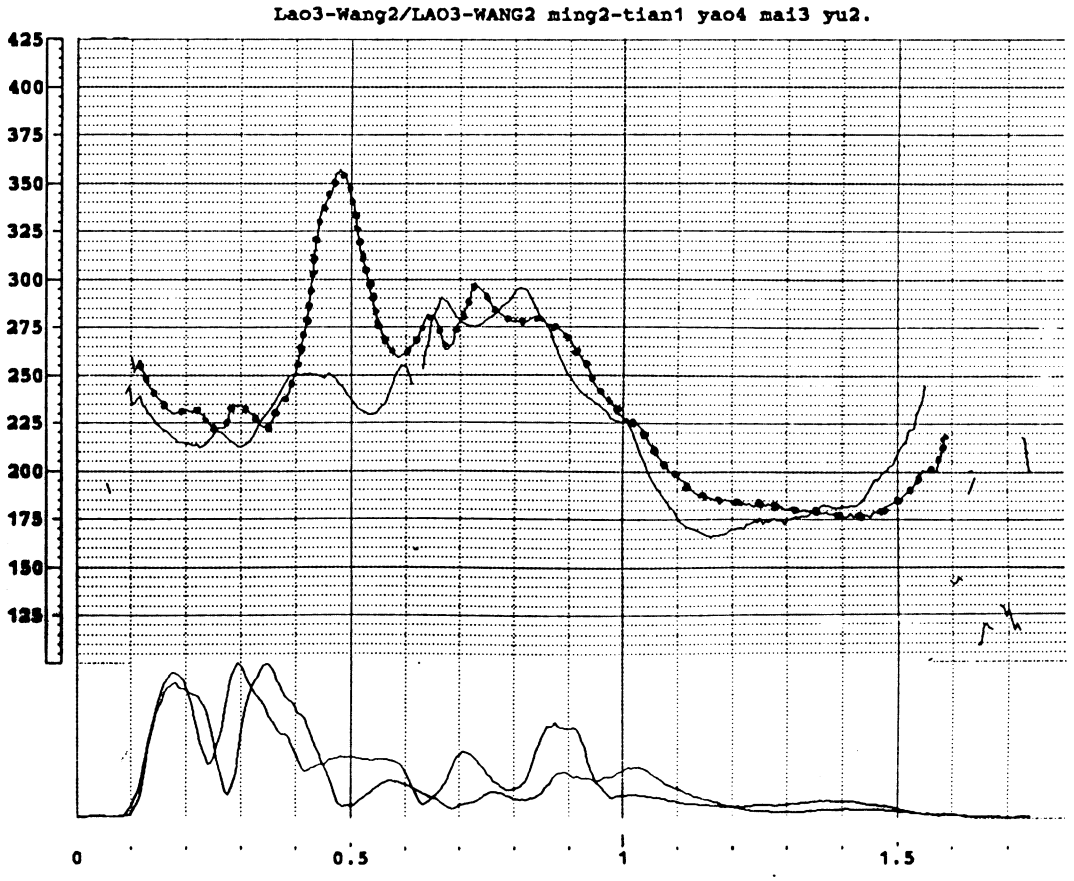


Figure (17)

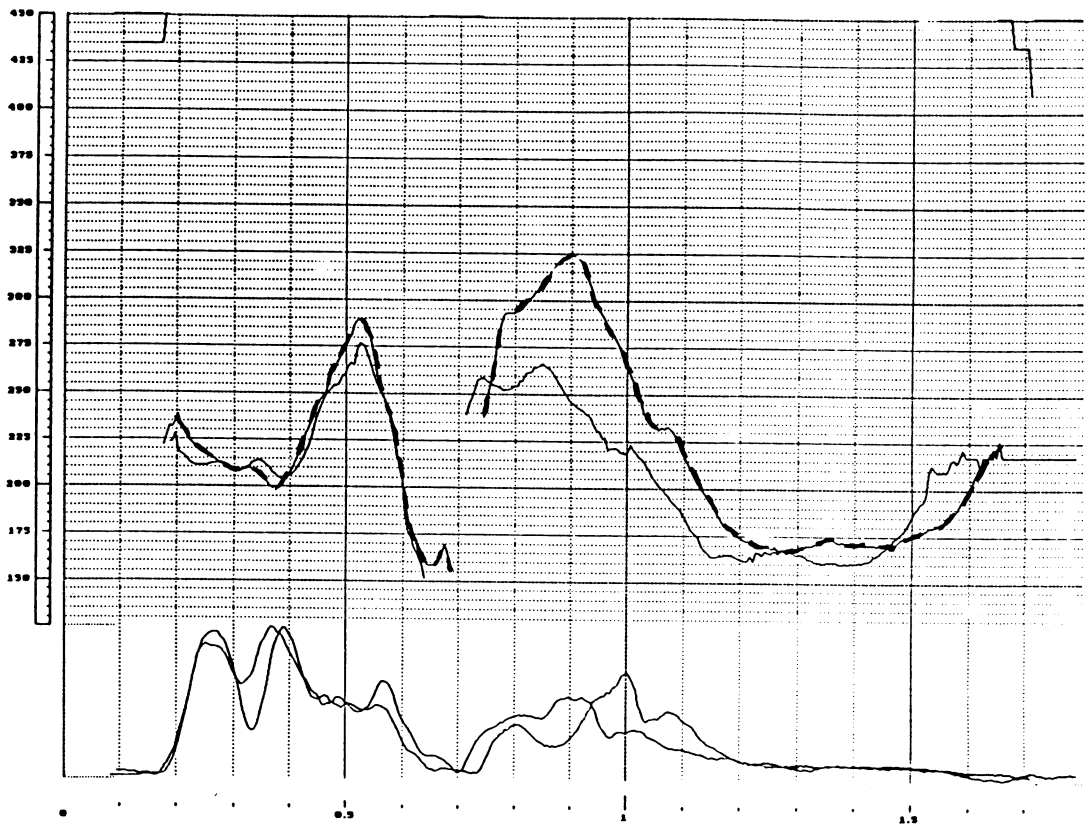


Figure (18)

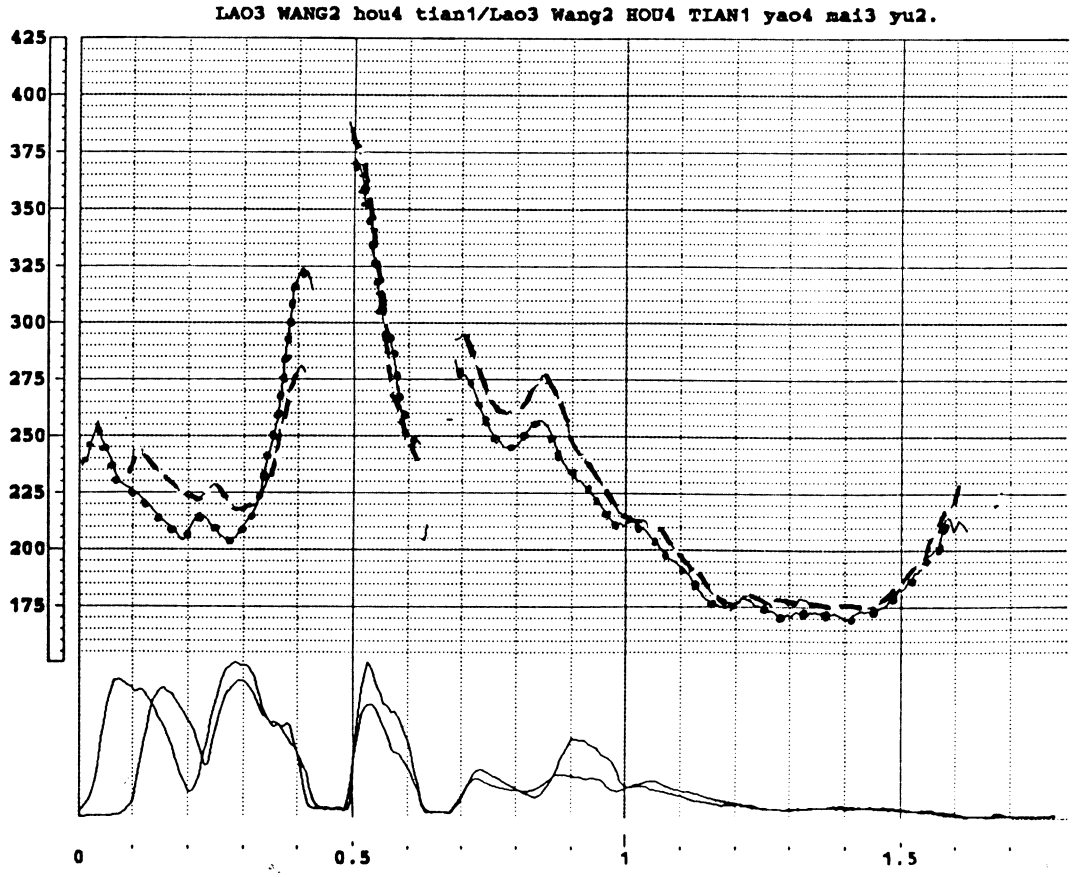


Figure (19)

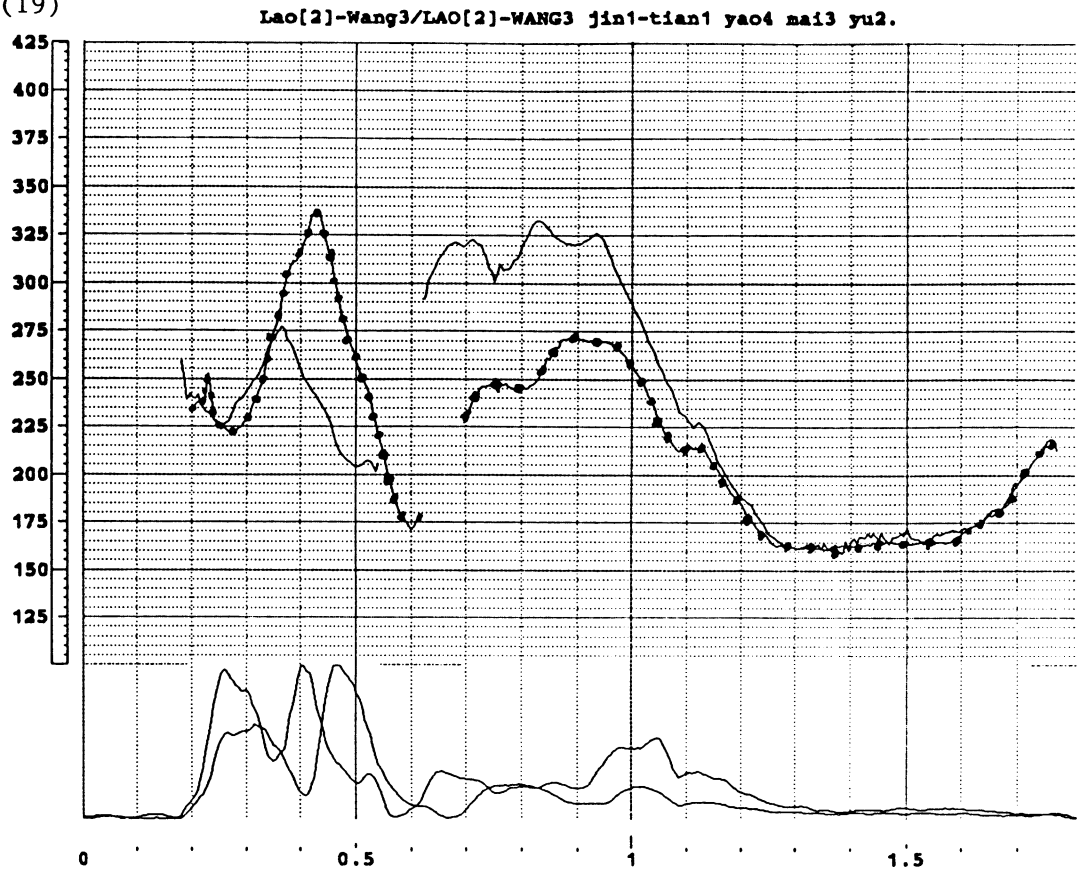


Figure (20)

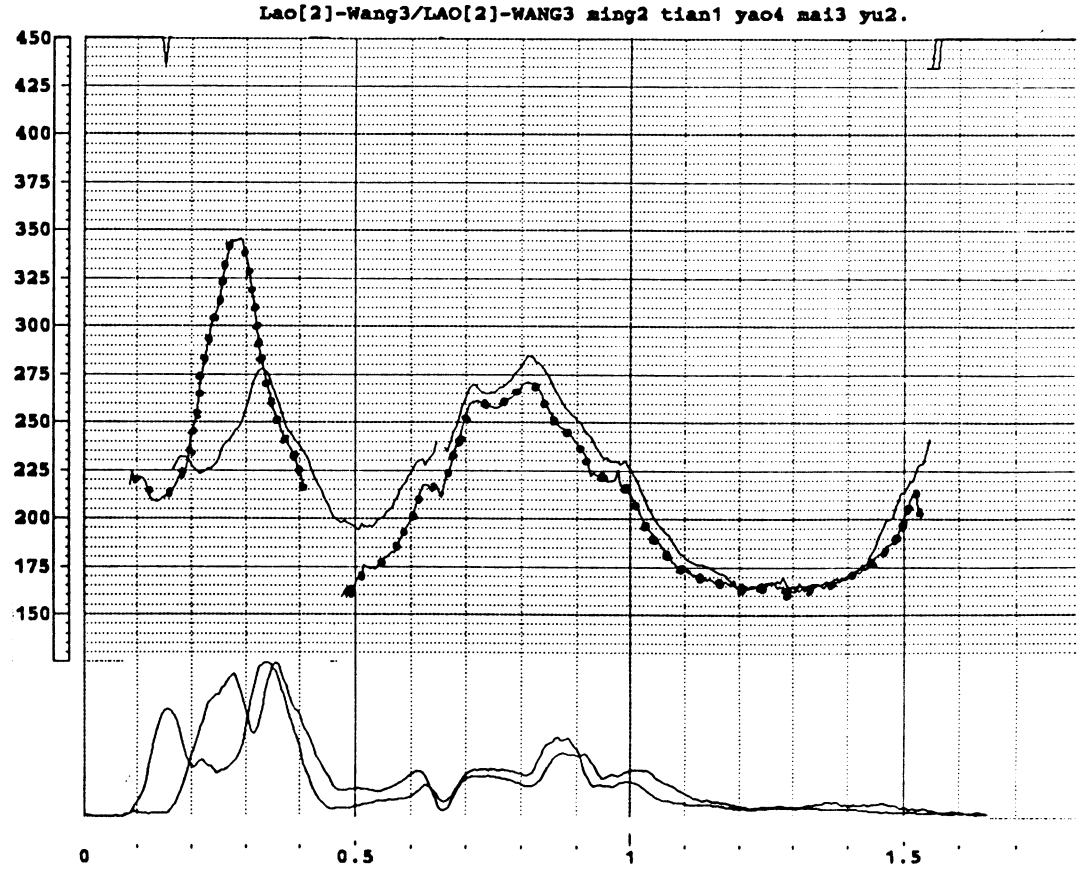


Figure (21)

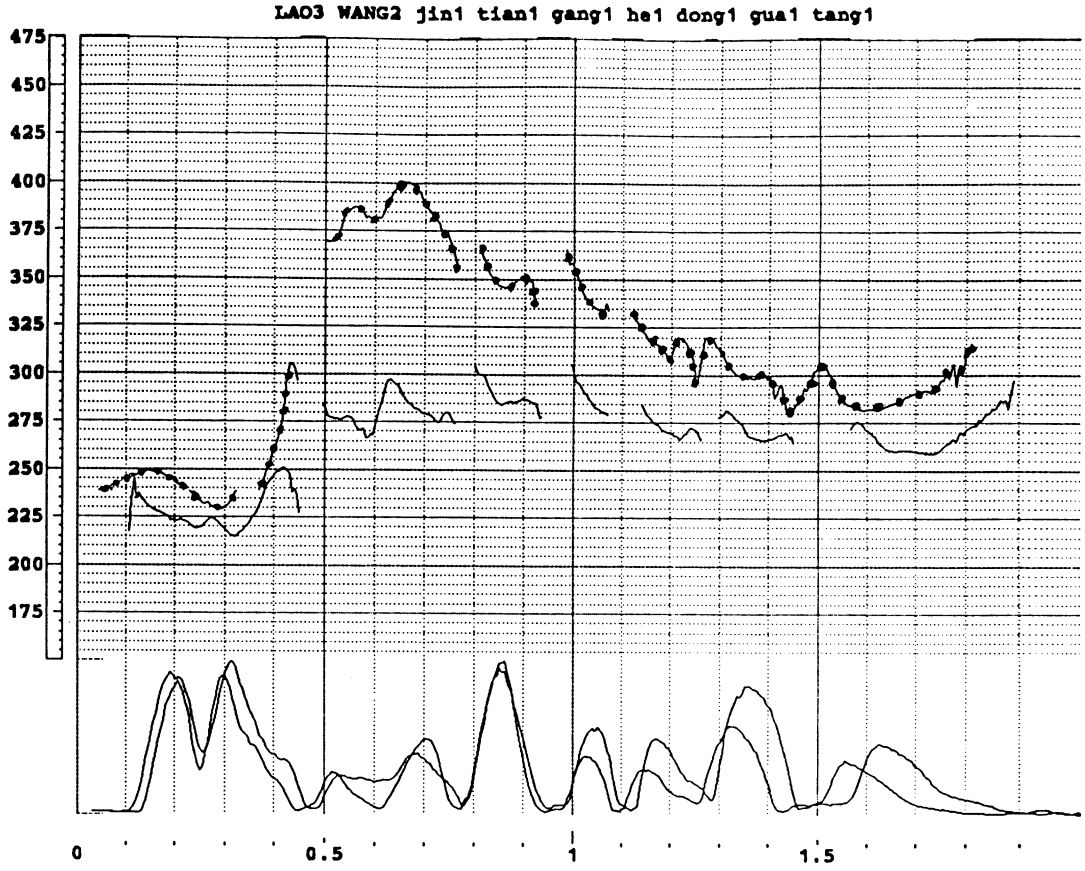


Figure (22)

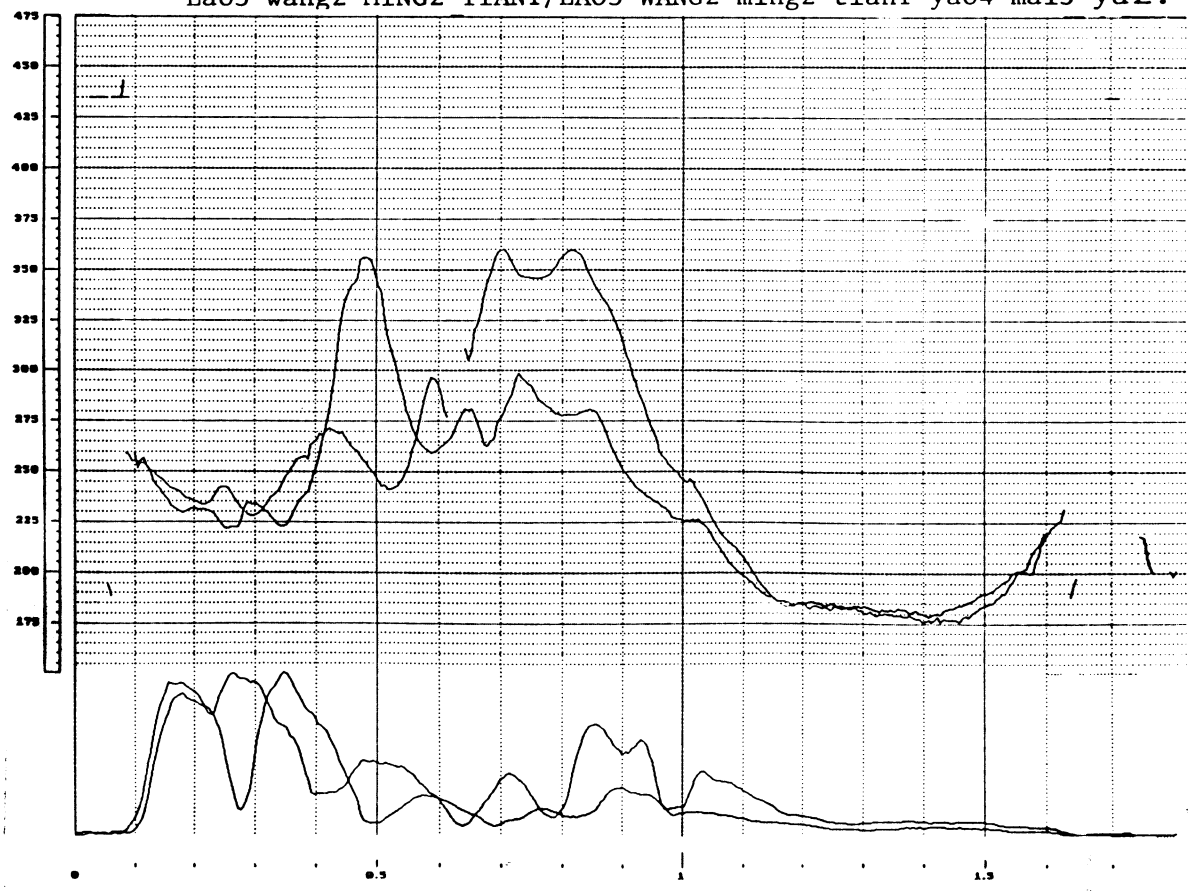
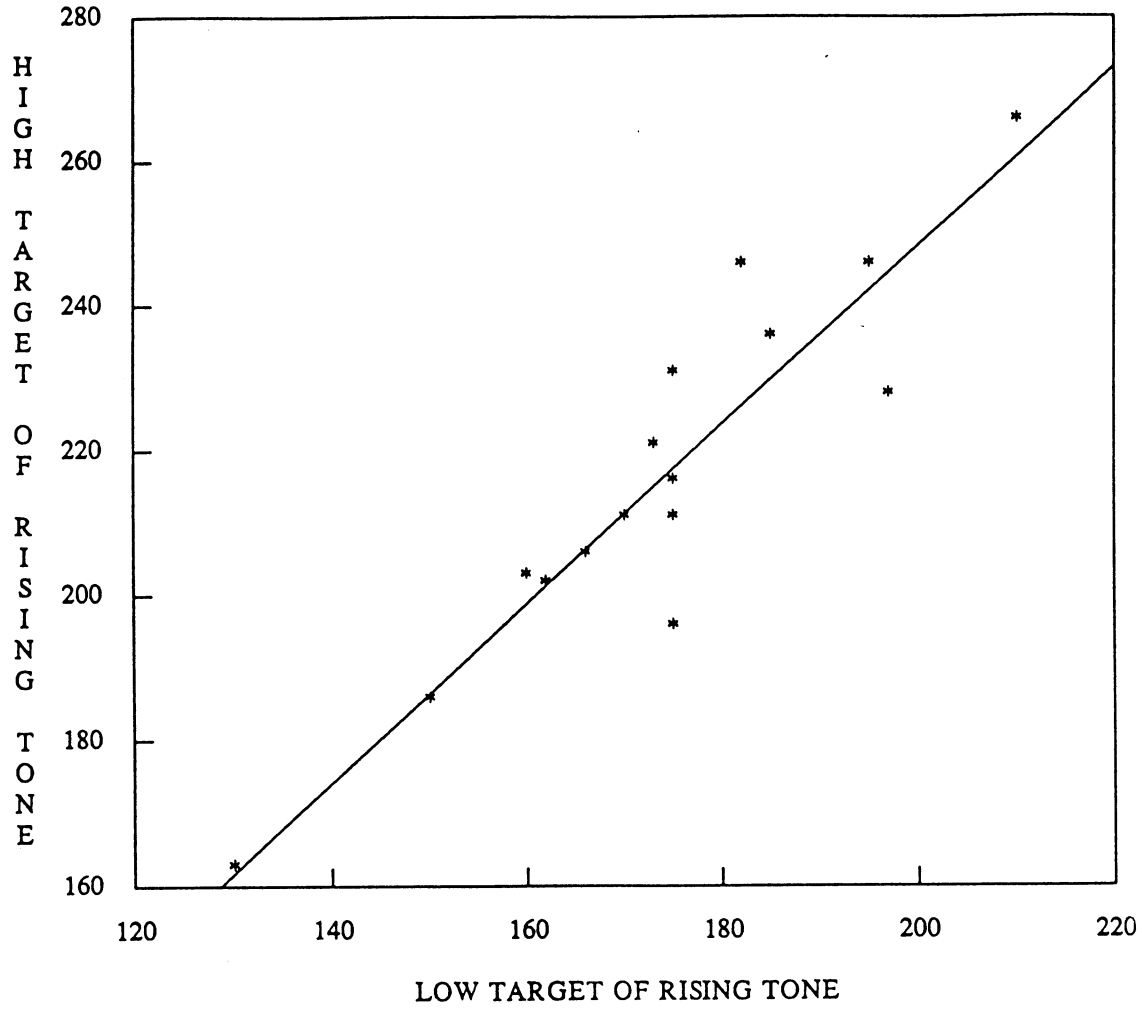


Figure (23)


HIGH TARGET OF RISING TONE
280
260
240
220
200
180
160
120
140
160
180
200
220
LOW TARGET OF RISING TONE

TONE AND INTONATION IN MANDARIN*

CHI-LIN SHIH

CORNELL UNIVERSITY and AT&T BELL LABORATORIES

This paper discusses the basic facts about Mandarin intonation. I start with a description of tone shapes and tonal targets in monosyllabic and disyllabic sequences, then proceed to discuss other factors such as catathesis, prominence, and pitch raising and lowering in discourse structure. Although we possibly can never give a exhaustive list of various factors that may influence the realization of F0, a step by step study will bring us closer to our goal.

1. MANDARIN TONES IN ISOLATION

Mandarin Chinese distinguishes four lexical tones: high level, rising, low falling, and falling, which are traditionally referred to as tone 1, tone 2, tone 3 and tone 4 respectively. A numeral following the segmental transcription refers to the designated tone. Figure (1) illustrate a set of minimal pairs with the syllable *ma*: *ma1* 'mother', *ma2* 'hemp', *ma3* 'horse', and *ma4* 'to scold'. The figure is a display of the time function of F0 values, with the y axis representing F0 in Hz, and the x axis representing time. In addition to four lexical tonal contrasts, a limited number of lexical items, mostly suffixes and sentential particles, do not have an underlying tone, and their F0 values are predictable from the surrounding tones. The absence of a tone on a syllable is traditionally referred to as neutral tone or tone 0.

A high level tone, or tone 1, starts in a speaker's high pitch range and remains high. As the name implies, there is no drastic pitch movement except a slight dip in the middle of the vowel, and a slight rise toward the end of the syllable. A rising tone, or tone 2, starts at a speaker's mid pitch range, remains level, or drops slightly, during the first half of the vowel and rises up to high at the end. A low tone, or tone 3, is phonetically a low falling tone. It starts at the speaker's mid range and falls to the low range. It is often accompanied by laryngealization at the second half of the syllable. A falling tone, or tone 4, usually peaks around the vowel onset, then falls to the low pitch range at the end. In syllables with initial voiceless consonants, the small rising slope is often invisible

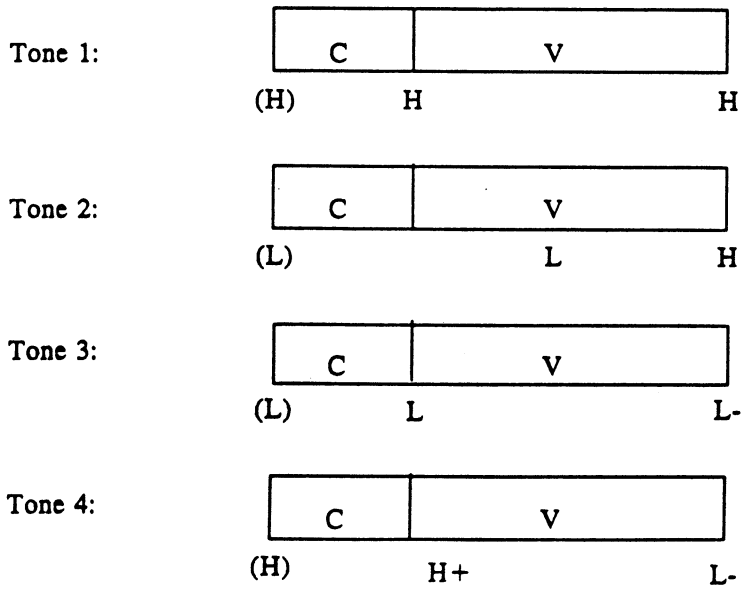
* Most of the experimental works reported in this paper is done while I was doing post-doc research at AT&T Bell Labs. I am grateful for the support that I received from Mark Liberman. This paper is greatly benefited from comments and suggests of the following people: Nick Clements, Mark Liberman, Janet Pierrehumbert, Kim Silverman, and Richard Sproat. I extend my appreciation to them.

and the pitch contour is a straight falling line. In syllables with an on-glide, the pitch is rising through the glide and gives the impression of a delayed peak.

The shape of tone 3 varies the most among all tones due to phonological processes. The low-falling pattern shown in Figure (1) has the highest frequency in speech. It is the pattern that occurs in non-final position. In isolation and in sentence final position, tone 3 may have a rising tail, and that is known as the falling-rising tone. Southern speakers often keep the low-falling pattern even in the final position in casual speech, and use the falling-rising pattern only in deliberate, emphatic speech, or in yes-no question. Northern speakers often use the falling-rising pattern sentence finally in all speech acts. The falling-rising pattern is considered the base form in much of the tonal literature, as in Woo (1969), Yip (1980), and Tseng (1981), because it is the citation form. Another complication at non-final position comes from a tone sandhi process which converts the first of two low tones into a rising tone, see Cheng (1973), Shih (1986). The falling-rising version of tone 3 has the longest duration among all tones, whereas the low-falling version has the shortest duration.

It is interesting to note that the beginning and end points of all tones fall on three distinct levels rather than scattering across a continuum. Tone 1 and tone 4 both start high, very close to where tone 1 and tone 2 end. Tone 2 and tone 3 both begin in the middle range, while tone 3 and tone 4 both fall to the low pitch range. More than 20 repetitions of each tone with various syllable structures confirm our observations so far. The following table summarizes our understanding of the relative values and the placement of tonal targets for each tone.

TABLE (1)



The table is obviously more complicated than a phonological representation consisting of only H and L. However, the additional variation is minimal. H+ represent a pitch level slightly higher then H, and that corresponds to the peak of tone 4. L- is lower than L, that is the end point of tone 3 or tone 4. The values of H, H+, L or L- have to be adjusted for individual speakers and for style of speech.

I included tonal targets in parentheses at the beginning of the consonantal region in order to account for the tone shapes of the consonant region in isolation and in sentence initial position. In non-initial position, the consonantal region is where the tonal transition occurs; see Figure (2). The F0 values there are derivable by interpolating surrounding tonal targets, and there is little evidence for the existence of a real target.

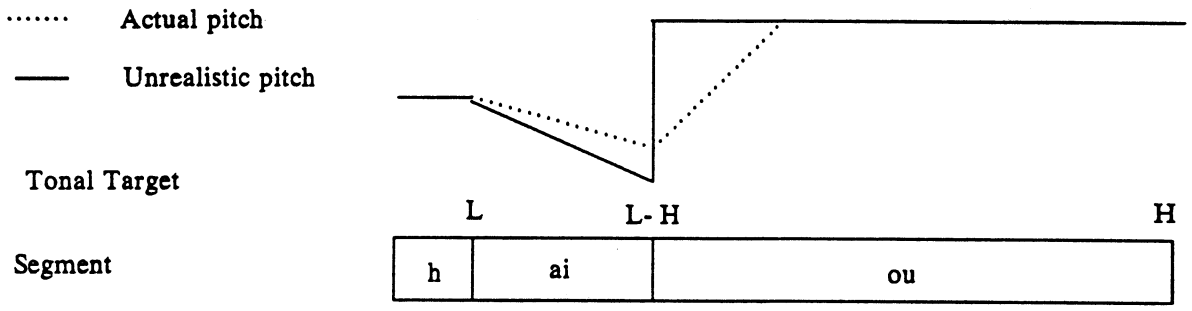
Contra Garding (1987), who proposes that the "turning points" of pitch contours are located at C/V boundaries, I find that a more flexible placement of targets as suggested in the above table gives us the best fit for tones in isolation as well as in connected speech. According to my data, tone 1 and tone 3 do have targets at the C/V boundaries, while tone 4 has slightly delayed target, furthermore, tone 2 doesn't have a target until the middle of the vowel. The delayed target of tone 2 ensures a relatively late rising contour, and predicts correctly that the first half of tone 2 varies according to the tonal transition in connected speech.

Table (1) enables us to generate stylized tonal contours that match tones in isolation. From there, we can proceed to study the more complicated interactions of tones, and tone and intonation. I should make it clear that I am not assuming that tones in isolation are the underlying tonal values, for monosyllables are subject to intonation effects just as longer sequences are. But by comparing tones in isolation to tones in longer sequence, and by controlling our test sentences, we have a good chance of isolating individual effects, and being able to predict pitch contours of uncontrolled sentences. To illustrate where we are and how far we need to go, Figure (3) compares tones generated by Table (1) to the natural pitch contour of the sentence *Mu4-diao1 ti2-cai2 you3 bu4-shao[2] qu3 zi4 min2-jian1 gu4-shi0*, 'The themes of wood carving are often taken from folklores'. I assign 250, 300, 200 and 150 to H, H+, L and L- respectively. The generated pitch countour requires improvements in many aspects. Firstly, some tonal targets need adjustment due to tonal co-articulation; secondly, the scaling of F0 values need more control. I turn to tonal co-articulation in the following section, and discuss two of the pitch scaling factors, catathesis and paragraph structure, in later sections.

2. TONE SEQUENCES

Firstly, we need to investigate tonal co-articulation. Pitch movement in natural speech is gradual. It takes time to reach a L target from a high point,

and possibly takes more time to rise to a H. Tonal targets within a syllable, say, H and L of a falling tone, are not adjacent to each other on the time scale: H is placed toward the beginning of the vowel and L at the end, and there will be sufficient time for the pitch to fall. The physical constraint limits the number of tonal targets in a syllable, and explains why a zig-zag tone is unheard of in natural languages. When different tones are strung together in words and sentences, the situation often arises that adjacent tonal targets have opposite values. That is where tonal co-articulation is expected. Some problems are resolved by not positing a tonal target at the beginning of the consonant. Doing that frees up the consonantal region for tonal transition. Figure (2), *mai3 mao1*, 'to buy a cat', illustrates this point. When there is no consonant on the second syllable, the tonal transition takes place at the beginning of the vowel, as if part of the vowel is interpreted as functioning as a consonant. The tone and segment template of *hai3 ou1*, 'seagull' is shown below. The solid line represents the unrealistic pitch by taking the face value of each tonal target. The dotted line gives the adjusted pitch contour that is closer to natural speech. The actual pitch track is shown in Figure (4).



Theoretically, there are several other possibilities to resolve the transition problem: a target may move backward; the value of H and L may be neutralized; or some targets may simply be deleted. In Mandarin, tonal targets rarely move back. Peak delay and adjustment of pitch level are quite common. In the template above and in Figure (4), the end of the syllable *hai3* is not as low as it would be in isolation.

The following set of data allow us to look into tonal co-articulation of all disyllabic tonal combinations. They consist of 16 minimal pairs that have the same segments "fu-ji" but contrast in tones. The whole set was recorded three times in random order. *Fu3-ji3* changes to *Fu[2]-ji3* as a result of tone sandhi, reducing 16 tonal combinations to 15 on the surface. Words/phrases in the *fu-ji* set differ in syntactic/morphological structures but are similar in prosodic structure: all are in a foot and with primary stress on the final syllable.

fu1-ji1	V N	'to hatch a chicken'
fu1-ji2	N's N	'husband's residence'
fu1-ji3	N N	'a hatchet and a spear'
fu1-ji4	Mod N	'apply-medicine, ointment'
fu2-ji1	$(V\ N)_{V/N}$	'planchette'
fu2-ji2	Mod V	'ambush'
fu2-ji3	V N	'to hold onto a spear'
fu2-ji4	Mod N	'the medicine that should be taken internally'
fu3-ji1	Mod N	'a spare engine'
fu3-ji2	Mod V	'dive to attack'
fu3-ji3	V N	'to assist oneself'
fu3-ji4	$(Mod\ V)_{N}$	'the assistant in a sacrifice ceremony'
fu4-ji1	Mod N	'abdominal muscles'
fu4-ji2	V N	'to carry the book case, to study at far away place'
fu4-ji3	V N	'to carry the spear, to fight'
fu4-ji4	Mod V	'attached-mail, to mail with an attachment'

Tables (2) and (3) below list the mean values of the tonal targets of *fu* and *ji* respectively. Both tables arrange the tones in columns and the tonal contexts in rows. Tones in isolation are given in the first row for comparisons. Pitch trackings of some sample sets are given in Figures (5)-(8).

Table 2 Tones in the Initial Position

	fu1	fu2	fu3	fu4
isolation	266-266	211-253	214-154	287-159
before tone 1	258-270	219-**245**	213-176	299-**218**
before tone 2	274-291	216-257	223-168	300-**227**
before tone 3	273-290	223-281	__	310-**238**
before tone 4	268-286	216-**243**	225-178	300-**227**

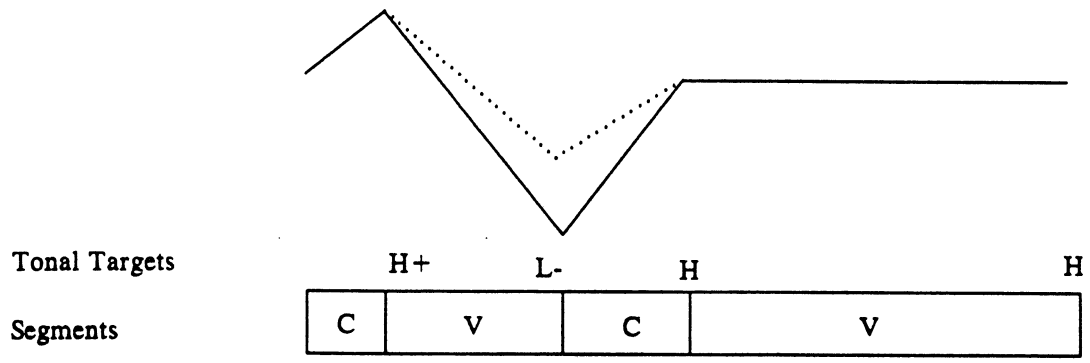
Table 3 Tones in the final Position

	ji1	ji2	ji3	ji4
isolation	259-258	209-262	211-153	291-162
after tone 1	269-276	209-235	219-144	296-176
after tone 2	266-273	217-249	**247**-144	281-177
after tone 3	263-272	**188**-258	_	284-175
after tone 4	262-266	207-252	201-141	278-174

There may be a reason behind every variation we see in the tables above. Some of the variations are not consistent across all syllable types as discussed in Shih (1987). They are possibly affected by consonant perturbation (Silverman 1987 and works cited therein) and intrinsic vowel height (Steele 1985 and works cited therein), two topics that I won't address in this paper. Details aside, I highlight the most consistent variations in boldface and discuss them briefly. The main issues could be accounted for by three most general tonal co-articulation rules below.

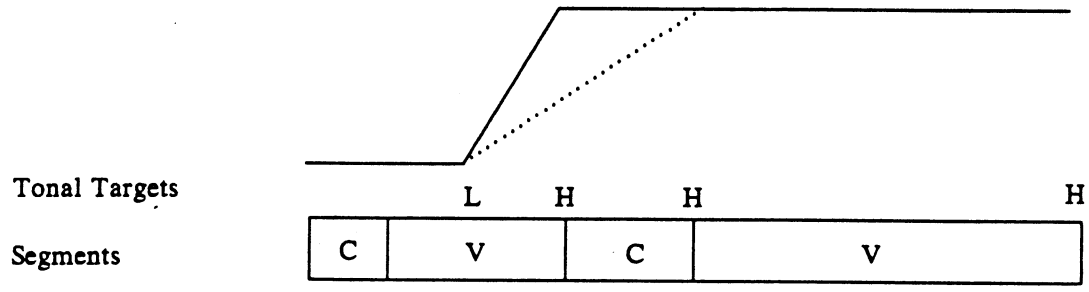
(1) Non-final tone 4 ends in L, rather than L-.
(2) The final H of tone 2 is deleted if the following tone starts with H.
(3) The beginning L of tone 3 tends to assimilate to the previous H.

In the initial position, as shown in Table 2, a tone 4 never reaches the L- target. Instead, the final value is comparable to the L value at the beginning of a tone 2 or tone 3 in isolation. Moreover, a following tone 2 or tone 3 starts exactly where tone 4 ends. I schematize this situation below. The dotted line shows the changes in pitch contour caused by coarticulation,

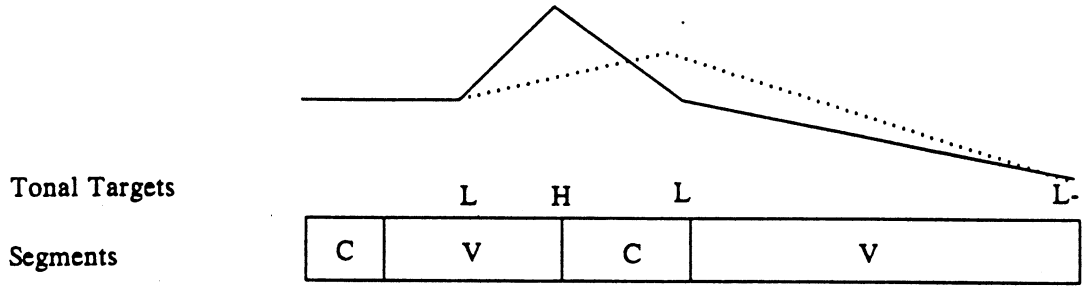


A tone 2 ends lower (245, 243) when the following target is H (tone 1 and tone 4); but ends higher (257, 281) when the following target is L (tone 2 and tone 3). The cause of this situation is not dissimilation. Rather, when a H target follows, the rising slope ends not at the end of the tone 2 syllable, but ends where the following H target is. This phenomenon suggests that the final H of a tone 2 is deleted in the presence of the next H target. Tone 2 seems to end lower in this context only because the value at the syllable boundary does not represent the H

target, but is a point on the rising slope. When the following target is L, the position of the final H is unaffected and the pitch value at the end of the syllable reflects the H value. The right-shift of the H target explains the surface dissimilation effect. The dotted line in the following picture shows the modification.



Everything being equal, tone 2 and tone 3 usually start with the same pitch height in isolation. We account for this by giving them the same beginning target L. However, when the preceding target is high, the beginning point of tone 3 is very often higher than that of tone 2. This situation is unexpected because the pitch value at the vowel onset of tone 2 should be somewhere between H and L, due to a delayed initial target, and be higher than a tone 3 with an earlier L target. The unexpected result arises from rule (3) above. The initial L of tone 3 is sometimes pulled up in the direction of the preceding H. This happens more frequently in fast speech across sonorant consonants. And the effect is more apparent when tone 2 precedes. It seems to me that the falling pitch (with absolutely no initial rise as in tone 4) and the very low ending is the characteristic of tone 3. The height of the initial target is less rigid. The co-articulation change is shown below.



Some other general points shown in the above tables include the following. An utterance initial H target has a much wider range than a L target. Initial H of tone 1 falls between 250-280 Hz, and H+ of tone 4 280-335 Hz. There is a difference of 30-50 Hz. The difference among utterance-initial L target is less than 20 Hz. The picture is reversed for the second syllable, where H target falls between 245-275 in tone 1, 260-300 in tone 4, but L target spreads from 200 to 255 in tone 3 and 175 to 225 in tone 2. The deviation of L targets is largely due to tonal co-articulation, for L target tends to be deleted or be assimilated to the

preceding target.

In the final position, a tone 2 starts lower when it follows a tone 3 (188 Hz). This situation is explained by the absence of a target at the vowel onset. The L- value of a preceding tone 3 will affect the beginning of tone 2, but not so much for other tones. Tone 4 does not have the same lowering effect because non-final tone 4 ends in L, the same as the initial tone 2 target, so the pitch at the first part of tone 2 would be level.

Figure (9) shows the improvement after implementing tonal co-articulation rules. What needs to be done at this stage is to study the factors that control F0 scaling. We discuss two of these factors below, catathesis and paragraph structure.

3. CATATHESIS

Catathesis (or down-step) refers to F0 lowering due to specific tonal combination. The most common case is in a sequence H L H, the second H is considerably lower than the first because of the intervening L (Liberman and Pierrehumbert, 1984). The following experiment was designed to test the difference in the catathesis effect of tones with a L target, namely, tone 2, tone 3, and tone 4. All test sentences are five syllables long and have tone 1 in the first, third, and final position. The intervening syllables have one of the four possible tones. The catathesis effect, if any, can be measured by comparing the F0 value of the tone 1 syllables.

The data

1. ji1 shi1 xiu1 tuo1 che1.	'The mechanic fixes the cart.'
2. gong1 ren2 shou1 fang2 zu1.	'The worker collects the rent.'
3. Jing1 li3 he1 guo3 zi1.	'The manager drinks juice'
4. shang1 dian4 chu1 jiu4 shu1.	'The shop publishes old books.'

The data set was recorded in random order, 4 times in natural speech, and 4 times in reiterant speech (Liberman and Streeter, 1978), in which all syllables in the target sentence are replaced by *da*, while maintaining the original prosodic pattern. The purpose of reiterant speech is to avoid interference of segmental effects from various syllables in the target sentence. Pitch values of the initial, medial, and final tone 1 were measured and averaged. The measurement was taken from the center area of tone 1, avoiding the consonantal effect at the beginning, and optional final raising at the end. Table 4 and Table 5 list the mean pitch value from reiterant and natural speech respectively. The values are arranged by sentential-position in columns, and by sentence types in rows. The sentences are mnemonically numbered after the conditioning tones in the second and fourth position. Figures (10)-(13) give a sample of each sentence.

Table 4 Reiterant Speech

	Initial	Medial	Final
1	271	260	249
2	278	246	227
3	287	247	217
4	283	247	229

Table 5 Natural Speech

	Initial	Medial	Final
1	284	265	252
2	286	255	239
3	299	260	236
4	280	259	241

In general, the final tone is the lowest and the initial tone is the highest. This situation is found in sentence 1 as well, where initial, medial, and final tone 1 scale down at a rate of 2% per syllable in reiterant speech, and slightly more in natural speech. Since there is no L target in sentence 1, the lowering effect could not have come from catathesis. I attribute it to declination.

In both reiterant and natural speech, the sentences with intervening tone 2, 3, or 4 have lower medial and final tone than the sentence that has only tone 1, suggesting that all tones with a L target have some catathesis effect. There is not much difference on tone 2 and tone 4, while tone 3 exhibits more lowering effect on the following tone.

Roughly speaking, the catathesis effect of tone 2 and tone 4 is 12% on the medial syllable, and 7% on the final syllable. The catathesis effect of tone 3 is 14% on the medial syllable, 12% on the final syllable.

There are two reasons why the medial syllable is lowered more than the final one in sentence 2, 3, and 4. 1. The medial syllable is surrounded by L tones, which might pull down the pitch level. 2. The medial syllable is prosodically weak,[1] and a weak syllable is more susceptible to the surrounding environment.

1. All medial syllables in the test sentences are monosyllabic verbs, which are usually weaker than nouns in Mandarin.

The effects of catathesis seem to be related to the actual pitch level of the preceding L. Tone 3 has the lowest pitch level, L-, in the non-final position, so it has the strongest catathesis effect. Tone 2 starts at the L level, and tone 4 only falls to L in the non-final position, therefore both have less catathesis effect. When tone 4 is followed by a neutral tone, in which pitch falls to a even lower point than tone 3, the catathesis effect is indeed much stronger than all our test sentences here (Shih 1987).

Figure (14) shows the result after implementing tonal co-articulation rules, catathesis effects, some other factors discussed in Shih (1987), and smoothing. Although the two pitch contours do not match perfectly, the re-synthesized speech using rule-generated pitch contour represented by the dotted line already sounds very natural.

4. PROMINENCE

This section investigates the interaction of tones and prominence. There are three questions I am trying to answer. 1. How is prominence realized on tones? 2. Would different combinations of tone sequences affect the realization of prominence? 3. What happens to the post-prominence constituents?

A set of time words *jin1-tian1* 'today', *ming2-tian1* 'tomorrow', *mei3-tian1* 'everyday' and *hou4-tian1* 'the day after tomorrow' is embedded in the context *Lao3 Wang2* ____ *yao4 mai3 yu2* 'Lao Wang wants to buy fish ____.' Each sentence is repeated three times in three ways: 1. plain statement; 2. statement with emphasis on the subject 'Lao-Wang'; and 3. statement with emphasis on the time words. The data are listed below.

Lao3-Wang2 jin1-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish today.'
Lao3-Wang2 ming2-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish tomorrow.'
Lao3-Wang2 mei3-tian1 yao4 mai3 yu2. 'Lao Wang wants to buy fish everyday.'
Lao3-Wang2 hou4-tian1 yao4 mai3 yu2.'Lao Wang wants to buy fish the day after tomorrow.'

There are three additional sentences to test the interaction of prominence and low tone; and the effect on post-prominence level tones.

Lao[2]-Wang3 jin1-tian1 yao4 mai3 yu2.'Lao Wang3 wants to buy fish today.'
Lao[2]-Wang3 ming2-tian1 yao4 mai3 yu2.'Lao Wang3 wants to buy fish tomorrow.'
Lao3-Wang2 jin1 tian1 gang1 he1 dong1 gua1 tang1. 'Lao Wang just drank winter melon soup.'

In the subject *Lao3-Wang2*, *Lao3* is a prefix, thus weak prosodically. When emphasized, the strong syllable *Wang2* receives more prominence effect. The reverse is found in time words. *Tian1* 'day' is a generic noun, which is a prosodically weak member in Mandarin compounds. When emphasized, the strong syllables *jin1*, *ming2*, *mei3* and *hou4* are the loci of the prominence.[2]

Figures (15) to (21) present samples from test sentences. When multiple pitch contours are presented in the same figure, the solid line shows the normal reading, the dotted line shows the sentence with the highest prominence on *Lao-Wang*, and dashed line has the highest prominence on time words. In the following discussion, I use upper case letters to represent a prominent syllable or word, and use lower case for words with less or no prominence.

It is apparent from the figures that prominence is reflected by expanding pitch range: high targets become much higher, while low targets remain at the same level or are slightly lower. Aside from the increased pitch range, more prominent forms also have longer duration and higher intensity.

While the above generalization is true, a closer look of the data reveals a number of surprising details. While longer duration and higher intensity always fall on the most prominent syllable or word, high pitch is sometimes realized on an adjacent but less prominent word. For example, in Figure (15), the most prominent *WANG2* ends at 300 Hz, while the following *jin1-tian1* is 325 Hz high. In Figure (18), we see no difference in the dotted and dashed pitch tracks of *hou4*, when the dashed *HOU4* should have the highest prominence, and the dotted *hou4* is post-prominence, thus a weak syllable.

The mirror image of this situation is also true. Figure (20) shows that the influence of the final target L of *WANG3* extends into the beginning of the next syllable *ming2*, causing a post-prominence tone 2 to begin at a much lower pitch than where *WANG3* ends.

This situation is a reflection of the tonal co-articulation rule (2), where I discussed the deletion of the final H target of tone 2 in the presence of a following H target. In those situations, a tone 2 takes the later H to be its target, and extends the rising slope beyond the syllable boundary. What we see in Figure (15), (18), and (21) is a more dramatic display of the same thing.

Taking the shifted H or L to be the real target, the pattern of prominence structures begins to emerge. Figure (22) compares sentences with focus at different loci. The high peaks reach the same level, even though the peaks may not coincide with the focused syllable. While it is widely accepted that prominence will raise a high tone, whether low tone will be lowered is more of a debate. It is not entirely clear from our data that the L target of tone 2 would be lowered under prominence. We have some cases that do and some others that don't. The clearer evidence of low tone lowering comes from the final target of tone 3, which is the lowest target among all tones. Figure (20) shows clearly the

---

2. Noun phrases consist of a modifier and a generic noun tend to have initial stress. Other examples include professions with *ren2* 'person', as in *gong1-ren2* 'work-man, worker', or country names with *guo2, country*, as in 'fa4-guo2, France-country'.

lowering of L target on *MEI3*. The effect is evident on the following syllable *ming2*. These figures combined provide support to the claim that prominence causes pitch expansion, rather than just pitch raising.

Following a prominent H or L, the pitch level goes back to normal. The first L tone after the prominent H, and the first H after the prominent L, would bring the pitch level back to normal immediately; see Figures (15), (16). However, a string of like tones changes gradually[3]. Figure (21) shows the gradual fall of tone 1 after a prominent *WANG2*. If we assume that pitch range expansion is done by scaling tonal targets away from the reference line, an abstract line corresponds to the mid pitch range, we would be able to explain automatically why post-prominence targets remember what the normal values are. When pitch range changes, the reference line remains unaffected. And after prominence, pitch range is scaled back in relation to the reference line, and in turn determines the normal values of H and L targets.

5. PITCH HEIGHT AND DISCOURSE STRUCTURE

In this section I will discuss briefly the difference between prominence effect and what is referred to as *initial raising* and *final lowering* of discourse structure.

Hirschberg and Pierrehumbert (1986) assume that discourse exhibits a hierarchical structure, and that hierarchical structure is reflected in intonation by varied pitch ranges: an increase in the pitch range signals the beginning of a discourse segment, while final lowering signals the end of it. The magnitude of increase corresponds to the hierarchical level of discourse structure. While adopting the main concept of Hirschberg and Pierrehumbert, I see strong evidence from Mandarin that initial raising and final lowering actually affects the level of reference line, but does not affect the pitch range.

To study the pitch scaling effect in discourse, I recorded a story without a text, repeated it until I could finish the whole story fluently without undesired pauses and interjections. The story was about thirty sentences long, and had four paragraphs. The story is pitch tracked and H and L targets of all tone 2 are measured for comparison. Tone 2 is chosen because previous study on Mandarin tonal co-articulation suggests that tone 2 is the only tone that will give us a reliable reading of both H and L target, provided that we take the measurement of the final H target in the following tone 1. It is difficult to obtain the value of a H target when a tone 4 follows, so all those samples were discarded.

---

3. The only case in Mandarin where a string of like tones occurs is in a sequence of tone 1, which has only H targets. There is no tone that is just plain L, the so-called low tone, or tone 3, is actually low-falling. Moreover, strings of tone 3 is subject to a tone sandhi rule that change some to tone 2.

Figure (23) plots all the *reliable* tone 2 from the story. The y axis represents the H target of a tone 2, while the x axis represents the L target. The plot shows a linear dependency between the value of H and L targets: the higher the H target, the higher the L target, and vice versa. The higher tone 2 occurs at the beginning of the story and at the beginning of paragraphs. The lowest tone 2 occurs at the end of the story. Unfortunately, the paragraph structure of the story is not represented in the plot, so the correlation between pitch height and discourse structure wouldn't be obvious. However, it disproves the claim that initial high pitch level is achieved by increased pitch range. If that is the case, we should see the samples gather around a vertical line that represents a more or less invariant pitch level for L targets.

This study sheds some light on the confusion between initial raising and prominence effects. While both have higher H targets, The Mandarin data suggests that the distinction between the two lies in the realization of L targets. Initial raising involves register shift, when both H and L targets are realized with higher pitch value. Prominence effects involve pitch range expansion, which causes H target to be higher and L target to be lower.

6. *CONCLUSION*

Several issues discussed in this paper differ from other intonational studies, I will address them briefly in the conclusion.

Han and Kim (1974) reports on the disyllabic tonal sequence of Vietnamese. They found very little tonal co-articulation. Basically, tone shapes, slopes, and the placement of tonal targets in their study are the same in monosyllabic and disyllabic forms. However, pitch height may be influenced by preceding or following tones. The difference of Vietnamese and Mandarin suggests that tonal co-articulation rules as described in this paper may be language specific.

The tonally triggered catathesis poses a problem for Garding's (1987) grid model for Chinese. The grid model assumes a separation of the speech act related intonation and the lexically defined tones. For statements like the test sentences in section 3, a falling grid is drawn first, and then tonal targets (Garding's turning points) will be placed on grid lines. Ideally, sentences of the same speech act/intonation should share the same grid. But a uniform grid fails to capture the variations caused by the nature of tonal targets. Sentences with only tone 1 take an almost level grid, while sentences with tone 3 require a more steeply falling grid. The sentences here are relatively simple. An uncontrolled sentence with freely combined tones would require even more individual adjustment of grid. At that stage, it is difficult to justify the complete separation of intonation and tones, at least for what a 'statement grid' represents. The catathesis experiment shows that F0 value of each target is not solely determined by speech act, but is largely determined by the tonal composition.

Garding (1983, 1984, 1987) proposes that prominence effect could be captured by setting a *jumping* grid, in which the grid line is broken and expanded at the location of a prominent constituent, and compressed afterwards for the de-accented post-prominence constituents. I agree in principal that prominence is related to the expansion and compression of pitch range, however, my data shows a great divergency in what a potential prominence grid should look like. Sequences of post prominence tone 1 suggest that grid line should compress gradually, while a L target would require a sudden compression of grid just where the L target is located. Even more problematic is the higher, but unfocused H target after tone 2. Apparently, pitch range can continue to expand when the grid line would suggest a return. In my view, expansion of pitch range is caused by prominence, but the actual implementation could be mechanical. Extension of rising slope into the following H tone is part of a tonal co-articulation rule, and it applies with no reference to meaning or intention. As a result, expanded pitch range can not be automatically translated back to prominence. Garding's model tries to relate speech acts and their functions directly to surface realization of intonation, represented by grids. But since the surface F0 values are affected by many factors simultaneously, it will be quite difficult to find a grid that is meaningful.

Several studies related prominence effect to high pitch. Eady and Cooper (1986) suggest that, among other acoustic correlates, higher F0 topline is a major manifestation of non sentence-final focus. Inkelas, Leben and Cobler (1986) suggest that prominence is represented phonologically by H register. The fact that a prominent L target either lowers or remains at the same height suggests that pitch range expansion is a more appropriate representation.

The confusion of what high pitch could represent may be a reason why Wells (1986) fails to find a correlation between *the highest peak* and prominence. Although there is a strong correlation between high pitch and prominence, this paper discusses two other possible interpretations: the highest peak in a sentence could be at the sentence/paragraph initial position, signalling the beginning of a discourse structure; it could also be a post-prominence peak, carrying no prominence in itself.

References

Cheng, C. C. (1970) Domains of phonological rule application. In *Studies presented to Robert B. Lees by his students.* ed by J. M. Sadock and A. L Vanek. Edmonton: Linguistic Research. p 39-59.

Eady S. J. and W. E. Cooper (1986) Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, 80(2), p 402-415.

Han, M. S. and K.-O. Kim (1974) Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2, p 223-232.

Hirschberg, J. and J. Pierrehumbert (1986) The intonational structuring of discourse. in *Proceedings of ACL, New York*, p 136-144.

Garding, E. (1983) A generative model of intonation, in *Prosody: models and measurements*, ed by A. Cutler and D. R. Ladd. Springer-Verlag.

Garding, E. (1984) Chinese and Swedish in a generative model of intonation. *Nordic Prosody* V 3, p79-91. University of Umea.

Garding, E. (1987) Speech act and tonal pattern in standard Chinese: constancy and variation. *Phonetica* 44: 13-29.

Inkelas, S. Leben, W. and M. Cobler (1986) The phonology of intonation in Hausa, unpublished ms., Stanford University.

Liberman, M. Y. and L. A. Streeter (1978) Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America*, 63, p 231-233.

Liberman, M. Y. and J. B. Pierrehumbert (1984) Intonational invariance under changes in pitch range and length. in *Language Sound Structure*, the MIT Press.

Pierrehumbert, J. B. (1980) *The phonology and phonetics of English intonation.* Doctoral dissertation, MIT.

Shih, C.-L. (1986) *The prosodic domain of tone sandhi in Chinese.* Doctoral dissertation, UC-San Diego.

Shih, C.-L. (1987) The phonetics of the Chinese tonal system. Technical memo, AT&T Bell Labs.

Silverman, K. (1987) *The structure and processing of fundamental frequency contours.* Doctoral dissertation, University of Cambridge.

Steele, S. A. (1985) *Vowel intrinsic fundamental frequency in prosodic context*, doctoral dissertation, the University of Texas at Dallas.

Tseng, C. Y. (1981) *An acoustic phonetic study on tones in Mandarin Chinese*, doctoral dissertation, Brown Univeristy.

Wells, W. H. G. (1986) An experimental approach to the interpretation of focus in spoken English. in *Intonation in discourse*, ed by C. Johns-Lewis. Croom Helm, London & Sydney.

Woo, N. (1972) *Prosody and phonology*, Doctoral dissertation, MIT, reproduced by the Indiana University Linguistics Club.

Yip, M. (1980) *The tonal phonology of Chinese.* Doctoral dissertation, MIT.

Figure (1)

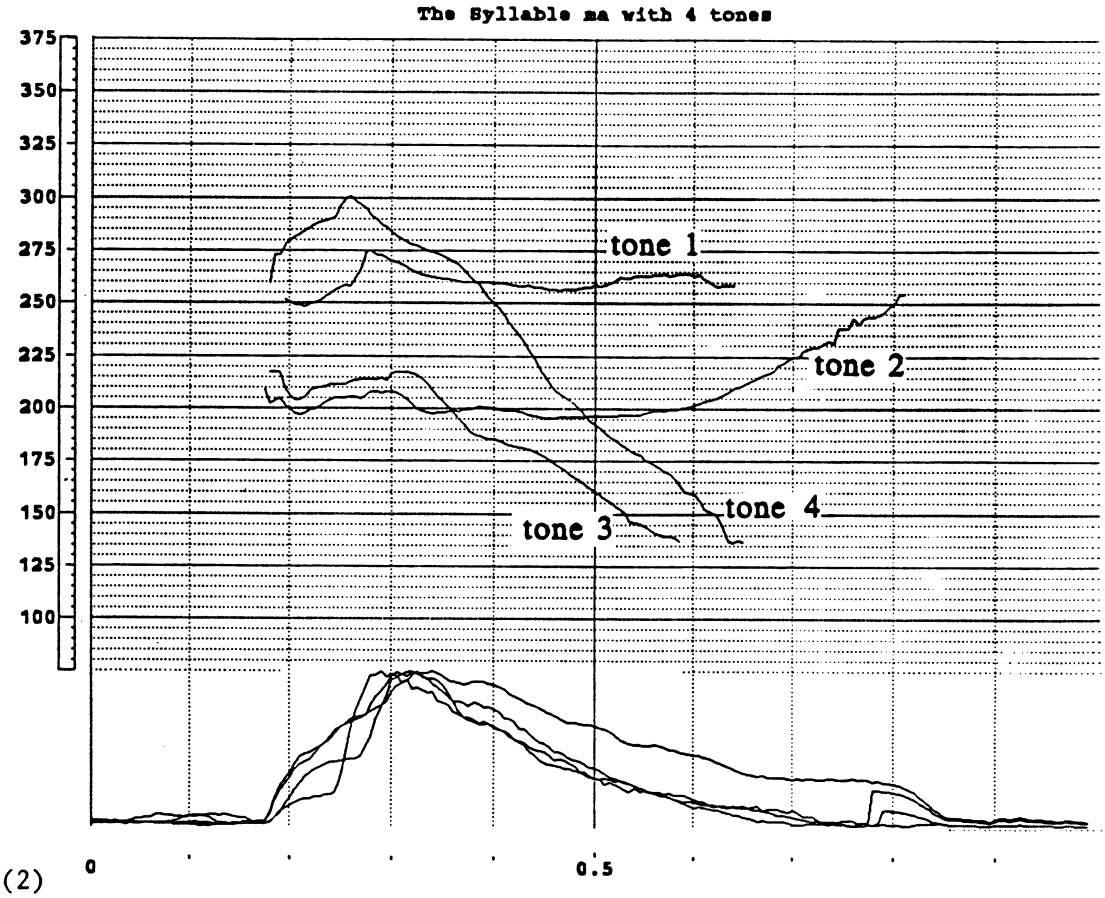


Figure (2)

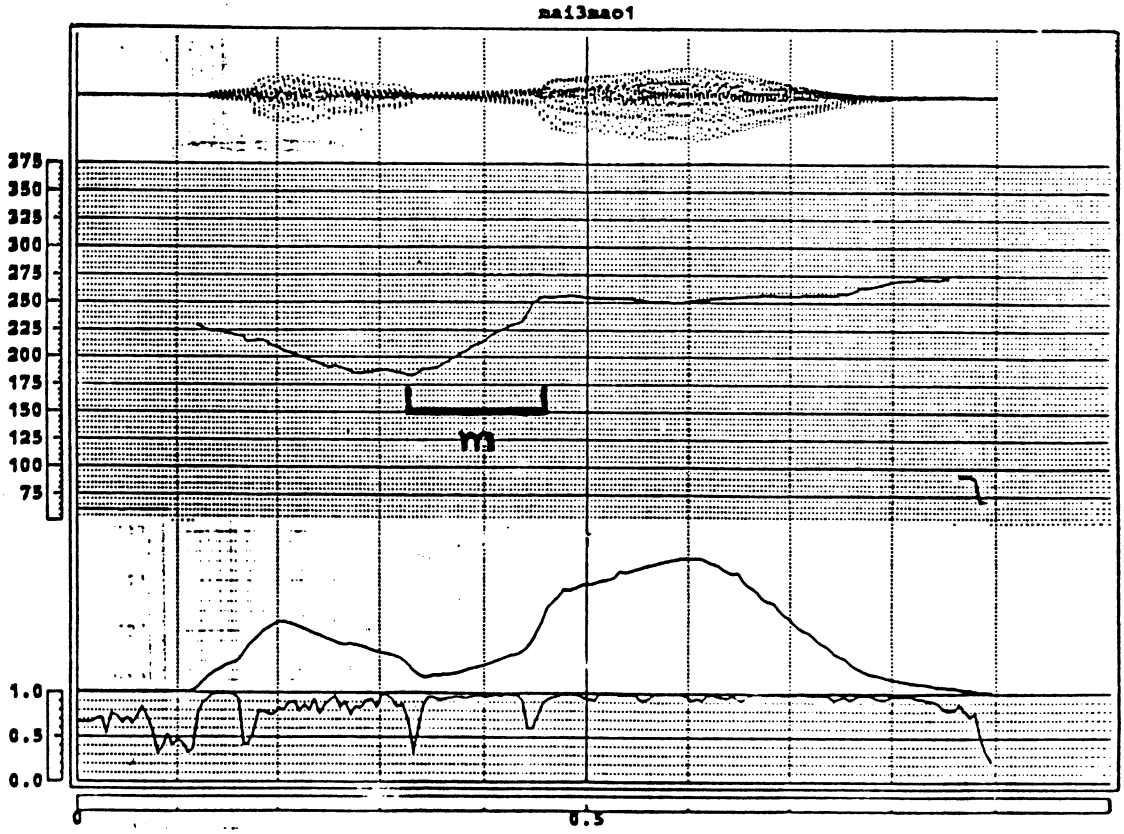


Figure (3)

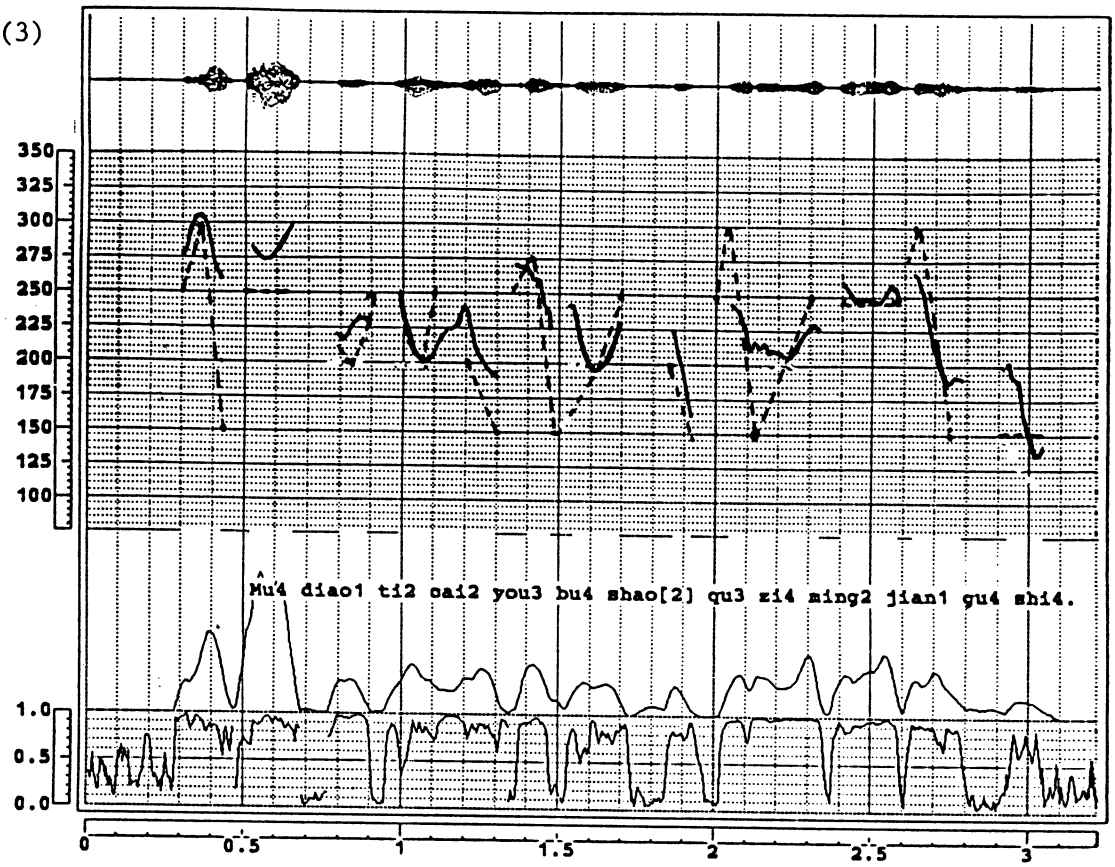


Figure (4)

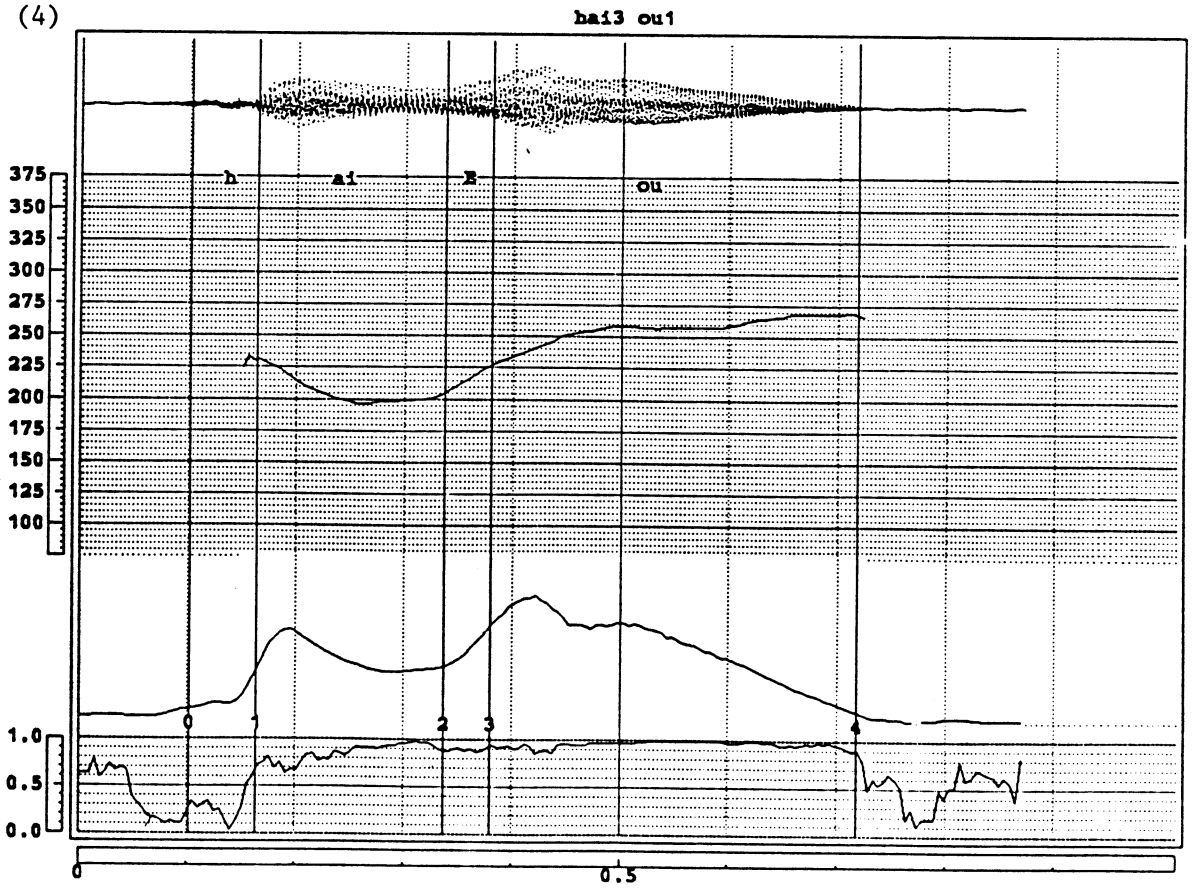


Figure (5)

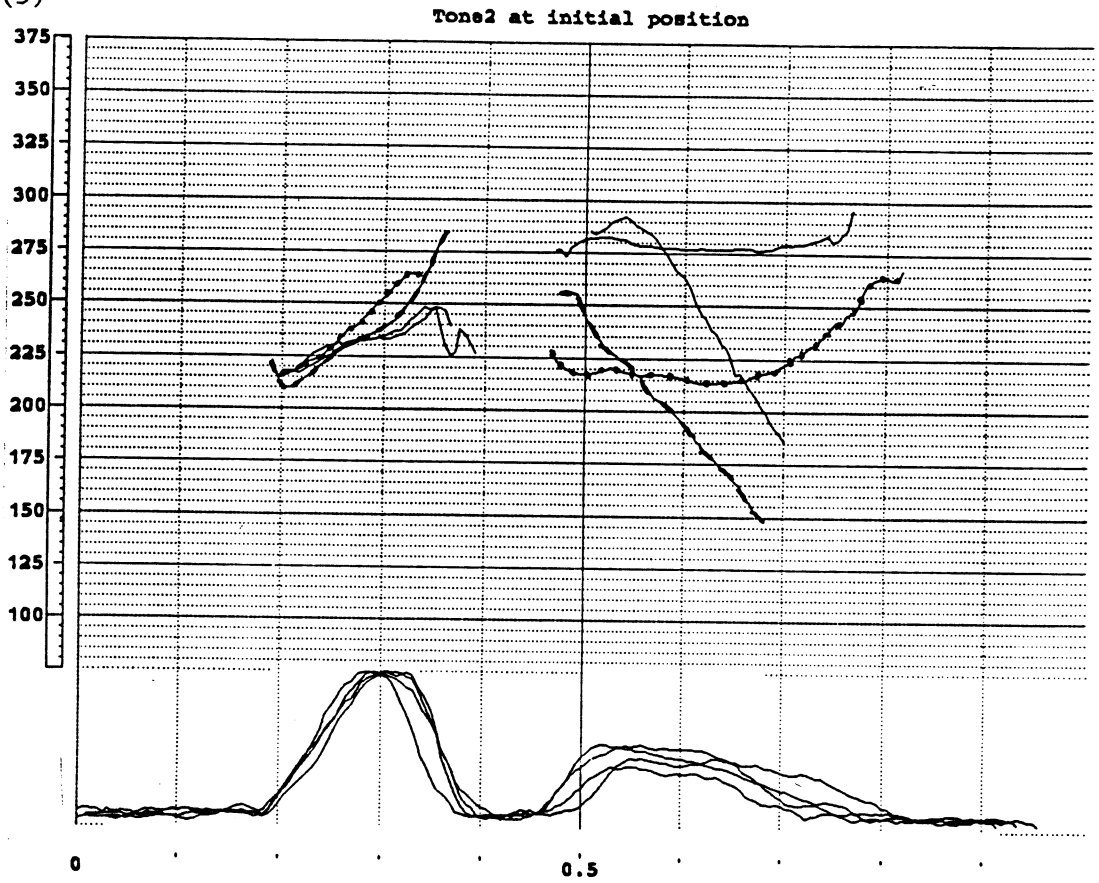


Figure (6)

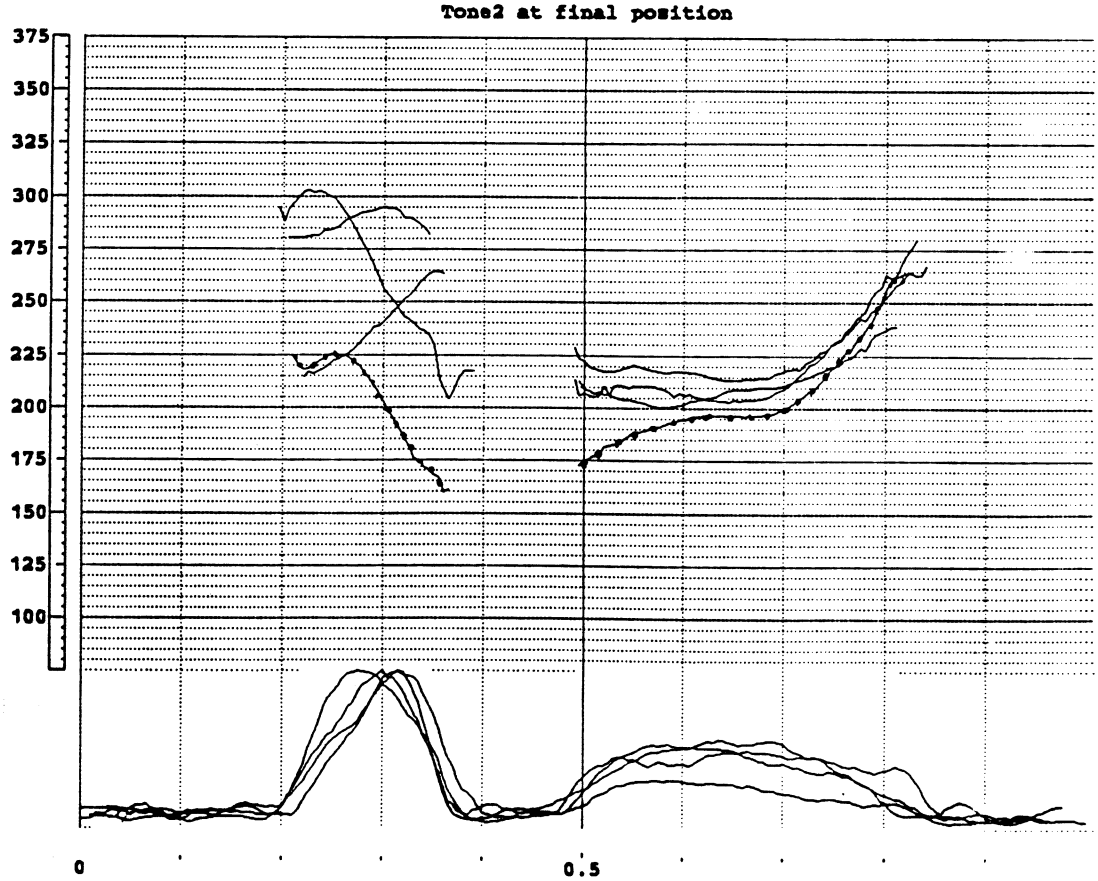


Figure (7)

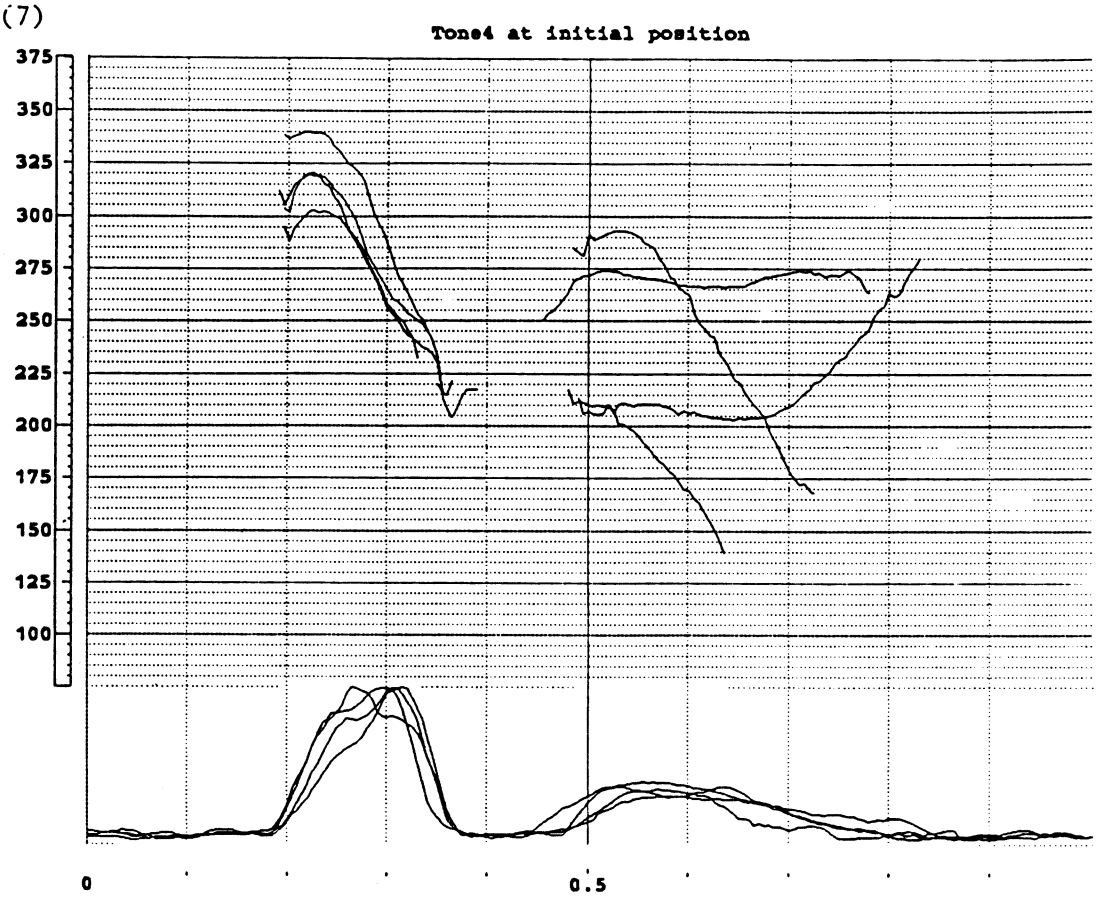


Figure (8)

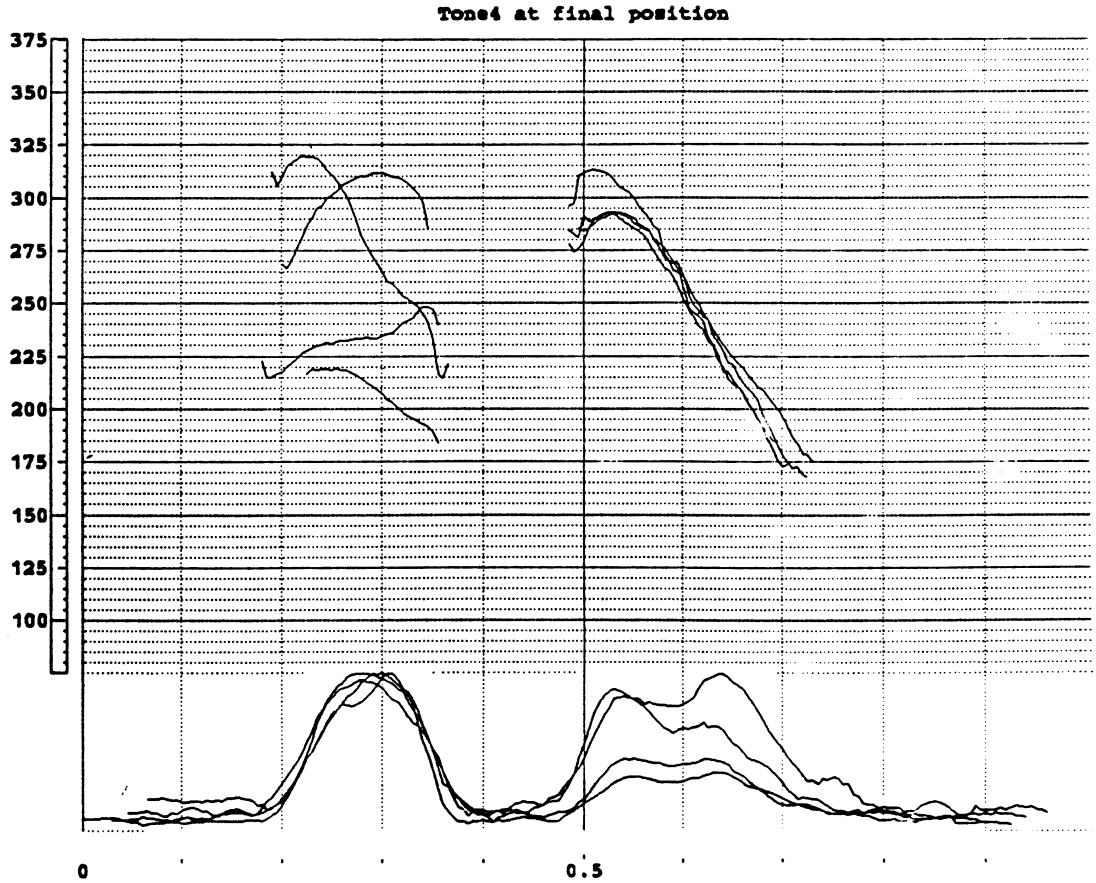


Figure (9)

Figure (10)

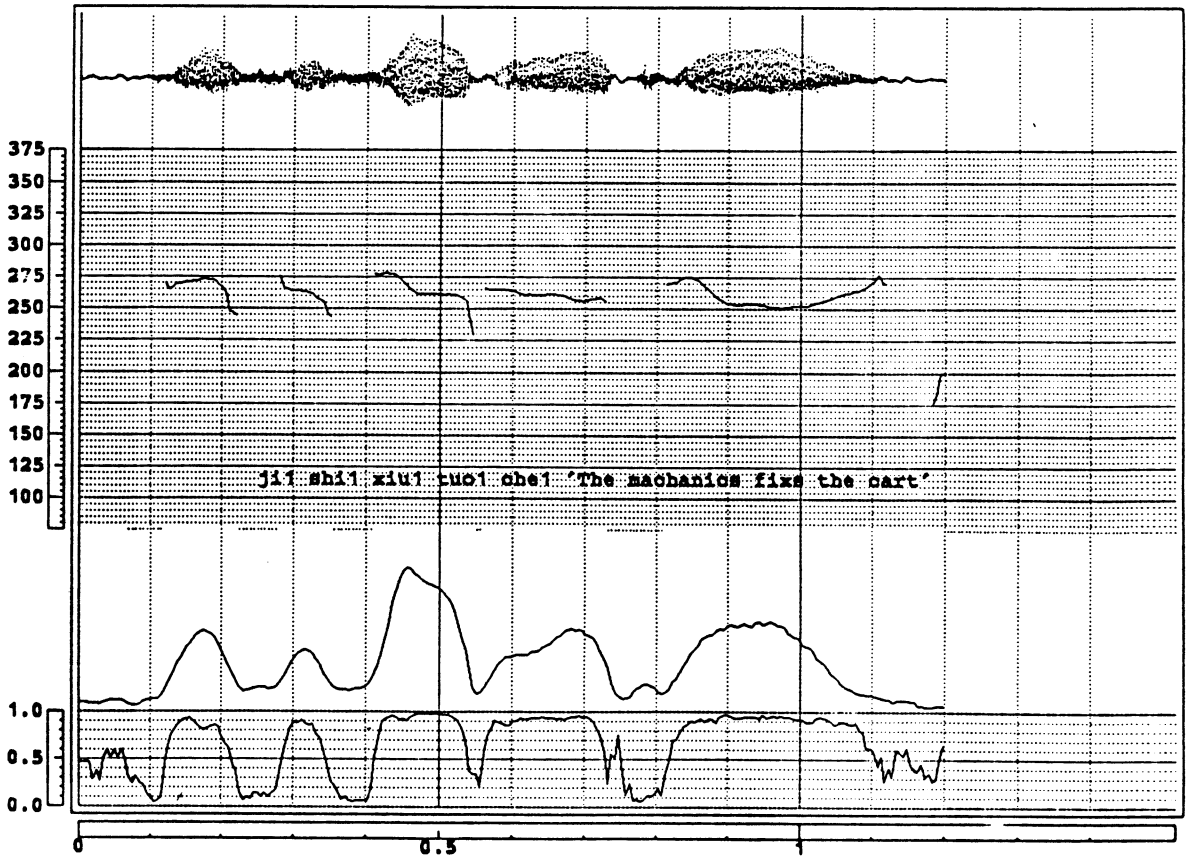


Figure (11)

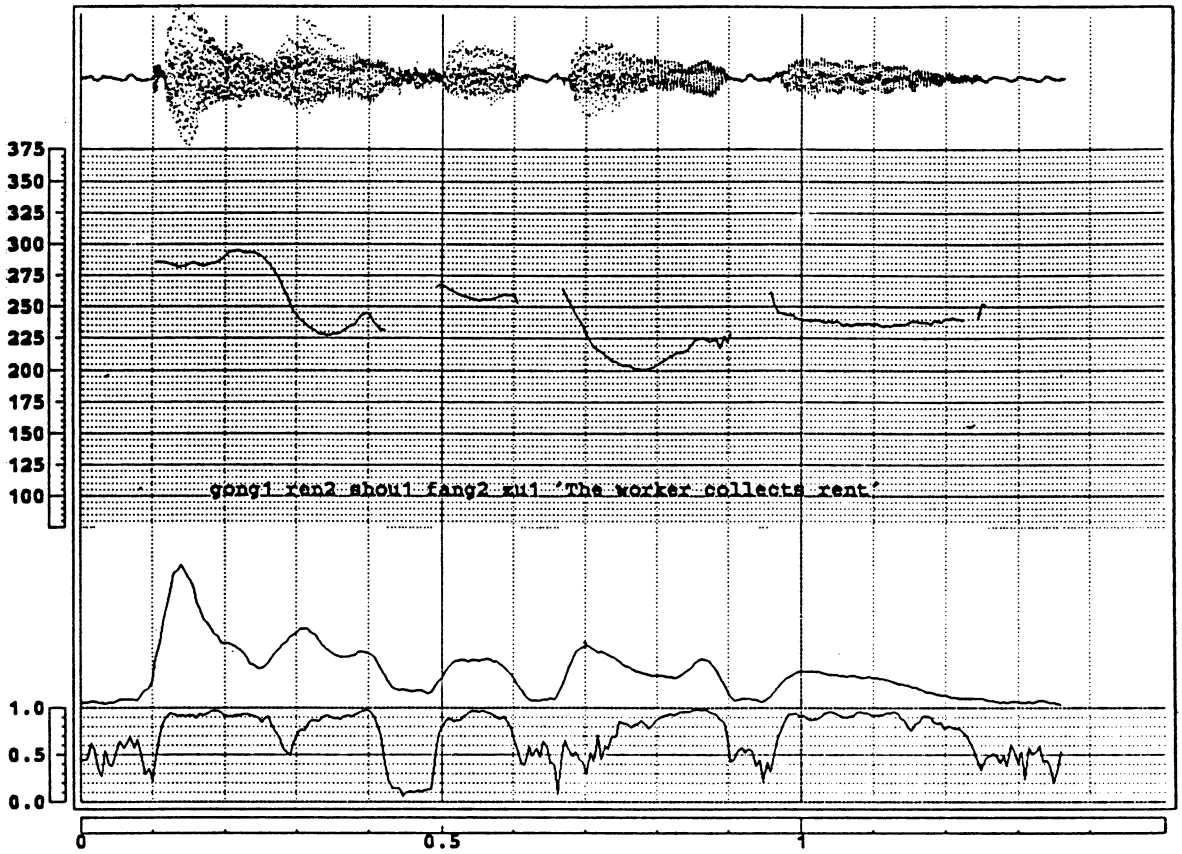


Figure (12)

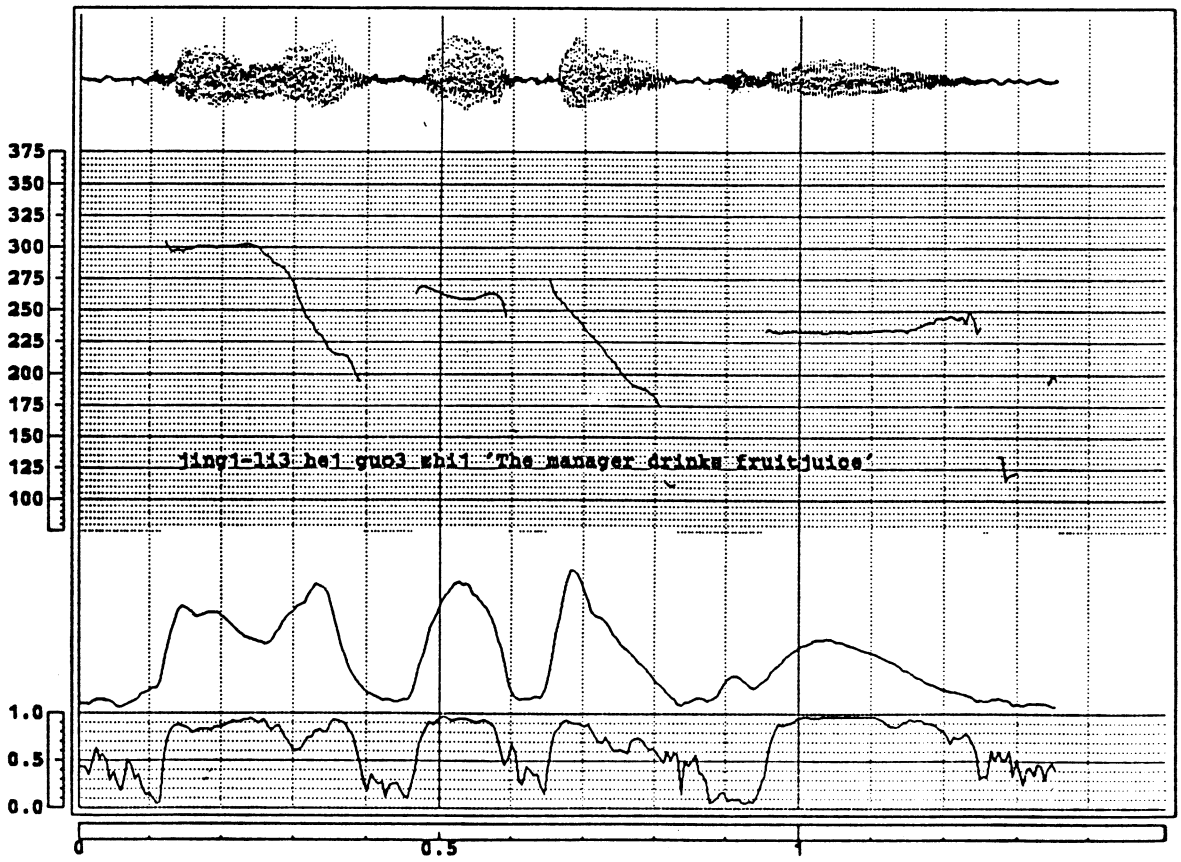


Figure (13)

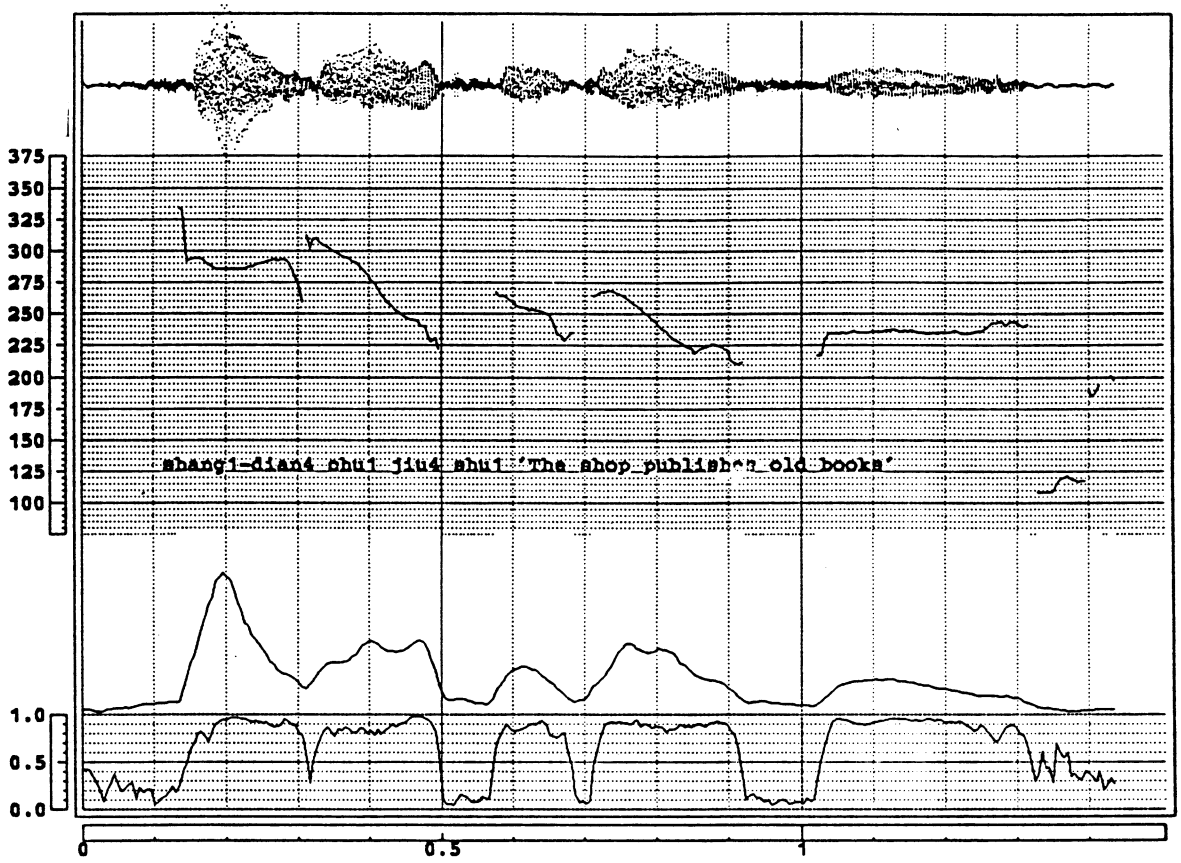


Figure (14)

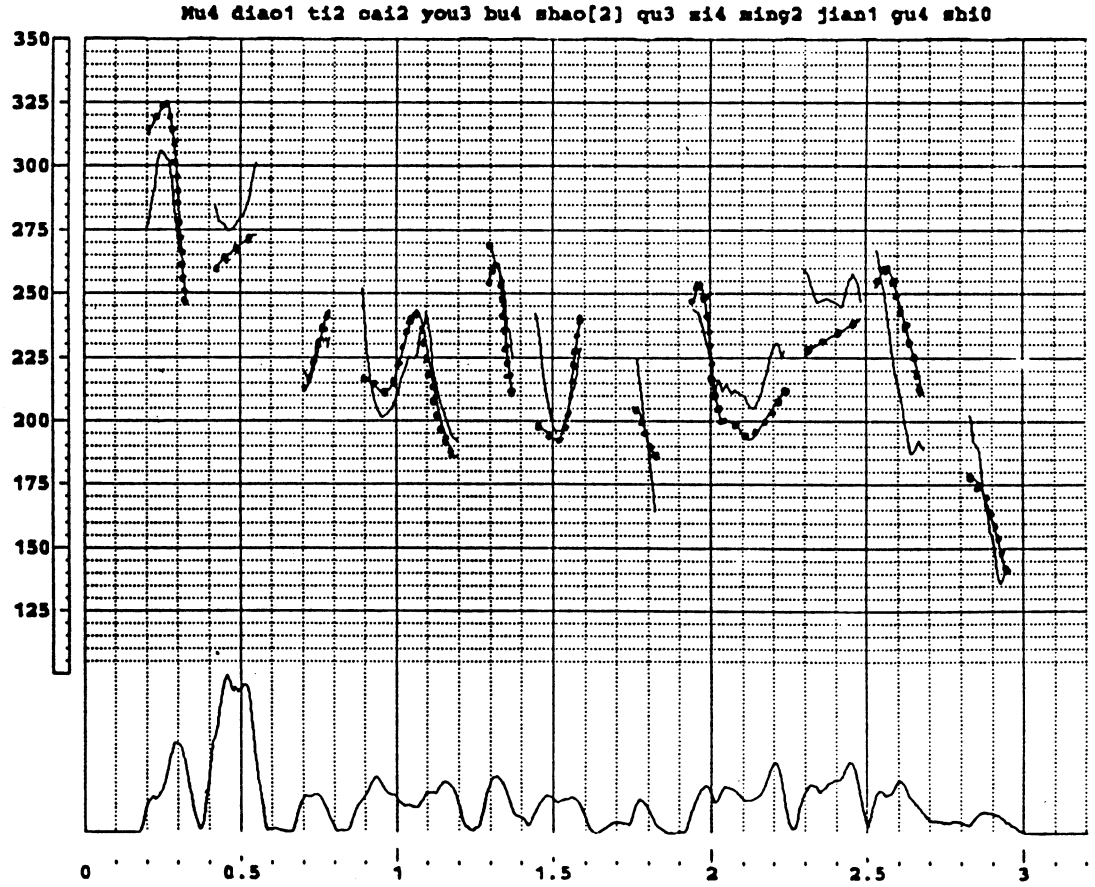


Figure (15)

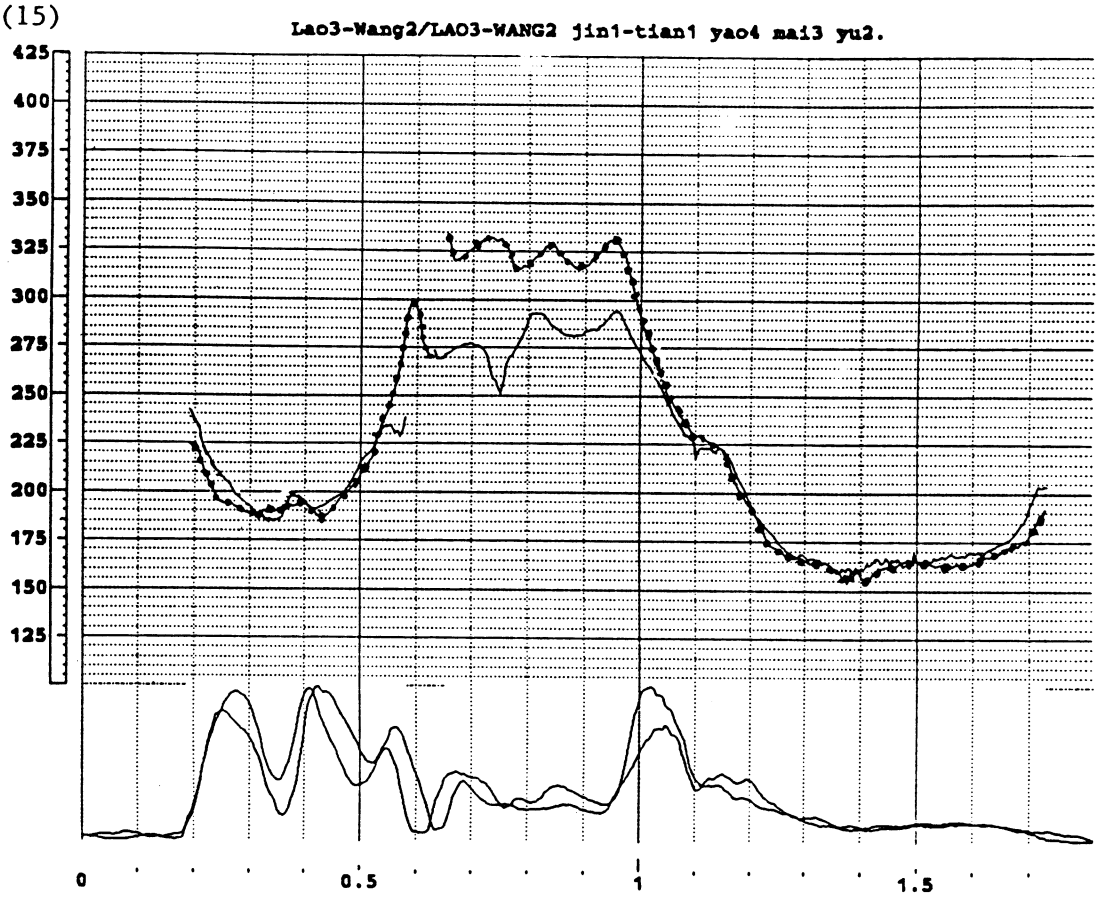


Figure (16)

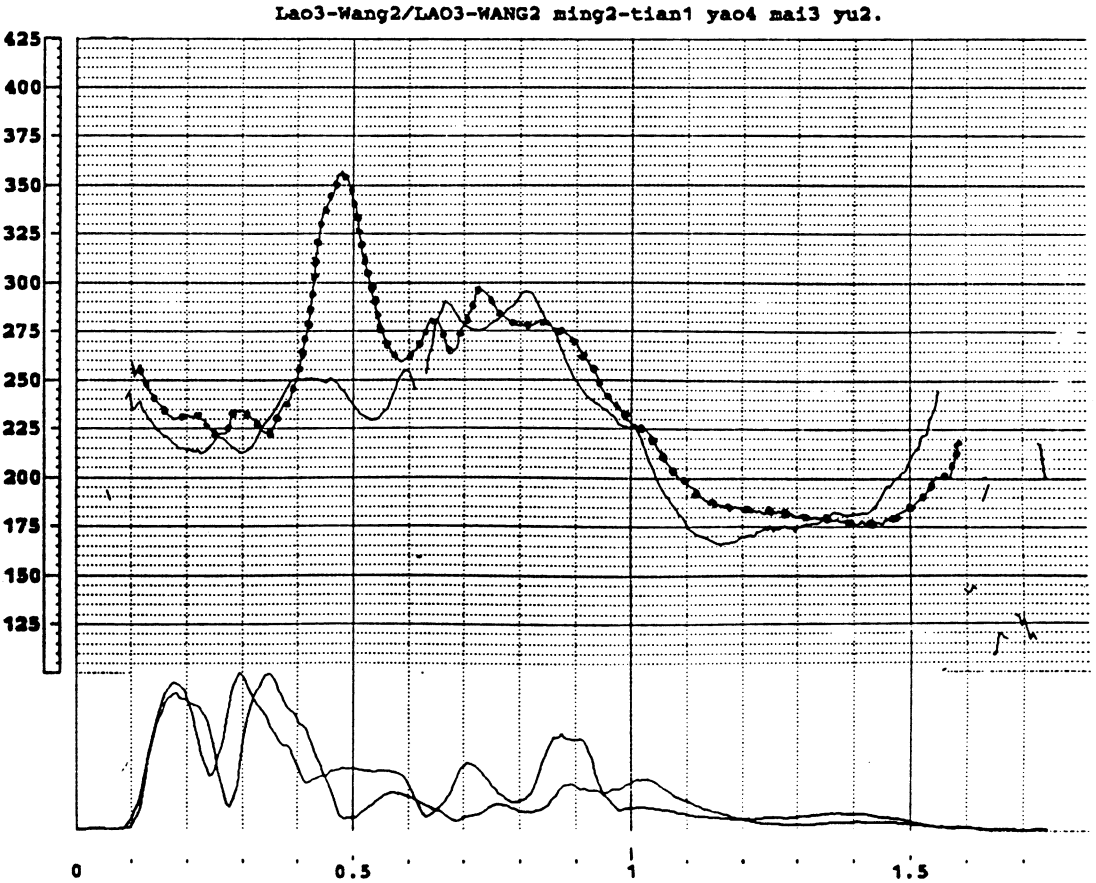


Figure (17)

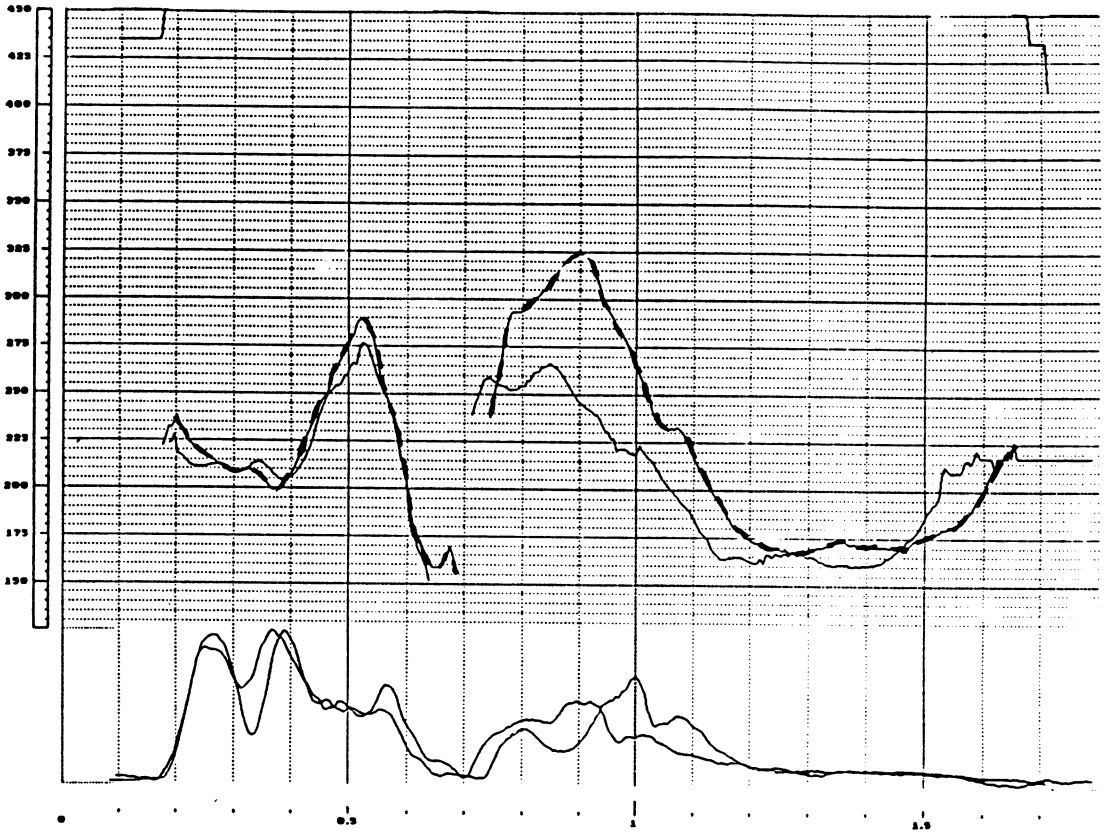


Figure (18)

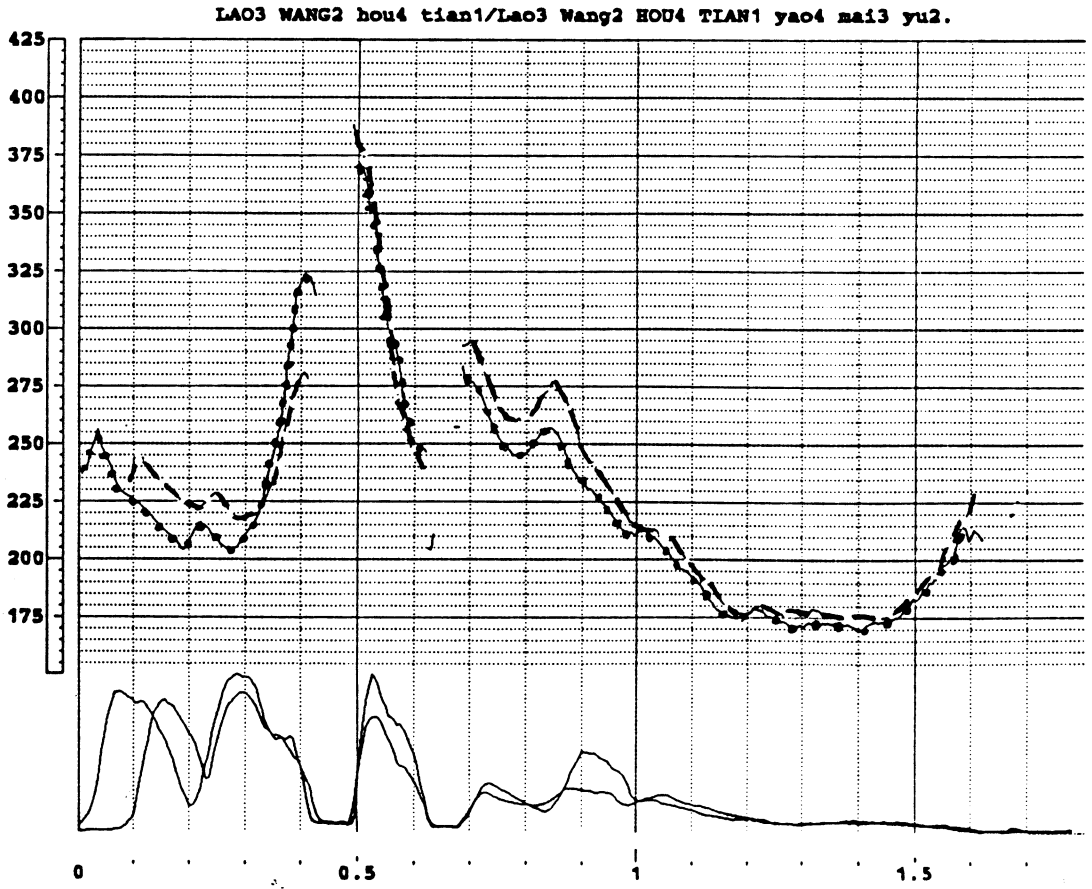


Figure (19)

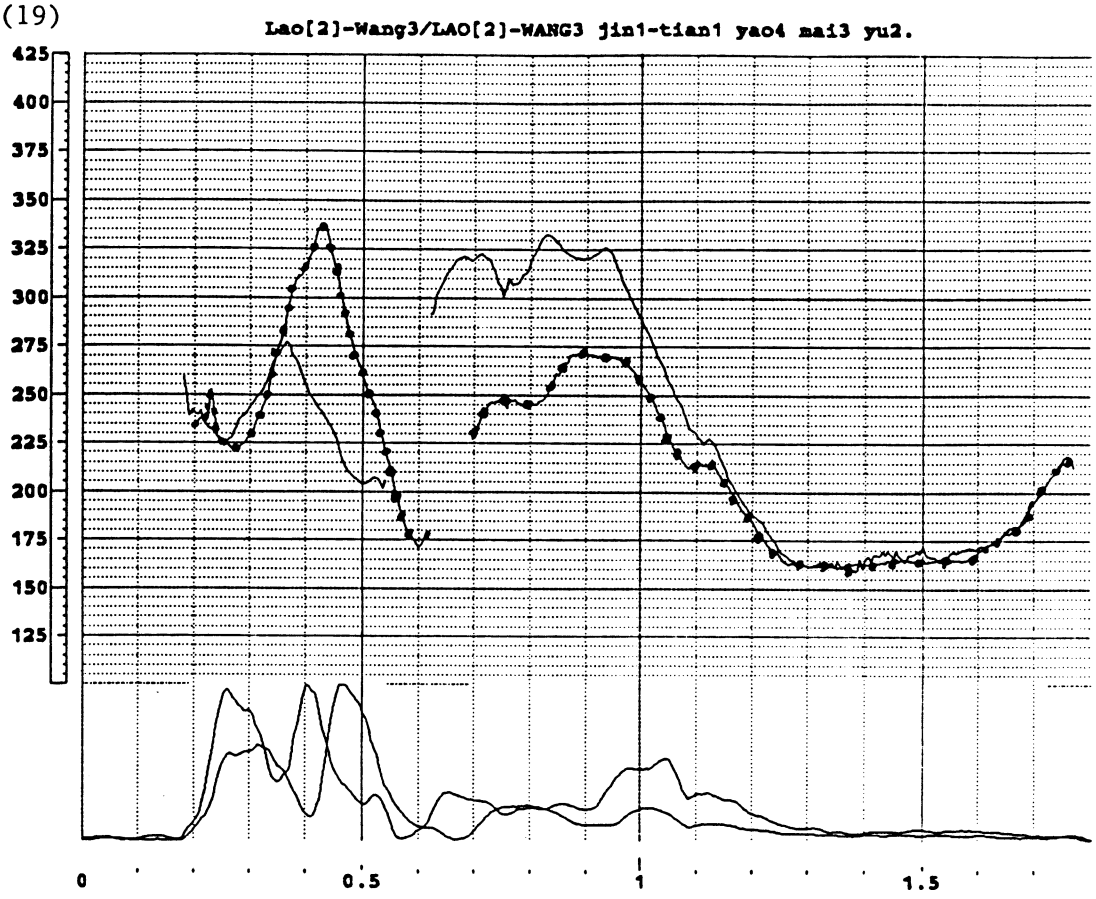


Figure (20)

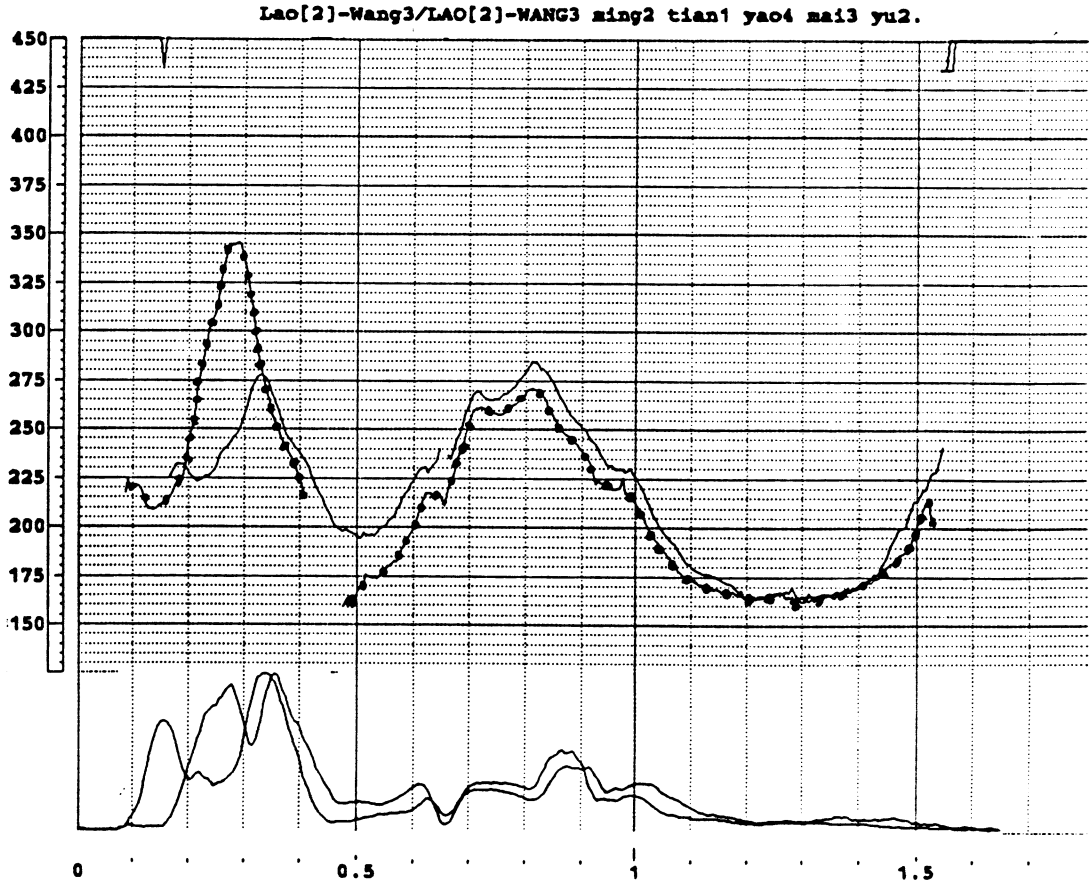


Figure (21)

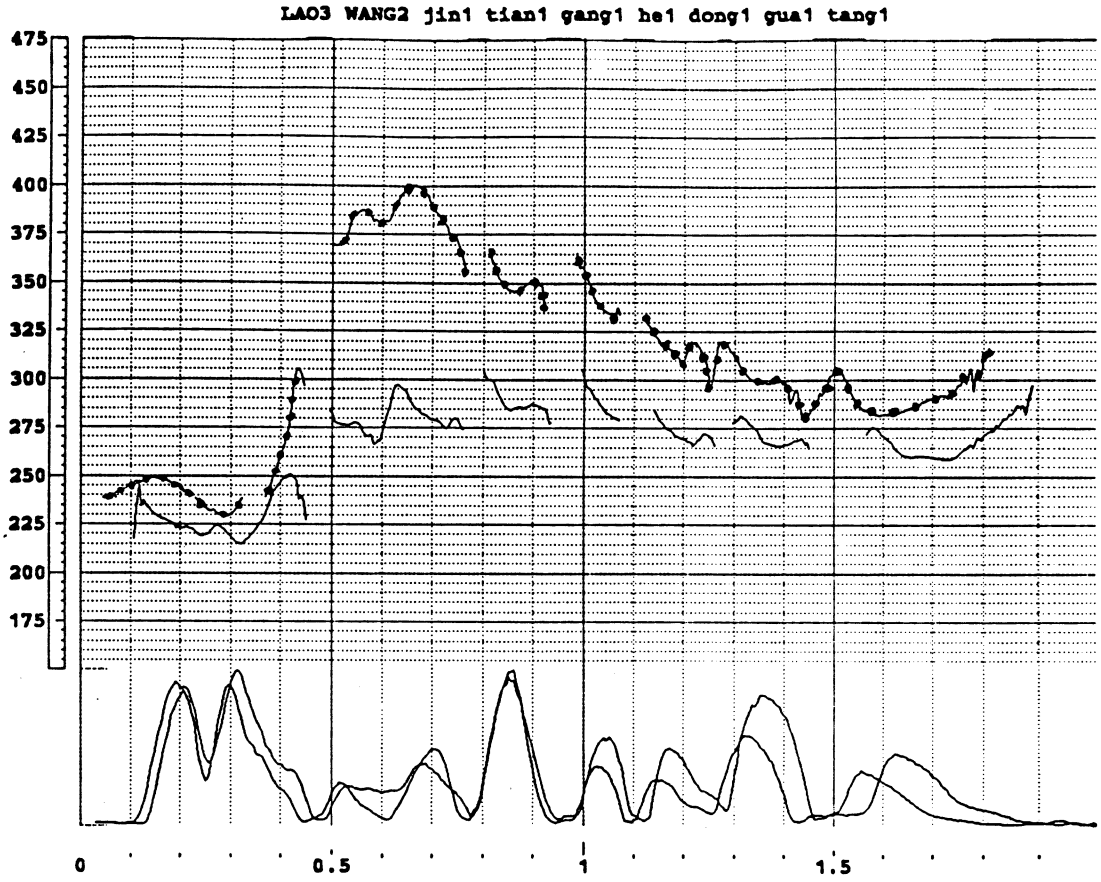


Figure (22)

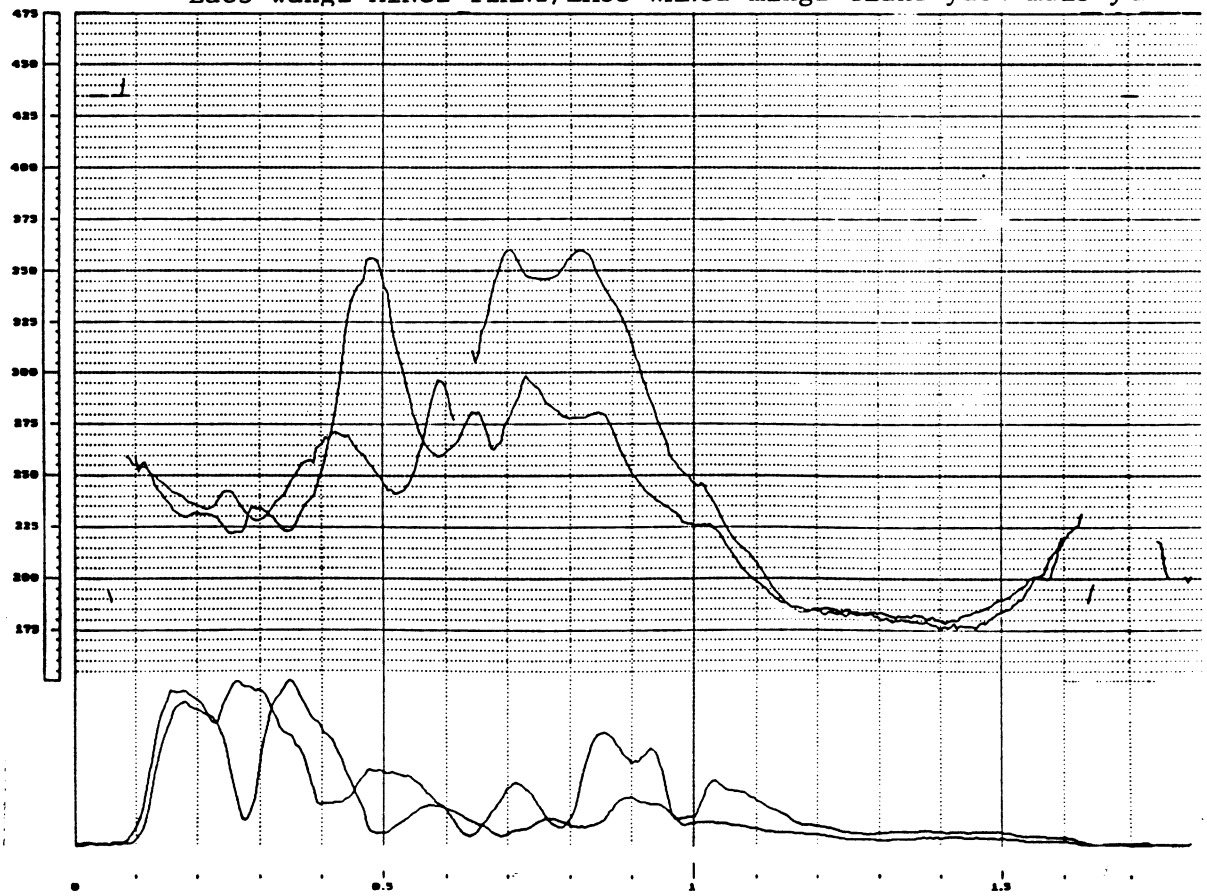
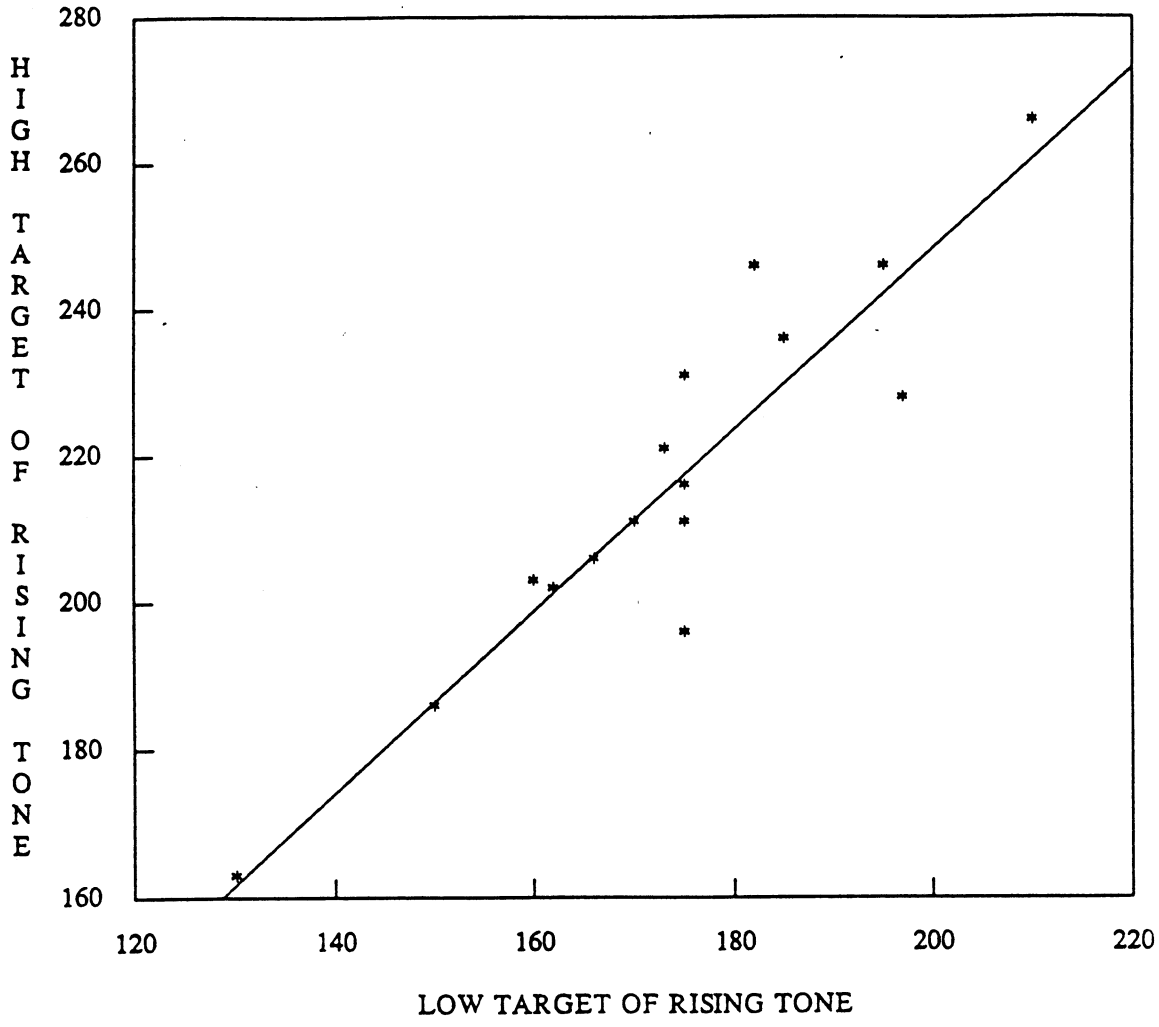


Figure (23)


HIGH TARGET OF RISING TONE
280
260
240
220
200
180
160
120
140
160
180
200
220
LOW TARGET OF RISING TONE