LANGUAGE AND SPEECH, 1989, 32(1), 25 – 44

CUES TO THE PERCEPTION OF TAIWANESE TONES*

HWEI-BING LIN
and
BRUNO H. REPP
Haskins Laboratories

A labeling test with synthetic speech stimuli was carried out to determine to what extent the two dimensions of fundamental frequency (F_0), height and movement, and syllable duration provide cues to tonal distinctions in Taiwanese. The data show that the high level vs. mid level tones and the high falling vs. mid falling tones can be reliably distinguished by F_0 height alone, whereas the distinction between tones with dissimilar contours, such as the high falling and low rising tones, is predominantly cued by F_0 movement. However, the other dimension of F_0 may collaborate with the dominant one in cueing a tonal contrast, depending on the extent to which the two tones differ along that dimension. Syllable duration has a small additional effect on the perception of the distinction between falling and nonfalling tones. These results are consistent with previous findings in tone languages other than Taiwanese in that they suggest that tones are mainly cued by F_0. While the primacy of F_0 dimensions as cues to tonal contrasts depends on the contrast to be distinguished, the present findings show that tones which nominally differ only in register (e.g., high falling vs. mid falling) may exhibit perceptually relevant contour differences, and *vice versa.*

Key words: Taiwanese, tone perception, fundamental frequency

INTRODUCTION

The primary acoustic attribute distinguishing linguistic tones is their fundamental frequency (F_0), although duration and amplitude of the syllable carrying the tone may also exhibit characteristic differences. This observation raises the question of whether, in perceiving a phonemic tone, listeners integrate all these acoustic cues, or whether they pay attention to F_0 alone.

Several studies have investigated the role of F_0 versus other properties as cues to tonal distinctions, in both synthetic and natural speech. For example, Abramson (1962) imposed artificial F_0 movements on natural Thai monosyllables by means of a vocoder and found that F_0 overpowered other concomitant features such as duration and

* This research, which formed part of the first author's doctoral dissertation, was supported by NICHD Grant HD-01994 to Haskins Laboratories. Hwei-Bing Lin was also at the Department of Linguistics, University of Connecticut, Storrs, CT 06268. Special thanks are due to Arthur Abramson, Carol Fowler, and Ignatius Mattingly for their helpful comments on an earlier version of this paper, and to Jackson Gandour and Eva Gårding for serving as reviewers for LANGUAGE AND SPEECH.

Address correspondence to: Dr. Hwei-Bing Lin, The Lexington Center, Inc., Research Division, 30th Ave. & 75th St., Jackson Heights, NY 11370.

amplitude in cueing tonal distinctions. The primacy of F_0 was confirmed in a later study using synthetic Thai speech (Abramson, 1975), though addition of natural amplitude contours improved identification further. The conclusion that F_0 carries sufficient information for conveying tonal distinctions has also been drawn by Howie (1976), Tseng (1981), and M.-C. Lin (1987), who investigated the tones of Mandarin Chinese. However, these studies either did not vary duration and amplitude at all, or they pitted these dimensions against unambiguous F_0 contours. In whispered monosyllables, where F_0 is altogether absent but duration and amplitude differences may be retained to some extent, tonal distinctions are resolved poorly, though above chance level (e.g., Abramson, 1972; Howie, 1976). It is conceivable that, in addition to increasing the naturalness of utterances (cf. M.-C. Lin, 1987; Rumyantsev, 1987), duration and amplitude have larger effects on tone identification when F_0 provides ambiguous information. Also, the relative informativeness of different acoustic cues for tonal identity may vary across languages.

The F_0 dimension itself may be decomposed into two aspects: height and movement.[1] The relative importance of these two aspects obviously depends on the nature of the tonal distinction to be made: If the distinction is between two tones differing primarily in register (e.g., "high" vs. "low"), F_0 height will be important; if two tones differ primarily in contour (e.g., "rising" vs. "falling"), F_0 movement will be the dominant cue. However, for tones that, according to linguistic nomenclature, differ only in register or in contour, the other aspect of F_0 might play a secondary role in cueing the contrast. Furthermore, for tones that differ in both register and contour (e.g., "high falling" vs. "low rising"), both F_0 height and movement may be relevant, though perhaps not equally important. Their relative importance may depend on what other tones there are in the language.

With these issues in mind, we conducted the present study to determine the relative importance of F_0 height, F_0 movement, and duration as cues to the tonal distinctions of Taiwanese. From traditional classifications by phonologists and from acoustic studies (Chiang, 1967; Zee, 1978; H.-B. Lin, 1988) it is evident that Taiwanese has five long tones, whose typical F_0 contours are illustrated in Figure 1.[2] They fall into two classes: Tones 1 ("high level") and 5 ("mid level") are fairly level or static, whereas tones 2 ("high falling"), 3 ("mid falling") and 4 ("low rising") are contoured or dynamic. Two pairs of tones, 1 vs. 5, and 2 vs. 3, have similar F_0 contours but differ in register. We would thus expect F_0 height to play a primary role in the distinction of these tone

---

[1] Throughout this paper, we refer to F_0 characteristics of phonological tones as F_0 *register* and *contour*, but to the corresponding phonetic dimensions as F_0 *height* and *movement*. Register and contour thus are characterized in discrete terms (high, mid, low; rising, level, falling), whereas height and movement are continuously variable and are described in acoustic terms. The term "register" is not intended to denote changes in vocal register.

[2] According to traditional classification, Taiwanese has five long tones and two short tones. The former occur in syllables ending with vowels or nasals, whereas the latter occur in checked syllables only (see Chiang, 1967). Because only an open syllable was used as the carrier in the present study, short tones were excluded from consideration.

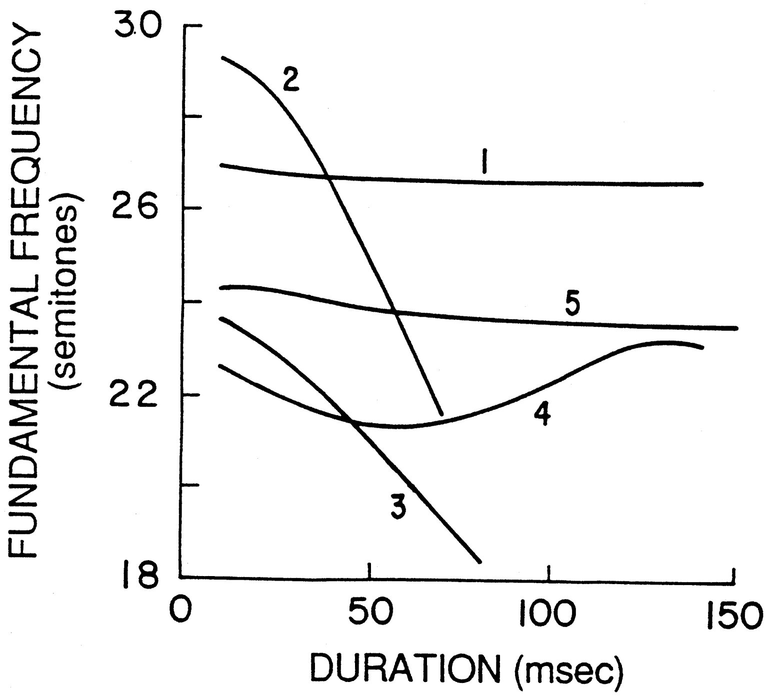

Fig. 1. Average F_0 movements of five Taiwanese tones produced by three male speakers on the syllable /do/ in sentence-final position. (Data from H.-B. Lin, 1988).

pairs. As for the tonal pairs with different F_0 contours, we expected that F_0 movement rather than height would be responsible for cueing the differences. However, we wondered whether the "other" aspect of F_0 would make a contribution to a distinction as well. For example, would the small difference in F_0 movement between the two level tones, 1 and 5, or the larger one between the two falling tones, 2 and 3, play any role in perception at all? And would F_0 height be relevant to the distinction, for instance, between tones 1 and 4, even though they differ in contour? The answers to these questions seemed less obvious.

With respect to duration, the five Taiwanese long tones fall into two groups: relatively long (tones 1, 4, and 5) and relatively short (tones 2 and 3). In isolated syllables, the respective average durations were 145 msec and 75 msec (H.-B. Lin, 1988; see Figure 1). Thus, in principle, duration could be a rather strong cue for the distinction between falling and nonfalling tones.

Our approach was to synthesize a variety of F_0 patterns by varying F_0 height, F_0 movement, and duration independently in isolated syllables.[3] In some of our stimuli,

[3] We did not vary amplitude characteristics of the stimuli, to keep the design within

the F_0 heights, movements, and durations of typical Taiwanese tones were juxtaposed in novel combinations, so the relative strength of these competing cues could be assessed. In addition, we synthesized F_0 patterns intermediate between those of the original tones in terms of F_0 height and/or movement. These relatively ambiguous stimuli provided the best opportunity to observe effects of secondary cues, such as duration or the "other" F_0 dimension, on perception of a given tonal contrast. Even though our stimuli were presented singly, we conceptualized the study in terms of pairwise tonal contrasts, for heuristic convenience. Listeners, of course, always had all five lexical alternatives in mind as they tried to identify our synthetic syllables.

METHOD

Materials

The syllable /do/ was modelled after a natural Taiwanese utterance on the software serial formant synthesizer at Haskins Laboratories. Since the Taiwanese /d/ is unaspirated and voiceless (i.e., [t]), only a 10-msec release burst preceded the onset of voicing. The onset frequencies of the first three formants were 450 Hz, 1160 Hz, and 2400 Hz. After 70 msec, they reached respective steady states of 560 Hz, 760 Hz, and 2000 Hz. The amplitude of the voicing source was kept at a constant value. The extreme F_0 values and durations of the five basic tones are shown in Table 1. The actual F_0 movements followed the natural models (cf. Figure 1) as closely as possible.

TABLE 1

Average F_0 onset and offset values (in Hz) and durations (in msec) of the five synthetic tones on /do/.
The "pivot" indicates the turning point of the F_0 movement in tone 4

Tone	onset	pivot	offset	duration
1	130		129	145
2	150		96	75
3	109		80	75
4	102	94	105	145
5	113		107	145

---

bounds. To confirm that Taiwanese tones could be identified in isolated syllables, tones produced on the syllable /do/ by a native speaker were presented to native listeners for identification in a pilot study. Although confusions occurred between tones 2 and 3, and between tones 1 and 5, average performance was 87% correct.

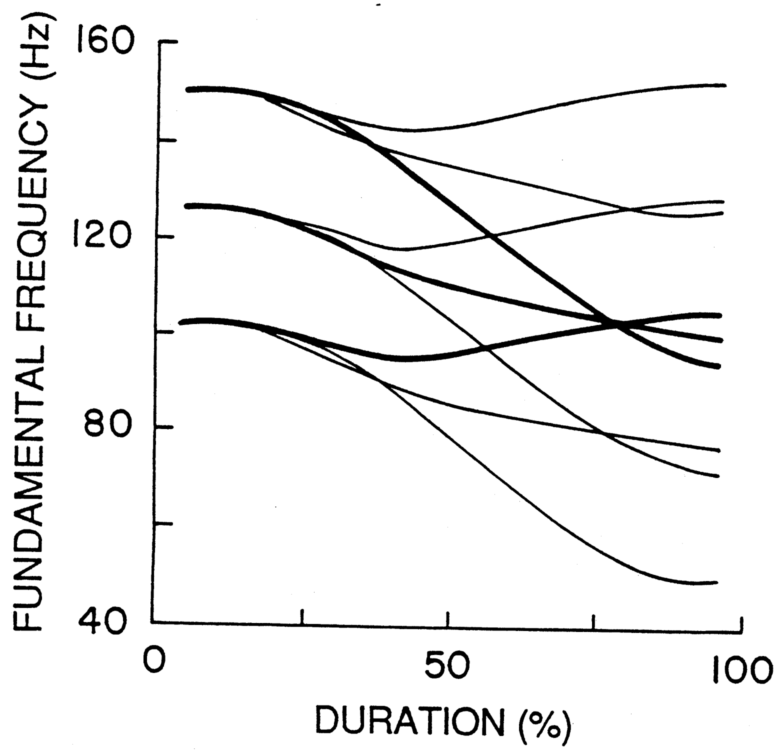

Fig. 2. Example of a stimulus set for a particular tonal contrast: tone 2 versus tone 4. The original (2 and 4) and intermediate (24) F_0 movements are shown by the heavy lines; the other patterns were obtained by changing the onset frequencies of these three F_0 movements. (See also Footnote 5.)

To assess the relative importance of F_0 height, F_0 movement, and duration cues in the perception of tonal distinctions, these properties were varied independently within each pairwise contrast. Thus we synthesized, in addition to the original tones, stimuli in which the F_0 movement of tone X was combined with the F_0 height of tone Y, and *vice versa.* This involved a translation of the whole F_0 movement up or down the linear frequency axis, by adding a positive or negative constant to all F_0 values in the synthesis specifications. We defined F_0 height operationally as the onset frequency of a tone.[4]

[4] As a result, variations in F_0 height were restricted for the tone 3–5 and 4–5 contrasts. Alternatively, we could have chosen the F_0 midpoint or the average F_0 as our measure of F_0 height. In that case, however, height variations would have been restricted for the tone 1–2, 2–3, and 4–5 contrasts. Our choice of F_0 onset frequency as the relevant measure is consistent with phonological terminology for tones, which usually mentions onset register and direction of contour, such as "high falling".

In addition, we created an F_0 movement intermediate between the two original tonal contours by averaging their F_0 values, and we chose an intermediate F_0 onset frequency as well. Thus we had three F_0 movements (X, Y, and intermediate) and three F_0 heights (X, Y, and intermediate), all combinations of which resulted in nine stimuli for any given tonal contrast. The stimulus set for the tone 2–4 contrast is illustrated schematically in Figure 2; tone 24 denotes the stimulus intermediate between tones 2 and 4 in both F_0 height and F_0 movement.[5]

Given five original tones, there were 10 pairs of tonal contrast. In four of these (2–3, 1–4, 1–5, 4–5), the durations of the two original tones were the same, and so all nine stimuli were synthesized at the same duration. In the six other contrasts (1–2, 1–3, 2–4, 2–5, 3–4, 3–5), the two original tones had different durations (cf. Table 1). For these, we synthesized the set of nine stimuli at three different durations, the two original ones (75 and 145 msec) and one intermediate one (110 msec), which resulted in 27 stimuli. When the duration of an original F_0 movement was changed, its onset and offset frequencies were maintained, but the F_0 trajectory was stylized by linear interpolation between the extreme values. In the case of tone 4, the location of the pivot was moved in proportion to the duration change, and two linear interpolations were performed.

In all, $6 \times 27 + 4 \times 9 = 198$ stimuli were created on the synthesizer, though a number of these were identical or closely similar to each other.[6] The stimuli were recorded in five different randomizations, each on a separate audio tape. On each tape, there were six blocks of 33 stimuli, separated by 8 sec of silence. There was a 3.5 sec interstimulus interval within blocks. Before the presentation of the test tapes, subjects had a chance to familiarize themselves with the original synthetic tones on the /do/ syllable. The order of test tapes was counterbalanced across subjects, and the experimental session lasted about 90 minutes.

Subjects

Four male and four female listeners were recruited from University of Connecticut and Yale University graduate students and were paid for their participation. All were native speakers of Taiwanese and reported to have no history of speech or hearing disorder. Like all educated Taiwanese, they were also fluent in English and Mandarin Chinese.[7]

---

[5] Figure 2 actually shows a tonal pair of different original durations, whose F_0 movements have been scaled to a common duration. The figure is slightly inaccurate because it shows curvilinear F_0 movements for both original tones. In reality, after two temporally contrasting tones had been scaled to a common duration, one or both of the F_0 movements were linear.

[6] For example, each original tone occurred four times because it occurred in four different stimulus subsets (i.e., in contrast with four other tones). The slightly different response percentages to physically identical stimuli in different tables (below) derive from the fact that each stimulus subset was treated separately in the data analysis.

[7] Note that /do/ is not a possible syllable in Mandarin, so second-language interference seemed unlikely.

Procedure

The stimuli were presented binaurally over headphones at a comfortable listening level. Listeners were instructed to identify each /do/ syllable as (1) 'city', (2) 'gambling', (3) 'envy', (4) 'map', or (5) 'surname' by writing down the number of their choice. The response choices were listed on the answer sheet in Chinese characters. Subjects were told not to leave any blanks.

RESULTS

The results for the 10 tonal contrasts will be discussed in the following order: First, the two contrasts that have nominally identical contours but differ in register (1–5, 2–3); then, two contrasts that have similar F_0 movements and differ in F_0 height (4–5, 1–4); finally, the remaining six contrasts which have very dissimilar F_0 movements (as well as differences in duration), grouped into three pairs exhibiting different magnitudes of F_0 height difference (2–1, 2–5; 5–3, 1–3; 3–4, 2–4). In naming the members of a tonal pair, the one with the higher onset frequency is always named first (hence 2–1 and 5–3).

Tables 2 to 11 show the response distributions for the stimuli pertaining to each of the 10 pairs of tonal contrast. In each table, stimuli are coded in terms of F_0 movement and height, with two-digit numbers standing for the intermediate values on these two dimensions (e.g., 15 is intermediate between the original tones 1 and 5). To simplify the tables, the results have been averaged across stimuli differing in duration; effects of stimulus duration will be discussed later. Thus, each stimulus set includes nine types. The average recognition rate for the five synthetic syllables modelling the original tones was 91% correct, which confirms that these stimuli were satisfactorily synthesized (cf. Footnote 3).

Tones with the same contour, differing in register

Presumably, the more pronounced the difference in an acoustic property between two tones, the more important it will be as a cue to the tonal distinction. If, moreover, differences along other psychoacoustic dimensions are small, it will emerge as the dominant cue. Thus, for pairs such as tones 1 vs. 5 and 2 vs. 3, which nominally have the same F_0 contour but differ in register, F_0 height was expected to be the dominant cue. However, tones 2 and 3, at least, also exhibit a difference in F_0 movement (see Fig. 1), and we wondered whether that dimension would contribute to the perceptual distinction.

Tone 1 vs. tone 5. Tones 1 (high level) and 5 (mid level) have almost identical, flat F_0 movements; the difference between them is in F_0 height. The data in Table 2 reveal that F_0 height indeed plays a primary role in the perception of this distinction. Movement 1 with height 5 was identified predominantly as tone 5, whereas movement 5 with height 1 was classified as tone 1. There were virtually no confusions with other tones. Did F_0 movement have any cue value at all? The stimuli with intermediate F_0 height, which should have been the most sensitive indicators of F_0 movement effects, suggest

TABLE 2

High level tone (1) vs. mid level tone (5)

Stimuli		Responses (%)				
Mov	Hgt	1	2	3	4	5
1	1	80				20
	15	57.5			2.5	40
	5	30				70
15	1	80			2.5	17.5
	15	62.5				37.5
	5	2.5		2.5		95
5	1	90			2.5	7.5
	15	52.5		2.5		45
	5	7.5	2.5			90

a negative answer. However, movement 1 with height 5 received only 70% tone 5 responses, whereas the original synthetic tone 5 received 90%. This difference notwithstanding, the effect of F_0 movement on the perception of this tonal distinction seems negligible simply because the difference between the tones is minimal on this dimension.

Tone 2 vs. tone 3. The falling tones 2 and 3 nominally have the same F_0 contour, though F_0 in tone 2 falls somewhat more steeply than tone 3. The main difference between them is in F_0 height. The data in Table 3 confirm that F_0 height is the major cue for the distinction: Movement 2 with height 3 was identified mostly as tone 3, and movement 3 with height 2 as tone 2. However, there was also some effect of F_0 movement: At each of the three F_0 heights, more tone 3 responses were obtained when the movement derived from tone 3 rather than from tone 2. Thus, even though both tones are considered merely "falling" in traditional phonological terminology, there is in fact a perceptually relevant difference in F_0 movement between them. The difference in F_0 height, however, is clearly the dominant cue.

Tones with similar contours, differing in register

The contour of tone 4, referred to here as "rising", is actually a complex falling-rising or dipping F_0 movement with a relatively limited range (cf. Fig. 1). Because of that

TABLE 3

High falling tone (2) vs. mid falling tone (3)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
2	2		92.5	7.5		
	23		82.5	17.5		
	3		22.5	72.5		5
23	2		82.5	15		2.5
	23		80	17.5		2.5
	3		17.5	80		2.5
3	2	7.5	65	27.5		
	23		67.5	32.5		
	3		7.5	92.5		

limited range, it looks somewhat similar to the flat contours of tones 1 and 5, though to the ear it may be quite dissimilar. In contrasting tones 5 vs. 4 and 1 vs. 4, which differ in both F_0 register and contour, we expected both dimensions of F_0 to be perceptually relevant. The question was which of them would carry more weight.

Tone 4 vs. tone 5. Tones 4 (low rising) and 5 (mid level) are relatively similar in F_0 movement during the first half of their durations, but during the second half tone 4 moves up and almost merges with tone 5. There is also a difference in F_0 height, tone 4 having a lower onset than tone 5. However, the data in Table 4 reveal F_0 movement to be the primary cue: High intelligibility of both tones was maintained when their original F_0 movements were presented at uncharacteristic heights. An effect of F_0 height emerged only when F_0 movement was intermediate. Evidently, F_0 movement is the dominant cue to this distinction. This is also suggested by the finding that tone 4 responses predominated for the intermediate F_0 movement: Detection of even a slight "dip" was sufficient to elicit tone 4 percepts.

Tone 1 vs. tone 4. Since tone 1 (high level) is very similar to tone 5 in F_0 movement but higher in register, the F_0 movement difference between tones 1 and 4 is similar to that between tones 5 and 4, just discussed, only the difference in height is larger. Does this imply a larger role of F_0 height in cueing the distinction? Table 5 suggests an

TABLE 4

Mid level tone (5) vs. low rising tone (4)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
5	5	5		5		90
	54	2.5		7.5		90
	4				2.5	97.5
54	5			2.5	65	32.5
	54			2.5	70	27.5
	4				95	5
4	5				100	
	54	2.5			97.5	
	4				100	

TABLE 5

High level tone (1) vs. low rising tone (4)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
1	1	82.5				17.5
	14	25				75
	4		2.5	2.5	10	85
14	1	75			22.5	2.5
	14	7.5			67.5	25
	4				100	
4	1	15			82.5	2.5
	14			2.5	97.5	
	4				100	

affirmative answer, but it is evident that F_0 movement is still the dominant cue: Movement 4 with height 1 was still predominantly identified as tone 4, and movement 1 with height 4 was identified as tone 5, which shares the contour with tone 1. A small effect of F_0 height is evident with the original tone 4 movement: At a very high F_0, some tone 1 responses did occur. A large effect of F_0 height was obtained for the intermediate movement stimuli, whose identification changed from tone 1 to tone 4 as the height was lowered. This shift is larger than that observed in Table 4, in accord with the larger height difference for the present contrast.

Tones with dissimilar contours

The falling F_0 contour (tones 2 and 3) is acoustically and perceptually very dissimilar from the level and rising contours; it is also carried on a shorter syllable, though differences in duration will be ignored for the time being. Since F_0 movement proved to be perceptually important even for distinguishing tones with relatively similar contours, we certainly expected that tonal contrasts involving falling tones would be primarily cued by F_0 movement, with secondary contributions of F_0 height depending on the amount of the difference. In fact, it will be seen that, because for each falling tone and each level tone in Taiwanese there is a similar tone differing in register, stimuli with altered F_0 height were often identified as tones other than those in the particular contrast under consideration.

Tone 2 vs. tone 1. Tones 2 (high falling) and 1 (high level), though they are both nominally "high", do show a difference in onset frequency. The most striking difference, however, is in their F_0 movements. As the data in Table 6 show, F_0 movement is indeed the dominant cue to the contrast: Movement 2 with height 1 was still mostly identified as tone 2, whereas movement 1 with height 2 was recognized as tone 1. The stimuli with the intermediate F_0 movement showed some effects of F_0 height on their identification as tone 1, but the effect was such that tone 1 responses increased as the height was raised, even though the higher onset frequency derived from tone 2; there was no effect of height on tone 2 responses. The paradoxical effect of F_0 height derived from a tendency to identify the intermediate F_0 movement as tone 3, which indeed has a contour intermediate between tones 2 and 1, and occasionally even as tone 5; both of these tendencies to give mid-register tone responses increased as F_0 was lowered, at the expense of tone 1 responses. Basically, F_0 height seems to be irrelevant to the perception of the tone 2 vs. 1 contrast.

Tone 2 vs. tone 5. The difference in F_0 movement between tones 2 (high falling) and 5 (mid level) is the same as that between tones 2 and 1 just discussed, but the difference in F_0 height (onset) in much larger. The data presented in Table 7 show that predictable confusions occurred as F_0 height was changed: Movement 2 with height 5 was often identified as tone 3, while movement 5 with height 2 was obviously tone 1. Stimuli with intermediate movement were predominantly identified as falling tones (tones 2 or 3, depending on F_0 height), though at a very high F_0 (above the characteristic height of tone 1) some tone 1 responses occurred. In general, however, F_0 movement remained the overriding cue for this falling versus level tone distinction.

TABLE 6

High falling tone (2) vs. high level tone (1)

Stimuli		Responses (%)				
Mov	Hgt	1	2	3	4	5
2	2		96.6	3.3		
	21		95	5		
	1		87.5	12.5		
21	2	25.8	55	18.3		0.8
	21	12.5	61.6	22.5		3.3
	1	6.6	59.1	25.8		8.3
1	2	96.6		1.6		1.6
	21	98.3				1.6
	1	92.5		0.8		6.6

TABLE 7

High falling tone (2) vs. mid level tone (5)

Stimuli		Responses (%)				
Mov	Hgt	1	2	3	4	5
2	2		96.6	3.3		
	25		83.3	14.1	0.8	1.6
	5		34.1	60		5.8
25	2	27.5	51.6	17.5	1.6	1.6
	25	4.1	61.6	28.3	0.8	5
	5		32.5	60.8	1.6	5
5	2	97.5	1.6	0.8		
	25	86.6		2.5		10.8
	5	6.6	0.8	5		87.5

TABLE 8

Mid level tone (5) vs. mid falling tone (3)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
5	5	11.6	1.6	3.3	0.8	82.5
	53	4.1		5.8	1.6	88.3
	3	1.6		7.5	2.5	88.3
53	5	1.6	10	49.1	1.6	37.5
	53		6.6	65.8		27.5
	3		7.5	65.8		26.6
3	5		23.3	70.8		5.8
	53		15.8	80	0.8	3.3
	3		20	67.5		12.5

Tone 5 vs. tone 3. Here we have another level versus falling contrast, at a lower F_0 height. Tones 3 (mid falling) and 5 (mid level) originate at almost the same frequency, so no effect of F_0 height was expected. The data in Table 8 confirm that the dominant cue for the tone 5 versus tone 3 distinction is F_0 movement, though the stimuli with the intermediate movement do show a small effect of F_0 height. The copies of the original tones were not identified very well in this pair of tones; this reflects in part their inherent confusability with tones having the same contour (tones 1 and 2, respectively) but, in addition, the neutralization of duration cues and stylization of F_0 movements may have increased the number of confusions. This general ambiguity may have made listeners extra sensitive to small differences in F_0 height.

Tone 1 vs. tone 3. Tones 1 (high level) and 3 (mid falling) not only differ in F_0 movement, as do tones 5 and 3, but also in F_0 height. As can be seen in Table 9, subjects' responses changed considerably with both changes in F_0 movement and in F_0 height. However, changes in height led to predictable "confusions": Movement 1 (level) with height 3 (mid) was largely identified as tone 5 (mid level), whereas movement 3 (falling) with height 1 (high) was most often classified as tone 2 (high falling). These responses thus merely reflect the fact that F_0 height cues the tone 1 vs. tone 5 and tone 2 vs. tone 3 distinctions. The intermediate movement stimuli, however, do reveal a genuine effect of F_0 height on the contrast between tones 1 and 3: Tone 1 responses decreased

TABLE 9

High level tone (1) vs. mid falling tone (3)

Stimuli		Responses (%)				
Mov	Hgt	1	2	3	4	5
1	1	85		1.6	1.6	11.6
	13	63.3		3.3		33.3
	3	5	0.8	6.6	9.1	78.3
13	1	31.6	23.3	12.5	0.8	31.6
	13	6.6	12.5	28.3		52.5
	3		7.5	56.6	0.8	35
3	1	3.3	69.1	24.1		3.3
	13		45	50		5
	3		15	81.6		3.3

and tone 3 responses increased as height was lowered, with an increase in tone 5 confusions in the middle. Thus it appears that both F_0 height and movement are important cues for the distinction between tones 1 and 3, just as we expected.

Tone 3 vs. tone 4. There is a striking difference in F_0 movement between tone 3 (mid falling) and tone 4 (low rising). However, the difference in F_0 onset is small. The data in Table 10 confirm that F_0 movement is the primary cue for the distinction between these tones: Small changes in F_0 height left the responses to the original F_0 movements unchanged. The stimuli with the intermediate movement did show an effect of F_0 height, despite the relatively small physical differences involved. However, the effect was less on identification of these stimuli as tones 3 or 4, but primarily on their identification as tone 5 (mid level). Indeed, the intermediate movement was relatively flat and thus could be mistaken for the mid level tone when its height was raised. For the tone 3 vs. tone 4 distinction, therefore, F_0 height seems to be of little importance.

Tone 2 vs. tone 4. Tones 2 (high falling and 4 (low rising) show the sharpest F_0 movement contrast of any tone pair, as well as the largest difference in F_0 height. As can be seen in Table 11, movement 2 with height 4 was identified as tone 3 (mid falling), which is not surprising. Interestingly, however, movement 4 with height 2 was identified

TABLE 10

Mid falling tone (3) vs. low rising tone (4)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
3	3		14.1	80.8		5
	34		9.1	85		5.8
	4		10	83.3		6.6
34	3	1.7	1.7	23.3	5	68.3
	34	1.7	0.8	38.3	23.3	35.8
	4	0.8		37.5	33.3	28.3
4	3			3.3	96.6	
	34			5	95	
	4			2.5	97.5	

TABLE 11

High falling tone (2) vs. low rising tone (4)

Stimuli Mov	Stimuli Hgt	Responses (%) 1	2	3	4	5
2	2	0.8	95	4.1		
	24		72.5	25		2.5
	4		11.6	82.5		5.8
24	2	37.5	28.3	20.8		13.3
	24	0.8	32.5	44.1		22.5
	4		3.3	86.6		10
4	2	86.6	1.6	0.8	10.8	
	24	7.5		4.1	88.3	
	4	0.8		4.1	95	

as tone 1 (high level), even though movement 4 with height 1 was not so identified (see Table 5). Thus, the movement barrier between tones 1 and 4 can be overcome by a sufficient raising of F_0 height. The intermediate movement stimuli (somewhat falling with a dip) were never labeled as tone 4, but were highly ambiguous at a high F_0 and perceived as tone 3 at a low F_0. Clearly, both F_0 movement and height are important for the tone 2 versus tone 4 distinction, though the actual weights of these cues are difficult to gauge because of the intrusion of other responses.

Overview of F_0 effects

In the preceding discussion of pairwise tonal contrasts, stimuli included in one particular subset were often also relevant to the perception of other tones, as evidenced by the various response intrusions. It is useful, therefore, to survey the pattern of predominant identification responses for the complete set of stimuli, still disregarding variations in duration. Table 12 provides such an overview. In its rows, F_0 heights are arranged in terms of decreasing onset frequencies, and in its columns F_0 movements are ordered in terms of decreasing differences between the onset and offset frequencies (with one minor, deliberate reversal). Each original height (movement) occurred with nine different movements (heights), whereas each intermediate height (movement) occurred with three different movements (heights). For each height-movement combination, Table 12 lists all tonal categories with more than 20% of responses, in rank order.

The first five columns show that strongly falling F_0 movements were perceived as either tone 2 or tone 3, depending on F_0 height. The secondary relevance of F_0 movement to this distinction can be seen in the fact that tone 3 responses increased as the steepness of the F_0 movement decreased. The next four columns show that moderately falling F_0 movements were perceived as tone 3 or tone 5 at the lower F_0 onsets; stimuli with high onsets were not well sampled here, but suggest tone 1 responses with tone 2 as the second choice. The next three columns show that shallow falling F_0 movements (with an onset-offset difference of 6 Hz or less) were invariably identified as level tones: as tone 1 or as tone 5, depending on F_0 height. Finally, the last three columns show that, when the curvature of tone 4 is imposed on a flat F_0 movement, the stimuli were mostly heard as tone 4. This salient F_0 movement cue was overridden only by very high absolute F_0 values, which favored tone 1 percepts.

Duration as a cue

Finally, we consider the possible role of duration as a cue to tonal distinctions. Because the two falling tones, 2 and 3, are characterized by shorter durations than the other tones (see Table 1), a short stimulus duration was expected to be a cue to the falling contour category. In addition, shortening the duration of a falling F_0 movement increases its slope, which may further enhance the "falling" percept. This may favor tone 2 over tone 3 responses for clearly falling contours, considering the steeper slope of the tone 2 movement (cf. Fig. 1).

The results were analyzed by comparing the response percentages for the short (75 msec) and long (145 msec) durations of each stimulus whose duration was varied. (The intermediate durations were not considered.) Table 13, which is arranged in the same

TABLE 12

Predominant response categories (with percentages exceeding 20%) for all combinations of F_0 movements and heights. Numbers in parentheses represent onset frequencies (for F_0 height) and the difference between onset and offset frequencies (for F_0 movement)

F_0 height Type/Onset (Hz)		F_0 movement Type and onset-offset difference (Hz)														
		2 (45)	23 (41.5)	25 (30)	3 (29)	12 (27.5)	24 (25.5)	35 (17.5)	13 (15)	34 (13)	5 (6)	15 (3.5)	1 (1)	45 (1.5)	14 (−1)	4 (−3)
2	(150)	2	2	2,1	2,3	2,1	1,2,3	–	–	–	1	–	1	–	–	1
12	(140)	2	–	–	–	2,3	–	–	–	–	–	–	1	–	–	–
25	(131.5)	2	–	2,3	–	–	–	–	–	–	1	–	–	–	–	–
1	(130)	2	–	–	2,3	2,3	–	–	1,5,2	–	1	1	1	–	1,4	4
23	(129.5)	2	2	–	2,3	–	–	–	–	–	–	–	–	–	–	–
24	(126)	2,3	–	–	–	–	3,2,5	–	–	–	–	–	–	–	–	4
15	(121.5)	–	–	–	–	–	–	–	–	–	1,5	1,5	1,5	–	–	–
13	(119.5)	–	–	–	3,2	–	–	–	5,3	–	–	–	1,5	–	–	–
14	(116)	–	–	–	–	–	–	–	–	–	–	–	5,1	–	4,5	4
5	(113)	3,2	–	3,2	3,2	–	–	3,5	–	–	5	5	5,1	4,5	–	4
35	(111)	–	–	–	3	–	–	3,5	–	–	5	–	–	–	–	–
3	(109)	3,2	3	–	3	–	–	3,5	3,5	5,3	5	–	5	–	–	4
45	(107.5)	–	–	–	–	–	–	–	–	–	5	–	–	4,5	–	4
34	(105.5)	–	–	–	3	–	–	–	–	3,5	–	–	–	–	–	4
4	(102)	3	–	–	3	–	3	–	–	3,4,5	5	–	5	4	4	4

TABLE 13

Response categories showing changes in excess of 10% following a change in stimulus duration from 145 to 75 msec. All relevant combinations of F_0 heights and movements are shown

F_0 height Type/Onset (Hz)		F_0 movement Type and onset-offset difference (Hz)										
		2 (45)	25 (30)	3 (29)	12 (27.5)	24 (25.5)	35 (17.5)	13 (15)	34 (13)	5 (6)	1 (1)	4 (–3)
2	(150)	–	–	–	+2	–1,+2,+3,–5	–	–	–	–	–	–
12	(140)	–2,+3	–	–	+2,–3	–	–	–	–	–	–	–
25	(131.5)	–	–	–	–	–	–	–	–	–1	–	–
1	(130)	+2	+2	+2,–3	+2	–	–	–5,+2,+3	–	–	–	–
23	(129.5)	–	–	–	–	–	–	–	–	–	–	–
24	(126)	–	–	–	–	+2,–5	–	–	–	–	–	–4,+3
15	(121.5)	–	–	–	–	–	–	–	–	–	–	–
13	(119.5)	–	–	–	–	–	–	–3	–	–	–1,+5	–
14	(116)	–	–	–	–	–	–	–	–	–	–	–
5	(113)	+3,–2	+2	+2	–	–	–5,+3	–	–	–5	–	–
35	(111)	–	–	–2,+3	–	–	–5,+3	–	–	–5	–	–
3	(109)	–	–	–	–	–	–5,+3	–5,+3	–5,+3	–	–5,+4	–4,+3
45	(107.5)	–	–	–	–	–	–	–	–	–	–	–
34	(105.5)	–	–	–	–	–	–	–	–5	–	–	–
4	(102)	–	–	–	–	–	–	–	–	–	–	+3

way as Table 12, lists all response categories in which a change of more than 10% occurred as stimulus duration was shortened. (The largest change was 35%.) Plus signs indicate increases, minus signs decreases, in order of absolute magnitude. The changes are often complementary for two tonal categories, with one response increasing at the expense of another.

Our expectations were that falling tone (2 and 3) responses would generally increase and level tone (1 and 5) responses would decrease as a consequence of stimulus shortening, but that for strongly falling F_0 movements tone 2 responses might increase at the expense of tone 3 responses. The overall pattern of changes supports the first prediction: There were 11 increases in tone 2 responses versus 3 decreases, 13 increases in tone 3 responses versus 3 decreases, 1 increase in tone 5 responses versus 12 decreases, and no increase in tone 1 responses versus 3 decreases. The second, more specific prediction was less well supported: In the left half of Table 13, increases in tone 2 responses are frequent, but there are also some decreases, and changes in tone 3 responses are inconsistent, though usually complementary to the changes in tone 2 responses. On the whole, it appears that duration did have a role as a secondary cue for the falling-nonfalling tone distinction.

DISCUSSION

From these results, it is quite clear that F_0 is the most prominent perceptual cue for tonal contrasts in Taiwanese, as in all other tone languages studied so far. The present data further show that either dimension of F_0 (height or movement) can emerge as the dominant factor in cueing a particular tonal contrast, depending on the tonal patterns to be differentiated. In general, the more pronounced an acoustic difference between two tones is, the more likely it will be an important cue in perceiving the contrast. Thus, F_0 height was found to be a main cue in the distinction between tones with highly similar contours but different registers, whereas distinctions between tones with dissimilar contours were mainly cued by F_0 movement.

On the whole, F_0 movement seems to be perceptually more important than F_0 height for Taiwanese listeners. This is especially evident in the contrasts between tones 1 and 4, and tones 5 and 4, which in principle could have depended on F_0 height. One reason for this finding may be that F_0 movement is a more stable dimension than F_0 height, which varies across speakers and is often ambiguous when no contextual reference is provided. (Note that tones 2 and 3, and tones 1 and 5, tend to be confused in isolated syllables.) However, it has also been observed that linguistic background can affect the perception of F_0 patterns (Gandour, 1978, 1983). Thus it could be that F_0 height and movement are given different weights in languages with different tonal inventories, even when similar tonal contrasts are being perceived. For instance, in Cantonese, F_0 height is rather important because four out of six tones are relatively similar in F_0 movement (Vance, 1977). On the other hand, all four tones of Mandarin are dissimilar in F_0 movement (Howie, 1976). Taiwanese represents an intermediate case. Gandour (1983) found that Cantonese speakers rely more heavily on F_0 height, and less on F_0 movement, than do speakers of Mandarin and Taiwanese when judging various F_0 patterns. Our finding of the relative importance of F_0 movement in the perception of Taiwanese tones is not inconsistent with Gandour's findings.

One purpose of perceptual studies such as the present one is to go beyond linguistic nomenclature, which omits acoustic detail, and beyond phonetic studies, which describe but do not establish the perceptual relevance of these detailed aspects. Thus we have shown that the Taiwanese high falling and mid falling tones, which nominally differ only in register, also exhibit some perceptually relevant differences in F_0 movement. We have also shown that the striking difference in duration between falling and nonfalling tones can provide distinctive information in ambiguous cases. Bailey and Summerfield (1980), in their thorough studies of the perception of segmental phonetic distinctions, have argued that any systematic difference in acoustic properties between segments can be shown to be perceptually relevant. This generalization also seems to apply to the perception of tonal distinctions.

(Received June 21, 1988; accepted April 4, 1989)

REFERENCES

ABRAMSON, A.S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments.* Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

ABRAMSON, A.S. (1972). Tonal experiments with whispered Thai. In A. Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 31–44). The Hague: Mouton.

ABRAMSON, A.S. (1975). The tones of Central Thai: Some perceptual experiments. In J.G. Harris and J.R. Chamberlain (eds.), *Studies in Thai Linguistics in Honor of William J. Gedney* (pp. 1–16). Bangkok: Central Institute of English Language.

BAILEY, P.J. and SUMMERFIELD, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 536–563.

CHIANG, H.T. (1967). Amoy-Chinese tones. *Phonetica*, **17**, 100–115.

GANDOUR, J. (1978). The perception of tone, In V.A. Fromkin (ed.), *Tone: A Linguistic Survey* (pp. 41–76). New York: Academic Press.

GANDOUR, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, **11**, 149–175.

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

LIN, H.-B. (1988). *Contextual Stability of Taiwanese Tones.* Doctoral dissertation, University of Connecticut.

LIN, M.-C. (1987). The perceptual cues of tones in standard Chinese. *Proceedings of the Eleventh International Congress of Phonetic Sciences, Vol. 1* (pp. 162–165). Tallinn, Estonia, U.S.S.R.: Academy of Sciences of the Estonian SSR.

RUMYANTSEV, M.K. (1987). Chinese tones and their duration. *Proceedings of the Eleventh International Congress of Phonetic Sciences, Vol. 1* (pp. 166–169). Tallinn, Estonia, U.S.S.R.: Academy of Sciences of the Estonian SSR.

TSENG, C.Y. (1981). *An Acoustic Phonetic Study of Tones in Mandarin Chinese.* Doctoral dissertation, Brown University.

VANCE, T.J. (1977). Tonal distinctions in Cantonese. *Phonetica*, **34**, 93–107.

ZEE, E. (1978). Duration and intensity as correlates of F_0. *Journal of Phonetics*, **6**, 221–225.