

How Can the Lyrics of a Song in a Tone Language Be Understood?

PATRICK C. M. WONG* and RANDY L. DIEHL

*MEZ 330, Department of Psychology, University of Texas, Austin,
TX 78712, U.S.A.*

pwong@uchicago.edu

diehl@psy.utexas.edu

Abstract

In a tone language, pitch variations are used to contrast word meaning. For example, the Cantonese syllable /si/ means “teacher” when spoken in a high pitch and “yes” when spoken in a low pitch. How is fundamental frequency (F_0) used to signal lexical tones that occur in songs? In an examination of Cantonese songs, it was found that songwriters abandon the ratio scale of F_0 differences that is applied to lexical tones in carefully read speech and instead use an ordinal scale. For example, a high tone that is normally 12% higher than a mid-tone in speech can be realised as *any* higher F_0 (but never a lower F_0) in songs. A perceptual experiment showed that native Cantonese-speaking listeners similarly apply an ordinal F_0 scale to arrive at the lexical meaning of the lyric. This ratio-to-ordinal mapping in Cantonese songs ensures the musicality of the melody while preserving adequate identifiability of lexical tones in the lyric.

Introduction

In a tone language, variation in pitch or its physical correlate, fundamental frequency (F_0), is used to contrast word meaning. Thus, in order to understand a word, listeners are required to extract its pitch pattern in addition to its consonants and vowels. Cantonese Chinese, with six lexical tones, offers a particularly rich example of the distinctive use of pitch pattern. Table 1 shows the different word meanings associated with these tones when carried by the syllable /si/. The numbers in parentheses describe the starting and ending pitch levels of each tone on a scale from 1 (the lowest pitch level) to 5 (the highest pitch level) (Chao, 1947). A difference of one pitch level on the scale is approximately one musical semitone (6% in F_0). Thus, the Cantonese high-level tone (Tone 1) is two musical semitones or one musical whole tone higher than the mid-level tone (Tone 3), an F_0 difference of approximately 12% (1.06^2). To produce a syllable with a high-rising tone (Tone 2), a talker starts at one pitch level and slides up by one musical whole tone (*e.g.*, from F to G).

Here we consider the question of how songwriters in a tone language reconcile the possible linguistic role of F_0 variation in expressing a lyric with the musical role of F_0 variation in specifying melodic intervals. There appear to be three options. The first is to ignore lexical tones and word meaning and to use pitch exclusively to mark the melody. This preserves musicality at the cost of reduced lyric

*Patrick Wong is now at Brain Research Imaging Center, Department of Neurology, the University of Chicago, 5841 S. Maryland Ave. MC-2030 Chicago, IL. 60637.

TABLE 1
The six Cantonese lexical tones.

<i>Tone</i>	<i>Description (Pitch Level)</i>	<i>Example</i>
1	High-level (5-5)	/si1/ "teacher"
2	High-rising (3-5)	/si2/ "history"
3	Mid-level (3-3)	/si3/ "to try"
4	Low-falling (2-1)	/si4/ "time"
5	Low-rising (2-3)	/si5/ "market"
6	Low-level (2-2)	/si6/ "yes"

intelligibility. The second option is just the reverse: to preserve the normal pitch variations of lexical tones while ignoring the melody, sacrificing musicality for intelligibility. Songs in this case would be very speech-like. The third option is intermediate between the first two: songwriters may attempt to preserve at least partially the pitch contrasts of lexical tones while not unduly restricting the melodic role of F_0 changes.

Chao (1956) investigated the relationship between melody (or tune) and lexical tone in Chinese songs of various styles. He found that in Chinese "Singsongs", a style of music that is intermediate between speaking and singing, each lexical tone is sung with a consistent pitch pattern throughout the song. For example, he found that one singer produces all the high-level tones in a song on the musical note A (440 Hz*). Lyric intelligibility is preserved because listeners can reliably use the pitch pattern to identify tones and consequently words. However, in contemporary Mandarin songs, Chao found that composers mostly ignore lexical tones in their compositions. Unfortunately, Chao reported only the names of the songs he analysed without providing the actual analyses. Furthermore, lyric intelligibility was not discussed in his article.

Yung (1983) found a melody-tone relationship in Cantonese Opera that was somewhat similar to what Chao (1956) reported in Singsongs. In the song that he analysed, the high-level tone is variously sung on E (659.3 Hz), G (784 Hz), or D (587.3 Hz), the mid-level tone is consistently sung on C (523.3 Hz), and the low-level tone is sung mostly on A (440 Hz) and occasionally on B (493.9 Hz). Thus, each lexical tone is sung on a relatively small number of musical notes, with no overlap in the mapping between musical notes and tones.

Songs of other tone languages have also been investigated. For example, List (1961) examined chants and songs of Central Thailand and found that "speech melody" (*i.e.* the pattern of pitch intervals of lexical tones) is mostly preserved, but the range of variation of the musical melody is limited.

Each of these earlier studies examined strategies used by composers and singers to reconcile linguistic and musical roles of F_0 , but they did not consider whether corresponding strategies are used by listeners to extract the lexical meaning of

*Throughout this article, F_0 values that are associated with musical notes are nominal values based on the musical scale, not actual measured values.

the lyric. In the present study, we first analysed the melody–tone relationship of contemporary Cantonese songs. We then conducted a perceptual experiment to determine whether listeners interpret the lyrics based on principles consistent with the results of our song analysis. Songs of Cantonese were chosen for the study because it is one of only 15% of the world’s tone languages having six or more lexical tones (Hombert, 1977). Reliance on F_0 to resolve word meaning is probably greater in tone languages with more tonal distinctions. Thus, a rule system governing the use of F_0 to signal both lexical tone and melody may be more likely to exist in languages such as Cantonese. Contemporary Cantonese songs were chosen because with the exception of Chao (1956), who did not find an association between tone and melody in contemporary Mandarin songs, previous studies in this area have examined only non-contemporary songs. We were interested in whether the evolution of music composition involved an abandonment or relaxation of rules governing melody–tone relations.

SONG ANALYSIS

Four contemporary Cantonese songs were chosen for analysis. Each was popularised in the early 1990s and was considered typical (non-avant-garde) music. The songs included: /sing1/ (*Star*) composed by Koo-Chuen Sun-Si, /zo6 yuen4 ye5 suen6/ (*In the Field*) by Kwan Ching-Kit, /yan4 zo6 lue5 tou4 sa2 lue6 si4/ (*When One Cries in the Journey*) by Koo Ka-Fai, and /hot5 cho2/ (*Cheering*) by Danny Chan. It was found that the six lexical tones are classified into three groups, high-pitch, mid-pitch and low-pitch, depending on their ending pitch levels. For the high-pitch group (the H group), including Tones 1 and 2, the tone ends with a “5” in terms of the scale proposed by Chao (1947). For the mid-pitch group (the M group), Tones 3 and 5, the tone ends with a “3”, while for the low-pitch group (the L group), Tones 4 and 6, the tone ends either with a “1” or “2”. Our analysis showed that, in general, an ordinal mapping between musical note and tone group occurs such that the direction of pitch change in two consecutive musical notes is the same as in the two consecutive tone groups attached to them. An example of this ordinal mapping is presented in Figure 1, which shows an excerpt from /zo6 yuen4 ye5 suen6/ (*In the Field*). The second musical note is D. It is followed by B-flat which is *eight* semitones *higher* than D. The two words that are associated with these two notes are /yuen4/, which belongs to the L group, followed by /ye5/, which belongs to the M group. Note that the M group is *higher* in pitch than the L group but only by *four* semitones. This excerpt demonstrates that whereas in carefully spoken Cantonese utterances, tone sequences have a characteristic F_0 ratio (*i.e.*, a constant percentage of F_0 change) (Chao, 1947), in contemporary Cantonese songs, corresponding tone sequences preserve only the *direction* of F_0 change. (Note also the contrast with Singsongs, in which each lexical tone is associated with a consistent musical note.)

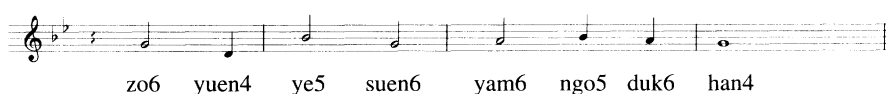


FIG. 1

An excerpt from /zo6 yuen4 ye5 suen6/ (*In the Field*).

Table 2 shows the number of times across the four songs a given type of musical note sequence was associated with a given type of tonal sequence, with both sequences expressed in terms of the presence and direction of pitch change (*i.e.*, Upward, Downward, Same). The numbers in parentheses are percentages of occurrence relative to the total number of cases included in the matrix. Across the four songs, an ordinal mapping between musical and tonal sequences occurred 91.81% of the time (sum of the diagonal cells in Table 2). Not included in the analysis are 20 instances in which a (monosyllabic) word was sung on two consecutive musical notes. In these instances, the first musical note was ignored in the analysis. For example, in one of the songs, the word /mei6/ “never”, which is normally spoken with a low-level tone (Tone 6), was sung on E followed by D. In such cases, the analysis was based on the second musical note (D) and the following musical note and their associated tones. (The following music-tone pair in this case was /nung4/ “able” sung on C.)

Across the four songs, there were 77 instances in which the interval between two consecutive musical notes was larger than two musical whole tones. Among these 77 instances, there were only 36 in which the difference between their two associated lexical tones was larger than two steps (*i.e.*, they either changed from L to H, or from H to L, but not from L to M or from M to H). This further demonstrates that the ratio scale of F_0 difference in speech is abandoned in music.

TABLE 2
Direction of two consecutive musical notes in relation to direction of the associated tone sequence.

		<i>Musical Sequence</i>		
		<i>Up</i>	<i>Down</i>	<i>Same</i>
Tone	<i>Up</i>	109 (38.79%)	3 (1.07%)	0
Sequence	<i>Down</i>	3 (1.07%)	120 (42.70%)	0
	<i>Same</i>	17 (6.05%)	0	29 (10.32%)

Sum of diagonal cells (% of time ordinal mapping occurred) = **91.81%**.

PERCEPTUAL EXPERIMENT

The aim of the perceptual experiment was to examine whether the ordinal mapping between musical notes and lexical tones implied by the above song analysis is used by listeners when identifying tones and words.

Method

Stimuli

Two different sets of three melodies were composed by the first author, who is a certificate holder of the Royal School of Music (United Kingdom). Within each set, the sequence of musical notes was the same except for the last note.

The lyric for all the melodies was a semantically neutral Cantonese sentence /ha6 yat1 go3 zi6 hai6 si3/ “The next word is to try”. Figures 2 and 3 show the stimulus melodies. The melodies were sung by a native Cantonese-speaking trained female amateur singer who had won numerous awards in singing competitions in Hong Kong. The recordings were made in a sound-attenuated chamber using a Shure SM 48 microphone; they were digitised and stored on a PC. The singer was provided with the musical score and was instructed to repeat the melodies played using the MIDI composition software. As a trained musician, the first author verified that the singer sang the melodies in pitch.

Listeners

Listeners were four male and four female native speakers of Cantonese with normal hearing (self-reported) and varying degrees of musical training. They were all students of the University of Texas at Austin and were paid for their participation in the study.

Experimental Task and Procedures

The experiment was conducted in a sound-attenuated chamber. Listeners were asked to identify the last (target) word of the lyric in each melody. Responses were made by pressing one of the three response keys labeled with the Chinese character of the following three words that differ only in lexical tone: “teacher” (/si1/), “to try” (/si3/) and “yes” (/si6/). The stimuli were output at 10 kHz via a 16-bit D/A converter, low-pass filtered at 4.9 kHz, and presented to subjects over Beyer DT-100 earphones at 72 dB SPL. There were three repetitions of each stimulus, with all items randomised.

Results and Discussion

Based on the written lyric alone, the singer was to produce a mid-level tone (/si3/) on the final (target) word of each melody. However, if listeners interpreted the sung words according to an ordinal mapping rule, they would be expected to identify the target tone by comparing the musical note on which it was sung to the immediately preceding tone and its musical note. For each stimulus, Figures 2 and 3 show predicted responses (underscored) based on an ordinal mapping rule. For example, in stimulus 1 the last note had an F_0 of 440 Hz (A), while the immediately preceding note had an F_0 of 329.6 Hz (E). Because this preceding tone belongs to the low-pitch (L) tone group, the target tone is predicted to be judged as either M or H (*i.e.*, /si3/ “to try” or /si6/ “yes”).

Table 3 displays the listeners’ responses to each stimulus target. Out of 144 cases (6 melodies \times 3 repetitions \times 8 listeners), 137* – or over 95% – of the responses were as predicted by an ordinal mapping rule†.

*Among the seven exceptions, five were associated with stimulus #5. This may be owing to the fact that the musical note B (247 Hz) of the target in that stimulus is lower in pitch than the preceding note C# (277.2 Hz), which is already a member of the L tone group. The additional pitch lowering of the target, which occurs uniquely with this stimulus, would be unexpected and may have caused some confusion among the listeners.

†The lack of a quantitative relationship between the stimuli and the expected responses (at least for some stimuli) effectively ruled out the use of inferential statistical analyses.

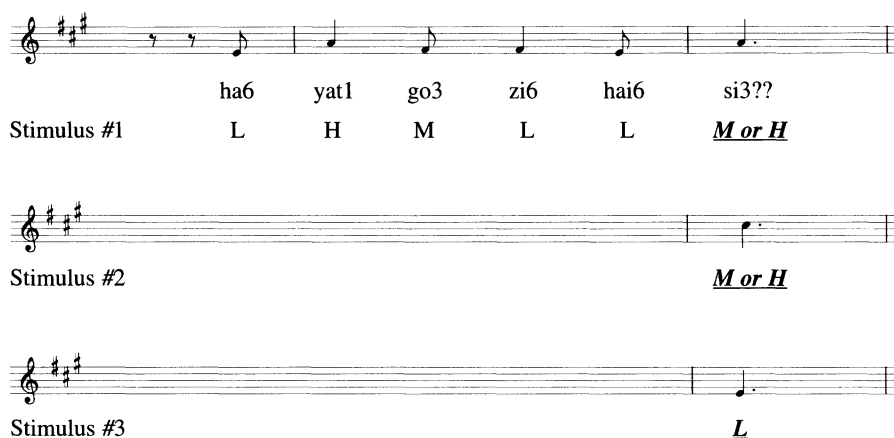


FIG. 2

Stimulus Set 1. For Stimulus #2 and #3, the initial five musical notes and their corresponding words are identical to those of Stimulus #1. The letters “L”, “M” and “H” refer to the pitch group of the words. The underscored text refers to the listener responses predicted on the basis of the ordinal mapping rule.

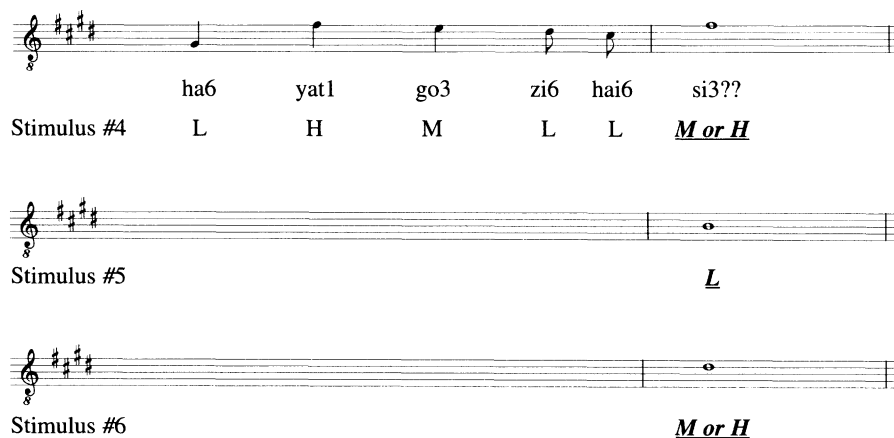


FIG. 3

Stimulus Set 2. For Stimulus #5 and #6, the initial five musical notes and their corresponding words are identical to those of Stimulus #4. The letters “L”, “M” and “H” refer to the pitch group of the words. The underscored text refers to the listener responses predicted on the basis of the ordinal mapping rule.

TABLE 3
Frequencies of subject responses to each stimulus. Expected responses are underscored.

<i>Stimulus number</i> (see Figs 2 & 3)	<i>Subject Response</i>			
	<i>*Musical Notes Diff.</i>	<i>/si1/</i>	<i>/si3/</i>	<i>/si6/</i>
1	+5	<u>12</u>	<u>11</u>	1
2	+9	<u>24</u>	<u>0</u>	0
3	0	0	0	<u>24</u>
4	+5	<u>22</u>	<u>1</u>	1
5	-2	0	5	<u>19</u>
6	+2	<u>0</u>	<u>24</u>	0

*This shows the difference between the target musical note and its immediate preceding musical note. +5 means the target is five musical semitones higher.

For four of the six stimuli (1, 2, 4 and 6), the ordinal mapping rule predicts that either of two responses – M or H – will be made following a tone from the L group, but the rule does not make differential predictions concerning the relative likelihood of either of those responses. With no additional assumptions, we might expect the observed frequencies of the M and H responses in those four cases to be roughly equal. As shown in Table 3, this expectation is confirmed only in the case of stimulus 1. In the remaining three cases, the responses lopsidedly favour one of the two tone categories. The relative probabilities of the M and H responses may be related to the distance in semitones between the penultimate L tone and the target tone (displayed in Figures 2 and 3). As this distance increases, the likelihood of an H response also tends to increase. Thus, the 9-semitone distance in stimulus 2 is associated consistently with H responses, the 2-semitone distance in stimulus 6 is associated just as consistently with M responses, and the 5-semitone distance of stimulus yields a mixed response pattern.

If, as these results hint, listeners use the pitch distance between consecutive syllables in making assignments of tone category within a melody, then it is necessary to supplement the ordinal mapping rule along the following lines: when ordinal mapping results in ambiguity of tone assignment, an interval or ratio scale of F_0 distances between consecutive tones may be used by listeners to help resolve the ambiguity. However, other aspects of the data are difficult to square with this proposal. In particular, both stimulus 1 and stimulus 4 have a 5-semitone distance between the penultimate L tone and the target, but the target in stimulus 4 was identified as an H tone much more consistently than the target in stimulus 1. The reason for this discrepancy is unclear, but in any case it does suggest that listeners do not consistently apply an interval or ratio scale of F_0 in judging tone category in a song lyric. Rather, listeners' tone assignments are most generally characterised by an ordinal mapping rule similar to that used by composers of contemporary Cantonese songs.

As mentioned in the Method section, listeners in this experiment had varying degrees of musical training. It is unclear how musical proficiency may have influenced the results. However, none of the listeners was a professional musician nor were any enrolled in a music degree program. Given the robustness of the results, it is not unreasonable to assume that use of the ordinal mapping rule is not strongly dependent on degree of musical proficiency.

GENERAL DISCUSSION

Unlike Chao (1956), we found that songwriters take into consideration F_0 differences among lexical tones in composing contemporary Cantonese songs. In carefully read speech, such F_0 differences may be characterised in terms of a consistent ratio scale. For example, Chao (1947) found that the Cantonese high-level tone (Tone 1) is 12% higher F_0 than the mid-level tone (Tone 3). Our song analysis indicated that songwriters abandon this ratio scaling among lexical tones in songwriting and adopt an ordinal mapping rule for adjacent tones such that, for example, Tone 1 must only be higher in pitch than Tone 3. This allows greater melodic freedom while preserving adequate intelligibility of the lyric.

Correspondingly, we found that when listening to songs, listeners apply an ordinal mapping rule in assigning tone categories. Although this is likely to reflect the listeners' experience with the phonetic realisation of lexical tones in songs, such a rule may be more generally applicable given certain processes observed in ordinary conversational speech, including downdrift (Wong, 1999) and tonal co-articulation (Shen, 1990). Downdrift refers to a cumulative pitch lowering of tones throughout the course of a phonological phrase, resulting in a falling intonation contour. Because in Cantonese the pitch lowering is relatively greater for higher tones (Wong, 1999), the effect is to diminish the pitch distances among tones near the end of an utterance. Tonal co-articulation can also influence the pitch distance between two tones depending on the tonal context in which the tones occur. For example, a high-level tone may have an F_0 closer to a mid-level tone due to a lower adjacent tone. In such cases, listeners may need to use something like an ordinal mapping rule in assigning tone categories to words in the utterance. Thus, the results of our perceptual experiment may reflect the listeners' experience with ordinary speech patterns as well as with the conventions of songwriting.

References

- Chao, R. C. (1956). Tone, intonations, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In: M. Halle, H. G. Lunt, H. McLean and C. H. Van Schooneveld (Eds), *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday, 11th October 1956*. The Hague, the Netherlands: Mouton & Co.
- Chao, R. C. (1947). *Cantonese Primer*. Cambridge, MA.: Harvard University Press.
- Hombert, J.-M. (1977). A model of tone system. *U.C.L.A. Working Paper in Phonetics*, **36**, 20–32.
- List, G. (1961). Speech melody and song melody in Central Thailand. *Ethnomusicology*, **5**, 16–32.
- Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, **15**, 281–295.
- Wong, P. C. M. (1999). The effect of downdrift in the production and perception of Cantonese level tone. *Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco*, **3**, 2395–2398.
- Yung, B. (1983). Creative process in Cantonese Opera I: The role of linguistic tones. *Ethnomusicology*, **27**, 29–47.