SOAS Working Papers in Linguistics and Phonetics Vol. 11 (2001): 3-29

Tonal coarticulation in Cantonese two-tone sequences

Choi-Yeung-Chang Flynn
flinzinchina@hotmail.com

1. Introduction

1.1. Defining tonal coarticulation

Coarticulation is defined by Ladefoged as the overlapping of adjacent articulations involved in utterances (1982: 82). In speech production, the vocal organs never move in separate steps. When a string of sounds is pronounced, the vocal tract is continuously in motion rather than exhibiting a sequence of steady states. The articulation is continually changing and sounds inevitably influence each other. As a result, sounds assimilate to each other. However, the concept of coarticulation is generally restricted to segments. For instance, Keating writes (1990: 452): "coarticulation refers to articulatory overlap between neighbouring segments, which results in segments generally appearing assimilated to their contexts." It is easier to grasp the process of coarticulation between segments as we understand that it takes time for one articulator, for example, the tip of the tongue, to move from one position to another. There is never a precise locus in gesture when producing more than one sound. Sometimes we can even feel the presence of such a movement physically. As regards pitch, we understand that the varying of the pitch of a sound depends on the number of vibrations of the vocal folds in a particular time, but we cannot feel the number of vibrations themselves. However, there is no reason to believe that the phenomenon of coarticulation happens only to segments and not to tones. Tone needs segments to carry it. During the coarticulation of segments, gestures may overlap in time (Browman and Goldstein 1986: 219). Tone, as a suprasegmental phenomenon associated with segments, may also participate in overlapping gestures. Moreover, the vocal folds may also need time to adjust their movement from producing one segment or one tone to another. These adjustments may only take some milliseconds or a few pulses of vibration to accomplish, but this is the crucial thing we need to look at. We presume that there must be a transition time for the vocal folds to adjust their movement from one tone to another. Such a simultaneous or overlapping process is called tonal coarticulation. As a result of tonal coarticulation, the phonetic value of tones is assumed to be varied. One objective of the present study is to examine the phonetic variation of tones when undergoing tonal coarticulation.

1.2. Directions of tonal coarticulation

Crystal (1985: 52) elaborates two types of segmental coarticulation. One is anticipatory coarticulation, which happens when "an articulator not involved in a particular sound begins to move in the direction of articulation needed for a later sound in the utterance". The other is the coarticulation "when a sound retains a characteristic deriving from an earlier articulation". (We call the latter carryover coarticulation.) It is reported that while English has a tendency to show larger anticipatory effects, French and Italian show larger carryover effects (Ladefoged 1982: 53). In Cantonese, Rao *et al* (1982: 294) reports that segmental assimilation effects are in both directions. Recent acoustic research on tonal coarticulation confirms the existence of bi-directional assimilation and its effect on both tone contour and tone height (e.g., Vietnamese, Han and Kim 1974; Thai, Gandour *et al* 1992 and 1994 and Potisuk *et al* 1997; Mandarin, Shen 1990 and Xu 1994). Interaction between tones is restricted to adjacent tones (Shen 1990). I am aware of no acoustic study on Cantonese tonal coarticulation. Hashimoto is probably the only one to observe that, in normal speech the onset and coda of the Cantonese tones "are modified, each according to its immediately preceding and following tones... for example, the rising tones in context tend to have a greater fall at the onset if preceded by higher frequency tones and a more prominent rise if followed by the same" (1972: 93). This indicates that anticipatory and carryover effects are both present and contours are affected positively in Cantonese. The present study attempts to provide a systematic exploration of the phonetic behaviour of individual tones in sequences.

2. Method

2.1. Material

After considering various factors which affect F_0 and examining several alternatives, it was decided to use real words rather than nonsense words, and words voiced throughout rather than words with voiceless onset or coda, for the investigation of tonal coarticulation in Cantonese. There are thirty-six combinations, each with two sets of words, one with /ji:/ as the first syllable and one with /ji:/ as the last. All possible combinations of disyllabic words from the six tones which consist of /ji:/ in two-tone sequences are used: for example, /**ji:**^{1} ji:^{1}/, /**ji:**^{1} jy:n^{2}/, /**ji:**^{1} wɐj^{3}/, /**ji:**^{1} li:w^{4}/, /**ji:**^{1} lɛŋ^{5}/, /**ji:**^{1} mi:n^{6}/ and /ɐw^{1} **ji:**^{1}/, /jɐw^{2} **ji:**^{1}/, /a:^{3} **ji:**^{1}/, /mɐŋ^{4} **ji:**^{1}/, /jy:^{5} **ji:**^{1}/, /ji:^{6} **ji:**^{1}/, etc, (the target syllable is in bold. For the word list, see appendix 1.) The target syllable /ji:/ is one of the few words which can carry all the six tones. It also has an advantage in that there is a similarity between the consonant and the vowel, as the approximant /j/ is known as a nonsyllabic version of the high vowel /i:/; and is described as usually consisting of a rapid glide from a high vowel to the following vowel when it occurs at the beginning of a syllable (Ladefoged 1975: 209). It is also known that the average F_0 of vowels shows a systematic correlation with vowel height. The higher the vowel, the higher the F_0 (Lehiste and Peterson 1961: 419, and Ohala 1978: 29). The difference in F_0 between high and low vowels may be as much as 25 Hz. The choice of /j/ and /i:/ sequences is to minimise any effect on F_0 due to vowel height.

/ji:/ is placed to follow or precede a syllable beginning and ending with sonorants in order to see what will happen in the course of voicing through the two syllables. What occurs at the transition of the two tones appears to me to be the crucial point of the coarticulation of the two tones. If we want to see the actual transition of the two tones through the F_0 contour, we have to let the vocal folds vibrate throughout two syllables. Sequences of /ji:/ with a syllable beginning and ending with a sonorant appear to be a desirable choice to maintain the vocal cord tension during the continuous voicing. All of the disyllabic words are represented with characters and are natural sequences, i.e. the words make sense to native speakers. Thus, the requirements of both voicing throughout and meaning place limits on the choice of possible combinations. Under these requirements, choosing absolutely identical phonetic elements and syntactic structure seems to be impossible. As the aim of this study is to investigate tonal coarticulation, only the target syllable is restricted to having the same phonemes. The disyllabic words used in this investigation are isolated words, i.e., not in carrier sentences, and are placed in a reading list in order to avoid tonal assimilation between the test words and other neighbouring words.

2.2. Speakers

The informants, JCCF, JHDG and LTHJ, are all native speakers of Cantonese born in Hong Kong and are currently students at SOAS. One female (JCCF) and two males (JHDG and LTHJ); two are aged 20+ (JHDG and LTHJ) and one 40+ (JCCF); one undergraduate (JHDG) and two postgraduates (JCCF and LTHJ).

2.3. Recording

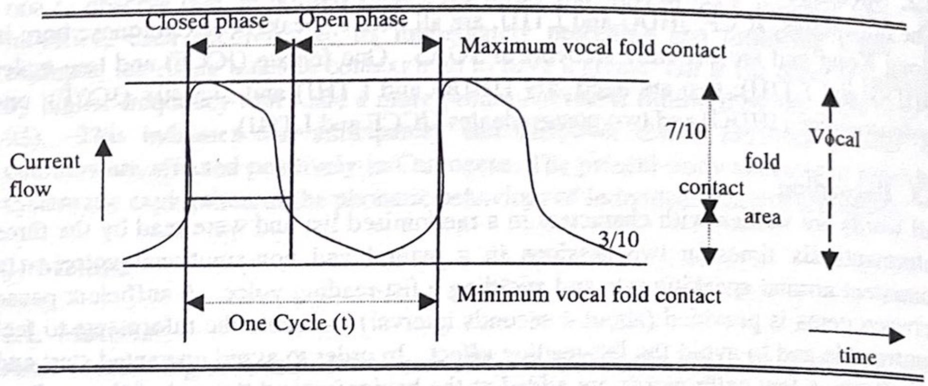
All words are written with characters in a randomised list and were read by the three informants six times in two sessions in a natural and non-emotional voice, at a consistent normal speaking rate and avoiding a list-reading voice. A sufficient pause between items is provided (about 4 seconds interval) to enable the informants to feel comfortable and to avoid the list-reading effect. In order to avoid unwanted start and end effects, a few extra words are added at the beginning and the end of the reading. The microphone-to-mouth distance was maintained at 30cm. A phone-hood covered the microphone to improve the quality of recording. The recording was made in a sound-proofed booth at SOAS using a laryngograph and a Bruel and Kjœr Electret condenser microphone with a measuring amplifier type 2609 into a Sony PCM. The PCM converts the analogue signal into a digital signal which is further adapted to a video signal and recorded onto a Sony Betamax Video Recorder. The recordings are then played into a laryngograph processor (this allows levels to be adjusted) and then into the PCLx SPG and the Speech Workstation (SW) software packages which are installed in a 486 Dx 66 16 M Ram PC computer.

2.4. Laryngograph

The laryngograph is a device enabling the vocal fold contact area to be investigated noninvasively (Fourcin *et al* 1971: 172). Two electrodes, with the help of a bondage wrapping on the neck, are placed on both sides externally at the level of the larynx. When the vocal folds vibrate, the current flow passing between the electrodes changes. With no vocal fold contact, the current flow between the electrodes is at the minimum; with full vocal fold closure, the current flow is at the maximum. The output of such changing current flow which passes between the electrodes is called the laryngograph waveform (Lx). In one cycle of vocal fold vibration (t) in a laryngograph waveform, the frequency (f) of the vibration is calculated as $f = (1/t)$.

Each waveform cycle (t) can be analysed into two parts: one corresponds to the closed glottis and is called closed phase and the remaining part is the open phase. The closed phase part of the cycle expressed as a percentage is known as the closed quotient (CQ). The closure is from the first vocal fold contact to the maximum vocal fold contact in the waveform. This is based on the assumption that the vocal folds snap together more rapidly than they part. The ratio between the closed phase and the open phase in this case is 7:3. This ratio is essentially arbitrary, but has been found convenient and reliable in previous studies (Davies *et al* 1986: 539). The area of the maximum vocal fold contact is used by the laryngograph analyser software to calculate the periodicity of the waveform. The closed quotient is calculated as CQ = [(closed phase / fundamental cycle of Lx) x 100]%) (Howard 1995: 165). Increases in the closed quotient are in response to the vocal folds abduction resulting in glottal closure.

Fig. 1. Illustration of laryngograph output waveform and the closed quotient measurement.



2.5. Measurement

The laryngograph recordings can be displayed as waveforms and spectrograms in both packages; F_0 contour, Lx waveform and closed quotient contour (Qx) can be derived from the laryngograph trace and displayed simultaneously in the SPG. The packages provide the optimum display for clarity of measurement. As each single vocal pulse can be clearly demonstrated in the Lx waveform, the duration of the disyllabic words is taken from the first vocal pulse to the last. All tokens are digitised at a 10 kHz sampling rate.

Time points of syllable boundaries are determined by inspection of the speech waveform, the wideband spectrogram, and the Qx and by listening to the signals. The of segmentation is conventional. The Qx is useful because it demonstrates the increase in the closed quotient of the glottal cycle associated with vocal fold abduction. Generally speaking, the higher the Qx, the lower the frequency - i.e., the Qx contour and the F_0 contour are in inverse direction. Any turning point in the F_0 contour more or less accompanies (but not necessarily) a turning sign in the Qx contour at the same time. Changing in segments is also signalled in the Qx contour by increasing or decreasing the closed quotient. The first F_0 value is taken as the tonal onset of the first syllable and the last F_0 value is taken as the offset of the last syllable. It is not abnormal to see the last few flaps of the vocal folds drop the F_0 contour. This signals that the vibration of the vocal folds is weakened when approaching the end of the vibrations. In this case, the last F_0 value is not necessarily reliable, so I use the F_0 value just before these drops (usually it sits approximately 20ms before the end). When the same phenomenon occurs at the beginning of the F_0 contour, i.e., the vocal folds have one or two weak vibrations at the beginning, and this would lead the F_0 contour to show a weak start, the first F_0 value after the weak start is used. It is observed that the F_0 contour is not necessarily a smooth contour. It has a few turning points. The turning points are observed to occur especially more or less a quarter of the duration of the syllable after the tonal onset or before the tonal offset. Therefore, it was decided to take the F_0 value at three points apart from the onset and offset: a quarter, half and three quarters of the total duration of a syllable. The values of these three points plus the values of the onset and offset are taken into computation to determine the F_0 contours. It is not uncommon to produce a creaky voice in low frequency words in which the F_0 contours become aperiodical. A handful of tokens which result from this and have a long period of aperiodical F_0 contour were excluded from the investigation together with some tokens judged to be unrepresentative. Subsequently many tokens had to be discarded in order to obtain an equal number of tokens for each set of the two-tone sequences. 648 tokens in total were involved: 6 tones x 12 preceding/following contexts x 3 repetitions x 3 informants. Thus a total of 3240 F_0 values (5 measurement locations x 648 tokens) were taken into computation.

2.6. Normalisation

For the reason that not all tokens of each tone are of equal duration, each syllable is normalised for duration on a percentage scale. F_0 contours are also normalised for each of the speakers by using the formula: $F' = 100^* (F_0 - B_{Hz}) / R$. (F': the normalised value; R: F_0 Range of each speaker = $T_{Hz} - B_{Hz}$; T_{Hz} and B_{Hz}: top and bottom Hz values of each speaker's F_0 range estimated from a reasonable sample of their speech). By such normalisation, F_0 contours may be compared across speakers. F_0 contours were smoothed by curve-fitting for display purposes only.

3. Results and discussion

3.1. Monosyllables

3.1.1. Shape and height in the six tones

First of all, I recorded eighteen monosyllables of /ji:/ with six tones, three different lexical items for each of the six tones and took their F_0 mean in order to give a brief description of the characteristics of the tones when they are pronounced in isolation.

The following figures 2-3 display the six tones in monosyllabic utterances. Figure 2 illustrates the 6 tonal patterns from the individual informants and figure 3 illustrates the 6 tones separately across the three informants.

Fig. 2 (a-c). Mean F_0 contours of the 6 tones pattern in monosyllables produced by JCCF, JHDG and LTHJ, respectively, on the following set of words: /ji:[1]/ (medical/ cloth/ aunt), /ji:[2]/ (chair/ rely on/ lean on), /ji:[3]/ (meaning/ Italy/ thought), /ji:[4]/ (son/ and/ move), /ji:[5]/ (ear/ already/ discuss), /ji:[6]/ (two/ righteousness/ easy).

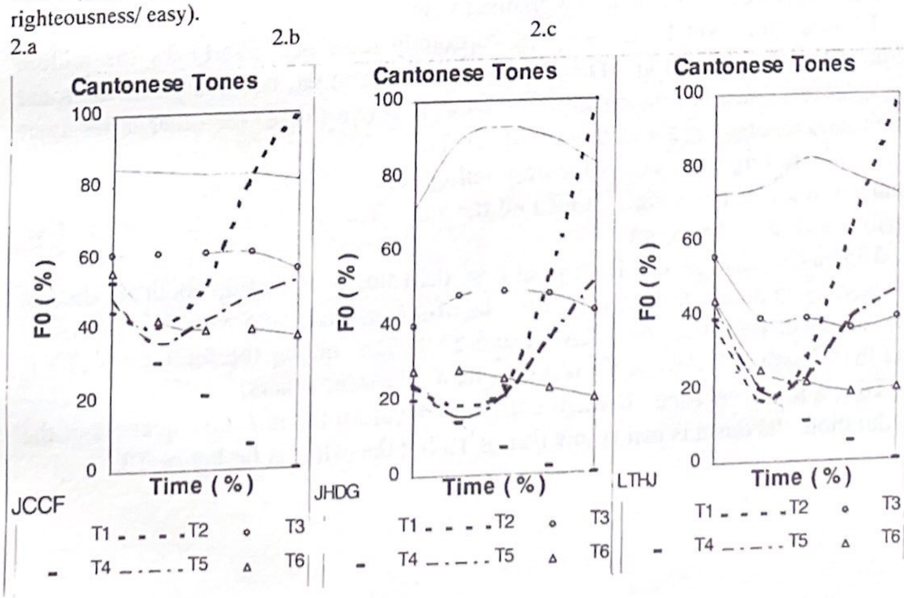
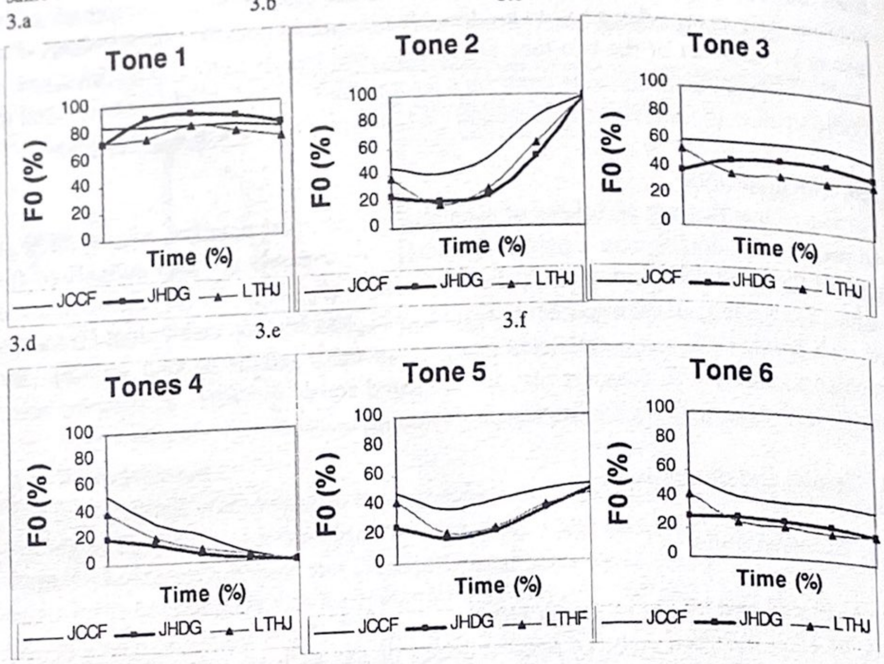


Fig. 3 (a-f). Mean F_0 contours of the six tones in monosyllables arranged by individual tones, on the same set of words as in figure 2.



T1 is a high level tone. Its shape is like a clothes-hanger: high in the centre but low at both ends. Informant JHDG has the most arched curve shape and JCCF the least. There is no falling variant among the informants. The onset is the highest compared with that of the other tones. It apparently stands out from other tones by virtue of its overall high pitch level.

T2 is a high rising tone. The rise is preceded by a slight dip. The turning point is about a quarter of the duration of the syllable from the onset. The rise starts slowly, and when it reaches half of the duration of the syllable, it suddenly jumps up sharply. The offset is the highest compared with other tones.

T3 is a mid level tone. It keeps constantly level particularly for the middle portion of the tonal duration. The slope in the first and the last quarter of the tonal duration differs from the informants. Its onset is above that of the other tones apart from T1, and its offset is above that of T4 and T6.

T4 is a falling tone, and is the only falling tone. Its onset is squeezed with T2, T5 and T6 in a mid low range. It falls all the way. The fall ends as the lowest offset compared with the other tones.

T5 is a low rising tone. It drops at first, then slowly climbs up about for the last three quarters of duration of the syllable. Its offset just meets up with T3. The order of the height of the offset between T3 and T5 differs among the informants (T3 is higher than T5 for JCCF while T5 is higher than T3 for the others).

T6 is a low level tone. It keeps constantly level for the last three quarters of the tonal duration. Its onset is just below that of T3 but the offset is far below T3.

3.1.2. Hierarchy of the height of the onsets and the offsets of the six tones

Shape and height are both important to Cantonese tones. Overall, the onsets are crowded in while the offsets spread over. The height of the onsets is particularly crowded in the lower part of the frequency range where T6, T4, T5 and T2 are squeezed, but T1 and T3 stretch above them. Incidentally, the onset of T6 just comes above the onsets of T5, T2 and T4 and the order of the last three is not consistent across the informants. The hierarchy of the height of onsets in this study agrees with that in Ching's informants (see figures in Ching 1981: 49) - the highest onset is T1, the next is T3, then the others' orders differ from one informant to another. The hierarchy of the height of the onsets of T2, T4, T5 and T6 also displays a non-uniform pattern in the data of Fok's informants (see figures in Fok 1974: 149-154). This implies that the F_0 values in onsets are not so important as constituent factors in the identity of tones (particularly T2, T4, T5 and T6).

The hierarchy of the height of the offsets is clearly located as T2 is the highest and T4 is the lowest; T1, T5, T3 and T6 (listed in order from the highest to the lowest) are in between; T3 and T5 are very close together and sometimes in reverse order across the informants. The hierarchy of the height of the offsets in this study agrees with that in Ching's informants but agrees with Fok's only where her T1 is a high level tone (see figures in Fok 1974: 149) (her T1 has a high fall variant). This implies that the F_0 values in offsets are not very important either, at least not T3 and T5, though more important than the F_0 values in onsets.

3.1.3. Summary

The level tones differ from each other in overall height particularly in the last three quarters of the tonal duration. The rising tones differ from each other both in height and in slope: the heights differ in the last two quarters of the tonal contour as a result of their different degrees of rising - T2 has a steep rise and T5 has a gentle rise. The falling tone differs from others both in height and in shape. The mid range tones T3 and T5 differ from each other primarily in shape - one is level, one is rising.

3.2. Disyllables

3.2.1. Coarticulatory effects on individual tones

3.2.1.1. Anticipatory effects

The figures 4-9 illustrate the six target tones as the first tone of the disyllabic words, displaying their phonetic variations as a result of the influence of their following tones.

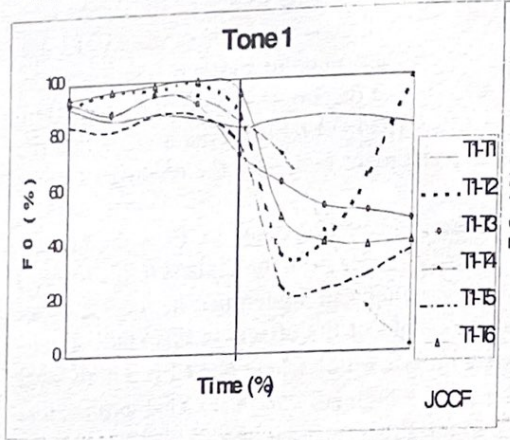
3.2.1.1.1. Phonetic variation in the height of individual tones

All informants show a greater height variation in T1 than in the other tones. T1 is significantly lower all the way through when followed by T1 than when followed by other tones. T1, on the other hand, remains at a higher level when followed by T5 or T6 than when followed by other tones. This can be ascribed to a dissimilatory effect on height and such an effect to a different extent can be seen in other tones.

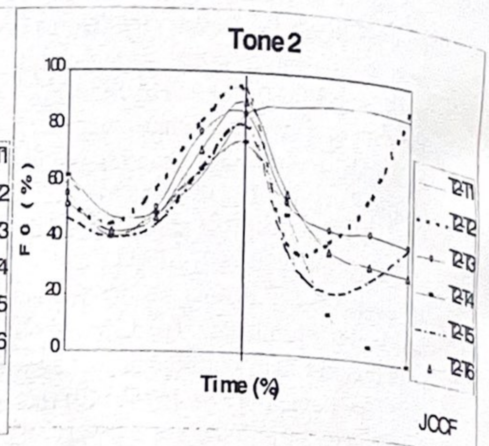
Fig. 4 (a-c). Mean F_0 contours of T1 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 5 (a-c). Mean F_0 contours of T2 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

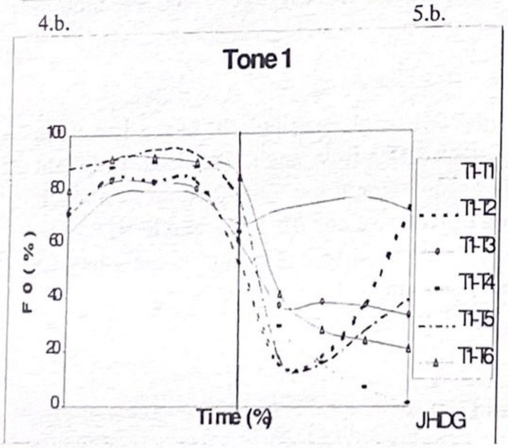
4.a.



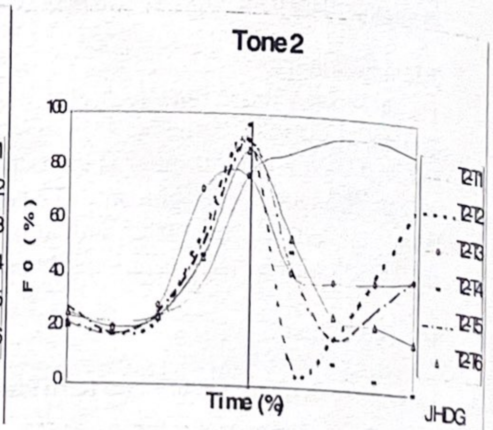
5.a.



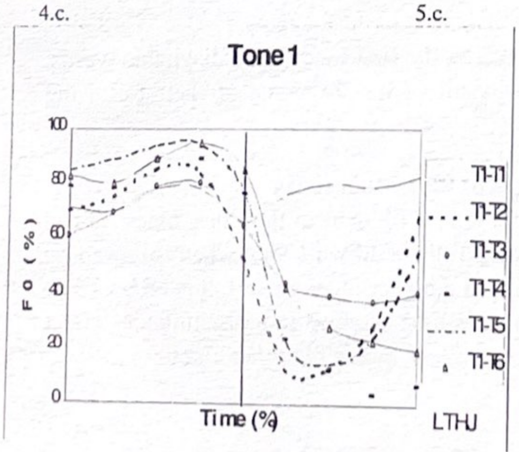
4.b.



5.b.



4.c.



5.c.

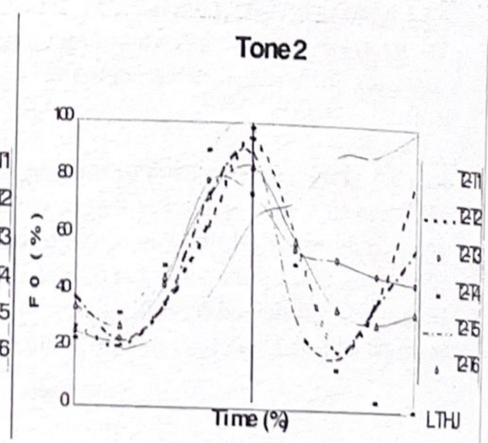
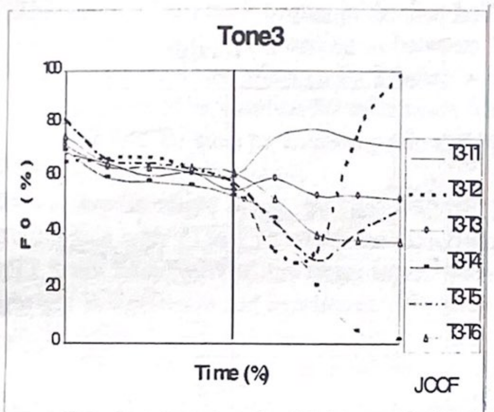


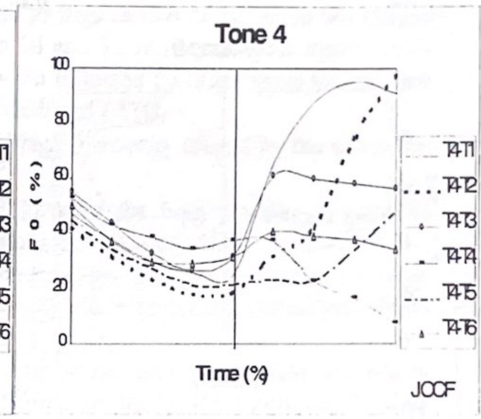
Fig. 6 (a-c). Mean F_0 contours of T3 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 7 (a-c). Mean F_0 contours of T4 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

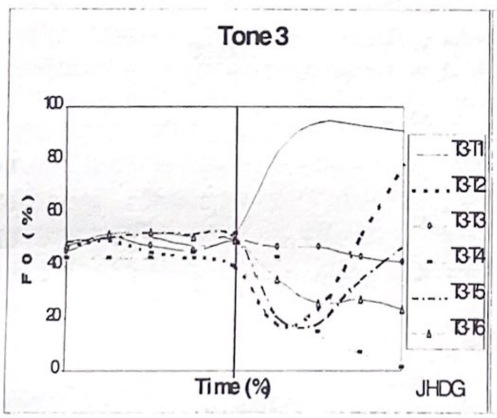
6.a.



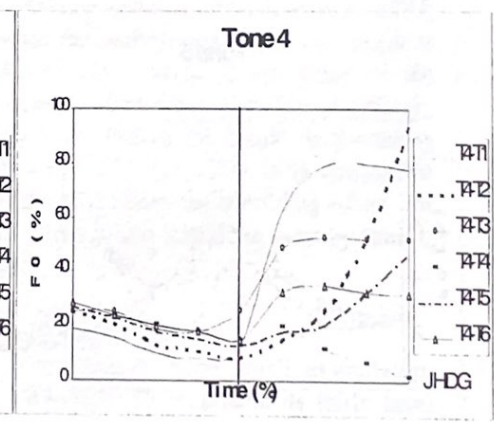
7.a.



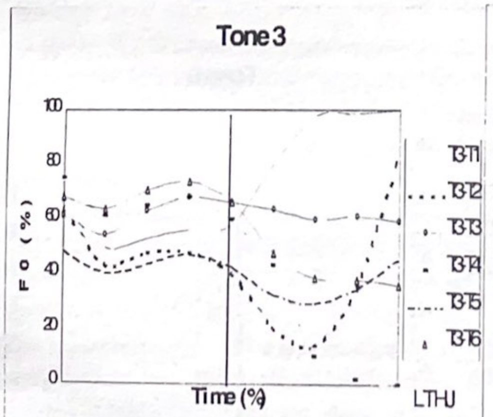
6.b.



7.b.



6.c.



7.c.

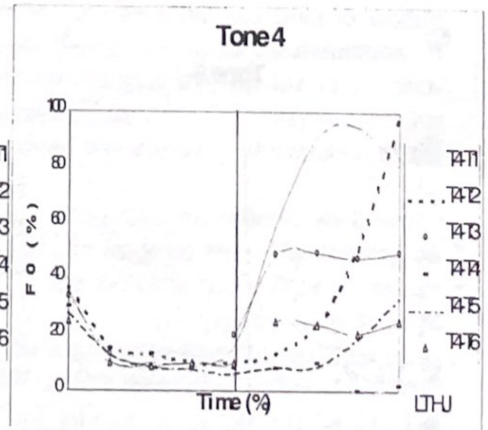
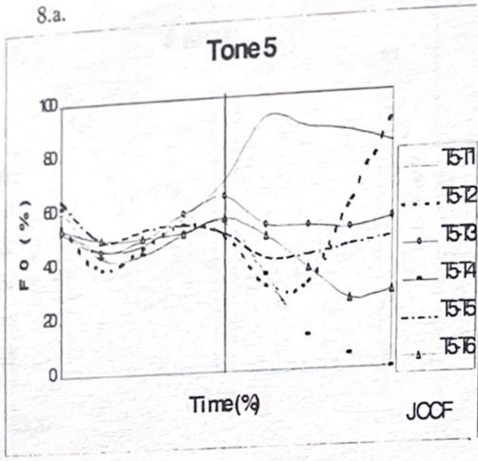


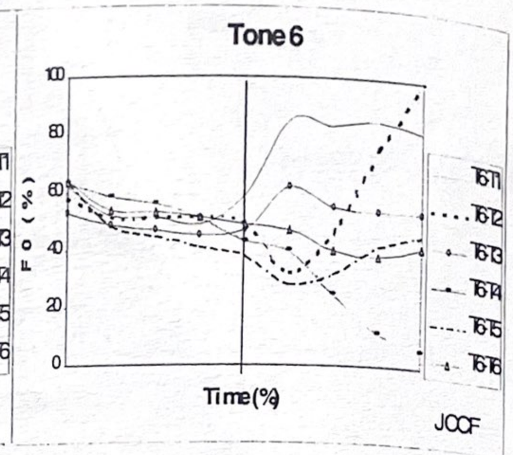
Fig. 8 (a-c). Mean F_0 contours of T5 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 9 (a-c). Mean F_0 contours of T6 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing anticipatory effects. Inside each figure, the vertical line indicates the tonal boundary.

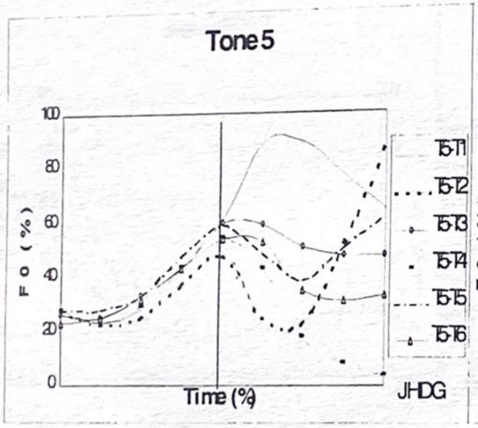
8.a.



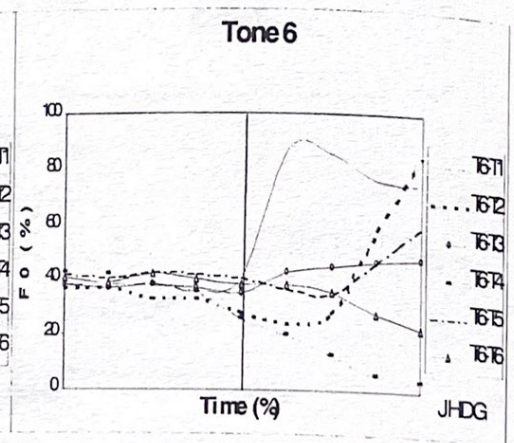
9.a.



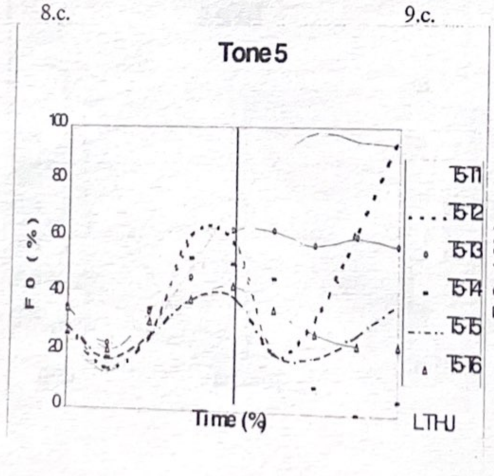
8.b.



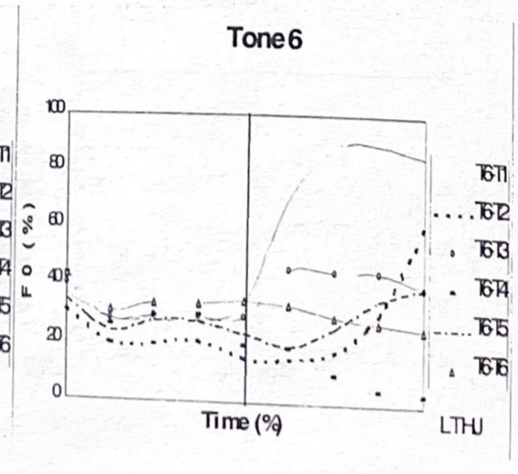
9.b.



8.c.



9.c.



T4 is apparently higher when followed by T4 than when followed by other tones. T2 and T5 are raised by different low onset tones across the informants. For informants LTHJ and JHDG, T2 is ostensibly higher at the last quarter of the tonal duration when followed by T4. For informant JCCF, T2 raises the preceding T2 much more than T4 does.

All informants show a smaller height variation in T5 than in the other tones. This may be because it begins in the first half as high as that of T2 but in the second half it rises less than T2, ending in between T6 and T2 in its canonical form. T5 is slightly higher when followed by T5 than when followed by other tones for the first three quarters of its duration for informants JCCF and LTHJ.

T3 and T6 have no uniform pattern of height hierarchy caused by the following tones among the informants.

Across target tones, T6, T5, T4 and T2 raise the height of the target tone compared with T1 and T3 which tend to set the target tone at a middle or low position. T1 is not found to raise any target tones. This indicates that the anticipatory effect on height is significant and constitutes a dissimilatory effect which can extend backwards to the onset of the target tone.

This finding regarding how far the anticipatory effect on height extends is similar to that in Mandarin, where "the effects are exerted on the whole tone" (Shen 1990: 284) but different from that in Thai, where "anticipatory effects extend backward to about 50% of the duration of preceding syllable" (Gandour *et al* 1994: 489). However, our finding regarding whether the anticipatory effect on height is assimilation or dissimilation agrees with that in Thai, where "a low onset of the following tone raises the height of the preceding target tones that have higher F_0 offsets whereas a high onset of the following tones lowers the height of preceding target tones with lower F_0 offsets" (Potisuk *et al* 1997: 33). This is in contrast to Mandarin, where a low onset lowers and a high onset raises the preceding offset, for example, "tonal offsets are lowered the most by tone Ø and raised the most by tone 4, and the next most by tone1." (Shen 1990: 285)

3.2.1.1.2. Phonetic variation in the slope of individual tones

All target tones preserve their canonical shape despite some small assimilatory modifications at the last quarter of the tonal duration. T1 maintains its fairly level contour with some fluctuation. The centre point (at 50% of its duration) seems to be a fulcrum of a scale which maintains the height but allows the two sides to modify slightly. T1 displays no uniform pattern for the first quarter of the tonal duration. It dips at the last quarter of its tonal duration in order to meet all other low onset tones. The dipping phenomenon is also seen when followed by a T1. This may suggest that the change in syllable boundary has an effect on the frequency of vibration of the vocal folds.

T2 maintains its small fall followed by a big rise, but reduces its dramatic `flying-up' tail at the last quarter of its duration to a moderate rise. The bending-off slope at the last quarter of the duration of T1 and the slow-rising slope at the last quarter of the duration of T2 exhibit a little anticipatory assimilation on the slope to the following tone. T2 exhibits a more steeply rising slope from the dip (25% of its duration) to the offset when followed by T2 and T4 than when followed by other tones. On the other hand, the slope of T2 is the least steep when followed by T1. The last quarter of the tonal duration even exhibits a small fall when followed by T3. This shows that dissimilation primarily affects the height (as discussed above) and hence also affects the slope. This finding on the behaviour of the T2 runs counter to Hashimoto's observation in Cantonese that the rising tones in context tend to have a more prominent rise if followed by higher frequency tones (1972: 93).

T3 maintains its fairly level shape but with a small downward slant. Even though there are traces of modifications at the offset, very little anticipatory change is found in T3.

The slope of T4 has different characteristics among the informants: for JHDG, T4 falls in a straight line; for LTHJ, T4 falls heavily at the beginning but levels out over the rest of the tonal body; for JCCF, T4 maintains its falling shape but dramatically reduces the extreme drop in its tail to a moderate falling or a levelling out or a moderate rising for the last half of its duration. The fall of T4 behaves in the same way when followed by any of the six tones.

T5 maintains its small fall followed by a gently rising shape. But for informant JCCF, the last quarter of the tonal duration of T5 sometimes (when followed by T2 and T5) changes its direction to a fall. The change of direction in the last quarter of the tonal duration is also found when followed by T4 and T5 for informant LTHJ. The rising of the last portion of T5 is significantly steeper when followed by T1 than when followed by other tones across the informants.

T6 maintains its fairly level shape with slightly downward slanting. The first quarter of its tonal duration maintains a gentle slant down in the data of informants JCCF and LTHJ while being level in the data of informant JHDG. At the last quarter of its duration, it turns from downward to upward only when followed by T1. This indicates that assimilation is present.

The phenomenon of changing direction of F_0 in the last portion of the target's duration is also seen in Vietnamese (Han and Kim 1974: 227), where the curve tone changes its upward tail to downward when followed by a level tone. In Mandarin, tone 1 changes from a shallow downward falling-rising contour to a falling contour when before tone Ø, while tone Ø generally has a falling contour across tonal environments (Shen 1990: 284). [Shen distinguishes tone sandhi from tonal coarticulation: the former is attributed to language specific morphophonemic constraints while the latter is attributed to language independent biomechanial constraints.] Across the six tones, anticipatory effect on slope in Cantonese is minimal and is assimilatory in nature extending backward to a quarter of the duration.

3.2.1.2. Carryover effects

The figures 10-15 illustrate the six target tones in the position of the second syllable of disyllabic words, displaying their phonetic variation as a result of the influence of their preceding tones.

3.2.1.2.1. Phonetic variation in the height of individual tones

Regarding the effect of coarticulation on height, two sides of the story are to be considered: first, the beginning part of a tone contour, which is about the first quarter of its duration; second, the main body of a tone contour, which is about the last three quarters of its duration. The beginning part of a tone contour tends to exhibit an assimilatory effect, whereas the main body of a tone contour tends to exhibit a dissimilatory effect. Generally speaking, the onsets of the six target tones are raised by the preceding high offset tones - T2 and T1, and pulled down by the preceding low

offset tones - T4 and T6. Whatever the target tone is, the height of its onset is ranked in order from high to low when it is preceded first by T2 or T1, then by T3 or T5 or T6, and then by T4 (with a very small number of exceptions, for example, the preceding T3 comes as the second highest onset of T1 for LTHJ). This order is as in that of the tonal offsets in monosyllabic words. This picture once again confirms that whatever modification in height or slope is made on the preceding tone, the resistant effect on the preceding tone is much stronger than that on the following tone. In other words, the carryover effect in Cantonese tones is much stronger than the anticipatory effect.

However, this assimilatory nature of the carryover effect in the beginning portion of a target tone changes in the main body of the target tone. All informants show a greater height variation in T1 than in the other tone. T1 remains at a low level when preceded by T1 (the lowest for JHDG and the penultimate lowest for JCCF and LTHJ). T1 is generally higher when preceded by tones other than T1.

T2 is apparently higher in the first half of its duration when preceded by T2 as a result of being raised by its high offset. However, it sinks down to the lowest offset soon after enjoying the highest position. Almost all onsets of the six individual target tones are the highest when preceded by T2 in comparison with other preceding tones. (Alternatively, the preceding T1 raises the following onset to the highest for T1, T5 and T6 for informant LTHJ, this may be due to tonal peak placement of the preceding T2 occurring much earlier). Even though the preceding T2 raises the onset of the following tones, it does not raise the overall contour of the following tones. What happens is that the target tone has to intersect with others in a quarter or a half of its duration when preceded by T2 but then settles in the lower level (the lowest for informant JCCF and the penultimate lowest for informants JHDG and LTHJ.) This finding is in contrast with Han and Kim's finding for Vietnamese, that "if the onset of a tone is raised under the influence of the preceding tone with a high overall pitch the end would also be raised, the amount of the rise being dependent on the amount of the rise at the onset" (1974: 228). This finding is also in contrast with Shen's finding for Mandarin that the entire tonal contours are shifted up or down by the surrounding tones (1990: 293).

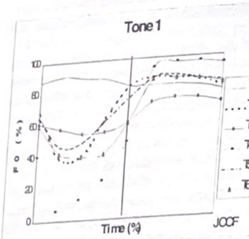
T3 is lower when preceded by T1 than when preceded by other tones. T4 is higher when preceded by T2 than when preceded by other tones for informants JHDG and LTHJ. It is higher when preceded by T1 for informant JCCF. There is no uniform pattern for T5 among the three informants. T6 is apparently higher when preceded by T4 than by other tones for informants JCCF and JHDG. It is higher when preceded by T6 for informant LTHJ.

Across target tones, T6 and T4 preceding the target tones tend to raise the height of the target tones when compared to T1 and T2. All these once again confirm that the carryover effect on height manifests itself as dissimilation in the main body of a target tone, which is similar to that for the anticipatory effect. This finding is in contrast with Han and Kim's finding for Vietnamese that a tone tends to be a high variant before or after a broken, rising, or level tone, and a low variant before or after a curving or falling tone. "Tones with a high overall pitch pull up the pitch of the tones in their immediate environment, and tones with a low overall pitch pull down the pitch of the tones adjacent to them" (1974: 226).

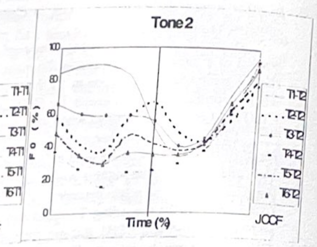
Fig. 10 (a-c). Mean F_0 contours of T1 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 11 (a-c). Mean F_0 contours of T2 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

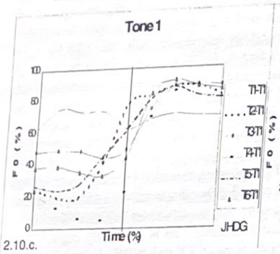
10.a.



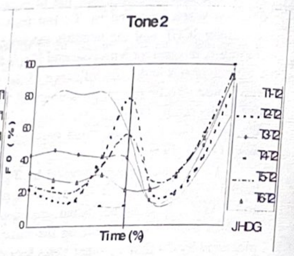
11.a.



10.b.



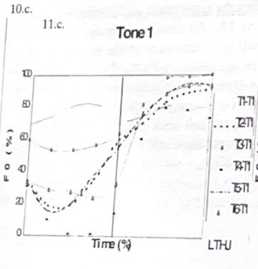
11.b.



2.10.c.

2.11.c.

10.c.



11.c.

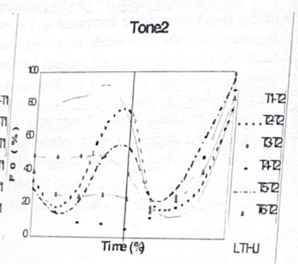
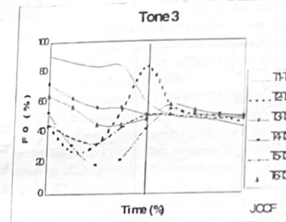


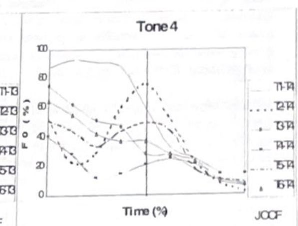
Fig. 12 (a-c). Mean F_0 contours of T3 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 13 (a-c). Mean F_0 contours of T4 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

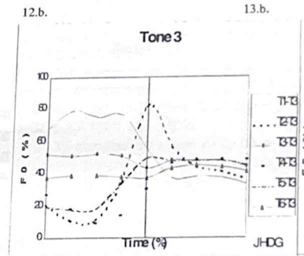
12.a.



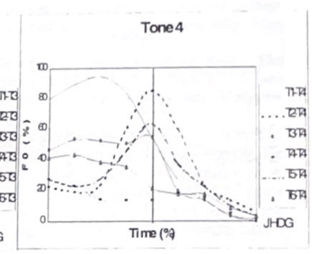
13.a.



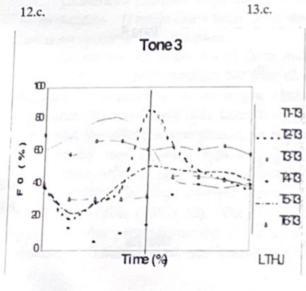
12.b.



13.b.



12.c.



13.c.

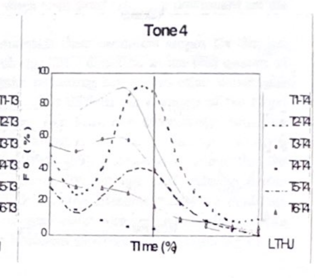
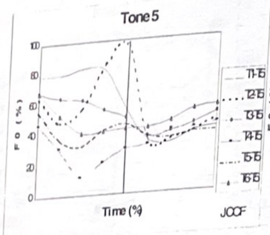


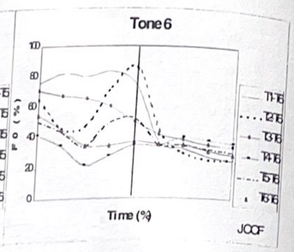
Fig. 14 (a-c). Mean F_0 contours of T5 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

Fig. 15 (a-c). Mean F_0 contours of T6 and its following tones produced by JCCF, JHDG and LTHJ, respectively, showing carryover effects. Inside each figure, the vertical line indicates the tonal boundary.

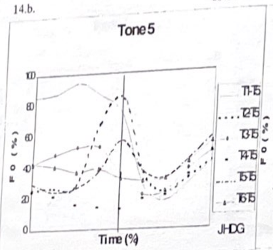
14.a.



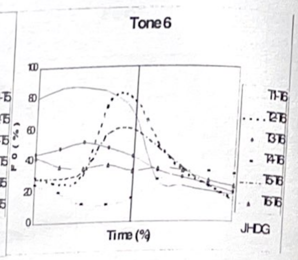
15.a.



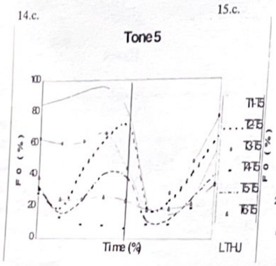
14.b.



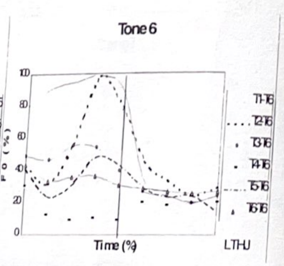
15.b.



14.c.



15.c.



Regarding the carryover effect on height, this finding is also in contrast with that for Thai claimed by Gandour *et al* that "carryover effects extend forward to about 75% of the duration of the following syllable" (1994: 489).

3.2.1.2.2. Phonetic variation in the slope of individual tones

The carryover effect on slope exhibits an assimilatory modification. All target tones maintain their canonical shape for the last three quarters of their duration. For the first quarter of their duration, the tone contours head up or head down from their onsets whose heights are different to their canonical position - the tonal onset of the second syllable is built on the offset of the first syllable which has small modifications towards the second syllable.

T1 remains level for the last three quarters of its duration and has a rising slope for the first quarter of its duration, even when preceded by T1 or T2. The occurrence of a small concave pattern in the juncture of the tone contour of the two-tone sequences of T1-T1 may suggest that change in syllable boundary is accompanied by change of frequency of vibration of the vocal folds.

T2 maintains its sharp rising slope in the last half of its duration. It has a falling contour in general (except when preceded by the very low offset T4) for the first quarter of its duration then starts rising gently from the beginning of the second quarter of its duration.

T3 maintains its fairly level shape with slight slanting downward for the last three quarters of its duration. It rises for the first quarter of the duration only when preceded by low offset T4 and T6.

T4 maintains its falling shape with different degrees of falling slope. The greatest steepness of fall is when T4 is preceded by the highest offset T2. It also reaches a very low offset in comparison with when preceded by other tones for informants JCCF and JHDG.

T5 maintains its gentle rising shape in the last three quarters of its duration. The beginning portion of T5 maintains a fall in general (except when following T4), the degree of the fall totally depending on the height of the offset of its preceding tone - a higher offset falls more and this results in a steeply downward slope; a lower one falls less and the result is a gentle downward slope. This partly agrees with Hashimoto's (1972: 93) observation for Cantonese that "... the rising tones in context tend to have a greater fall at the onset if preceded by higher frequency tones."

T6 exhibits a similar slope under the same conditions as T5 for the first quarter of its duration. It maintains a fairly level shape with gentle slanting downward for the last three quarters of its duration.

Across the six tones, all of them maintain their canonical slopes for the last three quarters of their duration, but may change their direction in the first quarter of the duration depending on the height of the preceding offset. In other words, the carryover effect on slope only extends to the first 25% of the duration of the target tones and the effect is assimilatory in nature. For Thai, two contrasting results are obtained by two separate experiments: Gandour *et al.* claim that in carryover assimilation, slope is relatively unaffected (1994: 489); Potisuk *et al.* claim that the carryover effect in slope extends forward about 70% of the total duration of the following syllable (1997: 32). They conclude that the different results are evidently caused by the presence or the absence of continuous voicing across the syllable boundary. Gandour *et al.* use intervening voiceless plosives at the beginning of the target syllables; Potisuk *et al* 's utterances are continuously voiced throughout. They think the effect of voiced onset on the F_0 contour manifests itself in stronger assimilatory effects. The purpose of the design of continuous voicing across the syllable boundary in this study is to enable a non-stopping F_0 contour throughout the adjacent tones.

Comparing the modification on slope in the two directions from the syllable boundary, the carryover effect is obviously greater than the anticipatory effect, even though both last for only about the first 25% of the tonal body (starting from the syllable boundary).

3.2.2. The pattern of the six tone contrast in different tonal environments

The twelve figures above (4 – 15) illustrate the six contrastive tones in the first or the second syllable of disyllabic words when they are adjacent to each of the six individual tones. They display the phonetic variations in the pattern of the six tone contrast despite the fact that they are modified to a great extent in terms of height and slope under the influence of either the preceding or the following tones. In particular the main body of the tone (i.e., three quarters of the duration away from the adjacent tone), clearly display the relatively constant contrastive pattern.

3.2.3. The impact of segments in tonal contours

The following figures 16 and 17 illustrate the six individual tones in identical tonal sequences of the disyllabic words carried by different segments (taking the mean of the same three informants).

a. In the figures 16 and 17, the solid line represents a tone carried by a /ji:/ segmental string for the first syllable and by other sonorants for the second syllable; the dashed line represents a tone carried by a /ji:/ segmental string for the second syllable and by other sonorants for the first syllable. The difference in height or in slope of the pairs of the identical tonal sequences in the identical tonal environment but carried by different segments strongly suggests that the difference may be caused by different segments.

Generally speaking, the tone contour carried by /ji:/ is considerably higher in F_0 and has less fluctuation than that carried by other sonorants for the same tone. This is clear because /i:/ is the highest vowel and the formants of /j/ are very similar to those of /i:/. A difference in height or in the slope of the pairs of the tonal contour may reflect the difference of segments which carry it.

This enables us to explain why in T2 the two tone contours differ in height in the first syllable but almost overlap in the second syllable. The segments in T2 are /ji: ŋɐj / and /ŋɐj ji:/. /ɐ/ is a low central vowel whose F_0 is lower than that of the high vowel /i:/, and therefore displays a lower F_0 and a steeper falling tone contour than that of /i:/ in the first syllable. For the second syllable, the tone contour carrying /ɐ/ falls more than the one carrying /i:/, and the two tonal contours meet up for the rest of the duration. The initial nasal consonant /ŋ-/ appears to have influence on the F_0 of the following vowel /ɐ/, owing to most of its energy being in the low frequencies.

T4 exhibits a similar phenomenon to T2. The two tone contours of T4 are carried by /ji: na:n/ and /jɐn ji:/ and end up with /jɐn/ lower in F_0 than /ji:/ in the first syllable and almost overlap in the second syllable.

Fig. 16. Mean F_0 contours of T1, T3, T4 and T6 occurring twice consecutively. The vertical line inside the figure indicates the tonal boundary. The "T" in the legend represents a tone with segments other than /ji:/. The segments are specified as: the solid lines represent the following syllables: /**ji:**^{1} ji:^{1}/, /**ji:**^{3} ow^3/, /**ji:**^{4} na:m^4/ and /**ji:**^{6} mej^6/; the dashed lines represent the following syllables: /ɐw^1 **ji:**^{1}/, /ji:m^3 **ji:**^{3}/, /jɐn^4 **ji:**^{4}/ and /ma:n^6 **ji:**^{6}/.

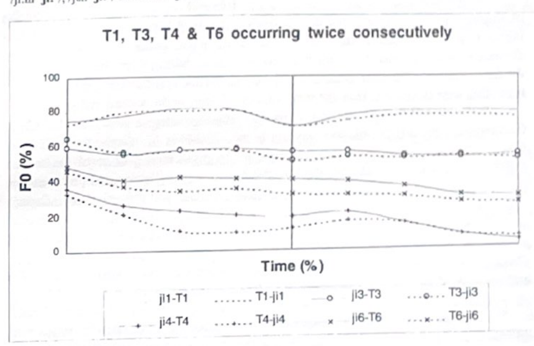
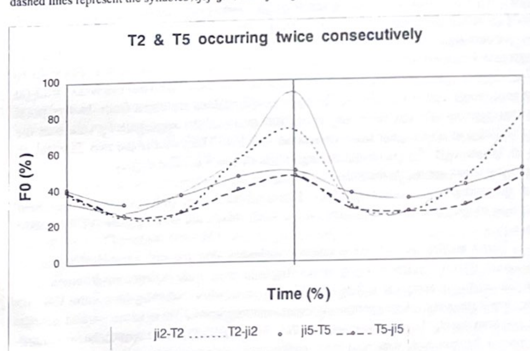


Fig. 17. Mean F_0 contours of T2 and T5 occurring twice consecutively. The vertical line inside the figure indicates the tonal boundary. The "T" in the legend represents a tone with segments other than /ji:/. The segments are specified as: the solid lines represent the syllables /**ji:**^{2} ŋɐj^2/ and /**ji:**^{5} jy:^{5}/; the dashed lines represent the syllables /ŋɐj^2 **ji:**^{2}/ and /jɐw^5 **ji:**^{5}/.



T5 is carried by /ji: jy:/ and /jɐw ji:/. The tone contour in the first syllable is similar to T2 - the tone carried by /ɐ/ has a lower F_0 and a steeper falling slope than that of /i:/. In the second syllable, the contour of /ji:/, beginning with a low onset is close to that of /jy:/ but they do not overlap because /y:/ is also a high vowel.

T6 is carried by /ji: mej/ and /ma:n ji:/. Its tonal contours have no overlapping or intersecting. The tone contour of /ji: mej/ consists of high vowels and occupies a higher F_0 position, whereas the tone contour of /ma:n ji:/, whose low vowel /a:/ keeps the tone contour in a low F_0 position, displays a deeper falling slope than that of /ji:/ in the first syllable. The tone contour of /ji:/ in the second syllable has less fall than its preceding tone contour or than the tone contour of /mej/ in the second syllable.

T3 is carried by /ji: ow/ and /ji:m ji:/. The tone contour of /ji: ow/ has slight fluctuation in the syllable of /ow/ whereas in the syllable of /ji:/ displays a fairly level line. Similarly, the tone contour of /ji:m ji:/ in the central portion of the second syllable keeps a fairly level line which is like that in monosyllabic words, whereas the tone contour in the first syllable of /ji:m/ displays a gentle fall reflecting the influence of the nasal coda.

T1 shows a different picture in which two tone contours (carried by /ji: ji:/ and /ɐw ji:/) are very close in terms of F_0 height and slope but diverge slightly in the /ɐw/. These findings strongly suggest that if we want to compare two tonal contours, it is more reliable if the two tonal contours bear the same segments.

b. No attempt is made in this experiment to control the intonation. When two syllables carry an identical tone, the pitch pattern clearly displays a declining contour from the beginning of the first tone to the end of the second tone. The absolute F_0 value of the tone height in the first syllable is higher than that in the second syllable. This may reveal that the universal declination operates in the disyllabic utterance and interacts with the two-tone sequences.

The three level tones display a gently slanting shape, as does T4. T6 declines more than T3, T1 the least. As a result of the continuing downward slope of the two identical tones when occurring consecutively, the first syllable is higher in pitch than the second syllable. This can be observed casually in figure 16. The mean of aggregate F_0 values at the first syllable and the second syllable of the two T1s is 78 vs 77 of the two T3s is 58 vs 42, of the two T6s is 41vs 33, of the two T4s is 22 vs 13 (at the percentage scale of the F_0). (The aggregate F_0 values are taken from the five point measurements). As for the rising tones, the mean of the aggregate F_0 values at the first syllable is also higher than the second syllable: 77 vs 47 for the two T2s and 36 vs 35 for the two T5s (at the percentage scale of the F_0). (The aggregate F_0 values of the rising tones are taken from the last four points of the measurements on the grounds that the assimilatory effect on slope heavily affects the first quarter of the tonal duration of the second syllable and this in turn raises the low F_0 onset of the rising tones.)

These results do not encourage a conclusion that the universal declination of intonation affects disyllabic words in reading lists more than in other environments, or that the universal terminal fall affects disyllabic words in reading lists more than in other environments. As experienced experimenters know, there is no perfect design for an experiment. Live conversation is the most real and natural speech but it is hard to control the material; with designed reading it is easier to control the material but it

is not natural everyday speech; reading disyllabic words in a carrier sentence is said to be more natural than in an isolated situation; nonsense words can be used but sound artificial. I am not denying that intonation and tonal coarticulation interact with each other, but in tonal coarticulation experiments it is best to minimise such interaction. The present study is aimed at examining the tonal behaviour in two-tone sequences. A reading list of designed disyllabic words avoiding surrounding tonal influence is expected to be ideal as long as it is not affected by any strong intonational interference.

3.2.4. Timing transition

3.2.4.1. Residual energy vs. energy run-out

For the illustration of the six tones when following T1 and T2 see figures 4 and 5.

It is reported that in Mandarin the overall tonal values of following tones are higher when after tone 2 than after tone 1 even though they both have high offsets, because the F_0 direction in tone 1 is downward, while that in tone 2 is upward. The overall tonal values are pushed higher by a preceding tone with a rising F_0 than by a preceding tone with a falling Fo, although the F_0 heights of the two tones are at the same tonal register (Shen 1990: 292). A similar phenomenon is reported in Vietnamese - that every individual tone when following a rising tone is higher by approximately 4-6 semitones than its corresponding tone in a monosyllabic utterance. This "seems to be due mostly to the higher onsets after a rising tone caused by the extremely high end of the preceding rising tone" (Han and Kim 1974: 230).

However, Cantonese seems to reveal a different picture of behaviour of the overall tonal pitch following a high offset tone. As discussed above, the onset of a Cantonese tone can be raised by a high offset tone or lowered by a low offset tone and lasts for about the first 25% of its duration before heading up or heading down to its own height and own shape. The two sets of data representing the pitch range of the main body of the contour of the six tones, i.e., when preceded by a high offset tone and by a low offset tone, are very close, because the elasticity of every individual tone is very small. This may be due to the fact that the Cantonese tones include three level tones. The three are very much alike in terms of shape and only differ in height, which does not leave much space for them to shift up and down dramatically.

Regarding the overall pitch height of the two sets of the following six tones, they are very similar when preceded by the high offset tones, T1 and T2. The only difference between these two sets is, one tends to stretch out a little for the first half of its duration when preceded by T1, whereas in the other one, tones tend to cling together for the first quarter of its duration when preceded by T2. The overall pitch values of the onset of the six tones when preceded by T2 is higher than the other set which are preceded by T1. However, the former set intersects with the latter set at the central portion of the tonal duration: the former set hangs down while the latter set stretches out. The slight stretching out in pitch range of the six tones when preceded by T1 may be due to the carry over of some residual strength from the preceding high level T1. The crowded clinging together in pitch range of the six tones when preceded by T2 may be due to no energy remaining from the preceding high rising tone T2 and this results in a `hanging down' shape. When the low rising T5 is the preceding tone compared with T3 which has a similar offset value, the hanging down shape of the following tones is not consistent among the informants. Only the high rising tone runs out its energy in the journey of steeply climbing up in order to maintain its identity and leaves almost nothing to the following tones. Moreover, it cannot resist being affected by the following tones: the tonal peak is pushed up to an earlier point before the tonal boundary.

On the other hand, a delayed peak can occur, for example, in a T1-T4 sequence. The high F_0 contour of T1 spreads over to the beginning portion of the following T4, where the F_0 reaches a peak after the offset of T1 and right in the middle of the initial consonant of T4 (see figure 18). Also, in a T4-T1 sequence, a creaky voice in a low F_0 T4 can spread to the initial consonant of the following high T1, or, in other words, a T1 carries over the laryngeal feature of the preceding T4 (see figure 19). When we talk about `spread over', we refer to a feature on the left segment or tone spreading onto the segment or tone on the right. The occurrence of a delayed peak may be due to the residual strength from the preceding T1, which is strong enough to spread over its high F_0 contour to the following low tone until its energy is finished - a delayed peak occurs.

Fig. 18. Delayed peak of T1. Extracted F_0 contour (Fx) of the disyllabic word /ji:^{1} la:n^{4}/ (lie on railing). The vertical line inside the figure indicates the tonal boundary. The highest F_0 value of T1 is 222Hz located at the offset. The highest F_0 value of the disyllable is 272Hz located after onset of the T4 (indicated by the arrow).

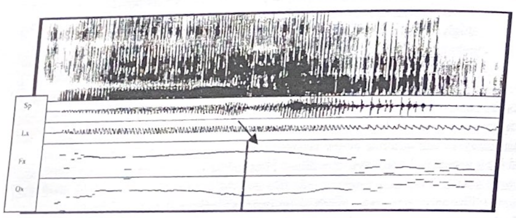
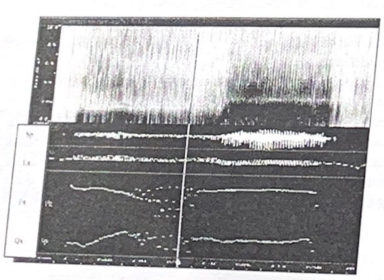


Fig. 19. The creaky feature spreading onto the following tone. Extracted F_0 contour (Fx) of the disyllabic word /ji:^{4} ma:^{1}/ (aunt), spoken by JCCF. The F_0 contour shows that the aperiodic cycles (indicated by the arrows) reflecting the mode of vibration of the vocal folds in the low falling T4 are across the syllable boundary and spread to the beginning of the following T1.



3.2.4.2. Earlier peak before a tonal boundary

The peak of a preceding high rising T2 is not necessarily at the far end of the offset phonetically, but can be located about 10-15% of the duration backward from the offset. We call this an earlier peak (see figures 20 and 21). This is different from when T2 is in canonical form where the highest F_0 is at the offset. In the case of the earlier peak, the F_0 of the actual peak is in between both of the last two measuring points (i.e., 75% and 100% of the duration). Similar cases can be found in low rising T5. The phenomenon of earlier tonal peak in T2 and T5 is particularly evident when they precede T2, T3, T4, T5 and T6. Also, in some cases of non-rising tones, a small peak can be created about 10-15% of the duration backward from the offset.

Fig. 20. Early peak of T2. Extracted F_0 contour (Fx) of /ji:n^{2} ji:^{6}/ (develop), spoken by JHDG. The dashed line indicates the location of the rising peak, the solid line indicates the offset of the T2. Both are labelled with arrows.

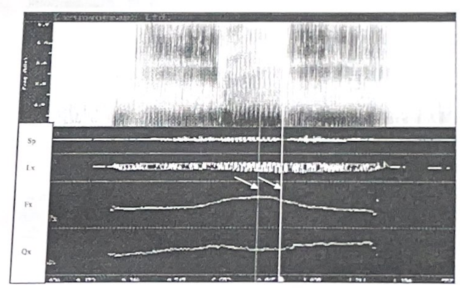
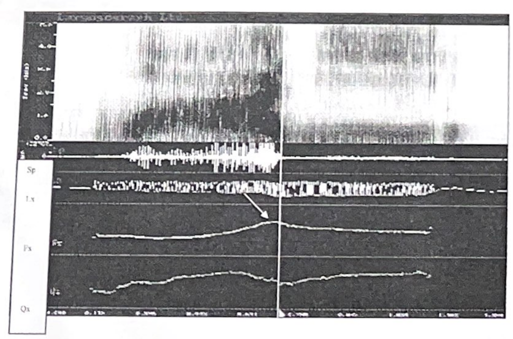


Fig. 21. Early peak of a T2. Extracted F_0 contour (Fx) containing the disyllabic word /wa:^{2} ji:^{3}/ (significance of a painting). The rising T2 reaches its peak (indicated by the arrow) 28ms before the syllable boundary (marked by the vertical line) and begins to fall at the syllable boundary.



In discussing the earlier peak placement in intonation, Bruce invokes a theory termed Gestural Overlap to explain some cases in Swedish and an alternative theory termed Tonal Repulsion to account for some cases in English. Bruce explains gestural overlap by saying that "an upcoming fall seems to overlap with the preceding rise giving an earlier peak location", whereas tonal repulsion: "implies a true temporal reorganisation of the accent gesture(s), which may result in anticipation of the first gesture and also delay of the second gesture." (Bruce 1990: 112) In the absence of understanding how the laryngeal gestures interact during tonal coarticulation, the present study does not attempt to test whether the earlier timing of a tonal peak happens because of gestural overlap, or because of tonal repulsion. But if we look at cases in non-rising tones where a peak is created, I doubt that an overlap model can explain how a small peak can be created in a fairly slanting slope.

The phenomenon of earlier tonal peak is also reported for Thai by Potisuk *et al* (1997). They see the earlier peak as a correlate of the dissimilatory nature of anticipatory effects. "[T]he lower the onset of the following tone, the earlier the point at which the peak of the rising F_0 contour, and conversely, the higher the onset of the following tone, the later the peak of the rising F_0 contour will be reached in the preceding tone." "[T]he peak F_0 of tones with rising F_0 trajectives is realised at an earlier point within the syllable in order to maintain tonal identity as well as to allow for a less abrupt change in acceleration that is required in making transition to a following tone with a low onset." (Potisuk *et al*. 1997: 34) This analysis seems to conflict with other findings on the dissimilatory nature of anticipatory effects on height in Thai tone, reported in the same paper, when it says that only high and rising tones are affected. If a preceding rising tone adjusted itself "for a less abrupt change in acceleration", it would not be allowed to be raised "greater in height when followed by a low or rising tone". Instead, it would be lowered in height in order to meet up with the low onset tones.

The earlier peak in T2 and T5 or the created peak in non-rising tones in the preceding syllable, seems to be explained by the tonal repulsion model: an oncoming fall follows too closely, so pushes the tonal peak to an earlier point or pushes up the level to a crease - a created peak. As discussed above, the carryover effect is greater than the anticipatory effect in Cantonese. The strong counteracting force of the oncoming fall from the low onset of the following tones not only pushes up the height of the entire preceding tones, but also presses up the tonal peak to an earlier placement.

3.2.5. Summary

In conclusion, the findings of this study support the view that tonal influence in two-tone sequences is bidirectional. Anticipatory and carryover effects on slope are both assimilatory in nature but on height are dissimilatory in nature. Both anticipatory and carryover effects on height encompass the whole of the duration of the target tones. An anticipatory effect on slope is minimal and extends for only the last quarter of the target tone's duration, whereas a carryover effect on slope extends for not more than the first half of the target tone's duration. The carryover effect is greater than the anticipatory effect. The contrastive pattern of the six individual tones remains despite the extensive phonetic variation in height and slope.

Appendix: list of disyllabic utterances

1. /ji:^{1} ji:^{1}/ aunt-aunt - Auntie
2. /ɐw^1 ji:^{1}/ Aw-aunt - Auntie Aw
3. /ji:^{1} jy:n^2/ medical-building - hospital
4. /ma:w^1 ji:^{2}/ cat-chair - cat-chair
5. /ji:^{1} wɐj^3/ lean-lean - lean on
6. /wa:j^1 ji:^{3}/ bad-idea - bad thought
7. /ji:^{1} li:w^4/ medical-treatment - medical treatment
8. /wɐj^1 ji:^{4}/ mighty-appearance - imposing appearance
9. /ji:^{1} leŋ^5/ cloth-collar - cloth collar
10. /mɐn^1 ji:^{5}/ mosquito-ear - mosquito ear
11. /ji:^{1} mi:n^6/ E-noodle – a kind of noodle
12. /ma:^{1} ji:^{6}/ twin-two - two pairs
13. /ji:^{2} ji:^{1}/ Ji-aunt - Auntie Ji
14. /jɐw^2 ji:^{1}/ pomelo-skin - pomelo skin
15. /ji:^{2} ŋɐj^2/ chair-short - short chair
16. /ŋɐj^2 ji:^{2}/ short-chair - short chair
17. /ji:^{2} jɐn^3/ chair-mark - chairmark (on floor)
18. /wa:^{2} ji:^{3}/ painting-meaning - meaning of a painting
19. /ji:^{2} la:n^4/ lean-railing - lean on railing
20. /wɐj^2 ji:^{4}/ zig-zag - zigzag
21. /ji:^{2} low^5/ rely on-old - old age arrogance
22. /ji:m^2 ji:^{5}/ cover-ear - cover ear
23. /ji:^{2} la:j^6/ rely-rely - rely on
24. /ji:n^2 ji:^{6}/ perform-righteousness - develop (human relationship)
25. /ji:^{3} ji:^{1}/ Ji-aunt - Auntie Ji
26. /a:^{3} ji:^{1}/(prefix)-aunt - Aunt
27. /ji:^{3} jeŋ^2/ idea-reflect - reflection (thought)
28. /a:^{3} ji:^{2}/(prefix)-two - servant
29. /ji:^{3} ow^3/ Italy-Austria
30. /ji:m^3 ji:^{3}/ bore-idea - bored-feeling
31. /ji:^{3} mɐn^4/ Italy-language - Italian language
32. /jɐw^3 ji:^{4}/ young-child - little child
33. /ji:^{3} mɐj^5/ barley-rise - barley
34. /ji:n^3 ji:^{5}/ swallow-look - swallow dance
35. /ji:^{3} mej^6/ idea-taste - imply
36. /ji:w^3 ji:^{6}/ need-easy - take things easy
37. /ji:^{4} ma:^{1}/ aunt-mother - maternal aunt
38. /meŋ^4 ji:^{1}/ famous-medical - famous doctor
39. /ji:^{4} ji:n^2/ suit-perform - suitable to perform
40. /lɵn^4 ji:^{2}/ wheel-chair - wheel chair
41. /ji:^{4} ji:n^3/ suit-feast - suitable to feast
42. /mɐn^4 ji:^{3}/ essay-idea - context meaning (of written pieces)
43. /ji:^{4} na:m^4/ query-difficult - difficulty
44. /jɐn^4 ji:^{4}/ person-(suffix) - little person
45. /ji:^{4} nɵy^5/ son-daughter - son and daughter
46. /mow^4 ji:^{5}/ model-mock - mock
47. /ji:^{4} mɐn^6/ query-ask - doubt
48. /lej^4 ji:^{6}/ separate-different - separate
49. /ji:^{5} loŋ^1/ ear-hole – ear-hole
50. /jy:^{5} ji:^{1}/ rain-coat - raincoat
51. /ji:^{5} wu:j^2/ discuss-meeting - assembly
52. /low^5 ji:^{2}/ old-two - number two (of brothers)

53. /ji:^{5} ɔ:n^3/ discuss-case - agenda
54. /jɛw^5 ji:^{3}/ have-idea - to desire
55. /ji:^{5} jy:n^4/ discuss-person – member of parliament
56. /nɵy^5 ji:^{4}/ daughter-(suffix) - daughter
57. /ji:^{5} jy:^{5}/ ear-speak - whisper (in ear)
58. /jɛw^5 ji:^{5}/ have-ear - listen
59. /ji:^{5} nɔ:j^6/ from-inside - within
60. /joŋ^5 ji:^{6}/ brave-righteousness - righteousness
61. /ji:^{6} na:j^1/ two-grand - mistress
62. /ji:^{6} ji:^{1}/ easy-medical - easily curable
63. /ji:^{6} wu:^{2}/ two-Wu - two-stringed musical instrument
64. /la:n^6 ji:^{2}/ broken-chair - broken chair
65. /ji:^{6} ji:m^3/ easy-bore - easily bored
66. /wu:j^6 ji:^{3}/ meet-idea - take the hint
67. /ji:^{6} ji:n^4/ easy-on fire - inflammable
68. /nɔ:j^6 ji:^{4}/ inside-move - move in
69. /ji:^{6} joŋ^5/ righteousness-brave - upright and brave
70. /wu:j^6 ji:^{5}/ meeting-discuss - conference
71. /ji:^{6} mej^6/ different-smell - strange smell
72. /ma:n^6 ji:^{6}/ ten thousand-two - twelve thousands

References

Browman, C & L Goldstein (1986) 'Gest: Towards an Articulatory Phonology'. *Phonological Yearbook* 3: 219-252.

Bruce, Gosta (1977) *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Chao, Yuanren (1968) *A Grammar of Spoken Chinese*. (1985 3rd ed.) University of California Press.

Ching, Chingyuk T (1981) *Communication of Lexical Tone Patterns in Cantonese*. Unpublished Ph.D Thesis, University College London.

Crystal, D (1985) *A Dictionary of Linguistics and Phonetics* (2nd ed.). Oxford: Basil Blackwell.

Davies P, Lindsey G A, Fuller H & A F Fourcin (1986) 'Variation in Glottal Open and Closed Phase for Speakers of English'. *Proceedings of the Institute of Acoustics* 8: 539-546.

Fourcin A J & Erm Abberton (1971) 'First Applications of a New Laryngograph'. *Medical and Biological Illustrated* 21: 172-182.

Fok, C Y-Y (1974) *A Perceptual Study of Tones in Cantonese*. University of Hong Kong.

Gandour, J, S Potisuk *et al* (1992) 'Anticipatory Tonal Coarticulation in Thai Noun Compounds'. *Linguistics of the Tibeto-Burman Area* 15: 111-124.

Gandour, J, S Potisuk *et al* (1994) 'Tonal Coarticulation in Thai'. *Journal of Phonetics* 22: 447-492.

Han, S & K-O Kim (1974) 'Phonetic Variation of Vietnamese Tones in Disyllabic Utterances'. *Journal of Phonetics* 2: 223-232.

Hashimoto, O-K Y (1972) *Phonology of Cantonese*. Cambridge: Cambridge University Press.

Hombert, Jean-Marie (1978) 'Consonant Types, Vowel Quality and Tone'. In *Tone: A Linguistic Survey*: 77-111. Ed. by V Fromkin, Academic Press.

Howard, David M (1995) 'Variation of Electrolaryngographically Derived Closed

Quotient for Trained and Untrained Adult Female Singers'. *Journal of Voice*: 9 (2) 163-172.

Keating, Patricia A (1990) 'The Window Model of Coarticulation: Articulatory Evidence'. *Papers in Laboratory Phonology 1:*. 451-470. Ed. by Kingston J and M Beckman, Cambridge University Press.

Ladefoged, Peter (1975) & (1982) *A Course in Phonetics* (1st and 2nd ed.). New York: Harcourt Brace Jovanovich.

Lehiste, I & G Petrson (1961) 'Some Basic Considerations in Analysis of Intonation'. *Journal of the Acoustical Society of America* 33: 419-425.

Potisuk, S, J Gandour & M Harper (1997) 'Contextual Variations in Trisyllabic Sequences of Thai Tones'. *Phonetica* 54: 22-42.

Rao, Bingcai & Wuji Zhao (1982) *Guangzhouhua Fangyan Cidian (Cantonese Dialect Dictionary)*. Hong Kong: Commercial Press.

Shen, Xioanan S (1990) 'Tonal Coarticulation in Mandarin'. *Journal of Phonetics* 18: 281-295.

Xu, Yi (1994) 'Production and Perception of Coarticulated Tones'. *Journal of Acoustical Society America*: 2240-2253.

SOAS Working Papers in Linguistics Vol. 11 (2001): 31-44

Pitch-phonation correlations in Sgaw Karen*

Justin Watkins
justin.watkins@soas.ac.uk

1. Introduction

This paper continues a line of research was initiated by the late Katrina Hayward, the first and principal author in 1998 of Hayward, *et al.* (to appear), which examined voice quality-pitch associations in Yorùbá. In that paper, we found evidence to suggest that voice quality, of which phonation type was one component, contributed to the phonetic character of the low tone in Yorùbá. Part of our conclusion was that, at least for Yorùbá, voice quality may help to distinguish between two tones in one register (low and mid) rather than emphasising the difference between registers. This finding was juxtaposed with the possibility that characteristic voice quality of the Yorùbá low tone was merely an artefact of its position near the bottom of the pitch range.

The main conclusion, however, was that much more research needs to be done on the relationship between voice quality and fundamental frequency, since the associations are not universal, despite the fact that distinctive voice qualities are reportedly associated with specific tones in many tone languages. The present paper is a step towards this end, and adopts a similar experimental approach.

1.1. Tone and syllable type

(Jones 1986) takes the view that the term 'tonal' has been 'rendered virtually meaningless' through being applied to virtually any language in which pitch is in some way lexically or morphologically distinctive. This suggestion was made within the context of South East Asian languages, but becomes doubly resonant in a broader forum including also East Asian and African languages, where one may encounters a bewildering diversity of contrastive pitch systems.

One may define a 'conventional' tone language as one in which any given syllable in any given context may be identified from its pitch or pitch contour as bearing one of a closed set of tones. Vietnamese may be such a language: it is cited by (Đờ Thế Dũng *et al.* 1998:398) as an example of a tone language in which tone *sandhi* phenomena are not observed (though coarticulatory effects may be observed in strings of adjacent tones). Properties typically encountered in the tone systems of African languages which elaborate on this basic type are so-called downstep and tone terracing (see Laver 1994:470 ff.), while those familiar with Asian languages will recognise tone sandhi as a common feature of tone systems in that region (Chen 2000). In such languages, the correct identity of a tone may be recovered only by interpreting the pitch of the tone through the filter of the sandhi or other phonological rules of the tone system.

---

* The author acknowledges the support of grants from the British Academy and the School of Oriental and African Studies which supported attendance at the *VIth International Conference on the Languages of the Far East, South East Asia and West Africa*, held during September 2001 at the University of St Petersburg, where an earlier version of this paper was presented. Thanks are due to Bernard Howard for assistance with recording and to Paul Boersma for help with writing scripts for the Praat phonetics program.

Special thanks are to Katrina Hayward, whose crystal clear thinking forms part of the basis for this work. I dedicate this paper to her memory.

There are also languages in which pitch is one component within a complex of phonetic features which make up contrasts which otherwise appear to pattern in the same way as tone. (Jones 1986):135 calls these systems "pitch register" systems, in which pitch is "minimally contrastive only between syllables of like type". The rationale for this is that in such languages the pitch element of the lexical pitch contrasts is so variable that it only makes sense to refer to pitch when comparing two syllables in actual juxtaposition if they have otherwise similar syllable structures.

2. Phonation type, voice quality and pitch in Sgaw Karen

Sgaw Karen is a Tibeto-Burman language spoken by about two million people, concentrated mainly in the Karen State of Burma and across the border into western Thailand, but also scattered across the predominantly Burmese-speaking delta region of Burma. In Karen orthography, there are six possible tonal contrasts, shown in Table 1.

Table 1 Karen names of Karen tones.

usual alphabetical order of tone	Karen orthography	name of tone in Karen	romanisation
1	–	အါ	a^{1}
2	–ၢ်	အၢသံ	er^{1}thi^{1}
3	–ၣ်	အၣ်သံ	a^{3}thi^{1}
4	–း	ဖျၣ်ဆံ	hpler^{5}hsi^{1}
5	–ၥ်	ဟးသံ	ha^{4}thi^{1}
6	–ၤ	ကၠၣ်ဖိ	ke^{5}hpo^{1}

Table 2 sets out various phonetic descriptions of Karen tones in some of the published sources. It is immediately apparent that there are serious contradictions, especially with respect to the descriptions of pitch characteristics.

(Haudricourt 1975) surveys the mechanisms by which various tone systems have developed in a range of East and South East Asian languages, including a brief mention of Sgaw Karen, using secondary data taken from the Linguistic Survey of India (Grierson 1903). This source suggests that Sgaw Karen is "an example of a language whose tonal system has passed from two phonemic tones to four." Haudricourt (1942) showed that the six tones of Sgaw (and Pwo) Karen resulted from the doubling of a three-tone system, but that one of these three tones was derived from syllables with final stops, so that only two proto-tones needed to be reconstructed. Haudricourt (1975:339) revises this position in light of the work of Luce (1957); Jones (1961) and Burling (1969) to posit the reconstruction of three Karen tones, rather than two.

Table 2 Summary of existing descriptions of Karen tones.

tone	Karen script	Description in Gilmore (1898)	Descriptions cited in Taylor (1927:95)	Transcription and description (Chao tone letters) from Jones (1961:64)	
1	–	with a rising inflection	high, level and long	´	55
2	–ၢ်	heavy falling inflection	begins at middle register and falls considerably; about medium length	`	21
3	–ၣ်	abruptly, at a low pitch	very slightly above middle register, level and short; sometimes abrupt; pitch lower than 1and 4 but same as 6	`ʔ	11ʔ
4	–း	abruptly, at an ordinary pitch	high, level and abrupt, ... has a glottal check	ʔ	33ʔ
5	–ၥ်	with a falling circumflex inflection	begins high, falls steadily to middle register, long	´ʔ	53ʔ
6	–ၤ	with a prolonged even tone	same pitch as 3, slightly high, level, and long		33

Jones (1961) is a descriptive and comparative monograph focussing on six dialects of Karen, including two of Sgaw Karen. The dialect which corresponds most closely to the speech of the language consultant used for the present paper is referred to by Jones as Moulmein Sgaw. In Jones's (1961) analysis, Moulmein Sgaw has three tones, each of which occurs in two allophones, with and without final [ʔ]; he revised this analysis in the light of his proposal of 'pitch register' languages to one with two 'pitch registers in each of which plain, breathy and stopped syllables occur' (Jones 1986:137). We shall see that this analysis is motivated more by phonological design than by phonetic reality.

2.1. Experimental procedure

Whether or not one adopts Jones's term, 'pitch register' systems are difficult to study and describe: if they are approached as conventional tone languages, the tones suddenly appear elusive as the pitches shift and bend around, influenced by phonological, morphological, syntactic or intonational contexts. The analyst can choose to try to account for and describe pitch phenomena for as many contexts as possible, in which case problems of data reduction dictate that detail in one will be at the expense of comprehensiveness in the other. Alternatively, one can opt to control the context as much as possible, allowing for more detailed description of the pitch phenomena, but running the risk that the frozen snapshot image of the pitch system does much resemble its real-life counterpart. One encouraging improvement in these bleak circumstances is the rapid expansion of the capacity to process large quantities of data quickly by computer.

Here, I have chosen to collect detailed phonetic data for Karen tonal contrasts in controlled-context suspended animation, as I have done elsewhere for Burmese (Watkins 2000).

The design of this experiment was constrained by the availability of a single Karen-speaking consultant. The word list consisted of three quasi-randomised sets of nonsense CV syllables, each containing three repetitions an /a/ vowel in each of the six orthographic Karen tones. Two of the sets of syllables began with an initial /m/

and the third began with an initial /k/. This yielded a total of 54 tokens. The syllables were placed in the frame sentence:

နတဲ ___ ဒ်အံၤ
/nətɛ¹ ___ di⁴ʔi⁶/
"You say ___ like this."

The consultant recorded for this study was a female speaker of Sgaw Karen aged 29, who did not report abnormal speech or hearing. She was brought up in Kyaukkyi (ကော်မူးရား) and Htihta (ကျောက်ကြီး) and spent her teenage years in Kawmuya (ထီးထာ). Though bi-dialectal in Karen, for the purposes of this study she read in a variety of Karen widely recognised as the standard, though the possibility of interference from her other dialect cannot be ruled out, as we shall see later. The materials were presented to the consultant in written form.

Simultaneous digital audio and laryngograph recordings were made on minidisk in a sound-proofed booth in the recording studio of the School of Oriental and African Studies using the following: a Bruel and Kjær condenser microphone (type 4165); a Bruel and Kjær measuring amplifier (type 2609); a Fourcin portable laryngograph and processor (Laryngograph Ltd.). (For general information about the laryngograph and its use, see Abberton *et al.* (1989) and Howard *et al.* (1990). The laryngograph and audio recordings were converted to computer sound files using the Goldwave digital audio editor, version 4 (*www.goldwave.com*), with the sampling rate set at 11025 Hz. The remaining analysis was carried using the Praat program, version 3.9.28 (*www.praat.org*).

2.2. Measurements

The vowels in the syllables of interest were segmented by hand using speech pressure waveforms and spectrograms, at which point the presence or absence of final glottal stops could be determined. Computer scripts were written ad hoc to direct the Praat program shell to make several series of measurements, using where possible the software's built-in capacity to do so automatically.

Fundamental frequency and closed quotient were derived from the laryngograph trace. Closed quotient (CQ) is defined as the ratio of the duration of the closed phase to the duration of one whole period of vocal fold vibration. In the algorithm in the script which was written to calculate this from the laryngograph waveform, the point of vocal fold closure – the beginning of the closed phase – was defined as the positive peak of the differentiated waveform, while the end of the closed phase was a threshold fixed at 30% of the total peak-to-peak amplitude of the waveform period in question. Other things being equal, higher CQ is diagnostic of relatively creakier phonation while lower CQ is diagnostic of relatively breathier phonation. In measurements of syllables in Wa (Watkins 1999), CQ contours were found to be a very clear indicator of glottal consonants: rising CQ is associated with glottal stops /ʔ/ while falling CQ is indicative of breathy-voiced /ɦ/.

The duration of each vowel was measured from the speech pressure waveform and derived spectrograms. If a final glottal stop was present, then the duration included the glottal stop.

The frequencies of the first three formants were measured at the mid-point of each vowel. There are many languages in which some kind of register or phonation type contrast interacts with vowel quality in some way, either in the form of a synchronic phonological contrast or diachronically, by splitting or doubling vowel systems (see Diffloth 1980:37 on Wa and Khmer).

The measurements of intensity were calculated automatically by the Praat program. Spectral tilt was measured using a variant of the 'resonance balance' measure described by Schutte and Miller (1985). The Praat program was instructed to resample the sound at 10 kHz and calculate the difference in energy present in two frequency bands (0-1 kHz and 3.5 – 4.5 kHz) of narrow-band spectra (40 Hz effective bandwidth). Higher values of this measure, indicating a greater difference in energy in the two frequency bands, indicate steeper spectral tilt and relatively more energy concentrated in the low frequency band, a characteristic of breathy voice, while lower values of this measure may indicate less spectral tilt, with relatively more energy concentrated in the 3.5 – 4.5 kHz frequency band. While measuring spectral tilt can yield an index of relative creakiness or breathiness of phonation type (Ní Chasaide and Gobl 1997:439), spectral tilt may also be reduced by other factors such as vocal effort and loudness, and measuring phonation type in this way is not straightforward (Watkins 1997).

Naturally, the configuration of the supraglottal vocal tract also influences the spectral profile, but since all the speech material for this investigation was taken from similar vowels, this unwanted effect is reduced to a minimum.

F0, CQ, intensity and spectral tilt were all measured at three points evenly spaced through the vowel, after one quarter, one half and three-quarters of the vowel's measured duration.

To sum up, the measurements taken from each vowel were:

- presence or absence of final glottal stop determined
- fundamental frequency (3 measurements)
- closed quotient (3 measurements)
- duration
- frequencies of F1, F2 and F3 (at midpoint of vowel only)
- intensity (3 measurements)
- spectral tilt (3 measurements)

Statistical analysis was carried out using the SPSS Statistics package (version 10.0). Means and standard deviations for all the measures were computed for each tone. A two-way ANOVA was carried out for each measure, with tone and syllable set as the independent variables. A one way ANOVA, with Scheffé post-hoc comparisons was also carried out for each measure, with tone as the single factor. In the presentation and discussion of statistical data, the significance or non-significance of a result is determined by setting the critical threshold at $p=0.05$.

2.3. Results

With a broad set of measurements such as there, the results are complex. The results of each measure are discussed individually below, after which the overall findings are discussed.

2.3.1. Summary of numerical data

Table 3 Means (n=9) and standard deviations of all measurements made of each Karen tone.

		Tone 1		Tone 2		Tone 3		Tone 4		Tone 5		Tone 6		Total	
		mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
f0 (Hz)	onset	241.046	6.48	195.27	9.09	186.54	7.80	239.83	12.24	193.91	10.62	217.85	3.46	212.41	23.73
	mid	240.5459	6.04	180.50	6.29	175.51	6.38	221.61	16.19	180.20	9.35	213.89	4.43	202.04	26.33
	end	229.8404	5.52	161.16	26.58	163.07	33.89	198.19	43.71	150.94	47.89	205.33	5.24	184.76	41.71
CQ (%)	onset	40.7125	6.55	46.49	7.43	55.93	6.66	41.80	5.23	46.07	6.95	35.17	10.84	44.04	9.34
	mid	39.39829	6.24	48.70	7.22	54.67	5.34	46.71	3.44	50.52	7.11	35.80	10.74	46.11	9.24
	end	40.95503	6.17	49.97	10.35	49.67	6.11	58.05	5.24	52.57	5.35	39.73	6.97	47.84	9.28
duration (s)		0.340	0.027	0.301	0.050	0.324	0.063	0.171	0.015	0.272	0.050	0.332	0.048	0.290	0.072
F1 (Hz)		990.32	99.23	949.36	88.89	944.56	101.87	1164.51	91.41	986.33	88.26	861.16	71.84	982.70	126.61
F2 (Hz)		1407.53	135.69	1547.82	106.41	1497.88	67.21	1686.18	104.17	1539.61	153.93	1562.24	71.13	1540.21	134.73
F3 (Hz)		2488.91	429.73	2371.19	445.15	2286.26	463.46	2618.12	324.80	2427.75	511.55	2877.53	255.73	2511.63	439.79
intensity (rel dB)	onset	80.78	1.84	79.82	2.38	79.79	3.53	84.66	2.17	80.25	2.73	78.81	2.96	80.68	3.16
	mid	80.46	2.05	74.60	2.87	77.61	2.76	84.69	1.41	77.46	2.55	79.22	3.13	79.01	3.96
	end	79.41	2.30	69.91	5.33	75.64	3.10	81.17	1.23	74.28	5.47	77.95	3.26	76.39	5.17
spectral tilt (rel dB)	onset	-26.10	4.43	-28.33	5.57	-27.83	4.07	-21.65	5.75	-27.16	4.02	-25.64	3.70	-26.12	4.96
	mid	-27.24	4.00	-26.63	4.14	-28.07	4.20	-17.95	6.67	-23.86	3.31	-26.42	3.52	-25.03	5.46
	end	-26.63	2.71	-24.46	4.44	-27.27	4.49	-16.56	4.36	-20.49	3.10	-29.43	4.26	-24.14	5.81

2.3.2. Statistical analysis of data

ANOVA test results

Table 4 Results of ANOVA tests performed with SPSS software on each of the measured correlates of tone with respect to the independent variables: 6 tones, 3 syllable sets. The tests were performed on 54 tokens. Significant results ($p < .05$) are marked with an asterisk. Values of p less than 0.0005 are recorded as 'near 0'.

		Main effects				2-way	
		TONE (d.f. = 5)		SYLLABLE SET (d.f. = 2)		TONE X SYLL. SET (d.f. = 9)	
		F	p	F	p	F	p
fundamental frequency	onset	93.79	near 0 *	8.71	0.001 *	2.75	0.021 *
	mid	140.33	near 0 *	13.14	near 0 *	4.34	0.002 *
	end	4.56	0.004 *	0.37	0.696	0.34	0.954
closed quotient	onset	11.09	near 0 *	3.12	0.061	3.50	0.006 *
	mid	11.56	near 0 *	0.32	0.731	3.62	0.005 *
	end	12.69	near 0 *	3.06	0.064	2.82	0.018
duration		12.49	near 0 *	4.93	0.015 *	0.68	0.718
F1		10.64	near 0 *	9.63	0.001 *	0.23	0.987
F2		7.84	near 0 *	17.00	near 0 *	0.49	0.865
F3		2.94	0.031 *	10.67	near 0 *	1.63	0.158
intensity	onset	3.81	0.010 *	3.97	0.031 *	0.96	0.493
	mid	11.51	near 0 *	1.77	0.190	1.21	0.330
	end	7.93	near 0 *	1.53	0.236	0.78	0.636
spectral tilt	onset	4.97	0.003 *	3.68	0.039 *	1.60	0.169
	mid	5.51	0.001 *	2.40	0.110	0.55	0.825
	end	8.31	near 0 *	1.25	0.303	0.46	0.890

Results of Scheffé post-hoc comparison

Table 5 Results of post-hoc Scheffé comparisons applied to each measure, calculated from one-way ANOVA tests carried out on each with tone as the single factor. As with the other statistical tests, the critical value of p is taken as 0.05.

dependent variable		homogeneous subsets by tone
fundamental frequency (Hz)	onset	(1, 4) > (6) > (2,5,3)
	mid	(1) > (4,6) > (2,5,3)
	end	(1,6,4) > (6,4,3,2) > (4,3,2,5)
closed quotient (%)	onset	(3,2,5) > (2,5,4,1,6)
	mid	(3,5,2,4) > (5,2,4,1) > (4,1,6)
	end	(4,5,2,3) > (5,2,3,1,6)
duration (s)		(1,6,3,2,5) > (4)
F1 (Hz)		(4) > (1,5,2,3,6)
F2 (Hz)		(4,6,2,5) > (6,2,5,3,1)
F3 (Hz)		none
intensity (rel dB)	onset	(4,1) > (1,5,2,3,6)
	mid	(4) > (1,6,3,5) > (3,5,2)
	end	(4,1,6,3) >(1,6,3,5) > (3,5,2)
spectral tilt (rel dB)	onset	none
	mid	(3,1,2,6,5) > (5,4)
	end	(6,3,1,2) > (1,2,5) > (5,4)

2.3.3. Final glottal stops

Glottal stops could be consistently be identified in the spectrograms of Tones 3 and 4, and were consistently absent from Tones 1 and 6, as illustrated in Figure 1 and Figure 2.

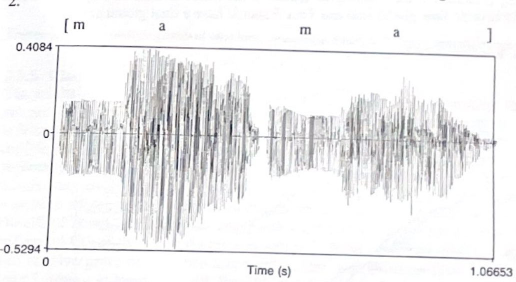


Figure 1 Waveform of syllables in Tones 1 and 6 with no final glottal stops: [ma].

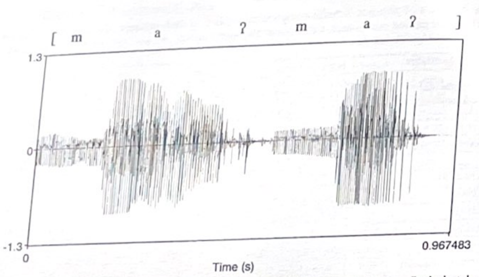


Figure 2 Waveforms of syllables in Tones 3 and 4, with irregular pulses indicating final glottal stops clearly visible: [maʔ].

In the case of Tones 2 and 5, there were mixed findings: four out of nine Tone 2 tokens and seven out of nine Tone 5 tokens had readily identifiable final /ʔ/. One interpretation of this, confirmed by the language consultant, is that the two tones may have been confused, and are interchangeable in her speech as a consequence of her dialect or mix of dialects. For the purposes of this paper, it is assumed that the reading produced in the majority of tokens is the intended one, namely that tone 2 should have no final glottal stop and Tone 5 should have a final glottal stop.

2.3.4. Fundamental frequency

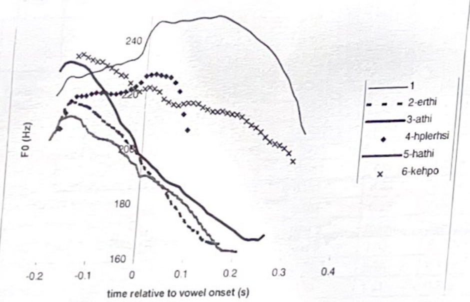


Figure 3 Fundamental frequency traces of a typical single syllable /ma/ in each of the six tones, aligned with the release of the nasal consonant at time = 0.

The F0 measurements reveal straight away one of the most interesting results of this study: that there is very little to tell tones 2,3 and 5 apart. In the restricted context of this experiment, all three have a mid to low falling contour, while Tone 1 is high, Tone 4 has a high to mid falling contour and Tone 6 is mid. The 'real time' F0 traces in Figure 3 may be compared with an illustration of the mean F0 measurements in Figure 4.

The significant effect of syllable set on the first quarter and mid-point measurements of F0 may be attributed to the predictable and often attested phenomenon of relatively higher frequency vowel onset after a voiceless stop /k/ consonant compared to a voiceless sonorant /m̥/ (see Hombert *et al.* 1979).

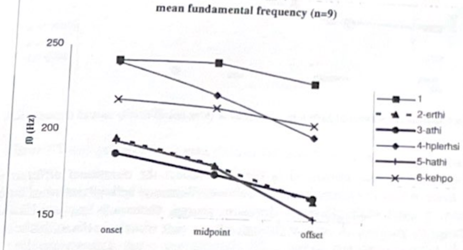


Figure 4 Mean F0 (n=9) in vowels of each tone, measured at three points evenly spaced through each.

2.3.5. Closed quotient

The results of the closed quotient measurements are particularly interesting, though not particularly clear, although tone is found to have a highly significant effect. Also at issue here is the presence or absence of final glottal stops. One can say with some confidence that Tones 1 and 6 are consistently relatively breathy, and that Tone 4 becomes progressively creakier as its final glottal stop approaches. Tone 3 appears to be relatively creaky, but CQ falls towards the final quarter. This unexpected finding is perhaps related to the preservation of a marked, creaky voice quality at a low F0. The CQ of Tones 2 and 5 is mid-range, with some divergence towards the end. If the mid-vowel CQ of these low F0 tones is said to indicate unmarked modal phonation, then the divergence of these two tones towards the final quarter – Tone 2 falling and Tone 5 rising – is consistent with Tone 5's mostly present final glottal stop and Tone 2's predominant lack of one.

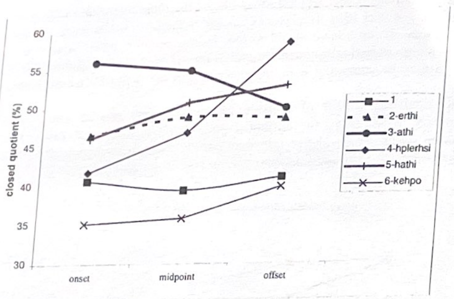


Figure 5 Mean CQ (n=9) in vowels of each tone, measured at three points evenly spaced through each.

2.3.6. Duration

The duration measurements illustrated in Figure 6 reflect the durational difference between the tones which is evident in Figure 3above. Tone and syllable set were both found to have a significant effect on duration, though there was no significant interaction between the two. A look at the data shows that vowels in the syllable set with initial /k/ were consistently shorter. The Scheffé test result shows very clearly that Tone 4 is shorter than the other five tones.

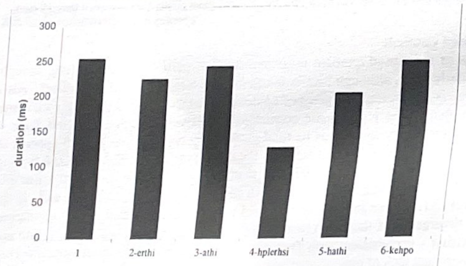


Figure 6 Mean duration (n=9) in vowels of each tone.

2.3.7. Vowel quality

The ANOVA test results show that both tone and syllable set have highly significant effect on all three formants. The syllable set effect can be accounted for by predictable formant transitions: after a bilabial initial we expect all three formants to start low, while after a velar initial we expect lower F1 and F2, and higher F2. The

Scheffé comparisons suggest that the vowel quality of Tone 4 clearly distinguishes it from the other tones, and that it is the F1 rather than the F2 difference which sets Tone 4 apart, and so we may conclude tentatively that the /a/ vowel in this tone in more open.

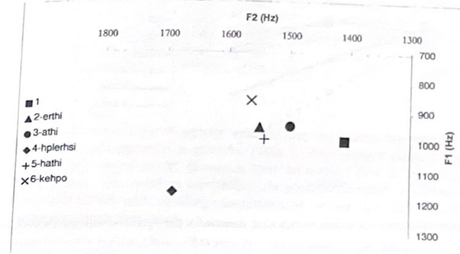


Figure 7 Vowel quality of Karen tones illustrated as mean F1 plotted against mean F2 (n=9).

2.3.8. Intensity and spectral tilt

The results of these two measures are inspected together, since it is expected that intensity and spectral tilt may be correlated, although no strongly significant statistical correlation is observed with these data. The ANOVA test found that syllable set has a significant effect on the first measurement of both intensity and spectral tilt, which can be accounted for by either by a positing that the phonation type at vowel onset can be expected to be relatively creakier after voiceless stops than after nasals, or that the sets of syllables were read at different speeds (syllable set had a significant effect on duration also).

Tone is a significant effect at all three points at which the measurements were made. The plots in Figure 8 and Figure 9 show a very clear result for Tone 4, which has greatest intensity and least spectral tilt, observations supported by the Scheffé comparisons in Table 5.

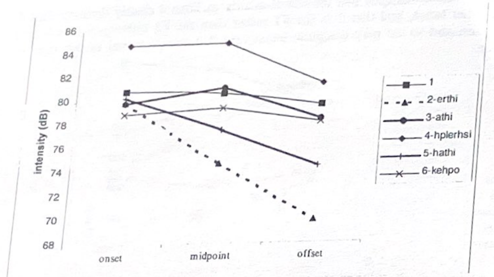


Figure 8 Mean (n=9) intensity of vowels in each tone, measured at three points evenly spaced through each.

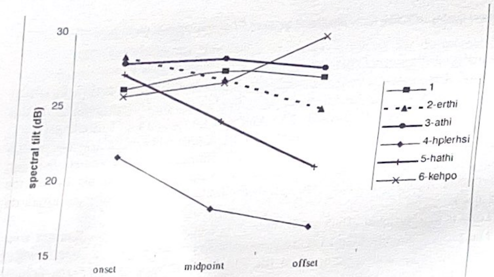


Figure 9 Mean (n=9) spectral tilt in vowels in each tone, measured at three points evenly spaced through each.

One unexpected finding here is that Tone 3 has been found to have creaky phonation but high spectral tilt; for the other tones, we find that spectral tilt correlates with phonation type in the predicted way.

2.3.9. Summary

From this set of measurements, we can propose a set of salient phonetic correlates, as set out in Table 6. Tones 2 and 5 remain problematic. The post-hoc Scheffé comparisons place them in homogeneous subsets for all the numerical variables and the status of the final glottal stop found in both of these tones is questionable.

Table 6 Sketch of proposed salient phonetic correlates of Karen tones in the context of this experiment.

tone	final /ʔ/	pitch	phonation	quantity	vowel quality	intensity	spectral tilt
1	no	High	breathy	long		mid	steep
2	?no	mid falling	modal	long		low	?
3	yes	mid falling	creaky	long		mid/low	steep
4	yes	high falling	creaky	short	more open	high	flat
5	?yes	mid falling	modal	long		low	?
6	no	mid level	breathy	long		mid	steep

3. Conclusions

How would Jones's 'pitch register' analysis play out against the phonetic data? Jones's own arrangement is given in Table 7. This array matches the phonetic data closely with regard to the phonation type measures. this arrangement tallies with Jones's own The problem surrounding the possible merging or confusing of Tones 2 and 5 will be left aside pending examination of further data, and we will concentrate on the remaining four tones.

Table 7 A 'pitch register' (Jones 1986) analysis of the six tones of Sgaw Karen.

syllable type	"high" register	"low" register
plain	5	2
breathy	1	6
stopped	4	3

How does this situation compare with other languages for which comparable data are available? In Yorùbá, voice quality (including phonation type) plays a rather minor supporting role in a predominantly pitch-led system of suprasegmental contrasts. In Wa, a language with a phonologically simple (though phonetically complex) register contrast, pitch is subordinate to a complex of voice quality features, with phonation type emerging as the most robust phonetic correlate of the contrast. In Burmese, some of the burden of contrast is carried by differences in syllable types (*killed tone*), and by duration: pitch and voice quality are usually required to indicate only binary contrasts.

In comparison with these three languages, we may observe that in the six tones of Karen, even after accounting for the cases where tonal contrasts are indicated by syllable type (final glottal stop), both pitch and voice quality seem to play a more substantial role in supplying potential acoustic cues to allow listeners to hear the difference between these tones. The final conclusion is that one might predict that within Jones's 'pitch register' model, pitch contours would be highly variable in the various contexts of connected or running speech, and that voice quality properties would remain robust.

A last comment is that all of these data are based on citation forms. To understand fully the implications of the present findings, it will be necessary to face up to the challenging prospect of analysing 'real-world' data. Another important component of future research should be perceptual tests to determine the relative robustness of pitch and voice quality contrasts in the ear of listeners.

References

Abberton, E., D. Howard, et al. (1989). "Laryngographic assessment of normal voice: a tutorial." *Clinical linguistics and phonetics* 3(3): 281-296.

Bradley, D. (1982). Register in Burmese. *Papers in South-East Asian Linguistics No.8: Tonation. Pacific Linguistics Series A-62*. D. Bradley. Canberra, Australian National University: 117-132.

Burling, R. (1969). Proto-Karen: a reanalysis. Occasional Papers of the Wolfenden Society on Tibeto-Burma Linguistics. A. L. Becker. Ann Arbor, MI, University of Michigan.

Chen, M. Y. (2000). *Tone Sandhi*. Cambridge, Cambridge University Press.

Diffloth, G. (1980). "The Wa Languages." *Linguistics of the Tibeto-Burman Area* 5(2).

Đỗ Thế Dũng Trần Thiên Hương and Georges Boulakia (1998). Intonation in Vietnamese. *Intonation Systems: a survey of twenty languages*. D. Hirst and A. Di Cristo. Cambridge, Cambridge University Press.

Gilmore, D. (1898). *A Grammar of the Sgaw Karen*.

Grierson, G. A. (1903). *A Linguistic Survey of India*. Calcutta, Office of the superintendent of government printing, India.

Haudricourt, A.-G. (1942). "Restitution du karen commun." *Bulletin de la Société de Linguistique de Paris* 42(1): 103-111.

Haudricourt, A.-G. (1975). "Le système de tons du karen commun." *Bulletin de la Société de Linguistique de Paris* 70: 339-343.

Hayward, K. M., J. W. Watkins, et al. (to appear). The phonetic interpretation of register: evidence from Yorùbá. J. Local, R. Ogden and R. Temple. Papers in Laboratory Phonology VI.

Hombert, J.-M., J. J. Ohala, et al. (1979). "Phonetic explanations for the development of tones." *Language* 55: 37-58.

Howard, D. M., L. G. A., et al. (1990). "Toward the quantification of vocal efficiency." *Journal of Voice* 4: 205-212.

Jones, R. B. (1961). *Karen linguistic studies: description, comparison, and texts*. Berkeley,, University of California Press.

Jones, R. B. (1986). Pitch register languages. *Contributions to Sino-Tibetan Studies*. J. McCoy and T. Light. Leiden, E.J. Brill. V: 135-143.

Laver, J. (1994). *Principles of phonetics*. Cambridge ; New York, NY, Cambridge University Press.

Luce, G. H. (1957). "Introduction to the comparative study of Karen languages." *Journal of the Burma Research Society* 42(1): 1-18.

Ní Chasaide, A. and C. Gobl (1997). Voice source variation. *The Handbook of Phonetic Sciences*. W. J. Hardcastle and J. Laver. Oxford and Cambridge (Mass.), Blackwell: 427-461.

Schutte, H. K. and R. Miller (1985). "Intraindividual parameters of the singer's formant." *Folia Phoniatrica* 37: 31-35.

Taylor, L. F. (1927). *Ethnographical survey of Burma : questions on the social structure, beliefs, customs and economic life of the indigenous races of Burma*. Rangoon, Government Printing and Stationery.

Watkins, J. W. (1997). "Can phonation types be reliably measured from sound spectra? Some data from Wa and Burmese. SOAS Working Papers in Linguistics and Phonetics 7: 321-339." *SOAS Working Papers in Linguistics and Phonetics* 7: 321-339.

Watkins, J. W. (1999). The Phonetics of Wa. Ph.D. dissertation. SOAS, University of London.

Watkins, J. W. (2000). *Phonation type phenomena in Burmese Tone*. Tone Symposium, Tromsø, Norway.

Phonology