Tonal Invariance and Downtrend in Cantonese

Ivan Yuen

Department of Theoretical and Applied Linguistics
University of Edinburgh, UK
{ivan}@ling.ed.ac.uk

Abstract

Studies of downtrend have reported that accents/tones are realised at progressively lower F0 values during an utterance. Listeners are reported to compensate for such F0 variations over time(Pierrehumbert 1979). Underlying the downtrend studies is the assumption that the phonological categories under investigation are invariant over time and that declination can be factored out of the quantitative characterisation of a given phonological category. In studying Vietnamese tones, Earle (1975) found that variable F0 values for tones can be resolved to distinctive tonal categories in relation to a normalised tonal space: after normalisation, stable and consistent tonal relations were uncovered. The current paper investigates whether such a normalisation is possible as the tonal space changes over time with declination. A production study based on read speech showed first that there is downtrend in Cantonese, and second tonemes exhibit quantitative invariance relative to a locally-determined tonal space.

1. Introduction

Intonation literature has extensively reported in many languages (for example, Dutch, English, Spanish, Yoruba, Hausa and Chinese) a phenomenon called downtrend, referring to a gradual overall F0 descent throughout the course of an utterance. A late-occurring tone or accent will be realised at a lower phonetic F0 than the same tone or accent occurring early in the same intonation contour. More generally, downtrend studies suggest unanimously that the phonetic F0 realisations of the same linguistic category vary over time. Various phonological phonetic accounts have been proposed for the phenomenon.

Underlying the attempts to model downtrend is the aim to retain and capture the phonological equivalence of tones/accents involved. Models of downtrend in a sense aim at explaining away the phonetic variations so that one can say an early-occurring accent/tone is phonologically the same as a late-occurring one. How do we recover the phonological equivalence from the variable phonetic realisations over time? Is there any systematic phonetic behaviour from the same linguistic category?

In work unrelated to the study of downtrends, it has been found that tonemes exhibit regular and systematic phonetic behaviour after normalisation. According to Ladd (1996), ‘a normalising model attempts to abstract away from differences between speakers, paralinguistic effects, and expresses the invariant characterisations of tones in terms of an idealised speaker range that results from the process of factoring out sources of variation.’ A simple normalising model can be found in Earle(1975) in his description of Vietnamese tonemes in citation form. He gave a value of 100point of a speaker and 0speaker-specific range. Relative to the range, the 6 tonemes were described quantitatively as percentage of the range. For example, Tone 1 has an F0 onset of 51down to 50normalised tonal contours exhibited a high degree of regularity and similarity between speakers. In other words, tonal invariance can be expressed relative to a normalising tonal space.

If the same toneme exhibits invariant phonetic characterisations across speakers relative to a tonal space, it is reasonable to ask whether the same would be true of the changes to the tonal space produced by downtrends.The current paper used a tone language, Cantonese, to study how the same toneme varies over time because of utterance-level downtrend and how the tonal invariance relative to a normalising tonal space for the same toneme behave over time. Is it possible to model downtrend as the gradual modifications of tonal space relative to which each toneme is scaled? Cantonese is selected for this study, because the explicit phonological identity of tonemes allows for close monitoring of F0 data points.

2. Experimental Design

The production experiment has two aims. First, it sets out to establish the presence of downtrend in Cantonese. Four tonemes, sharing a similar level/falling syllable contour and distinctive primarily in terms of F0 height, were used. These 4 tonemes occurred in a noun compound in a time-order sequence in an intonation contour. It is hypothesized that test tonemes occurring late in the time-order sequence will have lower F0 realisations than the early ones.

Second, the experiment investigates whether the scaling of a given toneme relative to a locally-defined tonal space will remain constant over time, despite the presence of downtrend. Three test tonemes embedded in the constant tonal context were used.

2.1. Materials

The experiment used the following 4 tonemes to test the presence of downtrend: High-Fall(HF), Mid-High(MH), Mid-Low(ML) and Low-Fall(LF).

Nonsense 2-syllable noun compounds were put together out of the 4 tonemes to form 3 tonal compounds, namely, HF-MH, HF-ML, and HF-LF. Each of the 3 test tonal compounds occurred in a time-order sequence in 3 utterance positions called here 1st serial position, 2nd serial position, 3rd serial position. Each of the test tonal compounds occurred in a carrier sentence in 2 sentence length conditions. The ‘short’ sentence length condition had 12 syllables and the 3rd serial position was utterance-final; whereas the ‘long’ sentence length condition had 13 and the 3rd serial position was not utterance-final. The experimental paradigm for the test tonal sequences is illustrated below:

- Test Tonal compounds: $HF-MH / HF-ML/ HF-LF$
- Test Utterance Positions: $X1/X2/X3$
- Carrier sentences:
- HF HR X1 LF LF X2 LF LF X3. (short sentence)
- ta se X1 lin mai X2 lin mai X3.
- HF HF X1 LF X2 LF X3 HR ML LF. (long sentence)
- ta choeng X1 tung X2 tung X3 se ha loi.
- He writes X1 and X2 and X3 (Gloss)

The constant tonal context in this experiment was so designed that MH, ML and LF in the test tonal sequences occurred between a preceding High tone and a following Low tone. This local tonal context defines a local tonal space. The High tone is assumed to reflect the top of the tonal space and the low tone the bottom. This tonal context makes it possible to examine the scaling relation between these 2 F0 extrema and the 3 test tonemes (MH, ML and LF).

Sonorants were used for all the test tonal sequences to elicit continuous F0 contours, as well as to avoid any F0 perturbation from consonants.

108 test tokens were designed and 4 speakers(2 male and 2 female) were recruited at University of Edinburgh and Hong Kong University of Science and Technology for this study.

2.2. Measurements

F0 values were measured for all the test tonal compounds embedded in the constant tonal context in all 3 utterance positions. The F0 peak from the High tone in the tonal context was recorded as the maximum and the F0 valley from the Low tone in the tonal context as the minimum in each utterance position. MHs, MLs and LFs in the tonal compounds were measured at the inflection point between a steep F0 drop from the preceding HFs and a levelling off of F0 towards the end of the syllable. When no inflection point was observed on MHs, MLs and LFs, F0 measurements were taken at the end of the test syllables.

3. Results and Discussion

Data from one male speaker was discarded because of the high incidence of disfluency. Data from three speakers are presented. As the short sentence length condition was complicated by final lowering, the current results focussed on the long sentence length condition.

It was found that the raw mean F0s for the tone tokens in the test tonal sequences (HF, MH, ML, LF) exhibit downtrend, as they move towards the end of an utterance, as seen in the following tables.

The first 3 rows of each table measured the HFs that occurred in 3 tonal compounds with MH, ML and LF. As seen from the tables, the mean F0 for HFs are largely unaffected by the identity of the following tone. Therefore, the first 3 rows were collapsed for each speaker for further statistical analysis.

Numerically, the tables indicate that the mean F0 values for HFs, MHs, MLs and LFs decline over time.

ANOVA with repeated measures was conducted on the recorded F0 values on 4 tonemes (HF, MH, ML, LF) in each of the 3 serial positions for each speaker to test for the effect of SERIAL UTTERANCE POSITION(3 levels) and TONE(4 levels). Significance (p¡0.01) was found for both factors for all 3 speakers. That means that all test tonemes are affected by the serial positions. Based on this analysis, we can establish the claim that downtrend is found in Cantonese. Interaction effect was also found for all three speakers (p¡0.01), suggesting that SERIAL UTTERANCE POSITION affects tonemes differently.

Table 1: *Mean F0s(hz) of Test tones in 3 Utterance Positions*

Test Tones	Speakers	1st Utt.Pos.	2nd Utt.Pos.	3rd Utt.Pos.
HF(MH)	1	135	115	109
	2	286	261	248
	3	261	225	212
HF(ML)	1	137	116	109
	2	288	262	249
	3	262	228	215
HF(LF)	1	138	115	109
	2	293	266	255
	3	272	234	218
MH	1	97	88	85
	2	197	188	183
	3	192	179	175
ML	1	91	83	80
	2	188	178	169
	3	187	177	173
LF	1	73	67	62
	2	157	154	148
	3	170	167	166

As mentioned under 'Experimental Design', the second goal of the experiment is to investigate if the realisation of the test tonemes relative to a tonal space also varies over time. To do this, we first computed a new variable RATIO to describe quantitatively the tonal relations between test tone tokens and the constant tonal context in each serial position, according to the following formula

$$Ratio = \frac{(testtone - localF0min)}{(localF0max - localF0min)}$$

Local F0max is the value taken from the F0 peak of the HF preceding each of the test tone tokens.

Local F0min is the value taken from the F0 valley of the LF following each of the test tone tokens.

Test-Tone is the value taken from MH, ML and LF in the test tonal compounds as described above in 'Measurement'.

On account of the lack of a following LF in the 3rd serial position, we separated the analysis into 2 sets. SET1 used a preceding HF and a following LF to define the local tonal space, as specified in the formula. Below are the mean values of RATIO for each of the 3 test tones in the 1st and 2nd serial utterance positions (called here SET1).

Table 2: *Mean F0 (hz) of RATIO of test tones in different utterance positions (SET1)*

Test Tones	Speakers	1st Ser. Pos.	2nd Ser. Pos.
MH	1	0.3509	0.3654
MH	2	0.2914	0.2901
MH	3	0.2604	0.2065
ML	1	0.2424	0.2435
ML	2	0.2102	0.1967
ML	3	0.1951	0.1788
LF	1	0.0019	-0.0017
LF	2	-0.0034	-0.0052
LF	3	0.0037	0.0044

As mentioned earlier, there is no following LF in the 3rd serial position. In this case, the local tonal space in the above formula was re-defined in terms of a preceding Low and a preceding High. Then the formula as above was applied to calculate RATIO for the test tokens in both the 2nd and the 3rd serial positions for comparison (called here SET2). Results are tabulated below:

Table 3: *Mean F0(hz) of RATIO in 3 different utterance positions (SET2)*

Test Tones	Speakers	2nd Utt. Pos.	3rd Utt. Pos.
MH	1	0.2936	0.3505
MH	2	0.2772	0.2744
MH	3	0.2026	0.1869
ML	1	0.167	0.2096
ML	2	0.1644	0.1301
ML	3	0.1378	0.1454
LF	1	-0.0091	-0.0018
LF	2	-0.0077	-0.0012
LF	3	0.0015	0.0041

The numerical data from RATIO (in both SET1 and SET2) suggested that each test toneme exhibits a stable and constant qunantitative scale within a local tonal space for all 3 speakers in all 3 serial positions, as long as the local tone space is defined consistently.

When we examined the RATIO values for the three tonemes, we noticed that the quantitative values for each toneme differ from speaker to speaker. For example, speaker 1 has 0.35 for MHs in the 1st serial position, speaker 2 has 0.29 and speaker 3 has 0.19 for the same toneme in the same utterance position. However, the quantitative values that characterise MHs and MLs for all three speakers are consistent in that MHs alwasy have higher values than MLs.

In fact, when ANOVA with repeated measures was conducted on the dependent variable RATIO with 2 independent factors - TONE and SERIAL UTTERANCE POSITION, the effect of SERIAL UTTERANCE POSITION disappeared in both sets. Only TONE reached significance level at p¡0.01.

The effect of SERIAL UTTERANCE POSITION was factored out, when the test tone tokens were normalised locally in relation to their adjacent tone tokens that defined a local High and Low as the tonal space. Tones appear to exhibit invarinat phonetic F0 scaling, despite the presence of downtrend. Whether downtrend results from pitch range compression or the lowering of pitch register, the local tonal space in this experiment in a sense captures the local variations of the pitch range. In other words, F0 variations over time can be compensated for if we assume a local normalising tonal space to maintain invariant tonal description.

The results in this experiment have made the first approximation of a quantitative characterisation of, in effect, 2 tonemes(MH and ML). As seen from the tables, the quantitative scales for the two tonemes vary from speaker to speaker in all three utterance positions. In order to arrive at a more general and explicit quantitative characterisation, we need to factor out the speaker variations. Further investigation is also needed to find out how to define the tonal space that allows us to characterise quantitatively the remaining tonemes.

4. Conclusion

The paper started out with the assumption that the same toneme will exhibit some sort of invariant phonetic characterisations. Evidence has been provided by Earle that such invariant phonetic characterisations can be expressed after normalisation. At the same time, intonation literature has documented extensively how F0s vary over time. With the aim of investigating the invariant phonetic characterisations of tonemes in Cantonese over time, the current study has made the following observations.

Though F0 values of each test tone vary over time as a result of downtrend, each test tone remains distinctive from one another in each corresponding time-order position. And in each time-order position the test tones appeared to maintain a relatively stable quantitative relation within a locally-determined tonal space over time. The experimental results suggest the presence of a normalising tonal space within which tones are scaled at some invariant ratios. It appears that downtrend can be re-expressed as a time-varying locally-determined tonal space.

5. References

[1] Shih, C., 2000. A declination model of Mandarin Chinese. In*Intonation:analysis, modelling and technology* Botinis, A. (ed.). Kluwer Academic Publishers

[2] Earle, M.A., 1975. An acoustic phonetic study of northern vietnamese tones, SCRL monograph.

[3] Inkelas, S.; Leben, W.R., 1990. Where phonology and phonetics intersect:the case of Hausa intonation. In*Papers in Laboratory Phonology 1,* Kingston, J.; Beckman, M. (ed.). CUP.

[4] Ladd, R., 1996. Intonational Phonology, CUP.

[5] Liberman, M.; Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In*Language Sound Structure*, Arnoff, M.; Oehrle, R.T. (ed.).

[6] Prieto, P.; Shih, C.; Nibert, H., 1996. Pitch downtrend in Spanish. *Journal of Phonetics*, 24, 445-473.

[7] Pierrehumbert, J., 1979. The perception of fundamental frequency declination. *Journal of Acoustical Society of America*.

[8] Shih, C., 2000. A declination model of Mandarin Chinese. In*Intonation:analysis, modelling and technology* Botinis, A. (ed.). Kluwer Academic Publishers.

6. Acknowledgments

I would like to thank my supervisors Prof. Bob Ladd and Dr. A. Turk for their insightful comments during this study. Any error is entirely my own.