F0 and phonation types in Nghe Tinh Vietnamese tones

Koichi Honda

Phonetics Laboratory, Linguistics (Arts)
Australian National University
Koichi.Honda@anu.edu.au

Abstract

This paper describes tonal F0 and phonation types in the unstopped tones of the hitherto little studied North Central Vietnamese dialect of Nghe Tinh. Field recordings of six female native speakers are analysed. It is shown that speakers differ in number of tones (5 vs. 4); tonal contour; and in amount of tonal laryngealisation, some of which correlate with sociological factors. Normalisation of F0 and duration reveals consistent characteristics of the contrastive tonemes. Phonation type is assumed to be used as a secondary cue.

1. Introduction

Vietnamese dialectal varieties are conventionally divided into three main groups: Northern, Central and Southern. Above all, the acoustics of Vietnamese tones have been intensively studied on the standard language of the North, where majority of works conclude that the use of phonation types is as important as pitch in signalling tonal contrast (Nguyen VL et al. 1997, etc.). Compared to the standard Northern Vietnamese (NV), little has been studied on dialectal varieties so far. Particularly, Central Vietnamese is known for preserving archaic characteristics in various ways, such as largest number of initial consonants, high amount of archaic words, sub-local tonal variations within the area, and so on (Hoang TC 2003). As to the tones, Central Vietnamese is reported (Vu 1982) to use narrower pitch range than other dialects in its realisation.

From a viewpoint of phonological tone, however, North Central variety, spoken in *Nghe An* (NA) and *Ha Tinh* (HT) provinces, hereinafter called Nghe Tinh Vietnamese (NTV), may be regarded as an isolated independent variation, which deserves special attention. First, it is often regarded by Vietnamese people as having 'peculiar accent', or being a difficult-to-understand dialect. Second, it is unique in its mode of tonal merger. Table 1 shows a summary of tone correspondences of unstopped open syllables across dialects. The tone numbers used in this paper, i.e. T1, T2…, correspond to their Vietnamese names as follows: T1 *ngang*, T2, *huyền*, T3 *sắc*, T4 *nặng*, T5 *hỏi*, T6 *ngã*.

Table 1: A summary of Vietnamese tonal merger across dialects (cited in Hoang TC 2003)

Dialect groups	NV	Central Vietnamese		SV
Sub-groups		*NTV*	others	
No. of tones	6	*5*	5	5
Mode of merger	-	*T6=T4*	T6=T5	T6=T5

As is shown, Vietnamese dialectal varieties, except for NV, represent five-tone-systems. The central point lies in the merger of T6 that is often called 'broken tone' of NV, which is realised with mid-syllable creaky voice and high rising contour. T6 of NV is merged with T4 in NTV, whereas in other dialects, including Thanh Hoa province which is located between Nghe An and Hanoi, T6 is merged with T5. The acoustic-phonetic study of tones in NTV promises to contribute to a deeper understanding of both synchronic and diachronic aspects of Vietnamese. Though the importance of a survey on NTV varieties was pointed out by linguists (Thompson 1987: 81, etc.) in the past, available literature is small and reflects only partial information on languages in Nghe An province (Alves et al. 1998, Alves 2002, Pham 2003 & 2005). Up till now, the acoustics of NTV tone have only been studied by Vietnamese scholars (Hoang CC 1989, Nguyen VL 2002, Pham 2005), who base their work on a single speaker from specified locations in Nghe An province. So far Ha Tinh province is yet to be surveyed.

This paper aims to describe NTV tones of both NA and HT provinces, by analysing fundamental frequency (F0) and phonation types in tones of isolated citation forms of unstopped open syllables. The subjects are six female speakers, who are from adjacent regions in both NA and HT provinces along the Lam River. The obtained F0 values are further normalised to extract common, linguistic-tonetic features of each tone.

2. Methods

Area: Fieldwork was done as part of my doctorial research from March to August 2006, in the northern HT province and the southern NA province, by basing in *Vinh* city, the capital of NA province. Since speech of my acquaintances from Vinh city and *Hung Nguyen* district (the suburbs of Vinh city) sounded to me similar to NV, such area was excluded from my research though it is located along the Lam River. Contrary to my previous knowledge of the existence of various sub-regional variations, my informants seemed to present a rather standardised NT dialect, except for minor differences in sub-regional lexicon and degrees of individual

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

laryngealisation. (In this paper, the term *laryngealisation* is used to refer to various degrees of glottal constriction.)

Informants: Six female speakers were chosen for analysis; they are from *Hong Linh* (1), *Duc Tho* (3), *Can Loc* (1) of HT province and *Nam Dan* (1) of NA province. They were chosen on the following criteria. 1) They were born and brought up consistently in one commune at least up to 18 years old. 2) They were confident about their native speech. 3) They came from adjacent areas. 4) They spanned a wide range of ages from 24 to 53. 5) They all sounded to have clear voices, suitable for the analysis of F0 and phonation differences. Profiles of the six informants are shown in the order of ID number, native district, province, age, occupation as follows:

- F2, Hong Linh, HT, 29, teacher
- F3, Duc Tho, HT, 30, teacher
- F4, Duc Tho, HT, 53, farmer
- F6, Can Loc, HT, 24, housewife, formerly farmer
- F7, Duc Tho, HT, 30, clerk
- F9, Nam Dan, NA, 53, restaurant owner, formerly farmer

Elicitation: Since all the informants also spoke standard NV, care had to be taken in eliciting the data. Unlike previous acoustic studies of Vietnamese tone, using an array of syllables with six tone diacritics and having them read out, was not employed, because such a method might include unnatural or even meaningless syllables in their dialect(s). More seriously, their pronunciation might be affected by the orthography. Instead, I wrote my wordlist in official NV orthography. It consisted of monosyllabic and polysyllabic common lexical items in the first column. In the second column, explanations of the items were given. The informants were asked to give the local words for the written words, and repeat them three times. In addition, local NTV words were given in a third column, in locally modified unofficial writing, that enabled speakers to imagine the objects easily. In this way, I took some trouble to elicit local items, with the assumption that they would have local pronunciation. From the recordings, only unstopped open syllables were chosen from the list as objects for the analysis of this paper.

Equipment: Informants were recorded at 44.1 kHz / 16bit / mono with a DAT recorder (Sony Walkman TCD-D8) and one-point Sony ECM-MS957 microphone. The speech was then downsampled to 22 kHz for making individual sound files. Analysis was with *Praat* (4-3-17).

Measurement: Tokens to be analysed were selected from the above-mentioned wordlist. In all, 1,679 tokens were analysed, comprising: T1/492, T2/259, T3/289, T4/260, T5/379 tokens by tone; F2/262, F3/309, F4/258, F6/250, F7/348, F9/252 tokens by speaker. Among various acoustic properties involving NTV tones, F0 and duration were chosen as primary cues. A sampling base was defined to extend from the onset of the Rhyme to its offset. For those tones with laryngealisation at their offset, such as T3/T4/T5, the sampling extended to the beginning of the F0 perturbation, and the laryngealised duration was not measured. F0 value was then sampled at every ten percent equalised point of the tonal duration (i.e. eleven points from 0% to 100%), which amounted to 18,469 sampling points (1,679 tokens x 11 sampling points).

3. Results

3.1. Tonal contrast

Five of the six NTV speakers appeared to contrast five tones, with both pitch difference and phonation type difference in their realisation. These can be exemplified with the speech of F2. Since it was difficult to find a complete set of minimal pairs from the wordlist, sub-minimal pairs are shown as follows: T(one)1 [mɛɛ 35] ‘calf’, T2 [mɛɛ 33] ‘carp’, T3 [ŋɛ̰ɛ 13] ‘young buffalo’, T4 [mɛ̰ 22] ‘mother’, and T5 [t̰ɛ̰ 31] ‘young’. The remaining speaker appeared to lack tone 4, and therefore only contrasted four tones.

At first, I found it difficult to identify T2 and T4; however, later I found there is a weak glottal constriction called stiff voice (Ladefoged and Maddieson 1996: 55) at the offset. It was easily imagined that for those having got accustomed to NV and SV, hearing an unexpected reversed pitch range in T3 and T2 would be astonishing: T3 is realised in a very low, while T2 in a quite high pitch range. The high pitch of T2 is also confusing in the identification of T1 and T2, because the beginning pitch of T2 usually exceeds that of T1, which is reverse to other dialects. The ending pitch of T3 is sometimes quite high, which may make identification of T1 and T3 difficult. The contour of T5 is similar to that of NV, however NTV T5 ends with a strong glottal stop, while NV T5 does not have a glottal stop.

3.2. Individual speaker’s mean F0 plots

Individual speakers’ mean fundamental frequency (F0 in Hz) is plotted as a function of raw mean duration in Figure 1. In order to show between-speaker differences, scales are fixed. A neat configuration can be seen consisting of two parallel pairs of F0 shapes: one rising or rising-falling, and one level or level-falling. For five of the speakers there is also a separate falling F0 shape.

The mean female F0 ranges from ca. 160 Hz to ca. 300 Hz. which is not a very narrow pitch range as Vu (1982) described as the characteristics of Central Vietnamese. The widest pitch range is F3’s 116 Hz, and the narrowest is F9’s 96 Hz; with T1’s peak highest and T5’s minimum lowest. It shows there are two groups of tones in F0 values: the higher T1/T2 and the lower T3/T4/T5 (see F2, F4, F7); the latter are usually accompanied by some degrees of glottal constriction in perception.

T1 has two obvious between-speaker variations: high rising contour (F2, F3, F7) and convex contour (F4, F6, F9) in the high pitch range. The difference appears to correlate with socio-economic background, since the former group are white- collar workers, and the latter are, or have been farmers (see above). It is possible therefore that the convex contour is more conservative. **T2** has relatively level pitch in the high pitch range and its duration is one of the longest of all the tones (F2, F7). Speakers show differing degrees of F0 decline, from none (F2, F7) through gradual (F3, F4) to substantial (F6, F9). **T3** has an overall rising contour in the low pitch range, however it should be noted that all start with falling contour at the beginning, with laryngealisation likely to occur at the lowest point. The ending point of T3 varies from one speaker to another: as high as T1 (F3, F9), near to T2 (F2, F4, F6), and as low as T4 (F7). **T4** has relatively level pitch in the low pitch range and its duration is one of the longest of all the tones (F2, F3), but some variations trail downward (F4, F7), and others show a dramatic fall at the end (F6, F9). Since

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

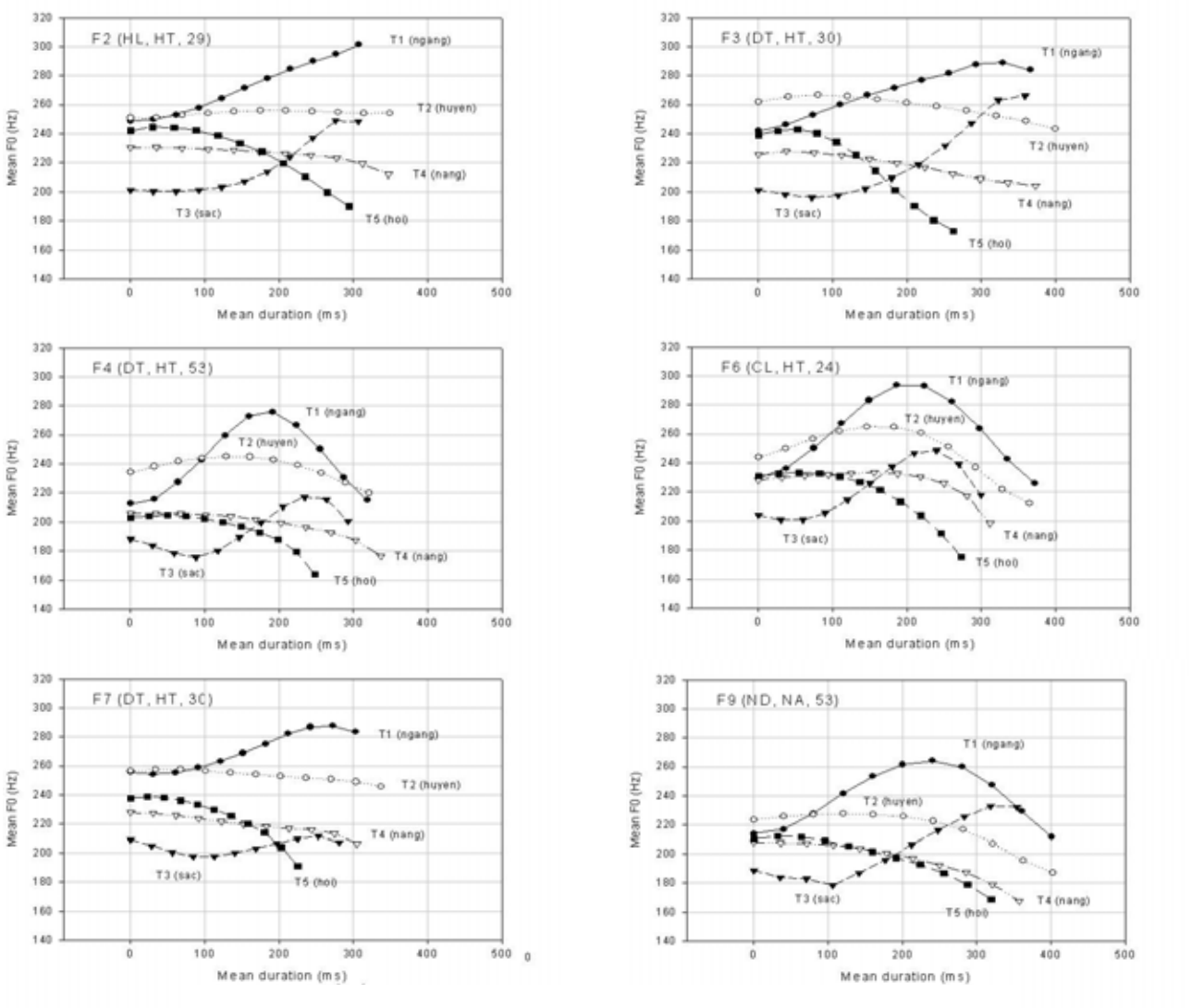


Figure 1: Individual speaker's mean F0 as a function of raw mean duration for their five tones on unstopped syllables

measured duration does not include the laryngealised portion, audible duration would be longer than the figure. **T5** falls more sharply than T4, consequently, its duration is the shortest and the lowest F0 is the lowest among the five tones. Besides, starting pitch range varies: higher than T4 (F2, F3, F7), or as high as T4 (F4, F6, F9). Noteworthy are F9's T4 and T5: they have practically the same F0 contour and duration, which corresponds to my perception that only one toneme is involved, thus forming a four-tone-system.

3.3. Phonation types

The lower F0 of T3/T4/T5 is normally accompanied by some degree of glottal constriction in mid-syllable (T3) and at the syllable end (T3, T4 and T5). The degree of constriction varies from speaker to speaker and there is within-speaker variation as well. Generally speaking, **T5** ends with a strong glottal stop and creaky voice. **T4** usually ends with weak stiff voice; otherwise with modal voice. The four-tone-system holder F9, to my perception, has within-speaker variation from stiff voice to glottal stop with T4/T5. F9 also has the most salient mid-syllable laryngealisation in **T3**. As I mentioned before, laryngealisation in T3 often occurs in the first half of the tonal duration where the pitch reaches its bottom, and the degree of constriction varies from stiff voice to creaky voice. As an example, the waveform and wide band spectrogram of F9's /kạ̃a 13/ 'fish' (T3) is shown in figure 2.

As can be seen, the peak-to-peak amplitude of the waveform shrinks in the first half of the Rhyme duration (ca. 0.15 second point), with a consequent decrease of intensity and F0. The obvious lack of jitter and shimmer is noteworthy. The offset of T3 is also often accompanied with creaky voice or glottal stop. It is of interest that the informants of *Thanh Chuong* and *Do Luong* districts (upper stream of the Lam River) tended to have stronger creak or glottal stop in mid-syllable that sounded like two syllables, and which made F0 extraction problematic. It may be possible to draw a picture of gradually changing laryngealisation along the River. (F2, a speaker from the most downstream of the river, does not have

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

laryngealisation but only a rising pitch.) This matter is left for future analysis.

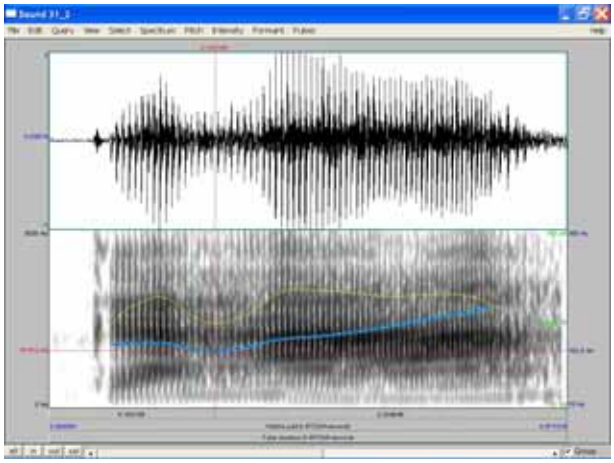


Figure 2: Waveform and wide band spectrogram of [kạa] T3 (F9) showing mid-syllable laryngealisation.

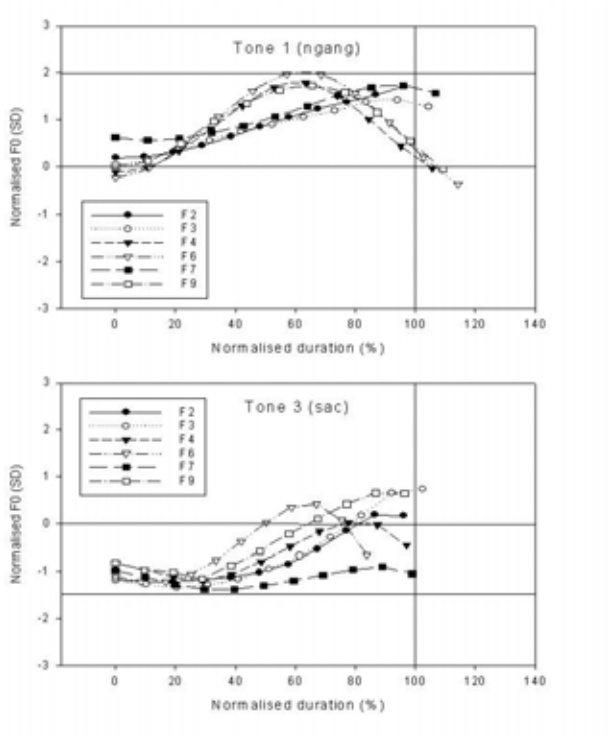
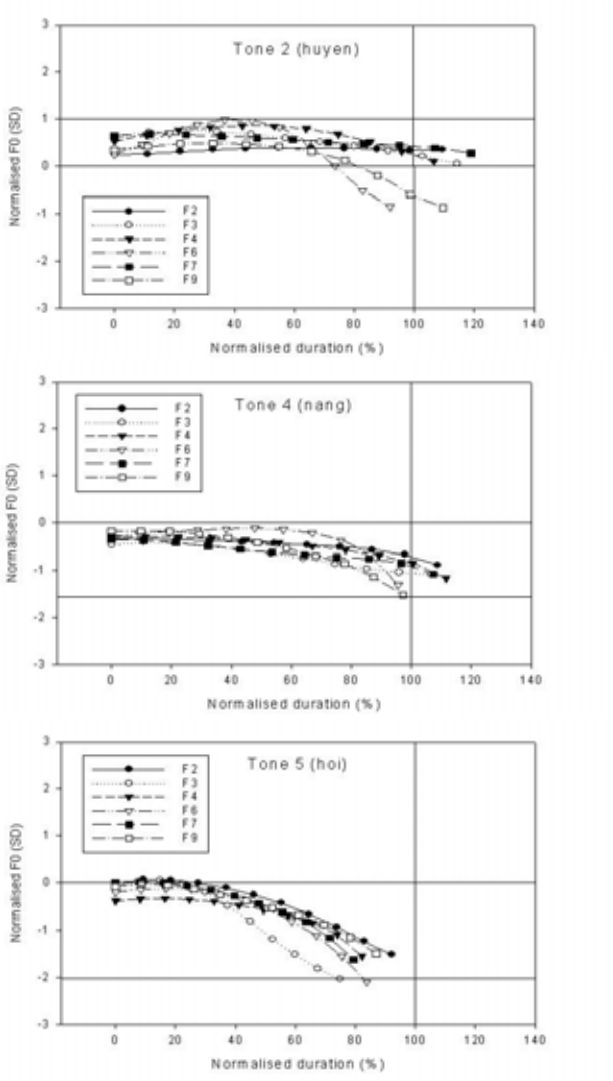


Figure 3: Normalised F0 as the function of normalised duration of six female NTV speakers by each tone.

It can be seen that the situation with respect to tonation (tonally relevant phonation – Bradley 1982: vi) in these dialects is complex, and shows more variation than in Standard Vietnamese. From these observations, it is likely that the phonation type is not contrastive, but rather correlates with the pitch: as the pitch falls to the lowest part of the speaker's range, the glottal constriction becomes stronger; consequently, phonation types change from modal voice, stiff voice, and creaky voice to glottal stop.

3.4. F0 normalisation

The aim of normalisation is to get rid of between-speaker difference and extract as much as possible the linguistic-tonetic content of the speech (Rose 2000). Each speaker's mean F0 in Figure 1 were z-score normalised (Rose 1987). In addition, in order to retain information on relative tonal duration, duration values were also normalised (Rose 2000). Each tone's normalised duration was calculated as a percentage of a speaker's mean duration value calculated from all tones. These values are illustrated by each tone in Figure 3, where it can be seen that the tones generally normalise very well, in the sense that their normalised contours cluster closely. Each tone is discussed separately in the next section.



Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

4. Discussion

Based on the results shown in Figure 3, five tones of NTV can be described as below. In addition, their tentative phonological representations using Yip's (2002) method, and Nguyen's (2002) analysis of Nghe An sub-dialects are also referred to. Yip represents tones phonologically with two registers (purely 'pitch range' in her use) [+/-U(pper)], and contours made up of sequences of tones [h, l, hl, lh], while Nguyen describes tones in NA province with contour and voice quality (phonation types). Among the sub-dialects Nguyen described, the F0 contour of a speaker from Thanh Chuong district, upstream of the Lam River, looks very similar to those of my informants.

Tone 1: T1 has a rising F0 contour in the upper pitch range, which rises from around the mean F0 of individual speaker to ca. 2.0 sds above the mean. Both shapes have the same normalised peak at ca. 2.0 sds above the mean, and differs only in its timing. The phonological representation of this is /+U, lh(l)/. Nguyen describes T1 as "đường nét lên, chất giọng thường [rising contour with modal voice]". This description agrees with the rising variant of NTV T1.

Tone 2: T2 has a level or a slightly convex contour in the upper pitch range, which lies within 1.0 sd above the mean. The duration, except F6, is 110% to 120% longer than the mean. Some speakers have falling contour (F6, F9) which looks parallel to the T1 convex contour, and they may be sub-regional variations (Can Loc and Nam Dan), however they maintain the T2 pitch range above the mean until 70-80% of the tonal duration, and the falling occurs at the last 20-30% of the duration. It is difficult to represent this tone with Yip's system, if the falling component has to be modelled. If not, its phonological representation of T2 is /+U, l/. Nguyen describes T2 as having "đường nét xuống, chất giọng chùng [falling contour with breathy voice]". Although I did not perceive breathy voice from my informants, T1 and T2 are often accompanied by h-like voiceless sound at offsets, which seems a sign of an opening glottis in contrast with the constricted glottis of T3, T4 and T5.

Tone 3: T3 is somewhat problematic. The overall F0 contour of T3 is regarded as a rising in the lower pitch range. In that sense, its phonological representation would have to be /-U, lh/. However, a question arises. I call T3 *negative rising*. It is because it involves lowering of F0 until it reaches ca. 1.4 sds below the mean at the first 20-40% of tonal duration which is often accompanied by glottal constriction. Normalised F0 values of all the speakers concentrate at this point, indicating some sort of a target. Due to the lowering of F0 in the fist half of the duration, the following tonal F0 becomes relatively high, but the degree of rising varies: the highest up to +0.6 sds (F3) and the lowest up to -1.4 sds (F7). Therefore the most important point of T3 seems to be at the bottom of F0 contour. An articulatory explanation may be that the rising pitch of the T3 is not produced by the lengthening of vocal folds in the latter half of the duration, but rather lowering of the pitch to the speaker's lowest pitch range in the first half of the tonal duration, in two ways: lowering the whole of the larynx and constricting the glottis. The pitch rises relatively in the latter half by the increase of air flow through the glottis after releasing the compressed subglottal air pressure when the glottal stricture is released. Due to this, initial stop consonants of NTV T3 words tend to become implosives with the decrease of air pressure in the supralaryngeal vocal tract due to the lowering of the whole larynx. Although it is beyond the scope of this paper, in two syllable utterances of T3+T2, the pitch of T3 does not rise but ends with glottal stop, which likely causes tone sandhi. This may be an evidence of it. Nguyen describes T3 as having a "đường nét xuống-lên, thanh quản hoá [falling-rising contour with laryngealisation]". This description agrees with mine.

Tone 4: T4 has a level or falling-off F0 contour in the lower pitch range, F0 of which lies mostly within 1.0 sd below the mean. The duration is normally longer than the mean, which is about 110% (except F6 and F9). The phonation at the tonal offset varies from stiff voice to modal voice. T4 parallels T2 in contour, and contrasts with it in its pitch range specification. Therefore, the phonological representation of T4 is assumed to be /-U, h/. Nguyen describes T4 as having a "đường nét xuống, kết thúc bằng thanh môn hoá [falling contour, ending with glottal stop]". This description, however, is more similar to my observation of T5 than to T4.

Tone 5: T5 has as a falling F0 contour in the lower pitch range, which falls near to 2.0 sds below the mean. The duration is shortest of all the tonemes, which is from 80% to 90%. T5 is the same as T4 for the one speaker (F9) who has a four-tone-system. (This may be a case of sub-regional variation or a personal variation. Others use shorter duration, lower offset F0 and/or higher onset F0 as acoustic cues to distinguish it from T4.) The phonation at the tonal offset varies from creaky voice to glottal stop. The phonological representation of T5 is assumed to be /-U, hl/. Nguyen describes T5 as having a "đường nét xuống, chật giođng kẹt thanh (thanh quản hoá) [falling contour with laryngealisation]". This description agrees with mine.

5. Conclusions

In this paper, the unstopped tones of six female speakers of NTV were analysed in terms of F0, duration and phonation type. NTV was shown to have a fairly simple tone system, at least on unstopped tones, consisting of two level, two rising, and one falling tone. However, some between-speaker variation was discovered in tonal offset. In spite of between-speaker variations in mean tonal F0 and duration of five contrasting tones among speakers, normalised F0 and duration showed homogeneity with five speakers, while one showed neutralisation of two of the lower tones (T4 and T5). As a result, pitch in perception is assumed to be used as a primary cue for the five contrasting tonemes. On the other hand, phonation types in three lower tones have between-speaker variation and even within-speaker variation as well. Contrary to my earlier expectations, therefore, phonation types are not contrastive in NTV and are assumed to be used as a secondary cue, possibly along with duration, in the realisation of each toneme.

6. Acknowledgements

I would like to thank my informants for their cooperation in the field. Thanks are given to Dr. Phil Rose, Dr. Nguyen Van Loi, Dr. Andrea Hoa Pham and my two referees for their valuable comments. The research project was conducted under the approval of the Local Ethics Committee, Arts, ANU.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

7. References

Alves, Mark J and Nguyen Duy Huong (1998). Notes on Thanh-Chuong Vietnamese in Nghe An Province. [Online]. Available: http://www.geocities.com/malves98/publications.html. [2006, November 10].

Alves, Mark J (2002). A Look at North-Central Vietnamese. [Online]. Available: http://www.geocities.com/malves98/publications.html. [2006, November 10].

Bradley, David (ed.) 1982. *Papers in South-East Asian Linguistics, No.8 Tonation*. Pacific Linguistics, Australian National University, Canberra.

Hoang, Cao Cuong (1989). Thanh điệu Việt qua giọng địa phương trên cứ liệu Fo. [The Fo values of Vietnamese tones in dialects]. *Ngôn ngữ [Linguistics]* 4: 1-17.

Hoang, Thi Chau (2003). Phương *ngữ học tiếng Việt [Vietnamese dialectology]*. Nhà xuất bản Đại học Quốc gia Hà nội [Hanoi National University Press], Hanoi.

Ladefoged, Peter and Ian Maddieson (1996). *The Sounds of the World's Languages*. Blackwell Publishing, Malden.

Nguyen, Van Loi and Jerold Edmondson (1997). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies* 28: 1-18.

Nguyen, Van Loi (2002). "Thanh điệu một vài thổ ngữ Nghệ An từ góc nhìn đồng đại và lịch đại" [Tones of some dialects in Nghe An from synchronic and diachronic perspectives]. *Ngôn ngữ [Linguistics]* 3: 1-12.

Pham, Andrea Hoa (2003). *Vietnamese Tone: A New Analysis*. Routledge, New York.

Pham, Andrea Hoa (2005). Vietnamese tonal system in Nghi Loc: A preliminary report. *Toronto Working Papers in Linguistics* 24: 183-201.

Rose, Phil (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6: 343-351.

Rose, Phil (2000). Hong Kong Cantonese citation tone acoustics: a linguistic tonetic study. In Michael Barlow (ed.) *Proceedings of the 8th Australian International Conference on Speech Science & Technology* (pp.198-203). Australian Speech Science & Technology Association, Canberra.

Thompson, Laurence C (1987). *A Vietnamese Reference Grammar*. University of Hawaii Press, Honolulu.

Vu, Thanh Phuong (1982). Phonetic properties of Vietnamese tones across dialects. In David Bradley (ed.) *Papers in South-East Asian Linguistics, No. 8, Tonation* (pp.55-76). Pacific Linguistics, Australian National University, Canberra.

Yip, Moira (2002). *Tone*. Cambridge University Press, Cambridge.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review