euronoise

A comparison of native speaker and American adult learner Vietnamese lexical tones

Allison Blodgett[a], Jessica Bauman[a], Anita Bowles[a], Lykara Charters[a], Anton Rytting[a], Jessica Shamoo[a] and Matthew Winn[b]

[a]University of Maryland College Park, Center for Advanced Study of Language, 7005 52nd Ave, College Park, MD 20742, USA
[b]University of Maryland College Park, Department of Hearing & Speech Sciences, 0100 Lefrak Hall, College Park, MD 20742, USA
mwinn@hesp.umd.edu

This study investigated native and non-native speaker tone production in Northern and Southern Vietnamese. Data analysis capitalized on normalization techniques for pitch and duration in order to allow direct comparison of individual speakers. The results revealed new insight into the relative starting positions of Northern Vietnamese tones. An analysis of non-native speaker tone errors indicated particular difficulty with low falling-rising tones, as well as difficulty with tone starting positions and changes in voice quality. We elicited all speech data using a dynamic carrier sentence task in which participants produced a series of three-word utterances in response to target words that appeared individually on a computer screen in one of four colors. In this way, participants actively described a changing event, while critical components of the utterance remained constant.

1 Introduction

Descriptions of Vietnamese lexical tones consistently report differences in pitch, duration, and voice quality [1, 2, 3, 4, 5]. However, only Vu [5] provides a description in which tones have been decoupled from speakers' pitch ranges and also from the duration of target words. The first goal of this study was to replicate Vu's normalized tone trajectories for Northern and Southern Vietnamese.

Initially, the second goal of the study was to contribute to discussions of whether certain tones are breathy. Although preliminary examinations suggested breathiness for huyền and hỏi, results of spectral tilt analyses were inconclusive and therefore will not be discussed. However, comparisons of individual speakers on a normalized scale did reveal new insight into the relative starting positions of Northern tones, which have been described inconsistently in the literature.

The third goal of the study was to identify difficulties in tone production among non-native speakers. This required us to modify a traditional carrier sentence task for use with fluent native speakers and non-fluent adult learners.

2 The dynamic carrier sentence task

In order to compare tones from adult learners to those from native speakers, we used a carrier sentence task in which all speakers produced the same utterances using a constant sentence frame and a changing target word.

Studies that investigate only native speaker production have few constraints on vocabulary and utterance length. However, with adult learners, vocabulary must be familiar, and the sentence frame must be short enough for learners to remember easily. Furthermore, target words cannot occur at the end of the utterance because this position is prone to phrase-final lowering for native English speakers [6].

For the current task, speakers generated short three-word utterances, using familiar color terms, with target words in utterance medial position. While traditional tasks involve reading word lists in sentence frames [1, 3, 5], the current task required participants to name and describe each target as it appeared on the computer screen. For example, if the word *bang* appeared in blue, the speaker said *Từ bang xanh* ("the word *bang* is blue"). From target to target, the color, vowel, and tone all changed, thus requiring participants to remember the general sentence frame, but to produce novel content for each trial within a constant syntactic and focal structure. The intent of this design was to reduce some of the repetitive nature of carrier sentence tasks, while still holding the word immediately preceding the target constant to control for segmental effects and tonal coarticulation [1].

3 Descriptions of starting points for Northern Vietnamese tones

Vietnamese orthography reflects six lexical tones: ngang, huyền, sắc, nặng, hỏi, and ngã. Each tone name contains its corresponding diacritic (or none in the case of ngang). Whereas Northern Vietnamese uses a six-tone system, Southern Vietnamese uses a five-tone system in which hỏi and ngã have merged in pronunciation. In Section 6 we provide three figures for Northern and one for Southern using native speaker data normalized for pitch and duration.

Studies of Northern Vietnamese vary in their descriptions of the starting points for tones. For example, Brunelle [1] describes two points: middle of the pitch range for ngang, hỏi, ngã, and nặng, and a position lower than ngang for sắc and huyền. Nguyen and Edmondson [7] describe a different pair of starting points: middle of the pitch range for all tones except huyền, which starts lower. Pham [3] describes three points: ngang, hỏi, and nặng as higher than huyền and ngã, which, in turn, are higher than sắc. With Vu [5], a figure summarizing the Northern data suggests that ngang starts higher than sắc and ngã, which starts higher than nặng, which starts higher than huyền and hỏi. However, Vu ultimately argues for two levels: ngang, ngã, and sắc starting higher than nặng, huyền, and hỏi.

<table>
<tr><th></th><th>Brunelle</th><th>N & E</th><th>Pham</th><th>Vu</th></tr>
<tr><td>Level 1</td><td rowspan="2">ngang, hỏi, ngã, nặng</td><td rowspan="2">ngang, hỏi, ngã, nặng, sắc</td><td rowspan="2">ngang, hỏi, nặng</td><td>ngang</td></tr>
<tr><td>Level 2</td><td>ngã, sắc</td></tr>
<tr><td>Level 3</td><td rowspan="2">sắc, huyền</td><td rowspan="2">huyền</td><td>huyền, ngã</td><td>nặng</td></tr>
<tr><td>Level 4</td><td>sắc</td><td>huyền, hỏi</td></tr>
</table>

Table 1 Summary of previous findings regarding starting points for Northern Vietnamese tones

Such variation is not surprising given that tone studies vary in their primary research questions, elicitation methods, speakers, and methods of analysis. However, one would expect generalizations to emerge about the trajectories of tones, including their relative starting and ending points. In this study, we apply pitch and duration normalization techniques to individual speakers and compare them on the same scale. This leads to the generalization for Northern Vietnamese (while excluding the two forms that appear in stop-final syllables) of a two-tier system in which ngang starts higher than all other tones, and the relative starting points of all remaining tones vary idiosyncratically.

4 Methods

4.1 Participants

Native speaker participants included 3 Northern dialect speakers (1 female, 2 male) and 1 Southern dialect speaker (female). All were originally from Vietnam and had been living in an English-speaking country for 6 to 26 years. They ranged in age from 42 to 64, and all had experience teaching Vietnamese as a foreign language to adults. Non-native speaker participants included 3 Northern dialect learners (all male) and 3 Southern dialect learners (1 female, 2 male). They ranged in age from 30 to 50. All had been studying Vietnamese intensively (i.e., at least 5 hours a day), but for varying lengths of time. Their weeks of training ranged from 14 to 32. All 10 participants resided in metropolitan Washington, DC, at the time of recording.

4.2 Stimuli

Targets comprised 102 real words and used 8 monophthongs: *i, ư, u, ơ, ô, a, â, ă*. Six vowels appeared with all possible tones for each of three syllable types: open (e.g., *ba, bà, bạ, bá, bả, bã*), stop-final (e.g., *bạt, bát*), and nasal-final (e.g., *bang, bàng, bạnh, báng, bảng, vãn*). Consistent with Vietnamese phonology, two vowels (*â* and *ă*) appeared with all possible tones in stop-final and nasal-final syllables only. To the extent possible, we matched targets for initial and final segments within syllable type and within vowel. We attempted to maintain consistent consonant place and manner, but, when necessary, sacrificed one or both in the interest of ensuring that all target stimuli were real words. Speakers were recorded in a sound-dampened room using Sound Forge 7.0 (22 kHz, 16 bit, mono), a Yamaha 01V96 digital mixing console with no effects settings, and a Neumann TLM 103 microphone.

4.3 Procedure

Participants produced 3-word sentences in response to individual target words that appeared on a computer screen in red, blue, black, or purple. For example, if the target word *bang* appeared in blue, the speaker said *Từ bang xanh* ("the word *bang* is blue"). Participants had access to the written color names as they completed 8 practice trials and then two lists of words. Each list contained all 102 targets, which were pseudo-randomized such that the vowel, tone, and color of the word always changed from one trial to the next. Four additional targets occurred on each list. Three were non-adjacent repetitions of existing targets but in a narrow contrastive context (i.e., in the same color as the immediately preceding word). This added one token each of *i, ư, u, ơ, ô,* and *a* to a separate vowel analysis (presented as a second paper in these proceedings), but none were included in the current tone analyses. The fourth addition (*ma*) occurred in list-final position and was never included in analyses. Targets that were paired with *xanh* and *tím* (purple) on List 1 were paired with *đen* (black) and *đỏ* (red), respectively, on List 2, and vice versa. Participants thus produced two repetitions of each target word, with each repetition produced within a novel utterance. In this self-paced task, participants could repeat any utterance before advancing to the next word. When speakers did repeat, we analyzed only the final repetition. In terms of the current analysis, each speaker produced 12 tokens of each tone in open syllables, 16 tokens of each tone in nasal-final syllables, and 18 tokens each of sắc and nặng in stop-final syllables.

5 Analysis

Using Praat [81], we marked tone region onsets and offsets based on auditory and visual inspection of each waveform and spectrogram. The beginning of the tone region coincided with vowel onset. The end of the tone region coincided with the end of vowel production in the open and stop-final syllables, and with the end of nasal production in the nasal-final syllables. Praat scripts automatically assigned nine evenly spaced points between each onset and offset to create time steps within the tone region in 10% increments. Scripts automatically returned the value in Hertz of the fundamental frequency (F0) or "pitch undefined" at all eleven points. We reviewed the output for possible pitch tracking errors and replaced suspicious values with hand measurements or "pitch undefined" as appropriate. This resulted in modifications to 3% of the data over time steps 10 – 90%. We excluded the first time step (0%) from analysis to exclude effects of initial segments (all non-nasal obstruents) on the tone contour. We excluded the final time step (100%) to exclude segmental effects in stop-final syllables and because pitch was generally undefined this late in the word in the open and nasal-final syllables. Following the procedure in Nolan [9], we converted Hertz to semitones using each speaker's average pitch as his or her baseline. We then calculated means for each speaker for each time step separately for the three syllable types.

6 Native Speaker Tone Results

6.1 Northern tones

In the native speaker data, we graphed means for every time step at which two thirds or more of data cells were defined. That is, when pitch was undefined in a minimum of 4 of the 12 open syllable targets or 5 of the 16 nasal-final syllable targets, at a given time point, the tone trajectory contained a gap. As shown in Figures 1 – 3, this gap appeared in the Northern speaker's nặng and ngã tones, which are known to have glottalization. The trajectory for nặng truncated about a third of the way into the tone region. The trajectory for ngã showed the expected gap in roughly its middle third.

The three Northern speakers each produced a mean falling-rising hỏi trajectory. Although there was some variation in whether the mean rise extended above or below huyền, none of the mean trajectories resembled hỏi's known falling variant [1, 2, 5, 6]. Speakers 08 and 09 did, however, show the falling variant on a few individual utterances.

Figures 1 – 3 (and Fig. 5) reflect data from open syllables only, but we observed similar patterns in the nasal-final data. Stop-final data showed the expected two-way contrast between two modal tones: a rising sắc and falling nặng [2]. Figure 4 provides a representative example. Note that stop-

final syllables are characteristically short [3, 5], a property that is not apparent when tones are normalized for duration. The mean tone region durations for sắc and nặng, respectively, averaged across long vowel targets from the native Northern speakers, were 366 ms and 284 ms in open syllables, but 152 ms and 148 ms in stop-final syllables.

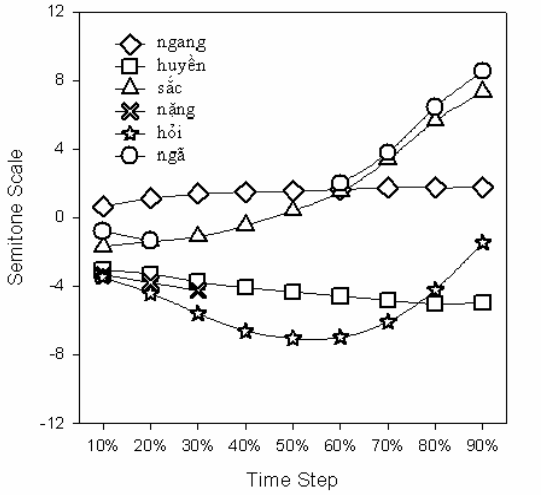

Fig. 1 Native Northern Speaker 01 open syllables

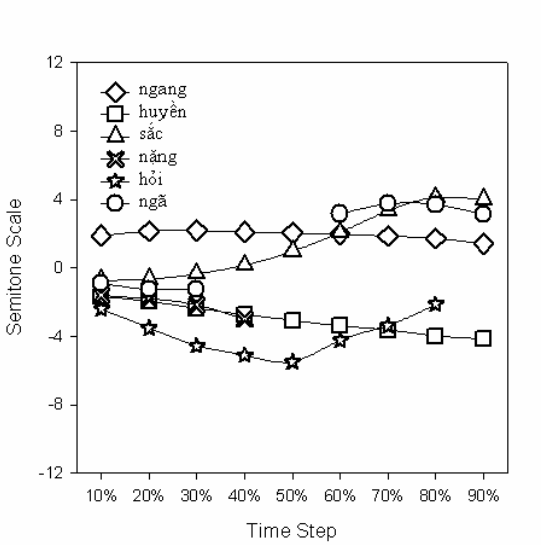

Fig. 2 Native Northern Speaker 08 open syllables

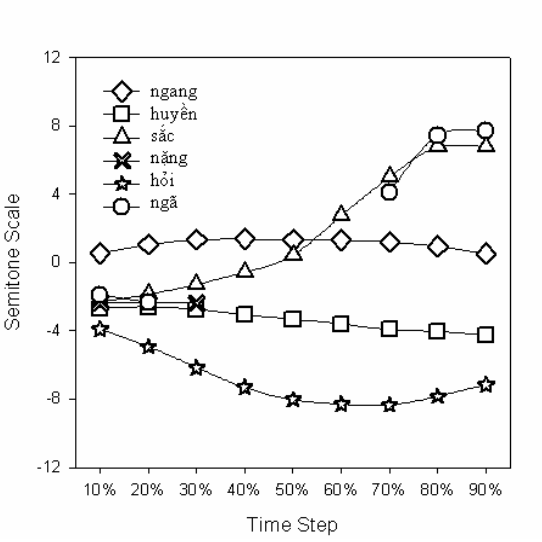

Fig. 3 Native Northern Speaker 09 open syllables

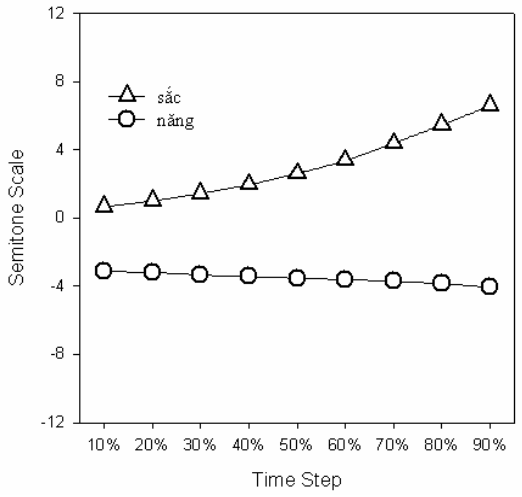

Fig. 4 Native Northern Speaker 01 stop-final syllables

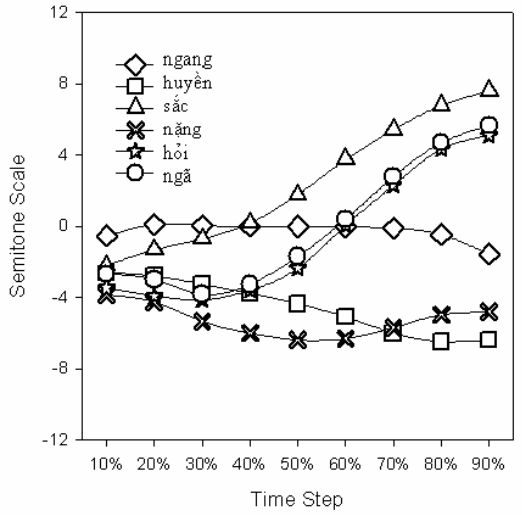

Fig. 5 Native Southern Speaker 10 open syllables

6.2 Southern tones

Figure 5 demonstrates the five-tone system of Southern Vietnamese in which hỏi and ngã have merged. This merger separates the sắc trajectory from the hỏi/ngã trajectory (whereas sắc shares a trajectory with ngã in Northern). The merger also provides space for nặng to shift to Northern hỏi's tone space and trajectory. This is consistent with reports that Southern nặng lacks the distinct glottalization that truncates the tone trajectory in the Northern form [4]. With one exception, the tone trajectories from our native Northern and Southern speakers replicated Vu 5]. The exception is that Vu's Southern sắc and ngang tones shared a starting point higher than the remaining tones. In contrast, our native Southern speaker started sắc below ngang. We did, however, observe Vu's pattern for one learner of Southern (Fig. 10).

6.3 Northern tone starting points

With respect to the six basic tones (i.e., setting aside those from stop-final syllables), two consistent starting points emerged for Northern: ngang vs. all remaining tones. All

three native speakers started ngang in highest position. Within the remaining tones, there was merely a general trend: immediately rising tones (i.e., sắc and ngã) showed origins near the top, while the tone leading to the lowest midpoint (i.e., hỏi) originated near the bottom. This trend is consistent with Vu's [5] figure of normalized Northern data. However, looking across the current three Northern speakers, no one rank order covers all non-ngang tones.

7 Non-native speaker tone results

7.1 Excluded data

With the non-native speaker data, we were primarily interested in tone production that was consistent and that approximated native speaker targets. For this reason, we first excluded obvious errors (e.g., a rising tone among otherwise level ngang tones). We then excluded forms that seemed inconsistent with the majority of productions, even if the majority did not approximate native speaker targets. Table 1 shows the number of exclusions for open syllables. (Exclusion data with nasal-final syllables revealed no consistent patterns across tones or learners.)

	ngang	huyền	sắc	nặng	hỏi	ngã
02 N	4	0	1	4	0	2
03 N	4	3	2	5	3	2
04 S	6	0	0	0	0	0
05 S	2	1	0	2	0	0
06 S	3	2	0	0	1	0
07 N	3	0	4	1	0	4

Table 1 Number of excluded tokens (maximum of 12 per cell) for open syllables by tone for adult learners (02 to 07) of Northern (N) and Southern (S) Vietnamese

7.2 Three common learner difficulties

Each of the adult learners showed a unique pattern of tones. Despite these idiosyncratic systems, there were several common problems. First, none of the adult learners produced native-like Northern hỏi or Southern nặng contours. Native speaker productions showed two characteristics: they dipped below huyền and never rose past ngang. Figure 6 shows an adult learner who succeeded with the first characteristic, but not the second. Two Northern learners (Figs. 7 & 11) and one Southern learner (Fig. 8) produced rises resembling ngã. Two Southern learners (Figs. 9 & 10) produced no rise at all. While the absence of a rise is a possible form for nặng, the speakers failed to produce the necessary dip below huyền.

Second, several learners failed to approximate native speaker starting points. One Southern learner (Fig. 8) started all tones from a central location. One Northern learner (Fig. 6) produced sắc with ngang's native speaker starting point, pitch height, and shape (and produced a lower parallel ngang). Two other learners may have adopted different strategies. Speaker 03 (Fig. 7) appeared to start low for tones that rise and high for tones that fall, while Speaker 05 (Fig. 9) appeared to start low for tones that fall and high for tones that rise.

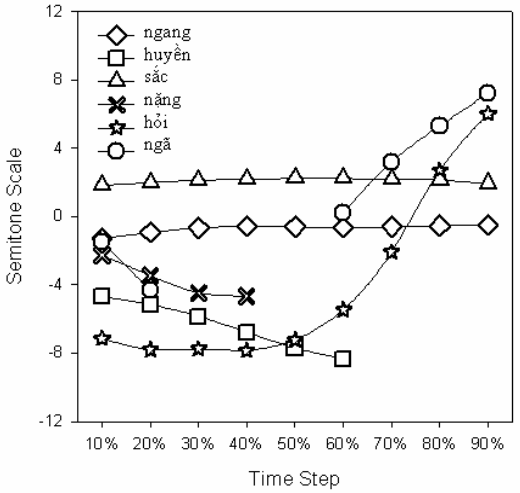

Fig. 6 Adult Northern Learner 02 open syllables

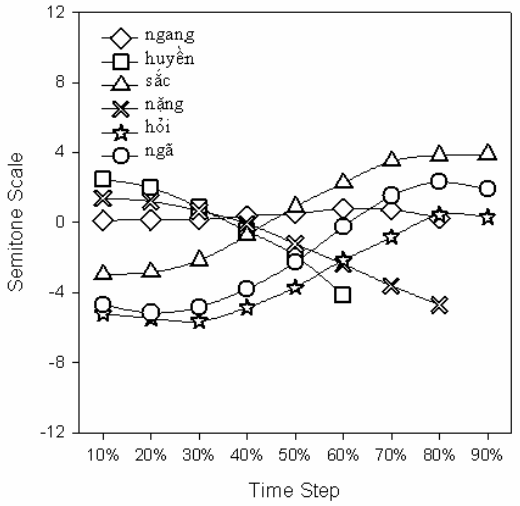

Fig. 7 Adult Northern Learner 03 open syllables

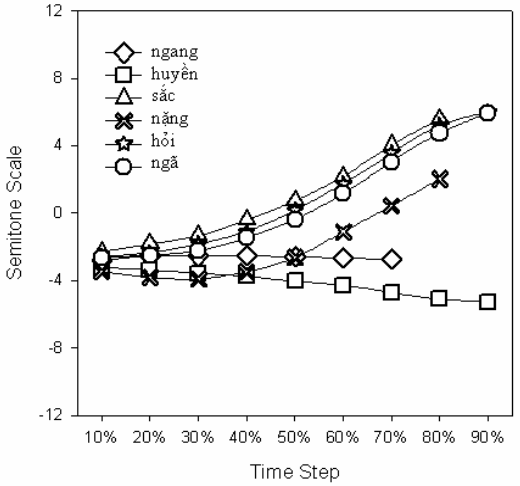

Fig. 8 Adult Southern Learner 04 open syllables

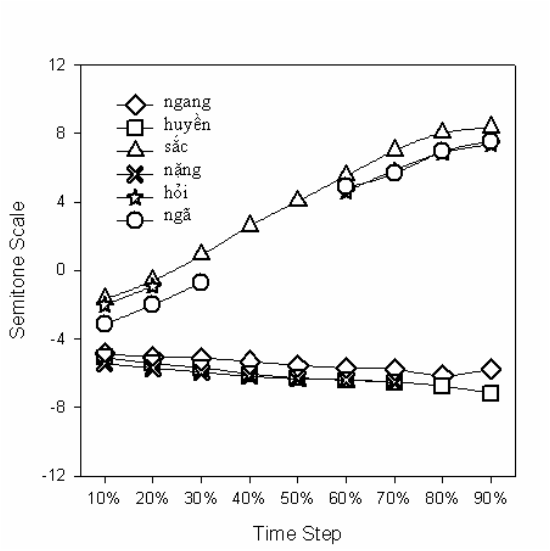

Fig. 9 Adult Southern Learner 05 open syllables

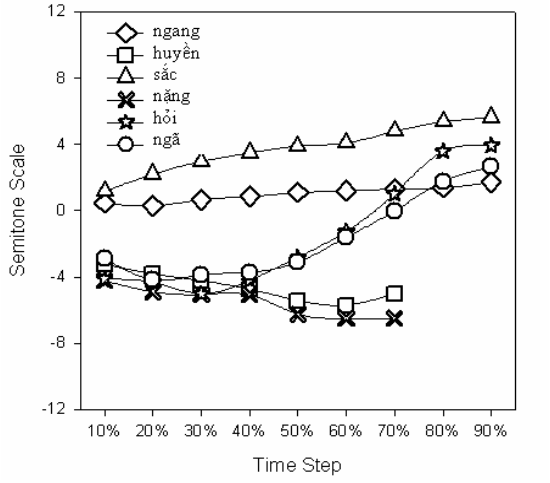

Fig. 10 Adult Southern Learner 06 open syllables

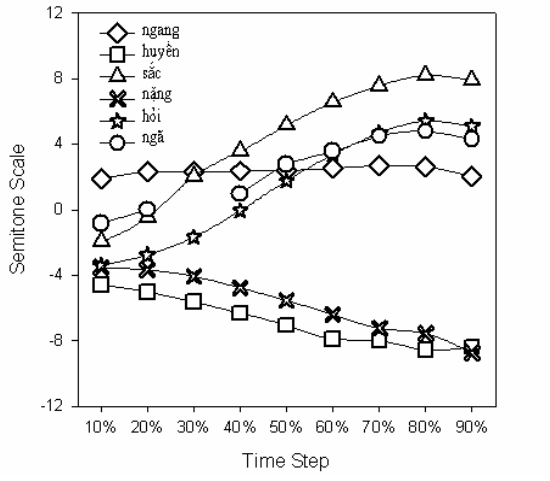

Fig. 11 Adult Northern Learner 07 open syllables

Third, several learners produced non-native patterns of glottalization. Whereas the ngã trajectories for two Northern learners were appropriately broken (Figs. 6 & 11), the trajectory was continuous for the third (Fig. 7). At the same time, the speakers in Figures 7 and 11 did not produce glottalized Northern nặng tones.

One of the Southern learners (Fig. 9) also produced ngã/hỏi with distinct glottalization. The native Southern speaker, however, distinguished sắc from ngã/hỏi with a difference in tone trajectory (Fig. 5). While the Southern learner in Figure 8 also did not produce a separate sắc trajectory, there was no Northern-style glottalization to distinguish the tones.

8 Conclusion

This study generally replicated the Northern and Southern tone trajectories from Vu [5]. However, by analyzing *individual* speakers on the same normalized scale, the current study revealed a new generalization about starting points of Northern Vietnamese tones, and identified common problems in seemingly dissimilar adult learner tone systems.

Acknowledgments

We gratefully thank our speakers for their participation.

References

[1] M. Brunelle, *Coarticulation effects in Northern Vietnamese tones*. Unpublished manuscript, Cornell University (2003)

[2] A. Michaud, "Final consonants and glottalization: New perspectives from Hanoi Vietnamese", *Phonetica*, 61:119-146 (2004)

[3] A. Pham, *Vietnamese Tone: A New Analysis*. New York: Routledge (2005)

[4] L. Thompson, *A Vietnamese Reference Grammar*. Hawaii: University of Hawaii (1965)

[5] P. Vu, *The Acoustic and Perceptual Nature of Tone in Vietnamese*. Unpublished doctoral dissertation, Australian National University (1981)

[6] M. Liberman, J. Pierrehumbert, "Intonational invariance under changes in pitch range and length", In M. Aronoff, R. Oehrle (Eds.), *Language Sound Structure*. MIT Press: Cambridge, 157-233 (1984)

[7] V. Nguyen, J. Edmondson, "Tones and voice quality in modern northern Vietnamese: Instrumental case studies", *Mon-Khmer Studies*, 28:1-18 (1998).

[8] P. Boersma, D. Weenink, *Praat: Doing Phonetics by Computer* (Version 4.4.28)

[9] F. Nolan, "Intonational equivalence: An experimental evaluation of pitch scales", *Proc. 15th ICPhS,* Barcelona, 771-774 (2003)