NORTHERN VIETNAMESE PERCEPTION OF NON-NATIVE TONES

Allison Blodgett, Alina Twist, Jessica Bauman, Anita Bowles, Melissa K. Fox, Phuongthao Luu, C. Anton Rytting, Jessica Shamoo Marx & Matthew B. Winn

Center for Advanced Study of Language, University of Maryland, USA
ablodgett@casl.umd.edu

ABSTRACT

We investigated native speaker perception of adult learner pronunciations of Northern Vietnamese *hỏi* tone contours to examine how listeners prioritize acoustic cues when they expect non-native speech. The non-native contours consisted of two adult learner renditions of the low falling-rising *hỏi* tone with sweeping final rises similar to *sắc* and *ngã*. One fell only as low as *huyền*, whereas the other began lower than *huyền*, as opposed to falling to a lower midpoint as in native speaker speech. We created two additional non-native renditions by crossing these contours with mid-tone creakiness.

Consistent with [4], listeners in a speeded tone identification task prioritized voice quality and F0 offset (or rise slope) over F0 onset. Results also suggested that listeners adjusted their perceptions of tones as they encountered more stimuli. Because listeners required a self-paced task in order to use F0 onset, adult learners would likely improve their intelligibility by attenuating the F0 offset (or rise slope) in their productions of *hỏi*. Learners may also be able to use the strength of the voice quality cue to overcome otherwise non-native F0 contours.

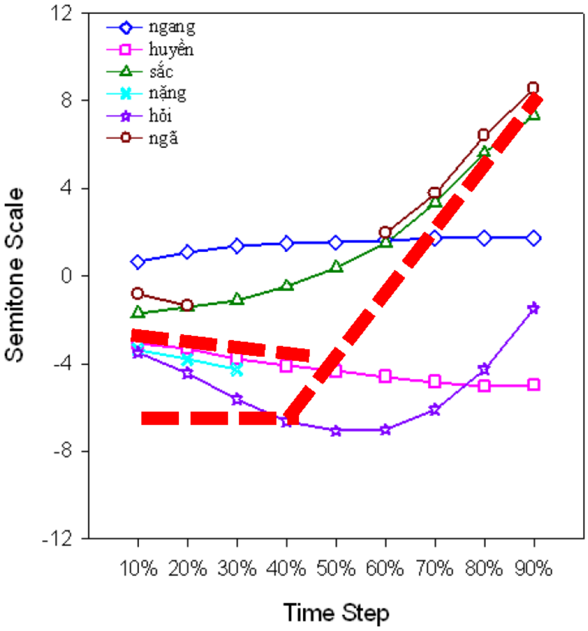
Keywords: Vietnamese, lexical tone, perception, second language acquisition

1. INTRODUCTION

Vietnamese orthography reflects six lexical tones: *ngang, huyền, sắc, nặng, hỏi,* and *ngã*. Each tone name contains its corresponding diacritic (or, in the case of *ngang*, no diacritic). Figure 1 illustrates the six tones of Northern Vietnamese using data from one individual. Each contour is an average of 12 open syllable tokens, and the tone system is representative of the four speakers analyzed [1].

Northern Vietnamese tones are distinguished by pitch, duration, and voice quality [3, 5, 8, 9, 11, 12], and the relevant contrasts are apparent in Fig. 1. Pitch is represented in semitones on the y-axis. Contrastive voice quality is represented by gaps in the tone trajectories for *ngã* and *nặng*. These gaps occur because creaky voice interferes with measurements at these points in the majority of tokens. In Fig. 1 *ngã* shares the same pitch contour as *sắc*, and the creakiness in *ngã* differentiates the two. Similarly, the contours for *nặng* and *huyền* align, until creakiness and duration make them distinct. Because pitch measurements have been taken at regular intervals within tone regions to normalize variation across tokens, duration is not technically a feature of Fig. 1. Fortunately, however, the creaky gap at the end of *nặng* correctly implies that this tone is contrastively shorter than the others [1, 12].

Figure 1: The dashed lines superimposed on the Northern Vietnamese tone system represent schematized versions of two adult learner contours for Northern *hỏi*. *Higher student hỏi* shares its F0 onset with the other low tones. *Lower student hỏi* starts lower. Both tones share an F0 offset with *sắc* and *ngã*.



In this study we investigated native speaker tone perception with a focus on non-native speaker pronunciations of *hỏi*. Little, if anything, is known about the extent to which native speakers perceive foreign accented Vietnamese as intended. While there are a few empirical studies of Vietnamese tone perception [4, 6, 7, 12], none specifies a particular listening context, and listeners may be assuming a native speaker environment. Given that non-auditory information, including sociolinguistic expectations, affects speech perception [10], a non-native speaker environment provides an alternative context in which to examine tone perception.

2. KEY STIMULI AND PREDICTIONS

For non-native contours, we began with two actual student contours for *hỏi*. These contours are schematized as the dashed lines in Fig. 1. *Higher student hỏi* shares an F0 onset with the other low-falling tones, *nặng* and *huyền*. *Lower student hỏi* starts lower. Both tones share an F0 offset with the rising tones, *sắc* and *ngã*. Because the student was male, we multiplied both sets of F0 values by 1.63 Hz to bring them into the pitch range of the female native speaker who produced the underlying syllables. We raised the F0 values of the original *higher student hỏi* by 50 Hz to align the contour more closely with *huyền* in order to approximate productions from other students. We created two additional non-native renditions by crossing these contours with mid-tone creakiness.

We base our predictions on Brunelle's [4] finding that native speakers prioritize voice quality and slope of the F0 rise over pitch contour (simple vs. complex) and F0 offsets. In his experiments with a wide range of synthetic contours, F0 onset did not emerge as a primary acoustic cue. Because our four non-native tones differed in voice quality (modal vs. mid-tone creakiness) and F0 onset (lower or equal to *huyền*), we predicted that listeners would treat the modal versions as *sắc* and the creaky versions as *ngã*, regardless of F0 onset.

3. SELF-PACED WEB-BASED LISTENING

3.1. Method

3.1.1. Participants

We present results from 20 self-reported native speakers of Vietnamese who were most comfortable speaking and listening to Northern Vietnamese. This Northern dialect group ranged in age from 36 to 80 and had spent 13 to 16 years living in Vietnam. 18 reported living in Vietnam at time of testing. One reported experience speaking or listening to English. Four reported experience listening to Vietnamese spoken by foreigners. They received no compensation or incentives.

3.1.2. Stimuli

The stimuli consisted of 12 nonsense syllables. Using Praat [2] we replaced the F0 of a native speaker's productions of "thôm" with the F0 from a non-native speaker, a different native speaker, or in one case, an artificial rising-falling-rising contour (intended to elicit "cannot determine" responses). We resynthesized new contours onto productions that matched in tone with the following exceptions. We resynthesized the native and non-native renditions of *hỏi* once onto *huyền* (for modal voicing) and once onto *nặng* (for medial creakiness). We could not resynthesize modal versions of *hỏi* onto the native speaker's production of *hỏi* (*thổm*) because she consistently produced *hỏi* with medial creakiness. We used *nặng* for versions of *hỏi* with medial creakiness because it resembled *huyền* in intensity as the speaker resumed modal voicing tone finally.

3.1.3. Procedure

We conducted the experiment in Vietnamese. Participants visited a website that described the study, provided informed consent, and indicated the need for speakers or headphones. Participants who indicated they were native speakers at least 18 years old progressed to demographic questions. The instructions indicated that they would be listening to single syllables that had no meaning in Vietnamese and that the speaker would be a foreign learner of Northern Vietnamese practicing her tones. The instructions asked participants to select the spelling that sounded closest to each syllable or "cannot tell." The page provided an audio clip for adjusting volume before initiating three practice trials (tokens with native *ngang*, *sắc*, and *ngã* contours in that order). The remaining 42 trials appeared in random order with the constraint that no two identical stimuli could appear back-to-back. Participants could repeat the practice set as many times as they liked, but they could play any individual trial no more than twice.

We expected non-native contours to be identified as *hỏi*, *sắc*, or *ngã*, and we presented an equal number (3) of the native contours (counting modal and creaky *hỏi* separately) and the (modal or creaky) non-native contours. See the *Tokens* column in Table 1, which sums tokens played across listeners, including practice trials.

Table 1: Summary of responses by experiment. The most frequent responses per stimulus are in bold. The number of non-responses in the Listening Under Time Pressure experiment equals Tokens minus the total number of responses per line.

	Responses							
Stimulus Contour	*th ôm (ngang)*	*thồm (huyền)*	*thốm (sắc)*	*thốm (nặng)*	*thổm (hỏi)*	*thỗm (ngã)*	*Cannot tell*	Tokens
			Self-Paced Web-Based Listening					
Native *ngang*	**114**	0	1	0	0	0	0	115
Native *huyền*	3	**100**	0	0	2	0	1	106
Native *sắc*	3	1	**57**	0	0	0	1	62
Native *nặng*	1	3	0	**102**	1	1	1	106
Native *hỏi* (modal)	2	5	0	3	**44**	0	0	54
Native *ngã*	0	1	0	0	7	**53**	1	62
Native *hỏi* (creaky mid-tone)	0	1	1	**39**	11	2	0	54
Higher student *hỏi* (modal)	0	0	**37**	0	9	7	0	53
Higher student *hỏi* (creaky mid-tone)	0	0	0	1	4	**49**	0	54
Lower student *hỏi* (modal)	0	0	13	0	**36**	3	1	53
Lower student *hỏi* (creaky mid-tone)	0	0	0	2	13	**40**	0	55
Rising-falling-rising	**43**	0	5	0	1	0	3	52
			Listening Under Time Pressure					
Native *ngang*	**114**	0	3	1	0	1	0	120
Native *huyền*	3	**130**	1	0	5	2	1	144
Native *sắc*	1	3	14	0	1	**28**	1	48
Native *nặng*	1	3	0	**103**	27	7	3	144
Native *hỏi* (modal)	0	5	0	3	**58**	5	1	73
Native *ngã*	0	0	0	1	4	**42**	0	48
Native *hỏi* (creaky mid-tone)	0	0	0	**36**	28	3	4	72
Higher student *hỏi* (modal)	0	3	**58**	0	3	5	1	72
Higher student *hỏi* (creaky mid-tone)	0	0	3	4	4	**60**	1	72
Lower student *hỏi* (modal)	1	2	**39**	0	14	13	2	72
Lower student *hỏi* (creaky mid-tone)	0	0	3	0	23	**44**	2	72
Rising-falling-rising	**41**	2	9	1	1	2	12	71

3.2. Results

The upper half of Table 1 summarizes the distribution of responses across stimuli. Listeners accurately identified the six primary native speaker contours, suggesting that the web-based data are valid. Contrary to predictions, however, listeners did not treat both modal versions of *student hỏi* as *sắc*. Although they did tend to identify the creaky versions of *student hỏi* as *ngã*, they tended to identify modal *higher student hỏi* as *sắc* and modal *lower student hỏi* as *hỏi*. This pattern suggests that in an offline judgment task listeners may prioritize F0 onset information over F0 offset. (A low tone-final amplitude may explain why listeners tended to identify native *hỏi* with mid-tone creak as *nặng*.)

4. LISTENING UNDER TIME PRESSURE

4.1. Method

4.1.1. Participants

38 university students participated in Hanoi in exchange for a non-monetary token of appreciation. Most reported always being most comfortable speaking and listening to Northern Vietnamese. Six listed more complicated dialect backgrounds, but each was still most comfortable speaking or listening to Northern Vietnamese at time of testing. Participants ranged in age from 18 to 25 and had always lived in Vietnam. 18 reported limited experience speaking or listening to English, and 22 reported experience listening to Vietnamese spoken by foreigners. We excluded data from eight participants for which the response window was erroneously set to two seconds instead of six and

from three participants with a response rate below 80%. Three additional data files were lost.

4.1.2. Stimuli

We modified the stimuli from the web-based experiment for use with speeded judgments. Although we had recorded the sound files in a sound booth in a single session, we further normalized them to equate peak sound level (-4.00 dB, SoundForge 6.0). We applied a 15 ms fade-in and fade-out to ensure an absence of transients, as well as 100 ms of silence at onset and offset.

4.1.3. Procedure

Participants began with six practice trials in a fixed order (native *ngang*, *sắc*, and *ngã* contours) before completing 42 additional trials in one order or its reverse. These trials consisted of blocks of 15, 15, and 12 trials separated by instructions to take a short break. Participants resumed when ready. Trials consisted of a one-second fixation cross followed by simultaneous presentation of the sound file and response options. Participants had a limit of six seconds to respond from the onset of the sound file before the next trial began. Providing a response initiated the next trial.

4.2. Results

The lower half of Table 1 summarizes the distribution of responses across stimuli. Consistent with predictions, listeners tended to identify the modal versions of *student hỏi* as *sắc* and the creaky versions as *ngã*, regardless of F0 onset. Although the native *sắc* contour elicited twice as many *ngã* responses as *sắc* responses, the practice trials tended to be identified as *sắc*. This suggests that listeners adjusted their perceptual thresholds over the course of the experiment.

5. GENERAL DISCUSSION

An F0 onset that was lower than *huyền* could cue listeners to perceive *hỏi* in spite of a high F0 offset unless listeners were making speeded judgments. Under time pressure the contrast between modal voice and medial creakiness on non-native rising contours elicited *sắc* and *ngã* judgments, respectively. The contrast in voice quality for the non-native contours may have led listeners under time pressure to treat these contours as *sắc* and *ngã*, respectively. If modal *lower student hỏi* had appeared by itself, it may have been more likely to elicit *hỏi* judgments. Evidence against this, however, comes from an investigation of unaltered student speech, demonstrating that modal *lower student hỏi* (without a creaky counterpart) elicited 64% *sắc* judgments and 22% *hỏi* judgments.

6. CONCLUSIONS

The fact that listeners required a self-paced task to use F0 onset as a tonal cue suggests that adult learners would likely improve their intelligibility by attenuating the F0 offset, which in this case would also attenuate the slope of the F0 rise, in their productions of Northern *hỏi*. Learners may also be able to use the strength of the voice quality cue to overcome otherwise non-native F0 contours for *sắc* and *ngã* and by extension *huyền* and *nặng*.

7. REFERENCES

[1] Bauman, J., Blodgett, A., Rytting, C., Shamoo, J. 2009. *The Ups and Downs of Vietnamese Tones: A Description of Native Speaker and Adult Learner Tone Systems for Northern and Southern Vietnamese* (Tech. Rep. No. E.5.3 TTO 2118). College Park, MD: University of Maryland Center for Advanced Study of Language.

[2] Boersma, P., Weenink, D. 2008. *Praat: Doing Phonetics by Computer* (Version 4.4.28).

[3] Brunelle, M. 2003. *Coarticulation effects in Northern Vietnamese Tones*. Unpublished manuscript, Cornell University.

[4] Brunelle, M. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37, 79-96.

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

[6] Mixdorff, H., Nguyen, H., Fujisaki, H., Lương, M. 2006. Quantitative analysis and synthesis of syllabic tones in Vietnamese. *Proc. of Eurospeech* 177-180.

[7] Nguyen, D., Kenny, D. 2009. Effects of muscle tension dysphonia on tone phonation: Acoustic and perceptual studies in Vietnamese female teachers. *Journal of Voice* 23(4), 446-459.

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

[9] Pham, A. 2003. *Vietnamese Tone: A New Analysis*. New York: Routledge.

[10] Strand, E. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18, 86-100.

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

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