SOCIAL EFFECTS ON THE PERCEPTION OF VIETNAMESE TONES*

Marc Brunelle and Stefanie Jannedy

University of Ottawa and Center for General Linguistics (ZAS Berlin)
marc.brunelle@uottawa.ca and jannedy@ling.ohio-state.edu

ABSTRACT

The rate of correct identification of tones in Vietnamese is influenced by the dialect of the stimuli to which the hearer is exposed (northern vs. southern). However, sociophonetic factors such as the dialect of the person administering the experiment (northern vs. southern) and, by extension, accommodation via length of exposure to the experimenter also play a role. Our results indicate that listeners adjust their interpretation of some tone-stimuli in accordance with the dialect of the person administering the experiment, strongly suggesting that both perceptual cues contained in the signal and also inferred social factors play a role in the categorization of tones in Vietnamese.

Keywords: Vietnamese, Tone, Perception, Cue shifting, Sociophonetic factors

1. INTRODUCTION

It has been shown that social information biases the perception and rating of naïve listeners [1, 2]. For example, the vowel merger between *near* /iə/ and *square* /eə/ in New Zealand English, which is more advanced in the working class, was identified more often when a picture of a supposed speaker in working class attire was shown in contrast to the same person wearing formal clothes [1]. In [2] a crossover effect was found whereby identical stimuli on a generated continuum from [sod] to [ʃod] were rated differently depending on the co-presentation of a male vs. female face.

The aim of the present study is to evaluate the effect of implicit social information on the identification of Vietnamese tones. Northern Vietnamese (NVN) has 6 tones and combines complex pitch contours with voice quality distinctions, whereas the distinctions between the 5 Southern Vietnamese (SVN) tones rely exclusively on pitch [3,4,5,6,7]. An important difference between these dialects is that the tones *C1* and *C2* have merged in SVN. The social information was implicitly introduced by two experimenters who conversed with the experiment participants in either SVN or NVN. Our hypothesis was that the identification ratings by the SVN speakers should be biased by the dialect of the experimenter.

Figures 1 and 2 illustrate the mean f0 values of the five utterances of each tone used as stimuli for this experiment as produced by a female (NVN) and by a male (SVN) speaker (measured on the vowel /a/).

Fig.1: The tone system of the northern speaker (NVN)

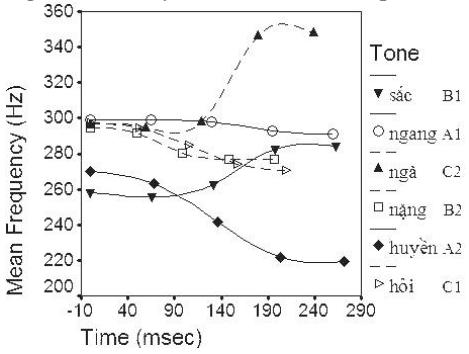
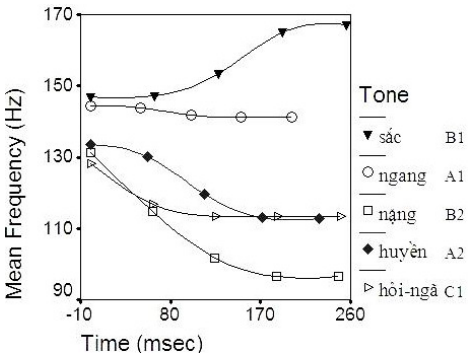


Fig.2: The tone system of the southern speaker (SVN)



The tones of the NVN speaker are representative of the standard Hà Nội dialect. The only exception is the rising tone *B1*, which is very low, a variant found in young female Northerners. The three NVN tones with a laryngealized voice quality are represented by a dotted line in Fig.1 (laryngealization is tone-medial in *C2* and tone-final in *C1* and *B2*). Laryngealization is much stronger in *B2* than *C1* and the latter is reported to be either breathy or laryngealized [5]. The contrast between these tones is based on these fairly salient differences.

The exact acoustic cues playing a role in the categorization of these tones have only been partly investigated [4,5,8], but it is clear that listeners tend to misidentify tones from other dialects more often than tones from their own dialect. More interestingly, it seems that despite dramatic differences in the acoustic realization of the tones across dialects, listeners are adjusting their perception to the dialect of the speaker to some degree. The aim of the present study is to evaluate the effect of the dialect (or the perceived dialect) of the speakers and experimenters on tone identification.

2. METHODS

Forty native listeners of SVN were asked to identify natural tones produced by two native speakers of NVN and SVN, respectively. A NVN experimenter administered the test to half of the subjects while a SVN experimenter administered the test to the other half.

2.1. Stimuli

Recordings of the Vietnamese stimuli were made with two speakers of NVN and SVN. The NVN speaker is a 25 year old woman who had been living in Hà Nội for eight years at the time of recording, while the SVN speaker is a 23 year old man who had been living in Hồ Chí Minh City for 4 years. Their speech is representative of the standard northern and southern dialects.

Ten recordings of each tone were made with a Marantz PMD-680 recorder and a AKG-C5900 microphone in Hà Nội and Hồ Chí Minh City. The speakers were asked to read the syllable /ma/ with the six tones (five for SVN) in a frame sentence where the target word was preceded and followed by words with the level tone *A1* so as to control for tonal coarticulation effects.

(1) *Tôi chào sư __ xem ông ấy có nhớ tôi không.*
/toj caw sɨ __ xɛm oŋ əj kɔ ɲə toj xoŋ/
"I greet monk __ to see if he remembers me"

Five utterances of each tone were selected. The target words /ma/ in these sentences (30 for NVN; 25 for SVN) were extracted to serve as stimuli (cf. audio files) for the identification task. Although the six syllables used as stimuli correspond to real words that have different lexical frequencies, frequency effects were not controlled for lack of a reliable frequency database.

2.2. Subjects

The experiment was administered to 20 male and 20 female college students in Hồ Chí Minh City, all native speakers of the SVN dialect, born and raised in the South.

2.3. Procedure

An identification task was administered in Praat 4.4.16 on a laptop with Sennheiser headphones. Each of the 30 natural stimuli recorded from the NVN speaker were played five times in isolation and in randomized order, for a total of 150 tokens. The subjects had to click on one of six boxes containing the six possible answers (in Vietnamese spelling) to identify the word they had just heard. The next stimulus was then played. This procedure was conducted for 5 repetitions of the 25 SVN stimuli, for a total of 125 tokens.

Two linguistically-aware female experimenters conducted the experiment independently with 20 subjects each. The first experimenter is a 26 year old native speaker of Hà Nội Vietnamese while the second one is a 31 year old SVN speaker with a perfect command of the Hồ Chí Minh City standard (spoken during the experiment). With each experimenter, 10 subjects carried out the task with the NVN stimuli, followed by the SVN stimuli, while 10 subjects were presented with the two blocks in the opposite order. The subjects had to take a five-minute pause between the two blocks during which they talked to the experimenter. Thus, three variables were controlled for in the procedure: 1. dialect of stimuli (SVN/NVN), 2. dialect of the experimenter (SVN/NVN), 3. order of presentation of the two blocks of stimuli (SVN first/NVN first).

2.4. Statistical Analysis

The identification results were analyzed using a General Linear Model (GLM) with a quasi-poisson distribution. The GLM was fitted to the entire set of identification ratings. The dependent variable is the number of identification mistakes made by the listeners. The fixed factors are *dialect* of the speaker, *order* of presentation, *experimenter* and *tone* of the stimulus. Interactions between the fixed factors were also included.

3. RESULTS

For lack of space, only the significant results of the GLM analysis are reported in Table 1. This model

explains 63.5% of the variance in our data. *Tone* is by far the most significant predictor ($p < 0.001$) of identification errors, accounting by itself for 51.3% of the variance. This means that some tones are better identified than others, but also that other factors and interactions vary considerably depending on the tone of the stimuli.

Table 1: Factors and interactions contributing to the number of incorrect identifications

Factor(s)/Interactions	Df	F value	p
tone	5	111.28	< 0.001***
tone*dialect	4	13.706	< 0.001***
tone*experimenter	5	3.6525	0.003**
tone*dialect*exp.	4	3.1856	0.014*
tone*exp.*order	5	3.2907	0.006**

*** < 0.001, ** < 0.01, * < 0.05

Because of the great importance of tone and of the difficulty of interpreting significant factor interactions with a 5-level variable, GLMs have been fitted independently for each tone. Significant factors for individual tones are reported in Table 2 and the estimated marginal means necessary to interpret the role of significant factors and interactions are reported in Table 3.

Tables 2 and 3 show that *dialect* affects the number of incorrect responses in the case of tones *A1*, *B1* and *C2*. However, while listeners make fewer mistakes with SVN stimuli in tones *B1* and *C2*, the opposite effect is found in tone *A1*. *Experimenter* also affects the identification of tones *B1* and *C1* in opposite directions. *B1* is better identified with the SVN experimenter, whereas C1 is better identified with the NVN one. *B1* is also better identified in the first block.

Table 2: Factors and interactions affecting the number of incorrect identifications, by tone class

Factor(s)	A1	A2	B1	B2	C1	C2
dialect	**		***			***
experimenter			**		*	
order			*			
dialect*exp.			**			
dialect*order			*			
exp.*order					*	
dialect *exp.*ord.		*				

*** < 0.001, ** < 0.01, * < 0.05

Interactions of factors also have an effect on some tones. With the SVN experimenter, SVN *B1* is much better identified than its NVN counterpart. Along the same lines, NVN *B1* is better identified with the NVN experimenter. Further, in the first block, *B1* is identified better when it is produced by the SVN speaker than by the NVN speaker. However, identification of NVN *B1* improves when it comes in the second block. As for *C1*, it shows an interaction of *experimenter* and *order*. With the NVN experimenter, *C1* is better identified in the second block. In this same second block, *C1* is more misidentified with the SVN experimenter. The last significant interaction is complex. In the second block, with the SVN experimenter, listeners are more successful at identifying NVN *A2* than SVN *A2*. Finally, SVN *A2* in the second block is better identified with the NVN experimenter.

Table 3: Sign. differences in estimated marginal means (# of identification errors/subject, out of 25)

Tone	Factor(s)	p value	E. M. Means
A1	dialect	0.001	NVN 0.55 SVN 2.875
A2	SVN exp * 2nd block * dialect	0.015	NVN dial. 0.5 SVN dial. 5.9
	SVN dialect * 2nd block * exp.	0.038	NVN exp. 1.3 SVN exp. 5.9
B1	dialect	< 0.001	NVN 3.3 SVN 0.2
	experimenter	0.007	NVN 0.65 SVN 2.85
	order	0.022	1st block 2.675 2nd block 0.875
	SVN exp. * dialect	< 0.001	NVN dial. 5.65 SVN dial. 5E-02
	NVN dialect * exp.	< 0.001	NVN exp. 0.95 SVN exp. 5.65
	1st block * dialect	< 0.001	NVN dial. 5.05 SVN dial. 0.3
	NVN dialect * order	0.003	1st block 5.05 2nd block 1.55
C1	experimenter	0.036	NVN 16.4 SVN 20.95
	2nd block * exp.	0.001	NVN exp. 12.6 SVN exp. 22.8
	NVN exp. * order	0.015	1st block 20.2 2nd block 12.6
C2	dialect	< 0.001	NVN 6.35 SVN 0.6

4. DISCUSSION

The results presented in the previous section can be better understood if we examine them tone by tone. We will start with *A1* and *A2*, which are especially difficult to interpret. The surprising result that tone *A1* is better identified when uttered by the Northern speakers even though the listeners are

Southerners is largely explained by the fact that one of the five SVN *A1* stimuli has a high f0 and is thus confused with *B1* 20% of the time. However, even if we exclude this stimulus, NVN *A1* is marginally better, which could be explained by idiosyncracies of the SVN speaker and by the overall similarity of *A1* in the two dialects. Finally, there is a surprising interaction between *dialect*, *order* and *experimenter* in tone *A2*. Since most of the errors in *A2* identification involve the incorrect choice of tone *B2* as a response, we suggest that a longer exposure to the SVN experimenter, who does not have voice quality distinctions in her tones, makes the NVN voice quality contrast comparatively more salient.

Results for other tones are easier to interpret. SVN *B1* is better identified by the SVN listeners than its NVN counterpart, which is expected due to the much lower f0 curve of the NVN variant. The reason why *B1* is better identified with the NVN *experimenter* (even SVN stimuli) is not entirely clear, but the main effect of *order* could simply be attributed to a learning effect in the presence of unfamiliar stimuli. The fact that SVN *B1* is much better identified than its NVN counterpart with the SVN experimenter suggests that listeners are reacting to the mismatch between the stimuli and the experimenter's production of this tone. This is also supported by the fact that NVN *B1* is better identified with the NVN experimenter. Finally, the better identification of *B1* in the first block when it is produced by the SVN speaker can once again be attributed to familiarity. The improvement in the identification of NVN *B1* in the second block is probably due a familiarization with the unexpected low rising pattern during the experiment.

Tone *B2* is perceived equally well in all contexts, which is probably due to its familiar shape in SVN and to its very salient final glottalization in NVN. Tone *C1* which was only produced by the NVN speaker is better identified with the NVN experimenter, which once again suggests that listeners are using the experimenter as a benchmark for their evaluation of the stimuli. An accommodation to the experimenter (i.e. a learning effect) is also visible in tone *C1*: with the NVN experimenter, listeners do much better in the second block than in the first one, suggesting that a longer exposure to the experimenter's speech helps the listeners to fine-tune to the stimuli. The last effect to be reported is the effect of *dialect* on the identification of tone *C2*. As expected, SVN listeners make fewer mistakes when identifying familiar SVN variants.

5. CONCLUSION

It does not come as a surprise that the *dialect* of the speaker plays a significant role in the accurate identification of tones. However, the interaction effects between *experimenter* and *dialect* can only be explained by listeners being aware of a mismatch between their dialect and the stimuli and resolving it by taking into account available patterns of contrasts that are different from their own. Listeners use their fresh experience of a different system with the implicit assumption that it will help them improve their identification strategies. Our hypothesis that the proportion of correct identification by SVN speakers would be influenced by the dialect of the experimenter is thus borne out: SVN listeners perform better on the NVN stimuli when the NVN tokens are presented by the NVN experimenter. The additional interactions of *order* suggest that listeners can rapidly build or fine-tune a model of 'the other' after a few minutes of exposure to a new or less familiar variety of speech. This result fits well in an exemplar theoretic explanation of phonological categorization and the storage of social information alongside tonal categories.

6. REFERENCES

[1] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress *Journal of Phonetics* 34, 4, 458-484.

[2] Strand, E. A. 1999. Uncovering the Role of Gender Stereotypes in Speech Perception. Journal of Language and Social Psychology, 18, 86-99)

[3] Vũ, T.P.1981. *The Acoustic and Perceptual Nature of Tone in Vietnamese*, Ph.D. dissertation, ANU.

[4] Vũ, T.P. 1982. Phonetic Properties of Vietnamese Tones across dialects. *Papers in Southeast Asian Linguistics.* D. Bradley. Sydney, ANU. 8 - Tonation**:** 55-75.

[5] Michaud, A. 2004. Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica* 61: 119-146.

[6] Gsell, R. (1980) Remarques sur la structure de l'espace tonal en Vietnamien du Sud (parler de Saigon). *Cahiers d'Études Vietnamiennes*(4): 1-26.

[7] Nguyễn, V. L. & Edmondson, J. (1997) Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies* 28: 1-18.

[8] Brunelle, M. 2006. *Tone perception in Vietnamese dialects*, pres. at TIE2, ZAS, Berlin.

* Many thanks to Hà Kiêu Phương and Nguyễn Thị Minh Tâm for conducting the experiment in HCM City and to Phạm Thị Lê Uyên for recruiting subjects.