

CATEGORICAL PERCEPTION OF TONES IN VIETNAMESE

Alain Content & Noémi Perwez

Laboratoire Cognition, Langage et Développement, Université libre de Bruxelles, Belgium

alain.content@ulb.ac.be; noemiperwez@gmail.com

ABSTRACT

Few studies have investigated the perception of tone contrasts and it is currently unclear whether tones are perceived categorically or continuously by listeners of tone language. In the present study, native speakers of French and Vietnamese received an identification test with several acoustic continua contrasting three among the six tones of the vietnamese language.

Results showed clear differences across groups, with native listeners of Vietnamese demonstrating categorical performance in the identification task. Further research is required to determine whether the difference generalizes to discrimination.

Keywords: tones, Vietnamese, categorical perception

1. INTRODUCTION

1.1. The tone system in the Vietnamese language

Like Chinese, the Vietnamese language is an isolating morphosyllabic language involving a system of contrastive tones. From a linguistic point of view, tones or tonemes are described as complex suprasegmental feature complexes [2]. Tone categories entail specific pitch and F_0 contour characteristics, but other acoustic and articulatory features such as duration, intensity, tenseness and breathiness also play a role [2, 4, 5, 13] and interact with syllable structure; there are even suggestions that visual cues may help tone identification in certain experimental conditions [11].

The standard Vietnamese language is generally considered to encompass six lexical tone categories, although dialectal variations affect the tone system as well as the other aspects of the phonology [4, 5]. The first tone is called *ngang* (“level”), and is produced with a modal voice quality and mid-range pitch. The *huyền* (“hanging”) tone starts lower and falls monotonously, whereas the third tone, *sắc* (“sharp”) also starts at mid-range and rises rapidly; the *sắc* rising tone also varies in duration, with a shorter variant in closed

syllables. The *nặng* (“heavy”) tone is also usually described as low and falling, but typically shorter, and ending in a glottal stop. Finally, *ngã* and *hỏi* (respectively, “to fall, tumble down” and “to ask”) are both falling-rising with a low-to-mid initial pitch; *ngã* also often involves a glottal stop synchronous with the pitch inflection. The last two tones are less frequent in the language, and appear to be less distinct in the southern dialect, with *hỏi* also often confused with *nặng* [5]. In sum, descriptively, it may be useful to distinguish between the *ngang*, *huyền* and *sắc* tones, which have simple contours (respectively level, falling and rising) and vary little across regional dialects, and the three others, which appear both less discriminable and more variable.

1.2. Categorical perception and tones

Categorical perception is classically (and ideally) defined when performance to a physical continuum exhibits the combination of an abrupt discontinuity in the identification function and the presence of a discrimination peak, corresponding to the category boundary [8, 12, 14]. Once thought to constitute a specific feature of phonetic perception and the hallmark of a special mode of processing, categorical perception or closely similar phenomena now appear ubiquitous, suggesting that it may rather constitute a general cognitive phenomenon crucial for the adaptation of the organism to its environment [8, 10].

Nevertheless, not all stimuli are perceived categorically, and it remains of interest to establish whether tone processing demonstrates the defining characteristics of categorical perception in native speakers of tonal languages, and whether perceptual performance differ as a function of language experience. In this regard, the investigation of Vietnamese speakers is of particular interest given that, contrary to Chinese, tones are explicitly and straightforwardly marked in the orthography through diacritic marks.

Previous studies with tone continua in tonal languages are scarce and have not led to a unanimous conclusion. Whereas an early study [6]

comparing Mandarin and American listeners claimed that tone perception was categorical, Abramson [1] tested Thai listeners and argued that perception was more continuous than categorical, based on the lack of discrimination peaks corresponding to category boundaries. Interestingly, more recent studies [7, 9] in Cantonese and Taiwan Mandarin produced a similar pattern, with relatively steep identification functions but no obvious discrimination peaks.

Here, we tested native speakers of Vietnamese and French on an identification task with three continua, using the simple tone patterns of Vietnamese (*ngang*, level, *huyền*, falling, and *sắc*, rising)¹.

2. METHODS

2.1. Participants

Fifteen native Vietnamese listeners (4 men, 11 women, mean age 26.7 yrs) participated in the identification and discrimination tests. All had Vietnamese as their first language. Nine were living in Vietnam and staying abroad at the time of the test for a study period. The six others were born in European countries. Eight originated from the South, six from the North and one from the Central region.

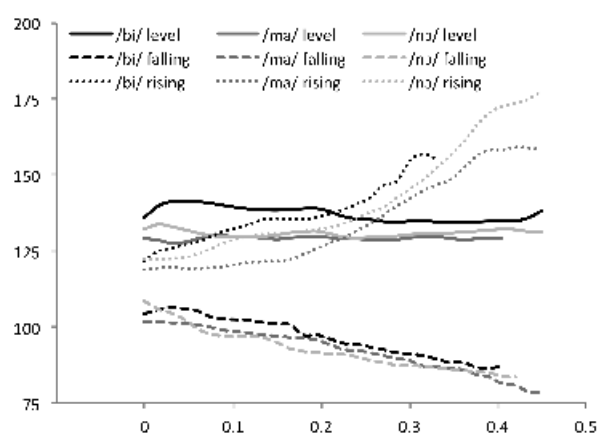
Nineteen native listeners of French (4 men, 15 women, mean age 18.6 yrs) constituted the non-tonal language group. None of them had ever been exposed to any tone language.

2.2. Speech materials

Four open syllables were selected such that they would constitute meaningful and familiar words with each of the three selected tones (/bi/ : *bi*, marble; *bì*, skin; *bí*, pumpkin; /ma/ : *ma*, ghost,devil; *mà*, but, where; *má*, mother or cheek; /nɔ/ : *no*, gorged, surfeited; *nò*, fish trap; *nó*, he/she/it/him; /t^hu/ : *thu*, autumn, to collect ; *thù*, hostile; *thú*, pleasure or animal). All 12 morphosyllables were recorded in a quiet room by a native speaker of Vietnamese of Southern origin. The 12 stimuli were mixed with fillers and recorded several times with different orders, so that exemplars of similar durations and intensity could be selected. The selected stimuli varied in duration between 0.59 and 0.69 sec and were very similar in terms of mean and peak intensity. As Figure 1 illustrates, the resulting stimuli conform very closely to previous qualitative and quantitative descriptions [2, 4, 5, 13].

Based on the natural stimuli (see Figure 1), a two- or three-point linear contour template has been estimated for each tone (level: 136-132Hz; falling: 108-85Hz; rising: 122-139-170Hz), and intermediate values were interpolated to obtain three eight-step continua (*Falling-Rising*, *Rising-Level* and *Falling-Level*). The obtained pitch contours were then applied to the natural stimuli using PRAAT [3] pitch manipulation tools.

Figure 1: Pitch contour (Hz) for the natural stimuli used as endpoints, aligned at consonant release.



As a preliminary test, the 3 (continua) x 8 (steps) x 4 (syllables) stimuli were mixed together with the 12 original exemplars and presented to a group of 15 uninformed Vietnamese speakers who were asked to assess the quality of the stimuli on a scale from 1 (natural) to 5 (artificial/manipulated). Overall, stimuli based on the syllable /t^hu/ were assessed as less natural (*mean* = 2.6) than the others (1.8 to 2.1), so they were not further used. Natural stimuli were judged only slightly and not significantly more natural than resynthesized ones (1.74 vs 1.98, respectively).

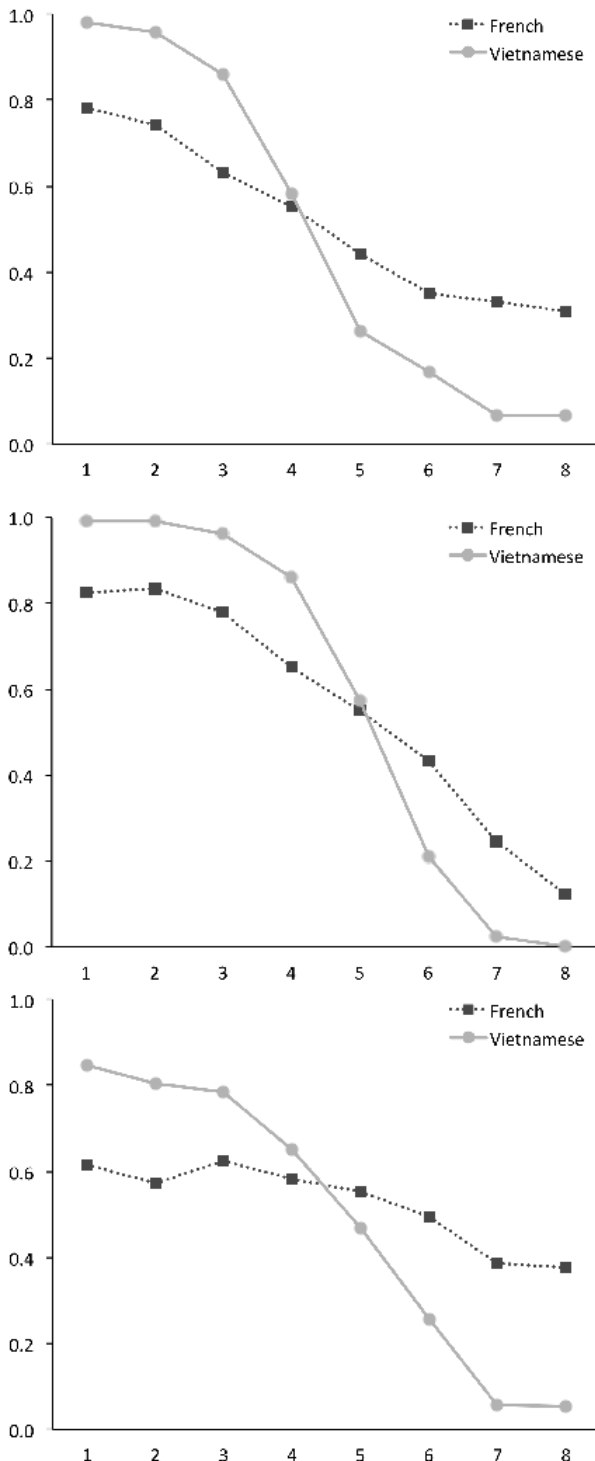
2.3. Categorical perception test

For the identification test, participants received three blocks, one for each continuum, in fixed order (*Falling-Rising*, *Rising-Level* and *Falling-Level*). Each block included five repetitions of each stimulus, for a total of 120 items. Stimuli were presented through headphones and participants responded by hitting labeled keys on the keyboard. A few examples, and a small number of training trials were run beforehand to convey the categories to the French participants. The total duration of the experiment was approximately one hour.

3. RESULTS

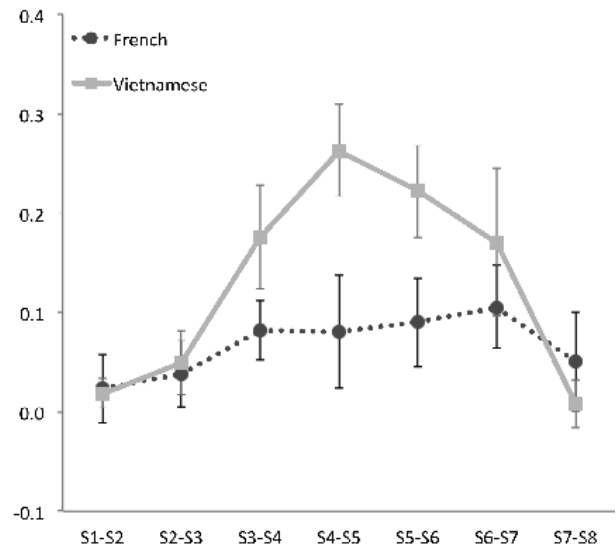
The average identification performance as a function of group (Vietnamese vs French) and step is illustrated in Figure 2 for each continuum.

Figure 2: Averaged response probabilities curves of the French and Vietnamese participants for the three continua. Top panel: Falling to Level tone; Middle: Rising to Falling tone; Bottom: Rising to Level tone.



Vietnamese participants display the expected abrupt switch from one category to the other with all three continua. In contrast, French listeners produce less steep response curves and also lower proportions of correct responses even at the extrema, suggesting that they had a hard time identifying the two contrasting categories, especially for the *Rising/Level* continuum. A mixed ANOVA, with Group as a between-subject factor and Step and Continuum as repeated measured, indicated that the difference was statistically reliable, as shown by a significant Group by Step interaction, $F(7, 217) = 28.6$, $\eta^2 = .48$, $p < 0.0001$.

Figure 3: Slopes of the identification curves averaged over the three continua (error bars correspond to .95 confidence intervals).



As a more direct test of categorical perception, we computed the differences between the response rates for successive steps. This calculation produces estimates of the slope of the identification curve along the steps of the continua, and permits a proper test of categorical perception, independently of the overall performance. As shown in Figure 3, the categorical prediction was clearly borne out: the Vietnamese group exhibits a peak between steps 4 and 5, whereas the French data show much less variation in slope, as expected with a continuous linear performance. Again, a mixed ANOVA indicated a significant interaction between Groups and Steps, $F(6, 186) = 8.0$, $\eta^2 = .21$, $p < 0.0001$. Bonferroni post-hoc comparisons indicated a significant difference in slope between step 1 to 2 and step 6 to 7 for the French group, while for the Vietnamese, the slope from step 3 to

4, 4 to 5, and 5 to 6 is significantly larger than the values near the ends of the continua.

4. DISCUSSION

The present study constitutes a first attempt at examining whether tone perception in Vietnamese is categorical. As far as the identification data is concerned, the response curves and cross-linguistic differences obtained provide support for a categorical mode of perception. For each of the three continua used, the identification responses of the Vietnamese participants showed the characteristic S-shaped pattern. French participants on the contrary exhibited more linear response fluctuations. They succeeded in identifying stimuli to the three tonal patterns, though to a lesser extent than native Vietnamese listeners, but showed no indication of an abrupt change from one category to the other, suggesting that their responses reflect non-linguistic, psychophysical mechanisms of auditory perception.

On the face of claims to the difficulty of recognizing tones for non-native speakers of tone languages, the fine performance of French listeners may appear surprising, given their lack of experience of lexical tones. Whether such successful performance would generalise in more natural settings is however unsure, as the stimuli were isolated syllables produced in a list context rather than in continuous speech, and at a rather slow speech rate.

Although the identification data clearly support the view that tones are perceived categorically by Vietnamese listeners, the lack of valid discrimination data clearly limitates the conclusions that may be drawn from the present data. Further studies will be necessary to examine whether Vietnamese tone perception entails all defining features of categorical perception, or whether it exhibits distinct patterns of identification and discrimination, as seems to be the case in other tonal languages [7, 9].

5. ACKNOWLEDGEMENTS

We thank Quoc Duy Tran and Mai Trang Huynh for their help at various stages of this project, and Nicolas Dumay for comments on a previous draft.

6. REFERENCES

- [1] Abramson, A. 1979. The noncategorical perception of tone categories in Thai. In Lindblom, B., Ohman, S. (eds.), *Frontiers of Speech Communication*. London: Academic Press, 127-134.
- [2] Alves, M. 1995. Tonal features and the development of Vietnamese tones. *Working Papers in Linguistics* 27. University of Hawaii, 1-13.
- [3] Boersma, P., Weenink, D. Praat. <http://www.praat.org/>
- [4] Brunelle, M. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37(1), 79-96.
- [5] Brunelle, M., Jannedy, S. 2007. Social effects on the perception of Vietnamese tones. *Proceedings of the 16th International Congress of Phonetic Sciences* Saarbrücken, Germany, 1461-1464.
- [6] Chan, S.W., Chuang, C.K., Wang, W.S.Y. 1975. Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America* 58, S119.
- [7] Francis, A.L., Ciocca, V., Ng, B.K.C. 2003. On the (non)categorical perception of lexical tones. *Perception and Psychophysics* 65(7), 1029-1044.
- [8] Goldstone, R.L., Hendrickson, A.T. 2010. Categorical Perception. *Interdisciplinary Reviews: Cognitive Science* 1, 65-78.
- [9] Halle, P.A., Chang, Y., Best, C.T. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32, 395-421.
- [10] Harnad, S. 1990. *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press.
- [11] Mixdorff, H., Luong, M.C., Nguyen, D.T., Burnham, D. 2006. Syllabic tone perception in Vietnamese. *Proceedings of TAL 2006* La Rochelle, France, 137-142.
- [12] Nguyen, N. 2005. La perception de la parole. In Nguyen, N., Wauquier-Gravelines, S., Durand, J. (eds.), *Phonologie et Phonétique: Forme et Substance*. Paris: Hermès, 425-447.
- [13] Nguyễn, Q.C. 2002. *Reconnaissance de la Parole en Langue Vietnamienne*. Thèse de Doctorat de l'Institut National Polytechnique de Grenoble, Grenoble, France.
- [14] Repp, B.H. 1984. Categorical perception: Issues, methods, findings. In Lass, N.J. (ed.), *Speech and Language: Advances in Basic Research and Practice*. New York: Academic Press, 10, 243-335.

¹ The study also included a ABX discrimination task. Unfortunately however, the results for both groups were at chance level overall and for all stimulus pairs. Hence they will not be reported in detail here.