

# **Can the tones (local function) influence the acoustic perception of the Vietnamese attitudes (global function) for French listeners (non tonal)?**

*Dang-Khoa Mac*<sup>1,2</sup>, *Véronique Aubergé*<sup>2</sup>, *Eric Castelli*<sup>1</sup>, *Albert Rilliard*<sup>3</sup>

<sup>1</sup> International Research Center MICA, HUST-CNRS/UMI 2954 Grenoble INP, Hanoi, Vietnam

<sup>2</sup> Laboratory of Informatics of Grenoble (LIG), CNRS, France

<sup>3</sup> LIMSI-CNRS, Orsay, France

{dang-khoa.mac, eric.castelli}@mica.edu.vn,  
veronique.auberge@imag.fr, albert.rilliard@limsi.fr

The attitudes are main part of the face-to-face interaction and linked to the language through the culture. These affects expressions are clearly social affects: they carry the intentions and points of view of the speaker (e.g: surprise, confirmation, etc.) and can give the social context on the interaction (e.g. intimacy, politeness). When the speaker does not express any attitude in his speech act (that is declaration or “simple” question), she/he expresses that she/he has no opinion in this utterance or that she/he does not want or cannot express any attitude [1]. Even if many are universal for their values or even for their prosodic forms, some prosodic implementation and even some attitudes values are specific to the culture the language [4,10]. Anyway, the attitudes are built inside each culture and language, and they must be learned by children inside the culture or by the learners of second language [3]. The understanding of this phenomenon may benefit from cross-cultural studies [3,4].

The attitudes or social affects can be supposed to be involved into voluntary cognitive controls, whereas emotions are involuntary controls [1]. The prosody has been shown in different languages as a main vector for expressing attitudes [1, 2]. The “classical” prosodic parameters (F0, intensity, timing), are strongly implied to carry attitudes [1,2,5]. According to [6], the voice quality is proposed as a 4th dimension of prosody; it has been also shown as a fundamental parameter for emotions [7] and used for attitudes [10]. Many different functions are implemented by prosody by using these same acoustic parameters (F0, intensity, timing and voice quality). In tonal language, a part of the lexical access function is implemented by F0. Moreover in Vietnamese language, tones can imply some voice quality such as creaky voice [8]. The domain of the tonal function is the segment, which is a local domain relatively to the utterance unit. The attitudinal function concerns the utterance unit, and the prosody of attitude can be proposed as a global contour related to the utterance [1]. The interaction inside F0 between the global attitude function and the “local” segmental tonal function seems clear for native tonal language speakers, but the question of the perceptive processing by non native, non tonal language, speakers, could inform on the deep mechanisms of cognition of this social signal.

This paper presents a study of Vietnamese attitudinal expressions under the light of their cross-cultural perception. This works aim to explore the eventual perturbation by the tonal system on the perception of Vietnamese attitudes in particular for French speakers that is non tonal language: will they be able to perceptively extract and separate, in the same signal parameters, the tones values from the attitudinal information? That is to process the lexical access function, attached to word domain, within the attitude function, attached to the whole utterance domain, but morphologically implemented by prominences (that is local vs. global cues, that can be explained by Gestalt approaches theories about prosody morphology [1]).

Based on some researches of attitudes in Vietnamese and other languages [3,4,12], 16 Vietnamese attitudes have represented in this corpus. The corpus contains 125 sentences without specific affective meaning, produced with all the 16 attitudes and balanced in terms of

tone position. These sentences were recorded with one male speaker, native of Hanoi (standard pronunciation). The whole corpus thus contained 2000 sentences corresponding to more than 90 minutes of signal after post-processing.

The perception experiment was carried out to study the influence of Vietnamese tones on the perception of the 16 Vietnamese attitudes. Nineteen sentences of two and three syllable length were chosen from the corpus for the test. The tones were set at varied positions (at the first, middle and last syllable), in order to check the effect of the tone in different positions of sentences. Twenty French listeners, who have not experience on Vietnamese language, were participated in this experiment. All subjects listened to each stimulus only one time. After each stimulus, they were asked, in a forced and closed choice, to indicate the perceived attitude among the 16 presented ones, and the level of confidence about this choice.

Table 1: *Output of the ANOVA on the percent of attitude recognition. Significant effects at the 1% level are set in bold face*

	df	% Reco.	
		F	p
<b>Attitude</b>	<b>15</b>	<b>13.652</b>	<b>.000</b>
Tone	7	1.094	.364
<b>TonePosition</b>	<b>2</b>	<b>6.408</b>	<b>.002</b>
<b>Attitude * Tone</b>	<b>105</b>	<b>1.546</b>	<b>.000</b>
<b>Attitude * TonePosition</b>	<b>30</b>	<b>2.027</b>	<b>.001</b>
Tone * TonePosition	6	1.784	.098
<b>Attitude * Tone * TonePosition</b>	<b>90</b>	<b>2.534</b>	<b>.000</b>

A previous cross-cultural perceptive experiment has been established with Vietnamese attitudes on utterances using only “neutral” tone. [9] This was carried out to have a reference for the comparison: in one hand with the Vietnamese listener’s performances, and especially on the other hand with the French listeners on utterances built with all the tonal complexity of Vietnamese.

The results of perception test were first analyzed with a repeated measure ANOVA, in order to evaluate the relative efficiency of the tones, the position of tones on the listener’s perception (see Table 1). Overall, the ANOVA results show that the non-native listeners can be separate the tonal effects (local) and the attitudinal effects (global). The tone has no significant effect on the perception result ( $p > 0.01$ ). However, the interaction between attitude and tone has significant effects on subjects’ perception.

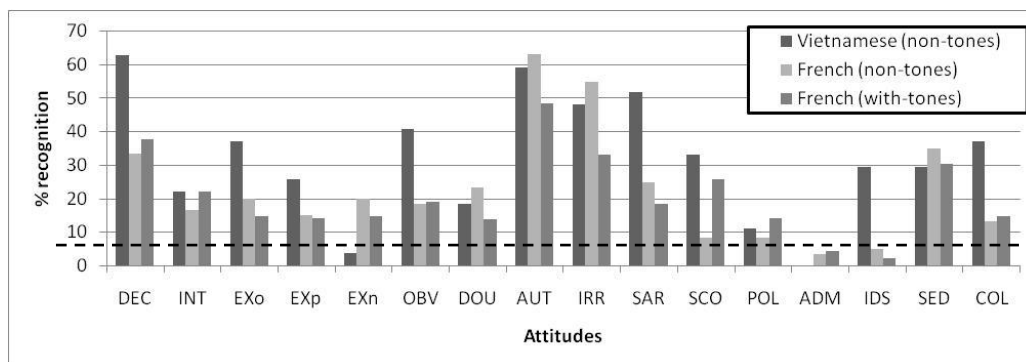


Figure 1. *Recognition rate of 16 attitudes by Vietnamese and French listeners (on the tone- and non-tone utterances). The dashed line indicates the chance level (6.25%).*

The attitude recognition perception (Figure 1) shows the differences in the perception of Vietnamese and French subject. Globally, most attitudes were recognized above chance level,

and native listeners have higher recognition scores than foreign ones. The result of French subjects on the utterances with tones and non-tone shows us the different influences of tone on the different attitudes. These results need to be further explained by studying the similarity in prosodic characteristics of Vietnamese tones and attitudes through the French prosodic patterns.

## References

- [1] Aubergé V. , “A Gestalt Morphology of Prosody Directed by Functions: the Example of a Step by Step Model Developed at ICP”, Proc. Speech Prosody 2002, Aix- en Provence, France, 151-154.
- [2] Wichmann, A. “Looking for attitudes in corpora”, Language and Computers, From the COLT's mouth ... and others'. Edited by Leiv Egil Breivik and Angela Hasselgren , pp. 247-261(15), 2002
- [3] Shochi T, Gagnié G., Rilliard A, Erickson D, & Aubergé V., “Learning effect of prosodic social affects for Japanese learners of French language”, SpeechProsody 2010, Chicago, USA.
- [4] Scherer, K. R., Banse, R., Wallbott, H. G., “Emotion inferences from vocal expression correlate across languages and cultures”, Journal of Cross-Cultural Psychology, 32(1), 76-92, 2001
- [5] Fónagy, I., “La vive voix”, Paris, Payot, 1991.
- [6] Campbell, N. & Mokhtari, P. “Voice Quality: the 4th Prosodic Dimension”, 15th International Congress of Phonetic Sciences, Barcelone, Espagne, 2417-2420, 2003
- [7] Audibert, N., Aubergé V. et al. “The Relative Weights of the Different Prosodic Dimensions in Expressive Speech: A Resynthesis Study”. Affective Computing and Intelligent Interaction: 527-534, 2005.
- [8] Do T.D., Tran T.H. & Boulakia G., "Intonation in Vietnamese", in Intonation systems: A survey of 22 languages, D. Hirst and A. Di Cristo, Eds.: Cambridge University Press, 395-416, 1998
- [9] Mac D.K., Auberge V., Rilliard A. & Castelli E. “Cross-cultural perception of Vietnamese Audio-Visual prosodic attitudes”, SPEECH PROSODY 2010, Chicago, USA
- [10] Shochi, T., Aubergé, V. & Rilliard, A. “Cross-Listening of Japanese, English and French social affect: about universals, false friends and unknown attitudes”, in ICPhS, 2097-2100, Saarbrücken, Germany, 2007
- [11] Gobl, C. and A. Ni Chasaide, "The role of voice quality in communicating emotion, mood and attitude." Speech Communication 40(1-2): 189-212, 2003
- [12] Le T.X., "Etude contrastive de l'intonation expressive en français et en vietnamien", PhD thesis of Linguistic and Phonetic, Université Paris 3, 1989