

Local vs. global prosodic cues: effect of tones on attitudinal prosody in cross-perception of Vietnamese by French

Dang-Khoa Mac^{1,2}, Véronique Aubergé², Eric Castelli¹, Albert Rilliard³

¹ International Research Institute MICA, HUST-CNRS/UMI 2954 Grenoble INP, Hanoi, Vietnam

² Laboratory of Informatics of Grenoble (LIG), CNRS, France

³ LIMSI-CNRS, Orsay, France

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

Abstract

Attitudes or social affects are strongly implied in interaction processing, and specifically to socio-cultural aspects of language. In face-to-face interaction, social affects are expressed through global patterns of prosody. This paper presents a cross-perceptual study of social affects in Vietnamese, a tonal language. Sixteen Vietnamese attitudes with tonal variation were used in a perception experiment with French listeners. The result of this perception test was also compared with that of a perception test on non-tonal Vietnamese stimuli. The results show that French listeners can globally process the local prosodic cues of tones together with the global cues of attitude patterns. The main effect of tone is to reinforce, but not change the confusion clustering.

Index Terms: attitude, social affect, tone, global prosodic patterns

1. Introduction

1.1. The social affects

The social affects are a main part of face-to-face interaction and are linked to language through culture. Whereas emotions are involuntarily controlled, the social affects, in particular attitudes, are voluntarily controlled and contribute to the speech act [1]. Social affects or attitudes carry the intentions and points of view of the speaker (e.g. surprise, confirmation, etc.), and they can give potency indices about the interaction (e.g. authority, politeness) as well as the social context of this interaction (e.g. intimacy, politeness). An utterance without any attitude (e.g., a declaration or “simple” question) can mean that the speaker has no opinion about this utterance or that she/he does not want to or cannot express any attitude [1]. Even if many attitudes are universal in terms of their values or even their prosodic forms, some prosodic implementations and even some attitudes are specific to a given culture or language [4, 11]. In any event, the attitudes are built inside each culture and language, and they must be learned by children inside the culture or by second language learners [12]. A better understanding of this phenomenon may benefit from cross-cultural studies [8, 10].

1.2. Local vs. global cues

Prosody has been shown to function as a main vector for expressing attitudes in different languages [1, 12]. The “classical” prosodic parameters (F0, intensity, timing) are strongly implied to carry attitudes [1, 13, 5]. [3] propose voice quality as a 4th dimension of prosody; it has been also shown as a fundamental parameter for emotions [7, 10] and used in the expression of attitudes [7, 11]. Many different functions

are implemented by prosody using these same acoustic parameters (F0, intensity, timing and voice quality) [7, 10]. In tonal language such as Vietnamese, a part of the lexical access function is implemented by F0. The Vietnamese language has 6 tones: level (1), falling (2), broken (3), curve (4), rising (5) and drop (6) as shown in Figure 1. Tone 5b and 6b correspond to tone 5 and 6 on a syllable ended by a stop consonant. Moreover the Vietnamese tonal system can employ some production of voice quality, within F0. That is the co-occurrence of glottalization during the production of tone 3 and tone 6: tone 3 is accompanied with harsh voice quality due to a glottal stop (or a rapid series of glottal stops) around the middle of the vowel; tone 6 has the same kind of harsh voice quality as tone 3; however, it is distinguished by dropping very sharply and it is almost immediately cut off by a strong glottal stop [5].

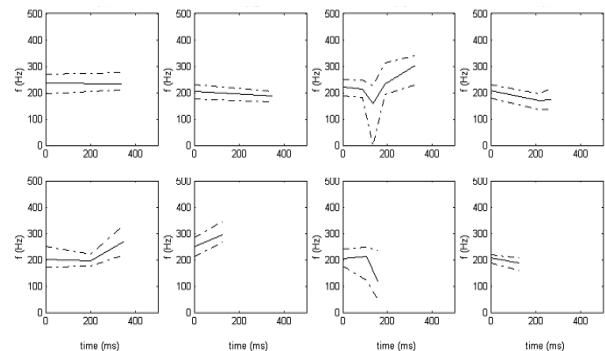


Figure 1: Examples of contours of 8 Vietnamese tone representations from a female subject [14]. From the left to right, top to bottom: tone 1, 2, 3, 4, 5, 5b, 6, 6b.

The domain of the tonal function is the vowel/syllable; the tone is used as phonological contrast for the minimal pair in lexical access for describing the double articulation. That means tone is a local domain relative to the utterance domain. 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 lexical access function attached to the word domain, within the attitude function attached to the whole utterance domain, is morphologically implemented by prominences (local vs. global cues) that can be explained by Gestalt approaches theories of prosody morphology [1].

The global attitude prosody is eventually organized by salient cues, and the “local” tonal prosody seems to be processed together without any interaction by native listeners. Because native listeners have cognitive lexicon representation, they could be separating local vs. global cues in the F0 signal. But for non-tone language speakers, is there any effect of tonal variation on the perception of attitudes?

This paper presents a study of Vietnamese attitudinal expressions in light of their cross-cultural perception. Because of the contrast of language characteristics (non-tonal vs. tonal) and the long distance of geography and culture (West European vs. East-Asian), French was chosen as the reference for this cross-cultural study of Vietnamese social affect. After presenting the construction of the Vietnamese attitude corpus, we describe the attitude perception experiment with French listeners. The perception results are analyzed and compared with previous results [9] of attitude perception on non-tonal (tone 1) Vietnamese utterances. The results allow us to answer the question of whether the non-tonal language listeners are able to extract and separate a tone's lexical F0 value from the attitudinal information,

2. The corpus

2.1. Variation of attitudes, tones and syntax

Based on research on attitudes in Vietnamese and other languages [4, 5, 6, 11, 15], 16 attitudes have been represented for Vietnamese in our corpus (Table 1).

Table 1: 16 selected Vietnamese attitudes, with their abbreviations

Declaration	DEC	Irritation	IRR
Interrogation	INT	Sarcastic irony	SAR
Exclamation of neutral surprise	EXo	Scorn	SCO
Exclamation of positive surprise	EXp	Politeness	POL
Exclamation of negative surprise	EXn	Admiration	ADM
Obviousness	OBV	Infant-directed speech	IDS
Doubt-Incredulity	DOU	Seduction	SED
Authority	AUT	Colloquial	COL

To observe the effects of tone and tonal co-articulation on attitudinal expression, the corpus contains 8 sentences of one-syllable length, corresponding to the 8 types of Vietnamese tone, and 72 sentences of two-syllable length, which correspond to all combinations of two tones among the 8 Vietnamese tones. The remainder of the corpus is based on 45 sentences of 3- to 8-syllable length and systematically varied in their syntactic structure: single word, nominal group, verbal group and a simple structure "subject-verb-object". That means the corpus is built from 125 sentences without specific affective meaning produced with all the 16 attitudes and balanced in terms of tone position. These sentences were recorded (both audio and video, but only audio is focused in this paper) by one male speaker native of the Hanoi dialect (standard pronunciation). The whole corpus thus contained 2000 sentences corresponding to more than 90 minutes of signal after post-processing.

2.2. Sub-corpus selection for tone variation

From the attitude corpus, a sub-set was selected with systematic variation of tones in different syntagmatic and paradigmatic locations. Nineteen sentences of 2- and 3-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, as shown in Table 2. The selection was done on 2- and 3-syllable length sentences, since they are short enough to avoid syntactic complexity.

3. The perception protocol

The perception experiment was carried out to study the influence of Vietnamese tones and the varied tonal location on

the perception of the 16 Vietnamese attitudes. Twenty French listeners who have no experience with the Vietnamese language were in this experiment. The testing program interface gave the label and the explanation of the 16 attitudes (in the native language of the listener). All subjects listened to each stimulus only one time. After each stimulus, they were asked to indicate the perceived attitude among the 16 attitudes and to indicate the intensity of its expressiveness (or the confidence about this choice) on a scale ranging from "hardly perceptible" (encoded as 1) to "very marked" (encoded as 100). The score 0 was assigned to the 15 other attitudes.

Table 2: Sub-set of tonal variation for 2 and 3 syllables length

Tone sequence	Utterance in Vietnamese	English
1_1	anh ta	him
2_1	người ta	them
3_1	đã xong	finished
4_1	thủy tinh	glass
5_1	chúng ta	us
6_1	chị ta	her
5b_1	héc ta	hectare
6b_1	tốp ca	choral
1_2	rau cần	celeri
1_3	dây kềm	steel wire
1_4	cây cảnh	home plant
1_5	y tá	male nurse
1_6	danh bạ	year book
1_5b	công tác	mission
1_6b	sa mạc	desert
4_1_1	bảy mươi ba	73
1_5_1	hai chúng ta	both of us
6_5b_3	hợp tác xã	cooperation
1_4_6	em bảo chị	you tell me

4. Result and discussion

To have a reference, a cross-cultural perceptual experiment has been established with Vietnamese attitudes on utterances using only "neutral" tone in [8]. This experiment was carried out to have a reference for the comparison: on the one hand with the Vietnamese listener's performances, and on the other hand with the French listeners on utterances constructed with all the tonal complexity of Vietnamese.

4.1. Effect of factors

The results of the perception test were first analyzed with a repeated measure ANOVA, in order to evaluate the relative effect of the tones and their position on the listener's perceptual responses. Firstly, the ANOVA result of neutral tone sentences (Table 3) shows the attitudes are well recognized by both of Vietnamese and French listeners without a sentence length effect. This verified the choice of the 2- vs. 3 syllable-length utterances for the experiment on the tonal sentences.

For the perception of the French subjects on the tonal sentences, the ANOVA results (Table 4) show that attitude has a significant effect on the perception result. There are also significant effects of the interactions between the attitude, tone and tone position. 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. That creates the appearance of the perturbation by tone prosody of some salient cues that are decisive information for some given patterns of attitudes. It must be further verified if it happens only when the local cues can be acoustically

confused with salient cues of another global pattern. However, the global confusions are not reorganized (see Figure 3 and 4).

Table 3: Output of ANOVA (on the percent of attitude recognition) for Vietnamese and French subjects and phrase without tone. Significant effects at the 1% level are set in bold face.

	df	Vietnamese		French	
		F	p	F	p
Attitude	15	47.804	0.000	33.100	.000
Sentence Length	2	3.735	0.024	1.655	.191
Attitude*SentenceLength	30	3.542	0.000	3.007	.000

Table 4: Output of ANOVA (on the percent of attitude recognition and level of confidence rating) for French subjects and phrase with tone. Significant effects at the 1% level are set in bold face

	df	% recognition		Confidence rating	
		F	p	F	p
Attitude	15	15.790	.000	15.419	.000
Tone	7	1.582	.136	1.321	.236
TonePosition	2	8.301	.000	10.007	.000
Attitude * Tone	105	1.976	.000	1.950	.000
Attitude * TonePosition	30	2.064	.001	2.005	.001
Tone * TonePosition	6	2.519	.020	2.038	.057
Attitude * Tone * TonePosition	90	3.528	.000	3.260	.000

4.2. Comparison of attitude identification for tone vs. non tone structures

Figure 2 shows the differences in the perception of Vietnamese and French subjects. Globally, most attitudes were recognized above chance level, and native listeners have higher recognition scores than French listeners (except the case of EXn), irrespective of tone variation or for the neutral tone sub-corpus. For AUT and IRR, the neutral tone utterances are better recognized than the non-neutral tone utterances; it is to be noted that the overall recognition rate of French subjects is slightly better than that of Vietnamese subjects.

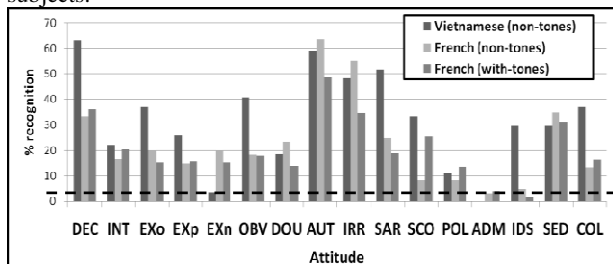


Figure 2: Recognition rate of 16 attitudes on non-tone and tonal sentences with Vietnamese and French listeners. The dash line: chance level

The attitude recognition result of the French listeners on the tone variation sub-corpus and the neutral tone sub-corpus are not so different. This is verified by the ANOVA result: globally, the tone variation has no effect on attitude perception. That means the non-native listeners can separate the tonal effects (local) and the attitudinal effects (global).

Figure 5 shows the recognition result of 16 attitudes for each tone located at the first and the last syllable of the sentences. As shown by ANOVA, there is no global effect of

the tone, but there is a significant effect of the interaction between the attitude, tone variation and tone position. For example, the DEC is poorly recognized if tone 4 is located on the first syllable and tone 5 on the final syllable. For AUT, the tone 2 (falling) inhibits recognition when located at the first syllable, but not if located on the final syllable, while tone 3 (broken) impedes recognition of this same attitude if located at the last syllable. EXn, DOU, SAR and SED are well recognized in the neutral tone sub-corpus, with a special effect of tone 6 for EXn in varied tone at the last syllable location. SCO is much better recognized in the varied tone corpus than in the neutral tone corpus, in particular tone 5b is efficient only if located at the first syllable, while the opposite is true of tone 5. INT is better recognized in the varied tone corpus, especially with tone 2 on the first syllable and tone 5b on the last syllable.

4.3. Comparison of confusion on attitudes for varied vs. neutral tone structures

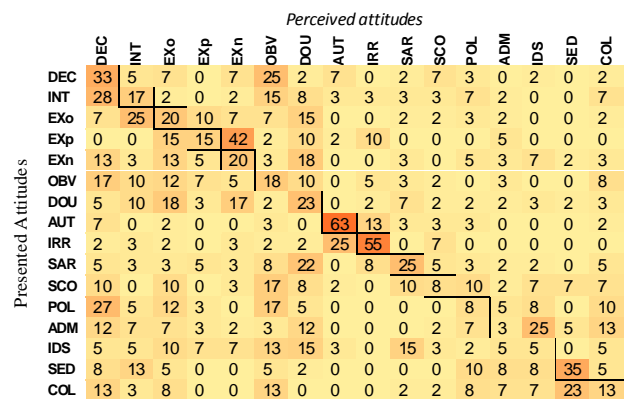


Figure 3: Confusion matrix for French on neutral tone variation Vietnamese attitudes (from [9])

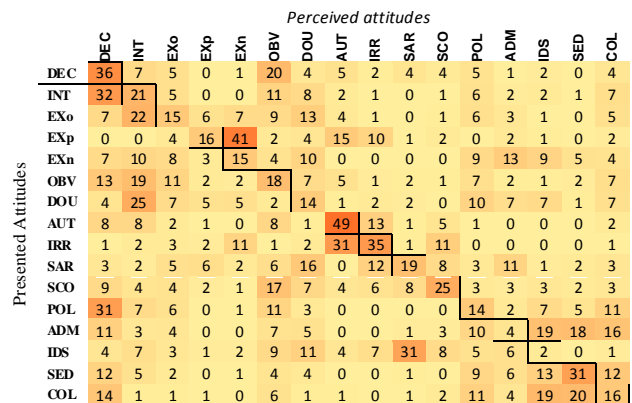


Figure 4: Confusion matrix for French on tonal variation Vietnamese corpus

Figures 3 and 4 show the confusions between attitudes for the varied tone sub-corpus and for the neutral tone sub-corpus. The two clear results are: (1) on varied tone stimuli, the degree of all the confusions increases; (2) the confusions are shared the same tendencies in both sub-corpus. Only one new confusion from DOU to INT (that is conceptually closed) appears quite clearly for the varied tone stimuli. It means that the local perturbation by tones increases the complexity of the global cues processing, but does not imply re-organization, or specific clear misunderstanding by perturbing salient global cues. These result needs to be further explained by studying

the similarity in prosodic characteristics of Vietnamese tones and attitudes through the French prosodic patterns.

5. Conclusions

This work aims to study the cross-cultural social affect of Vietnamese, a tonal language where a “neutral tone” can be used. The question can be asked of the prosodic influence of the tones local cues on the global processing of attitudinal prosody. Some varied tones stimuli of attitudes were presented to French listeners, who have no experience with lexical tone processing. The main experimental result is that the French listeners can globally separate the tone (local) processing from the attitude (global) processing. The tone processing can be considered as an increased cognitive load for French listeners that reinforces the degree of confusions between attitudes. However, interactions between tone type, tone location, and attitude indicate that the local cues of tones and the salient cues of global patterns [1] could be confused. The results need to be verified by further appropriate acoustic analysis to find out the acoustical parameters that lead to the perception of these social affects.

6. 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] 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
 [3] Campbell, N. & Mokhtari, P. “Voice Quality: the 4th Prosodic Dimension”, 15th International Congress of Phonetic Sciences,

Barcelone, Spain, 2417-2420, 2003
 [4] Diaféria, M.-L., “Les Attitudes de l’Anglais : Premiers Indices Prosodiques”, Master thesis, INP Grenoble, France 2002.
 [5] 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
 [6] Fónagy, I., “La vive voix”, Paris, Payot, 1991.
 [7] 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
 [8] Mac D.K., Auberge V., Rilliard A. & Castelli E. “Cross-cultural perception of Vietnamese Audio-Visual prosodic attitudes”, SPEECH PROSODY 2010, Chicago, USA
 [9] Mac D.K., Auberge V., Rilliard A. & Castelli E. Vietnamese multimodal social affects: How prosodic attitudes can be recognized and confused, SLTU 2010, Malaysia, 24-28
 [10] 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
 [11] 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
 [12] 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.
 [13] 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
 [14] Pham, T. N. Y., Castelli, E., and Nguyen, Q. C., “Gabarits des tons vietnamiens”, in JEP, 23-26, Nancy, France, 2002
 [15] Le T.X., “Etude contrastive de l’intonation expressive en français et en vietnamien”, PhD thesis of Linguistic and Phonetic, Université Paris 3, 1989

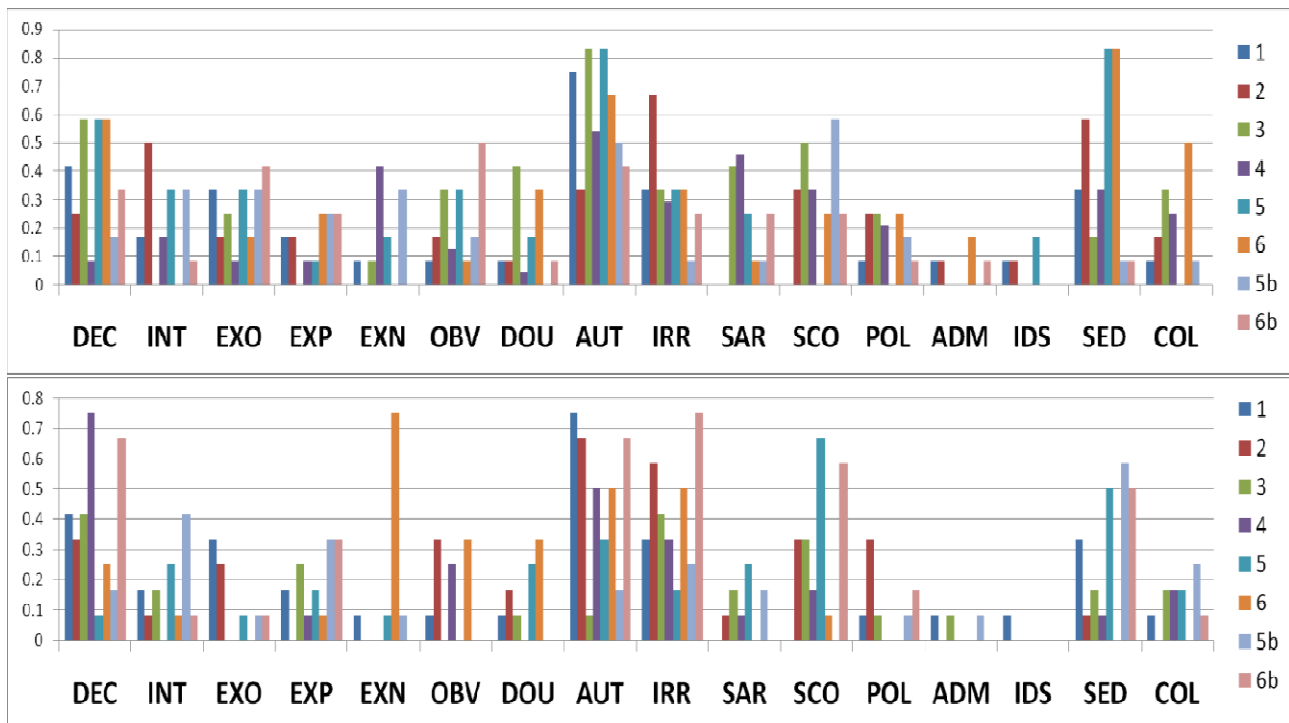


Figure 5: Recognition rate per attitude for each tone (1,2,3,4,5,6, 5b and 6b) located at the first (top) and the last (bottom) syllable of the sentences. Others syllable in sentences are bear the neutral tone (tone 1)