



HAL
open science

Tonal Influences on the prosodic Cross-linguistic Perception of Mandarin Social Affects by French and Vietnamese listeners

Yan Lu, Véronique Aubergé, Albert Rilliard

► **To cite this version:**

Yan Lu, Véronique Aubergé, Albert Rilliard. Tonal Influences on the prosodic Cross-linguistic Perception of Mandarin Social Affects by French and Vietnamese listeners. TAL 2012 - Third International Symposium on Tonal Aspects of Languages, May 2012, Nanjing, China. pp.n/c. hal-00744703

HAL Id: hal-00744703

<https://hal.science/hal-00744703>

Submitted on 23 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tonal Influences on the Prosodic Cross-linguistic Perception of Mandarin Social Affects by French and Vietnamese listeners

Yan Lu¹, Véronique Aubergé¹, Albert Rilliard²

¹GIPSA Lab, CNRS, Stendhal University, Grenoble France

²LIMSI-CNRS, Orsay, France

{Yan.Lu, veronique.auberge}@gipsa-lab.grenoble-inp.fr, albert.rilliard@limsi.fr

Abstract

Social affects play an important role in the face-to-face interaction and are implied in the realization of speech acts. The prosody is a main vector of social affects and its cross-language variability is a challenge for language description as well as for foreign language teaching. The present paper aims at examining the influences of tones on the perception of 19 prosodic social affects of Mandarin Chinese by French and Vietnamese subjects. A corpus of attitudinal speech was designed with a variation of tones values and location. The test results show that most social affects were globally recognized over chance by the two groups of foreign-language subjects. For French subjects, “contempt”, “irony” and “confidence”, and “irony” for Vietnamese, were not identified – independently of tones distribution. A significant effect of tones for the two groups was found, with a higher magnitude for French than for Vietnamese. Moreover, a significant attitude & tone interaction is observed for both languages.

Index Terms: prosodic attitudes, cross-linguistic perception of tones, Mandarin Chinese, Vietnamese, French

1. Introduction

Whereas emotions are involuntarily expressed, the expression of social affects results of a voluntary cognitive process which is an integral part of spoken interaction [1]. A privileged vector of social affects is the audio-visual prosody [2]. Such expressions need to be learned during infancy and would benefit to be explicitly taught in foreign language teaching. Within social affects, the speaker informs his interlocutor of different communication functions, which can be differently described [3, 4]. In this study, we selected some values of social affects which are supposed to reveal the attitude/opinion of the speaker or to signal some social or situational cues about the interaction [5], and these values are classified into three categories: firstly, the attitude, intention or opinion of the speaker about what he says, even if he does not express any attitude by performing a simple declaration or question, it is considered as an attitude to not give information on his attitudes [1]; secondly, some characteristics of the social relation implied in the interaction, e.g. politeness, authority; thirdly, the socio-cultural context of interaction, typically for intimacy, infant-directed speech and seduction.

After Martins-Baltar [6] had developed a repertory of expressive prosodic patterns for the purpose of language teaching, many studies of attitudinal prosody have been undertaken in different languages [7, 8, 9, 10, 11], and some of them had a focus on the cross-cultural comparisons [12].

Mandarin Chinese (also referred to as Putonghua or Standard Chinese) has four tones which were defined customarily according to the characteristics of their fundamental frequency curve as: high level (tone 1), rising (tone 2), dipping (tone 3) and falling (tone 4, cf. fig. 1).

Belonging to different families of languages, Mandarin Chinese, Vietnamese and French have their own specific linguistic structures. Both Mandarin and Vietnamese are tonal languages; French is not tonal (and not stressed). Some similarities can be found between Chinese and Vietnamese, e.g. according to [13], they have salient intonational phenomena, which have a marked influence on the phonetic realization of lexical tones, but do not modify their phonological identity. Compared to French, and from the prosodic and cultural point of view, Vietnamese could be considered as closer to Chinese. Previous studies on Vietnamese [14] and on French [1] have been published and can give some references about the social affects in these languages.

The present study is dedicated to investigate: (i) how prosodic social affects in Mandarin Chinese can be perceived by French and Vietnamese; (ii) what are the confusion patterns between the 19 proposed social affects; (iii) whether the effect of tones can be shown on the perception of social affects outside of any morpho-syntactic and semantic influences.

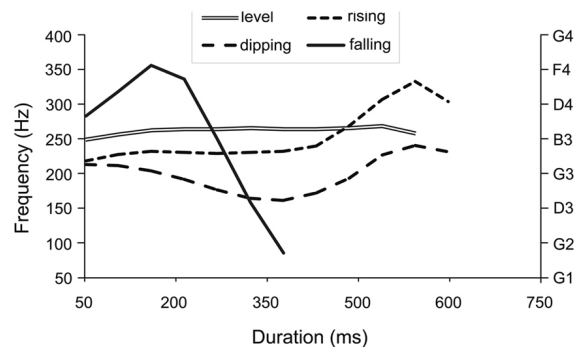


Figure 1: Fundamental frequency contours of the four Mandarin tones presented on the syllable [ma], spoken by a native female speaker.

2. Corpus

2.1. Speech corpus design

In order to compare the parameters implied in the variability of prosody, a dedicated and controlled corpus was built to convey different social affects.

The corpus was designed with consideration of utterances' length (in syllables), of tones location and of syntactic structure of the utterances, which were systematically varied in order to analyze further the variation of one parameter in the same context for the others. As the social affects could not be produced without reference to context, a dedicated context was described for each social affect, in order to help the speaker to express them as naturally as possible. All utterances were constructed to bear a literally neutral meaning (i.e. not containing any words which imply a specific social affect nor

emotion) but in the same time could be expressed with all the social affects studied.

Meanwhile, the complete corpus contains 152 utterances performed with 19 attitudes, i.e. 2888 stimuli, which cannot be evaluated in a single test. A sub-corpus of 16 utterances was selected with the special and unique consideration of tones values and location. There are not any morpho-syntactic and semantic variations in the sub-corpus, i.e. the utterances are nominal groups, verbal groups and numbers. The sub-corpus contains four 3-syllable words and twelve 4-syllable phrases. To vary the value of tones, each tone is placed at the first and the last syllable of 4-syllable utterances, while other syllables were always in the first tone. The tone of unstressed syllables was used in the last syllable of 3-syllable words. Hence the sub-sample contains 304 stimuli, (i.e. 16 utterances produced with 19 different attitudes).

2.2. Selected social affects

19 social affects which are commonly encountered in daily conversation were selected after [7] and [1, 10, 11, 5]. Table 1 shows the 19 social affects and their abbreviations in three categories.

Table 1. *Classification of social affects and their abbreviation*

	Social affects and abbreviation
Attitudes	Declaration (DECL) Question (QUES) Admiration (ADMI) Confidence (CONFI) Irritation (IRRI) Resignation (RESI) Contempt (CONT) Irony (IRON) Doubt (DOUB) Obviousness (OBVI) Disappointment (DISA) Neutral surprise (NEU-S) Positive surprise (POS-S) Negative surprise (NEG-S)
Social parameters	Politeness (POLI) Authority (AUTH)
Social context	Seduction (SEDU) Infant-directed speech (IDS) Intimacy (INTI)

2.3. Corpus recording

One native Mandarin female from Shaanxi province of China took part in the recording. She is teacher of French as a foreign language in a Chinese college, and speaks unmarked standard Mandarin Chinese. The recording was conducted in a sound-proof room of GIPSA-Lab in Grenoble of France, both in video and audio modalities. To make the attitudinal expression consistent, the speech sharing the same attitude was recorded in one session after the speaker had understood and had got familiar with the situational context of the given affect. 19 social affects were conveyed one by one in the same way. Another native Chinese from the same area of China as the speaker was also present during the recording to supervise the performance of the speaker.

3. Perceptual Experiment

The perception test was aimed to study the recognitions capabilities of French and Vietnamese listeners towards the Chinese social affects and a potential interaction between attitudes and tones. The listeners were 15 French (9 males and 6 females, average age of 33 years) and 15 Vietnamese (7 males and 8 females, average age of 27 years). Both of them work or study in Grenoble of France. None of the 30 subjects reported any listening disorder.

All 304 target stimuli were presented to the subjects through headphones in a quiet room and were introduced by a presentation of the experiment and a description of each social affect with examples of situations in which such a social affect can happen. The listeners had the instructions written in their mother languages (French for French listeners and both French and Vietnamese for Vietnamese listeners) at their disposal during the experiment. They listened only one time to each stimulus and had to choose the perceived attitude from the 19 proposed labels written in French. The presentation order of the stimuli was randomized for each subject.

4. Analysis and Results

4.1. Analysis of variance

An analysis of variance (completely randomized three-factorial design) was carried out on the data. The three fixed factors were the presented attitudes (A, 19 levels), the sequence of tones (T, 16 levels) and the native language of subjects (L, 2 levels). The significance level was set at 0.01. Table 2 shows the general results of the analysis of variance for each factor.

Significant effects of Attitude and of Tones sequences are observed as well as the significant interaction between Attitude & Tones. Both Attitude and the Attitude & Tone interaction have the most important effect size (*cf.* the η^2 column of table 2), and thus have the most important influence on the listeners' answers. These results show that the listener's recognition of attitude strongly depends on the presented prosodic attitudes and on the sequence of tones underlying these attitudes. Moreover, the global recognition of attitudes depends somehow (but in a less important proportion) on the language background of judges. The factor Language, the interaction between Language & Tones, and between Attitude & Language & Tones did not elicit any significant effect on listeners.

Two separated ANOVAs on French and on Vietnamese subjects were run (table 3, bottom). Results show that the effect of Tones is significant for French subjects while it is not significant for Vietnamese subjects (*cf.* mean results on fig. 2), although there is not a significant interaction between Attitude & Tones.

Table 2. *Global ANOVA results – significant effects in bold.*

	Sum Sq.	Df	F value	P	η^2
A	81.85	18	35.8	0.000	0.070
L	0.11	1	0.8	0.362	0.000
T	4.61	15	2.4	0.001	0.004
A*L	10.36	18	4.5	0.000	0.009
A* T	68.53	270	2.0	0.000	0.060
L* T	3.11	15	1.6	0.058	0.003
A*L* T	37.72	270	1.1	0.131	0.034

Table 3. *Separated ANOVAs by language – significant effects in bold.*

	Df	French		Vietnamese	
		F	p	F	P
Attitude	18	20.2	0.000	20.1	0.000
Tones	15	2.4	0.002	1.6	0.054
Attitude * Tones	270	1.4	0.000	1.4	0.000

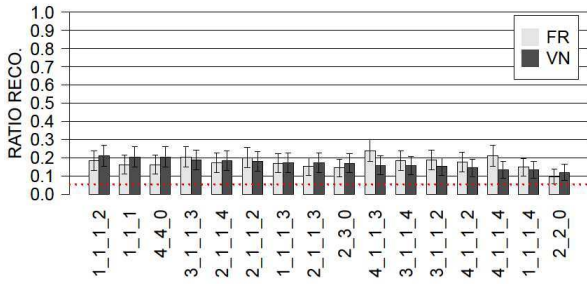


Figure 2: mean recognition rate for each tone sequence, per language background.

Figure 3 shows the mean recognition of the 19 social affects by French and Vietnamese subjects. It is to be noted that almost all of the social affects were identified above chance, except “contempt” “irony” and “confidence” for French subjects and “irony” for Vietnamese subjects.

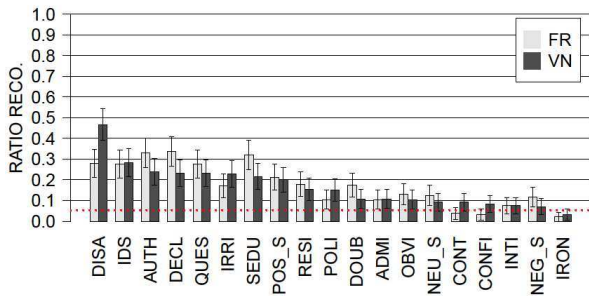


Figure 3: Mean recognition for the 19 social affects, per language background.

Globally, for French subjects, “declaration” (33.8%) is the best recognized attitude, while for Vietnamese subjects, it is “disappointment” (46.7%). For “seduction”, “authority” and “obviousness”, French subjects recognized better than Vietnamese subjects; in the case of “confidence”, Vietnamese subjects received a higher score. The two subjects groups had almost the same recognition percentage for “admiration”. It is worth to note that “irony” was poorly perceived both by French and Vietnamese subjects. [15] found that French “irony” was the most difficult to perceive for French subjects. Moreover, in the test of [14], Vietnamese “irony” was also difficultly recognized by French listeners but well recognized by Vietnamese listeners. Such a result may be due to the conceptual and cognitive differences showed by this attitude in different language, but more investigations are required.

The attractivity of attitudes – the sum of all the confusions from other attitudes to a given attitude (cf. fig. 4) – shows some interesting results. In comparison with the data gathered in previous work [5], for all listeners, including native ones, the attitude attracting most answers is “declaration”, which is mainly used when judges cannot identify any attitude. This result is coherent to common behaviors of perceiving his own language [4, 9, 10 12]. French and Vietnamese listeners show, to a lesser degree, the same preference for “declaration”, but with quite clear second choice: “question” for Vietnamese and “obviousness” for French judges. “Irony” was not well recognized by Vietnamese judges, nor did it attract any attitude.

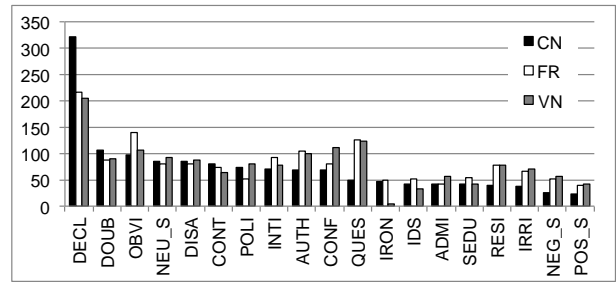


Figure 4: The attractivity – cumulated percentage of confusions from others attitudes to each one, per language background.

4.2. Clustering of attitudes

In order to measure the perceptive distances between each stimulus and identify the higher perceptual categories for French subjects and Vietnamese subjects, as well as the perceptual differences between the two groups, a hierarchical cluster analysis was run on the dispersion matrix and the calculation was carried out by using the correlation between the rows (1-r is used as the distance, where r is the correlation between two rows). From this matrix of perceived distances, a hierarchical clustering algorithm was applied, which allowed the observation of the main groups of attitudes for each subject group (cf. figure 5).

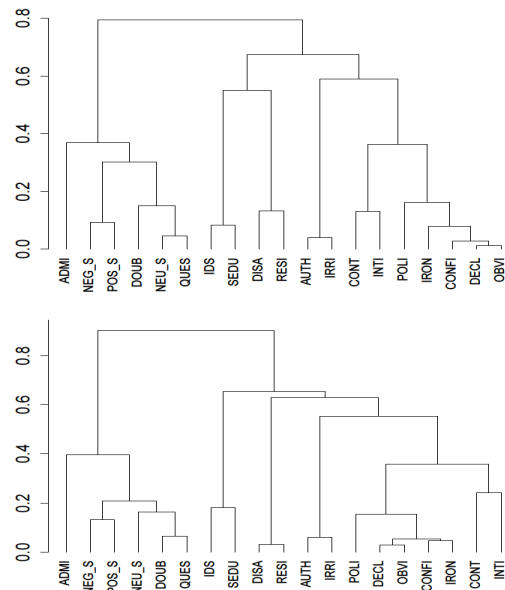


Figure 5: Hierarchical clustering of perceived social affects, based on R complete grouping criterion. The grouping done by French subjects is shown on top while the grouping of Vietnamese is on the bottom.

Firstly, the two groups of listeners have regrouped the perceived social affects almost in the same way: they have clustered the attitudes in eight groups (see the table 4). In comparison with the hierarchical clustering of the same social affects done by native subjects in the previous study [5], some differences were found: the native grouped the social affects in seven groups. The difference is due to the fact that the native subjects have regrouped attitudes of neutral and negative surprise (but not “positive-surprise) together with “question” and “doubt”, which can be considered as linked with some neutral “unexpected, uncertain” mental states; while the

French and Vietnamese subjects separated “negative-surprise” and “positive-surprise” from the others. The second finding is that most social affects have been grouped by both native subjects and foreign subjects in the same way (except “positive-surprise”, “negative-surprise”, “admiration”, “irony” and “intimacy”). “Admiration” was not well recognized by French and Vietnamese and it was neither consistently related to other attitudes, so it is isolated in the dendrograms. “Intimacy”, which was classified with “infant-directed-speech” and “seduction” in our introduction, was regrouped on the contrary with “contempt” which is considered as an impolite attitude. For French and Vietnamese subjects, “irony” is the attitude whose recognition score is the lowest. It was very poorly recognized and did not attract other attitudes. According to the dispersion matrix, it was linked to “declaration”.

Table 4. *Main clusters obtained from the hierarchical clustering of social affects, for each language groups.*

	Attitudes	
	Chinese	French / Vietnamese
Group 1	ADMI, POS-S	ADMI
Group 2	NEG-S, NEU-S, QUES, DOUB	NEG-S, POS-S
Group 3	IDS, SEDU	DOUB, NEU-S, QUES
Group 4	AUTH, IRRI	IDS, SEDU
Group 5	DISA, RESI	AUTH, IRRI
Group 6	IRON, CONT	DISA, RESI
Group 7	DECL, CONF, POLI, OBVI, INTI	CONT, INTI
Group 8		DECL, CONF, POLI, OBVI, IRON

5. Discussion and conclusion

The present work has investigated the prosodic perception of social affects in Mandarin Chinese – a tonal language with four tones – in order to examine the potential relationship between the recognition of attitudes and tones. A Mandarin Chinese corpus of attitudinal speech was designed for a perceptual experiment with 15 French and 15 Vietnamese listeners.

The majority of attitudes have been recognized over the chance level, except “contempt” (3.8%), “irony” (2.1%) and confidence” (3.3%) for French subjects and “irony” (3.3%) for Vietnamese subjects. For all listeners, “declaration” attracted most answers and was mainly used when the listeners cannot identify any attitude. “Irony” was the least recognized attitude for both French and Vietnamese listeners. Concerning these less identified attitudes, they have been supposed to be strongly visually coded in speech communication [15]. Hence, another cross-language multimodal perceptual experiment will be carried out to investigate how the 19 audio-visual prosodic attitudes will be perceived by native and non native subjects. The results of analysis also showed that the tones have some influences on the perception of several social affects and the tonal effect is more significant for French subjects than for Vietnamese ones. As it was commonly accepted that there are cross-culturally similar uses of F_0 to signal affect, intention, or emotion [16], in order to validate the findings, our future work will involve an acoustic analysis of the production of social affects with emphasis on F_0 contour of tones which is the primary acoustic parameters to characterize Mandarin tones [17].

According to a hierarchical clustering, French and Vietnamese listeners have grouped the perceived social affects

in the same eight clusters, that differ to some extent from the seven groups made by Chinese subjects. This result is contrary to our hypothesis in which there should be more similarities between Chinese and Vietnamese listeners in respect to cognitive processing of social affects. More researches are required to explain this phenomenon.

6. Acknowledgements

The corpus recording and analysis couldn't have been conducted without the assistance of Christophe Savariaux and Lionel Granjon. This paper is partially supported by Chinese government within a PhD thesis grant.

7. References

- [1] Aubergé, V., Gestalt, A., “Morphology of Prosody Directed by Functions”, *Speech Prosody 2002 Proc.*, 151-154, Aix en Provence, France, 2002.
- [2] Fónagy, Y., “La Vive Voix”, Paris, Payot, 1991.
- [3] Wichmann, A., “Looking for attitudes in corpora”, in L. E. Breivik and A. Hasselgren [Ed], *Language and Computers, From the COLT's mouth ... and others'*, 247-261(15), 2002.
- [4] de Moraes, J. A., Rilliard, A., Alberto, B. and Shochi, T., “Production and perception of attitudinal meaning in Brazilian Portuguese”. *Speech Prosody 2010 Proc.*, Chicago, USA, 2010.
- [5] Lu, Y., Aubergé, V. and Rilliard, A., “Do you hear my attitude? Prosodic perception of social affects in Mandarin”, *Speech Prosody 2012 Proc.* (accepted), Shanghai, China, 2012.
- [6] Martins-Baltar, M., “De l'énoncé à l'énonciation: une approche des fonctions intonatives”, Didier, Paris, 1977.
- [7] Fujisaki, H. and Hirose, K., “Analysis and perception of intonation expressing paralinguistic information in spoken Japanese,” *ESCA Workshop on Prosody Proc.*, 254-257, Lund, Sweden, 1993.
- [8] Mejvaldova, J., “Expressions prosodiques de certaines attitudes en thèque et en français: étude comparative”, Université Paris 7 – Denis Diderot, Paris, 2000.
- [9] Diaferia, M. L., “Les Attitudes de l'Anglais : Premiers Indices Prosodiques”. Master thesis in Cognitive Science. National Polytechnique Institut of Grenoble, France, 2002.
- [10] Mac, D. K., Aubergé, V., Rilliard, A. and Castelli, E., “How prosodic attitudes can be recognized and confused: Vietnamese multimodal social affects”, *SLTU, Penang, Malaysia*, 2010.
- [11] Gu, W., Zhang T., and Fujisaki H., “Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes”, *Proceedings of Interspeech 2011, Firenze, Italy*, 1069-1072, 2011.
- [12] Shochi, T., Rilliard, A., Aubergé, V. and Erickson, D., “Intercultural Perception of English, French and Japanese Social Affective Prosody”. in S. Hancil [Ed], *The role of prosody in Affective Speech*, 31-59, Linguistic Insights 97, Peter Lang AG, Bern, 2009.
- [13] Michaud, A., “Tones and Intonation: Some Current Challenges”, keynote speech, 8th International Seminar on Speech Production (ISSP'08), pp. 13-18, Strasbourg, France, 2008.
- [14] Mac, D. K., Aubergé, V., Castelli, E. and Rilliard, A., “Local vs. global prosodic cues: effect of tones on attitudinal prosody in cross-perception of Vietnamese by French”, *Speech Prosody 2012 Proc.* (accepted), Shanghai, China, 2012.
- [15] Shochi, T., “Prosodie des affects socioculturels en japonais, français et anglais: à la recherché des vrais et faux-amis pour le parcours de l'apprenant”, PhD thesis in Language Science. University of Grenoble III- Stendhal, France, 2008.
- [16] Ohala, J. J., “The frequency codes underlies the sound symbolic use of voice pitch”, In L. Hinton, J. Nichols, & J. J. Ohala [Ed], *Sound symbolism*, 325-347, Cambridge University Press, Cambridge, 1994.
- [17] Allard, J., Wang, Y., Corinne, B. M. and Joan, A. S., “Perception and Production of Mandarin Chinese Tones”, In E. Bates, L.H. Tan, and O.J.L Tzeng [Ed], *Handbook of Chinese Psycholinguistics*, Cambridge University Press, Cambridge, 2006.