



Universiteit
Leiden
The Netherlands

Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch

Zhu, Y.

Citation

Zhu, Y. (2013, December 12). *Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch*. LOT dissertation series. Retrieved from <https://hdl.handle.net/1887/22850>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/22850>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/22850> holds various files of this Leiden University dissertation.

Author: Zhu, Yinyin

Title: Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch

Issue Date: 2013-12-12

Summary

Since Darwin published his book *The Expression of the Emotions in Man and Animals* in 1872, there has been an increasing number of studies on perception and production of emotions. Earlier studies were mainly conducted in the fields of psychology, physiology, biology; later they were extended into other areas, such as sociology, linguistics, pathology, computer science, neuroscience, musicology and second language acquisition. In addition, there have always been studies on perception or production of emotion cross-culturally or cross-linguistically. It is claimed by some researchers that perception of emotion is universal. However, other researchers believe that it is partly universal and partly cultural-or-language specific. Previous studies showed that emotion is generally better recognized when expressed by a speaker of the same cultural group as the listeners (in-group advantage). These studies also indicated that automatic recognition of human-produced emotions can reveal some of the acoustic cues that humans use to perceive and produce emotions. Although previous findings of perception and production of emotion are abundant, previous studies do not give us a clear picture of how well listeners of a non-tonal language can perceive emotions produced in a tonal language (especially through the audio channel). Neither do earlier studies give us a clear view of how well non-native speakers of a language can vocally produce emotion in the second or foreign language (L2) compared to native speakers, especially when the L2 is a tonal language but the L2 speakers' L1 is not. It is also not clear whether a speaker can vocally produce emotions in his L2 as well as he does in his native language (L1).

Therefore, the first aim of this dissertation was to investigate experimentally how well native (Mandarin) and non-native (Dutch) listeners of a tonal language (Mandarin) perceive vocal emotions portrayed in the tonal language. Non-native listeners included both naïve listeners and advanced L2 learners of Mandarin who shared the same L1 (Dutch) as naïve listener group. Secondly, I investigated whether Dutch L2 speakers of Mandarin are able to vocally produce emotions in the L2 as well as they do in their non-tonal L1; at the same time, I studied how well naïve native listeners and advanced learners of Mandarin perceive vocal emotion expressed by L2 speakers of the tonal language. An acoustic analysis was later conducted to analyze vocal correlates that speakers and listeners use in the production and perception of the vocal emotions. Finally, I investigated whether the in-group advantage reported by other researchers (claiming that listeners generally better recognize emotional prosody produced in their L1 than in an unknown language) is universal.

From a theoretical point of view, this dissertation aims to test a functional hypothesis, claiming that the prosodic space which languages may use, is finite. Therefore, if a language uses duration to mark a segmental contrast between long and short vowels, the duration parameter will not play a role (or a less important role) in the marking of stress – which in other languages depends rather heavily on duration cues (Berinstein 1979, Potisuk et al. 1999, Remijsen 2002a, b). By the same token, if a language such as

Mandarin uses pitch for lexical purposes (i.e. lexical tone), less room will be left for the signaling through pitch of paralinguistic contrasts, such as the expression of emotion. As a consequence of this I predict that native listeners of Mandarin will have limited exposure to clear exemplars of prosodically expressed affect. More generally, I predict that native listeners of a tonal language will be less intent on (and in fact less experienced in) decoding paralinguistic use of prosody than listeners of a non-tonal language.

In the introductory **Chapter 1** I summarize what has been done by previous studies, and then identify some unsolved issues. Background knowledge on tonal and non-tonal language, specifically Mandarin Chinese and Dutch, tone and (emotional) prosody, and the acoustic aspects of emotional prosody is provided in this chapter, too. I then itemize my research questions and derive specific predictions from the functional view mentioned above. I also motivate the choice of six emotional prosodies ('neutral', 'happy', 'angry', 'surprised', 'sad' and 'sarcastic') for the present study and propose feasible research methods needed to answer my the research questions. The end of this chapter provides an outline of the dissertation.

Chapter 2 reviews previous studies in more detail. There are many on the perception of emotional prosody within the same cultural group; however, research on cross-cultural perception of vocal emotion is relatively scarce. There are even fewer cross-linguistic studies on the production of vocal emotion. And there seem to be no earlier studies which directly compare the production of emotional prosody in a speaker's L1 and L2. Therefore, the production part in this dissertation is pioneering.

Many researchers have claimed that listeners can recognize emotional prosody above chance level both within-culturally and cross-culturally. However, emotion is generally better identified between speakers and listeners who are from the same cultural groups. The accuracy of recognition decreases as the cultural distance between two cultural groups is bigger. Previous studies also found that successful communication of vocal emotion depends on both the speaker and the listener, although the role of the listener seems to be more important than the role of the speaker. Some researchers indicate that production of emotional prosody cross-culturally may be universal to some extent; however, production of some emotions is cultural-specific.

In this chapter, I also provide information on the methodology used. There are three judgment studies in this dissertation. The first judgment study includes one perception experiment (Exp. 1), in which native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese perceived and identified the six Chinese emotional prosodies (see above) portrayed by native Chinese speakers. This experiment aimed to find an answer to research question (i) *How well can native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese perceive Chinese emotional prosodies vocally portrayed by Chinese native speakers? What will be the confusion patterns of the three listener groups?* The second judgment study included two perception experiments: in the first the same three listener groups listened to the same six Chinese emotional prosodies but produced by Dutch L2 speakers of Chinese (Exp. 2A). This experiment was designed to find answers to research question (ii) *How well can native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese perceive Chinese emotional prosodies vocally portrayed by Dutch L2 speakers of*

Chinese? What will be the confusion patterns of the three listener groups? In the second perception experiment (Exp. 2B), Dutch native listeners listened to the same six emotional prosodies portrayed in their native language (Dutch) by the same Dutch L2 speakers of Chinese. This was to test how well the same Dutch L2 speakers of Chinese produce the emotional prosodies in their L1. The results of this perception experiment were compared with the results obtained in the first perception experiment of the second judgment study to answer research questions (iii) *Can Dutch L2 speakers of Chinese produce emotional prosodies in the L2 as well as they do in the L1- Dutch? What will be the similarities and differences between these two types of production?* and (iv) *Does L2 limit the expression of emotional prosody, especially when the native language of L2 speakers of the tonal language is a non-tonal language?* Research question (v) *Is the functional view true, predicting that listeners of a tonal language might be less intent on (and in fact less experienced in) decoding the paralinguistic use of prosody than listeners of a non-tonal language?* was answered after the first and the second judgment study, questioning whether the functional view is true. An acoustic analysis was made of the stimuli used in the two judgment studies. The results answered the research question (vi) *What acoustic parameters contribute to differentiate between emotional prosodies in general? What acoustic correlates do speakers and listeners use to produce and perceive the vocal emotions in their L1 and in an L2? Do Dutch L2 speakers of Chinese use L1-transfer to produce emotional prosody in Chinese? To what extent automatic recognition reflects the perception of the emotional prosodies by the human listeners?* The third judgment study was conducted in a reciprocal way. It included two perception experiments in which Chinese and Dutch novice listeners perceived the six emotions vocally portrayed in their L1 and in the other language (Exp. 3). This experiment was designed to test whether the in-group advantage claimed by other researchers is universal, which was the research question (vii) *Is the in-group advantage universal, claiming that listeners are better in recognizing emotional prosody produced in their native language than in their L2 or an unknown language? Moreover, is the perception of vocal emotion cross-culturally symmetrical between Chinese and Dutch listeners, i.e., will Dutch and Mandarin listeners have similar abilities of identifying emotional prosody expressed in the other language?* The three judgment studies altogether answered the research question (viii) *Are perception and production of emotional prosody universal? Or are they more language-specific and culture specific?*

Chapter 3 reports the results of the first judgment study (Exp. 1), which was used as a baseline for the later studies. Twenty Chinese native listeners, 20 naïve Dutch listeners and 20 advanced Dutch L2 learners of Chinese recognized the six Chinese emotional prosodies (neutrality, happiness, anger, surprise, sadness and sarcasm). The results show that advanced Dutch L2 learners of Chinese recognized Chinese emotional prosody significantly (54 % correct) better than Chinese native listeners (46 % correct) and Dutch naïve listeners (46 % correct). The results also indicate that naïve non-native (Dutch) listeners could recognize emotions in an unknown language (Mandarin) as well as the natives did. Chinese native listeners did not show an in-group advantage for identifying emotions in Chinese more accurately and confidently. ‘Neutrality’ was the easiest emotion for all the three listener groups to identify and ‘anger’ was recognized equally well by all the listener groups. The prediction made in the beginning of the study is confirmed: listeners of a tonal language will be less proficient in the paralinguistic use of prosody than listeners of a non-tonal language. The results in this chapter provide the baseline for Chapter 4.

Chapter 4 investigates the differences between perception of six Chinese emotional prosodies (neutrality, happiness, anger, surprise, sadness and sarcasm) produced by Dutch L2 speakers of Chinese and those encoded by native Chinese speakers (control group). This chapter compares the results of Exp. 1 and Exp. 2A. Twenty Chinese native listeners, 20 naïve non-native listeners (Dutch) and 20 advanced Dutch L2 learners of Chinese listened to the Chinese emotional prosodies expressed by both L1 and L2 speakers of Chinese. The results show that the three listener groups recognized emotional prosodies encoded by Chinese natives significantly better (49% correct) than those produced by L2 speakers of Chinese (39% correct). Also, the naïve non-native listeners could recognize the emotions in the unknown language (46% correct) as well as the natives did. In terms of perceiving L2-produced Chinese emotional prosody, although the advanced Dutch L2 learners of Chinese performed slightly better than the other two listener groups, there was no significant effect among the three listener groups in terms of identifying L2-produced emotional prosody. The functional view is once again confirmed, which claims that listeners of a tonal language will be less proficient in the paralinguistic use of prosody than listeners of a non-tonal language. Therefore, in some cases at least, linguistic use of a particular acoustic feature in spoken language limits its use for the communication of emotion.

Chapter five gives the complete picture of how Dutch L2 speakers of Chinese produced the six emotional prosodies in their L2 (Chinese) and how they expressed the same emotional prosodies in their native language (Dutch). This chapter reports the results of the first and the second judgment study together; it was written from the production point of view. The results show that emotional prosodies produced by Dutch L2 speakers of Chinese were overall less recognizable ((39% correct by Chinese listeners) in their L2 than those encoded by Chinese natives (46% correct). Dutch L2 speakers of Chinese were better at vocally producing emotions in their L1 (57% correct by Dutch listeners). The prediction made in the beginning of the chapter is confirmed: speaking in an L2 limits the speaker's communication of emotion. The results also show that the naïve Dutch listeners were able to recognize the emotions in the unknown language as well as the natives Mandarin listeners did. Moreover, naïve Dutch listeners showed an in-group advantage: they identified emotions in Dutch more accurately than they did in Chinese.

Chapter 6 presents an acoustic analysis of three types of production of emotional prosody: L1 Mandarin, L2 Mandarin and L1 Dutch (the latter two were produced by the same individuals). Eight acoustic correlates were examined: tempo, mean fundamental frequency (pitch or F0), the standard deviation of the pitch (SD_F0), rate of change of the F0 (slope_F0), compactness of the spectral energy distribution, the standard deviation of the intensity (SD_int), jitter (cycle-to-cycle variation of the glottal pulses) and HNR (harmonics-to-noise ratio in the vocal signal). I also performed an automatic recognition of the six emotional prosodies portrayed by the three speaker groups, using Linear Discriminant Analysis (LDA). The acoustic analysis shows that fundamental frequency, including mean F0, SD_F0 and slope_F0, is an influential variable in the production of vocal emotions by the three groups of speakers. This finding confirms the study of Scherer (1996), who claimed that F0 plays a crucial role in the production of emotional prosody. Jitter and standard deviation of the intensity did not contribute much to differentiating between emotions in the present study. 'Tempo'

and ‘compactness’ were only sensitive to Mandarin L1 speakers. Slope of the F0 indicates that Chinese uses rising intonation to express surprise, which confirms previous studies, claiming that many tonal languages use rising intonation to express surprise (Yip 2006). Moreover, HNR can clearly distinguish ‘sad’ from ‘neutral’ in L2 Mandarin and L1 Dutch (produced by the same individuals), but not in L1 Mandarin. In summary, fundamental frequency is a very influential variable in the production and perception of vocal emotion in general. Other parameters studied in this chapter also contribute to differentiating between emotional prosodies, but they are more emotion-specific or speaker-type specific.

The results of the LDA show that the human-produced emotional prosody can be automatically recognized from the acoustic measures well above chance level (50% overall correct). There was significant correlation between confusions obtained by the automatic recognition and by the human listeners in the present study, so that automatic recognition reflects the perception of emotional prosody by human listeners to some extent. However, there may also be some other acoustic parameters contributing to the production and the perception of emotional prosody in general, which have been missed in this dissertation.

The acoustic analysis indicates that Dutch L2 speakers use some acoustic parameters in the production of emotional prosody in the L2 (Chinese) the same way they do in their L1 (Dutch), e.g. SD_F0 and SD_Int. Therefore, we may conclude that L1-transfer is a strategy for L2 speakers to vocally produce emotions in the L2. However, this strategy may not work for all the emotions, e.g. not for ‘surprise’ and ‘sarcasm’. Moreover, the results suggest that the L2 speakers did not automatically acquire the native approach to vocally produce emotional prosody in the target language during their L2 learning process. It seems that these advanced L2 speakers of Chinese have developed a hybrid system of producing emotional prosody in the L2. The hybrid system approximates the Chinese native manner of portraying vocal emotion to some extent (the way it involves tempo, mean F0, slope of the F0, compactness and jitter), but exploits the variability in F0 and intensity that the L2 speakers use to produce emotional prosody in their L1.

Chapter 7 investigates the perception of emotional prosody by native and novice listeners in a reciprocal way (Exp. 3). Twenty Chinese and 20 Dutch native listeners who do not have any knowledge of Dutch and Chinese, identified the emotional prosodies in these two languages. The results showed that novice Dutch listeners (46% correct) could recognize emotional prosody in the unknown language (Chinese) as well as natives did (46% correct); and they performed significantly better in identifying emotional prosody expressed in their native language (Dutch, 57% correct). In contrast to this, Chinese novice listeners were only able to recognize emotional prosody in their L1 reasonably well (46% correct) but failed to identify vocal emotion in the unknown language (Dutch, 15% correct) above chance level. This finding confirms the existence of the in-group advantage found by other researchers, claiming that listeners generally better recognize emotional prosody produced in their L1 than in an unknown language. Moreover, the results suggest that perception of vocal emotion cross-culturally is not symmetrical, so that some cultural group might be generally better than some other cultural group at perceiving emotional prosody. This, again, lends credibility to the

functional view which predicts that listeners of a tonal language will generally be less proficient in the perception of vocal emotion than listeners of a non-tonal language.

Chapter 8 reviews the main findings of this dissertation, and uses these to answer the research questions asked in Chapter 1. The chapter is concluded by a discussion of aspects that can be improved in the future.