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

Zhu, Y.

Chapter Four

Perception of Chinese Emotional Prosody Produced by Dutch Learners and Native Speakers of Chinese

Abstract

This chapter investigated 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).[11] Twenty Chinese native listeners, 20 naïve non-native listeners (Dutch) and 20 advanced Dutch L2 learners of Chinese participated in each of the perception experiments. The results showed that the three listener groups recognized emotional prosodies encoded by Chinese natives significantly better than those produced by L2 speakers of Chinese. Also, the naïve non-native listeners could recognize the emotions in the unknown language as well as the natives did. Chinese native listeners, therefore, did not show an in-group advantage (i.e., identifying emotions in Chinese more accurately). Moreover, advanced Dutch L2 learners of Chinese could recognize native-produced Chinese emotional prosody significantly better than the Chinese native listeners themselves. A functional view is confirmed, which claims that listeners of a tonal language will be less intent on the paralinguistic use of prosody than listeners of a non-tonal language. Furthermore, it seems that, in some cases at least, the linguistic use of a particular acoustic feature in spoken language limits its use for the communication of emotion.

[11] This chapter will appear as Y. Zhu (2013b) Perception of Chinese emotional prosody produced by Dutch learners and native speakers of Chinese. *Chinese as a Second Language Research*.

4.1 Introduction

Perception and production of emotion is an essential part in human/animal communication (Darwin 1872). Scherer (2000) claimed that emotion needs to be viewed as a multicomponent phenomenon that should be studied simultaneously from biological, cognitive, physiological, cultural and linguistic perspectives. Each of these aspects may contribute to the shaping of emotions, and affect the way in which emotions are expressed and perceived within and across cultures. In this paper I will study the perception of emotions that are expressed vocally, concentrating on the question how successfully emotions produced by speakers of one language are perceived by listeners of a different language. Earlier findings obtained in such cross-cultural and/or cross-linguistic studies have borne out that the perception of emotion is partly universal and partly language/culture-specific. Some emotions, such as 'anger', 'sadness', 'neutrality', are produced and perceived through universal means of expression, meaning that these emotions are generally recognizable even by different cultural groups. To the extent that the vocal expression of emotions depends on general biological and physiological mechanisms shared by all humans, some emotions are distinguished by specific properties that are shared across languages and cultures. Non-native listeners will be able to recognize these emotions even if they are expressed by speakers of another language. However, for some emotions, for example 'sarcasm', 'disgust' or 'shame', may well be expressed in different ways depending on the native language and culture of the speaker, and may therefore not be successfully identified by listeners from a different linguistic or cultural background. As a case in point, Van Bezooijen (1984) studied ten emotional prosodies: neutral, disgust, surprise, shame, interest, joy, fear, contempt, sad, and angry. Her study aimed to find out how (Taiwanese) Chinese and Japanese listeners without any knowledge of Dutch, perceived the Dutch emotional prosodies. All three listener groups recognized the Dutch emotional prosodies well above chance level, with scores of 66, 37 and 33% correct for Dutch, Taiwanese and Japanese listeners, respectively. The Asian listeners' identifications correlated at $r = .6$ with the Dutch identification percentages but correlated somewhat more strongly between Japanese and Taiwanese ($r = .7$). Each of the emotions was identified better by the native listeners than by the Asian listeners, which points to a strong and consistent in-group advantage. Yet, the native and non-native identifications were relatively close together for 'sadness', 'fear', 'surprise' and 'anger' (< 30 percentage points difference) whilst other Dutch emotions were identified quite poorly: e.g. 'joy' and 'shame' (both 22% correct against 76 and 61% correct for the native listeners). We assume that the communication of first group of vocal emotions very much relies on a universal code whereas the latter two depend largely on language-specific cues. Thompson and Balkwill (2006) conducted a different experiment in which 20 English-speaking listeners judged the emotive intent of utterances spoken by male and female speakers of English, German, Chinese, Japanese, and Tagalog. Identification accuracy was above chance for all emotions expressed in all languages. Across languages, 'sadness' and 'anger' were more accurately recognized than 'joy' and 'fear'. The (English) listeners showed an in-group advantage for decoding emotional prosody, with highest recognition rates for English utterances and lowest rates for Japanese and Chinese utterances. It would also indicate that, again, emotional prosody is decoded by a combination of universal and culture-specific cues. Pell et al. (2009) carried out a similar study, in which they compared how monolingual speakers of

Argentine Spanish recognize basic emotions from pseudo-utterances ('nonsense speech') produced in their native language and in three foreign languages (English, German, Arabic). Results indicated that vocal expressions of basic emotions could be decoded in each language condition at accuracy levels exceeding chance, although Spanish listeners performed significantly better overall in their native language ('in-group advantage'). These findings suggest that the ability to understand vocally-expressed emotions in speech is partly independent of linguistic ability and involves universal principles, although this ability is also shaped by linguistic and cultural variables.[12]

In addition, a few previous studies investigated the perception of native produced vocal emotion by L2 learners. And some of them also studied the correlation between the learner's ability to recognize emotions in the L2 and his/her L2 proficiency. For example, Graham et al. (2001) examined the ability of native and non-native speakers of English to identify emotions being portrayed by English speakers. They concluded that the ability to accurately identify emotions being portrayed through vocal cues in a second language (L2) may not be acquired by L2 learners without extensive exposure in a native context or without special attention to developing these skills in an instructional context. Moreover, an analysis of judgments made by learners of English as a Second Language (ESL) at different proficiency levels did not show an increase in ability to judge the emotional content of English speech with increased language proficiency. Chen (2005) studied how L2 English learners and L2 Dutch learners perceive emotional prosody in English and Dutch. She found that L1-transfer is an important strategy in interpreting pitch variation in L2. However, L2 learners may also activate their knowledge of intonational universals embodied in the biological codes. L2 learners at different levels seem to have acquired different degrees of understanding of the differences between their L1 and L2 and adjust their interpretation of pitch variation in L2 accordingly, with advanced L2 learners being more successful than beginning and intermediate learners. Shoshi and Gagnié (2010) investigated the differences in the perception of six culturally encoded French social affects through audio and visual channels for French native listeners, naïve Japanese listeners and trained Japanese learners of French. The trained Japanese learners of French recognized the emotions better than the naïve Japanese listeners; however, culture-specific attitudes (i.e. 'suspicious irony' and 'obviousness') were confused by Japanese listeners (including trained listeners). Facial information cues seemed to be more salient than auditory cues.

However, previous studies on perception of vocal emotion by different cultural groups mainly concentrated on how native and non-native listeners perceived emotion produced by native speakers. There was little attention for perception of emotions encoded by L2 speakers, especially encoded by L2 speakers of a tonal language (e.g. Mandarin). Moreover, previous studies on vocal communication between native and non-native speakers of Chinese have mainly been carried out in the area of perception

[12] Pell et al. (2009) report a significant in-group advantage but omitted the responses to one of the emotions ('neutral'). However, when the Pell et al. data are aggregated over all six emotional categories, there is no significant in-group advantage for the Argentinean Spanish listeners.

or production of Mandarin lexical tones by L2 learners of Chinese (Flege 1997, Gandour 1983, Leather 1990, Stagray & Downs 1993, Wang et al. 1999). Therefore, the present study has the following aims:

(1) Investigate to what extent (i) native, (ii) naïve non-native and (iii) advanced second-language learners of Chinese can perceive Mandarin emotions encoded vocally by L2 speakers and also to find out what would be the differences between these listener groups in perceiving Chinese emotion vocally produced by native speakers.

(2) Test whether an in-group advantage really exists, which means native listeners should get a significantly higher recognition rate than non-natives.

In order to avoid terminological inconsistency I only use the term 'emotional prosody' in this chapter, and use it to refer to both vocally produced emotions (e.g. happiness, sadness, anger, fear, disgust) and attitudes (e.g. sincerity, irony, sarcasm). In this study six Chinese emotional prosodies have been studied by using the discrete-emotion approach: 'neutrality', 'happiness', (hot) 'anger', 'surprise', 'sadness' and 'sarcasm'.

4.2 Methods

Two perception experiments were conducted. The first perception experiment aimed to test how native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese perceive Chinese emotional prosody produced by native Chinese speakers. The second perception experiment was designed to test how well the three listener groups recognize the same Chinese emotional prosodies when encoded by Dutch L2 speakers of Chinese.

4.2.1 Speakers

Four native Chinese speakers (2 males, 2 females, mean age = 45 years) whose mother tongue was standard Mandarin voluntarily took part in the recording of the stimuli for the first perception experiment. Four Dutch L2 speakers of Chinese (2 males, 2 females, mean age = 33 years) voluntarily participated in the recording of the stimuli for the second perception experiment. These four Dutch L2 speakers of Chinese, whose mother tongue was Dutch, were teachers in the Chinese department of Leiden University in the Netherlands. None of them were early bilinguals. They had learnt Chinese for 6 to 10 years, and they had been teaching Chinese for 2 to 10 years when the recording was made. All spent at least one year living or studying in mainland China or Taiwan.

4.2.2 Listeners

Twenty native Mandarin listeners (10 males, 10 females, mean age = 24 years), 20 naïve Dutch listeners (10 males, 10 females, mean age = 33 years) and 20 advanced Dutch learners of Chinese (10 males, 10 females, mean age = 20 years) voluntarily participated

in each of the perception experiments. The Chinese listeners were bachelor and master students at the University of Science and Technology Beijing who hailed from different parts of China. The naïve Dutch listeners were mainly bachelor students at the Humanities Faculty of Leiden University in the Netherlands and volunteers with variable education backgrounds. None of the naïve Dutch listeners spoke any Mandarin. The advanced Dutch learners of Chinese were mainly third-year BA students in the Chinese Program of Leiden University; the others were MA students and some outstanding second-year BA students. Early bilinguals were excluded; therefore, all students had learnt Mandarin after the age of puberty. There was no special course in the curriculum designed for training these students to recognize emotions in Chinese.

4.2.3 Materials and procedures

The first perception test used six Mandarin statements as vocal stimuli (e.g. *She is three months pregnant; He has been to Xiao Ge's place once*). Some of the sentences may be associated more readily with some emotions than with others but on aggregate the lexico-syntactic materials will not be biased towards specific emotions. Generally, speakers find it easier to pronounce meaningful sentences with specific emotions than they do with meaningless materials. This method has been widely used by other researchers in the vocal emotion study (e.g. Banse & Scherer 1996, Li et al. 2009, Van Bezooijen 1984, You et al. 2005, Zhang et al. 2006). We did not resort to the recording of meaningless *pseudo-utterances* (which has been proposed as an alternative solution by e.g. Castro & Lima 2010, Pell et al. 2009, Scherer et al. 1991) as these would be too difficult for L2 speakers of Chinese to vocally produce. The list of stimulus sentences is shown in Table 4.1.

Table 4.1. *Stimulus list in Chinese (Pinyin) with English glosses.*

1.	**Shì nǐ.* 'It is you.'
2.	*Xièxiè nǐ.* 'Thank you.'
3.	*Xiǎo wáng wánquán bù zhīdào zhè jiàn shì.* 'Xiao Wang does not know about this matter.'
4.	*Jīntiān xiàwǔ tā bùnéng lái cānjiā zhège huì.* 'He cannot attend the meeting this afternoon.'
5.	*Tā huáiyùn sān ge yuèle.* 'She is three months pregnant. '
6.	**Tā qùguò xiǎo gě jiā yì cì.* 'He has been to Xiao Ge's place once.'

Note: '*' means sentences were excluded in the second perception experiment. Macron '¯' = high-level tone, acute accent '´'= rising tone, haček 'ˇ' = falling-rising tone, grave accent '`' = falling tone; a syllable without tone mark has neutral tone.

Each of the six statements was expressed in six different emotions – neutrality, happiness, (hot) anger, surprise, sadness and sarcasm – by the four native Chinese speakers. The stimuli were digitally recorded (44.1 KHz, 16 bits) in a sound-proofed booth through a Logitech desk-top microphone. This procedure resulted in a stimulus set that consisted of 6 Chinese statements × 4 Mandarin speakers × 6 emotions = 144 discrete emotional utterances.

For the second perception experiment, the four Dutch L2 speakers of Chinese were asked to express the same six emotional prosodies in Chinese. The stimuli were digitally recorded under the same conditions as in the first perception experiment. Two sentences were discarded from the stimulus set (see Table 4.1), as these two sentences were less well perceived by the three listener groups in the first perception test. Therefore, the final stimulus set for the second perception experiment consisted of 4 Chinese statements × 4 Dutch L2 speakers × 6 emotions = 96 discrete emotional utterances. It made the second experiment shorter than the first one. In the comparison between the two experiments, I only used the shared materials.

In both perception experiments, all the participants including native Chinese listeners, naïve Dutch listeners and advanced learners of Chinese were asked to make a forced choice of the speaker's intended emotion, from the six given emotions, immediately after they heard a stimulus. They also gave a confidence rating to each choice they made. A three-level confidence rating scale was used, with the following interpretation: 3 = 'The speaker expressed the intended emotion well. I am very confident in my answer', 2 = 'The speaker reasonably expressed the intended emotion. But I am not so sure about my answer' and 1 = 'The speaker did not express the intended emotion well. I made the choice only by guessing.' The confidence scale was introduced in order to obtain a potential weighting factor such that responses given with greater confidence would be weighted more heavily than responses that were largely based on guessing. The first experiment lasted 25 minutes and the second one lasted 15 minutes, including the time for the participants to read the instructions in their native language before they started the test and a 6-second pause in between the emotional utterances for the participants to make a choice.

Each participant did the experiment individually in the presence of the experimenter. The stimuli were presented to the subject over closed headphones (but remained inaudible to the experimenter).

4.3 Results

4.3.1 Identification of emotions

The results proved insensitive to any weighting based on response confidence. Therefore, I report unweighted identification results only. Tables 4.2 (which repeats Table 3.2) and 4.3 are confusion matrices of intended versus perceived emotions in the two perception experiments by the three listener groups, i.e., native Chinese listeners, Dutch naïve listeners and advanced Dutch learners of Chinese.

Table 4.2 (= Table 3.2). *Perception of emotional prosody produced by native Chinese speakers: Confusion matrix of intended and perceived emotions by Chinese (upper panel), naïve Dutch (middle panel) listeners and advanced Dutch learners of Chinese (lower panel). Correct responses are located on the main diagonal (shaded).*

Intended	Ang	Hap	Neu	Sar	Sad	Spr
	Responded emotion by Chinese native listeners					
Angry	**56.3**	4.8	10.2	5.2	10.0	13.5
Happy	12.1	**37.3**	34.8	1.7	.8	13.3
Neutral	7.3	7.3	**73.5**	2.5	4.2	5.2
Sarcastic	11.7	17.5	34.0	**17.3**	3.1	16.5
Sad	13.3	8.1	32.7	4.4	**37.1**	4.4
Surprised	12.9	4.0	10.6	13.3	5.0	**54.2**
	Responded emotion by Naïve Dutch listeners					
Angry	**52.6**	4.0	15.1	9.5	7.9	10.9
Happy	35.7	**20.4**	14.1	5.8	3.6	20.4
Neutral	4.0	4.2	**71.2**	9.9	7.3	3.4
Sarcastic	13.3	6.5	15.7	**28.8**	16.1	19.6
Sad	5.2	2.4	28.6	10.5	**49.2**	4.2
Surprised	9.9	12.3	5.0	6.3	15.1	**51.4**
	Responded emotion by advanced Dutch learners of Chinese					
Angry	**53.3**	2.7	16.5	10.0	5.0	12.5
Happy	30.2	**25.2**	14.2	3.5	2.1	24.8
Neutral	5.2	.8	**80.2**	2.9	9.6	1.3
Sarcastic	11.9	8.3	17.3	**31.9**	10.4	20.2
Sad	5.8	1.3	19.6	6.7	**65.4**	1.3
Surprised	6.0	9.4	2.5	10.0	3.8	**68.3**

Table 4.3. *Perception of emotional prosody produced by advanced Dutch L2 speakers of Chinese: Confusion matrix of intended and perceived emotions by Chinese (upper panel), naïve Dutch (middle panel) listeners and advanced Dutch learners of Chinese (lower panel). Correct responses are located on the main diagonal (shaded).*

Intended	Ang	Hap	Neu	Sar	Sad	Spr
	Responded emotion by Chinese native listeners					
Angry	**25.6**	6.3	34.4	12.8	8.1	12.8
Happy	3.4	**37.8**	21.3	12.2	3.1	22.2
Neutral	2.5	8.4	**63.1**	3.8	18.8	3.4
Sarcastic	13.1	15.9	27.2	**21.3**	11.9	10.6
Sad	8.4	2.5	27.8	5.3	**47.2**	8.8
Surprised	7.8	19.4	16.6	9.7	8.1	**38.4**
	Responded emotion by Naïve Dutch listeners					
Angry	**38.4**	4.7	14.4	17.2	12.5	12.8
Happy	14.1	**29.4**	12.8	11.3	8.4	24.1
Neutral	5.0	4.7	**60.0**	10.3	16.6	3.4
Sarcastic	13.8	15.6	18.4	**19.1**	14.7	18.4
Sad	6.9	1.9	25.9	9.7	**42.2**	13.4
Surprised	7.5	20.6	14.1	9.4	11.9	**36.6**
	Responded emotion by advanced Dutch learners of Chinese					
Angry	**33.8**	8.1	23.4	13.8	8.4	12.5
Happy	6.3	**33.1**	17.5	12.8	3.8	26.6
Neutral	2.8	6.3	**59.1**	5.3	24.7	1.9
Sarcastic	9.4	16.3	19.7	**22.5**	12.5	19.7
Sad	5.9	3.4	22.8	6.6	**51.6**	9.7
Surprised	5.0	17.8	18.1	7.8	7.8	**43.4**

The confusion matrices show that native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese perceived the six Chinese emotional prosodies produced by native Chinese speakers (mean recognition rate: 48.4%) substantially better than those encoded by Dutch L2 speakers of Chinese (mean recognition rate: 39.0%). For the perception of native-produced emotional prosody show quite different confusion patterns in identifying native produced emotional prosody. For instance, Chinese listeners tended to mistake 'happiness' mainly for 'neutrality' (34.8%) while naïve Dutch listeners massively misidentified 'happiness' as 'anger' (35.7%). In the perception of non-native emotional prosody, native Chinese listeners and naïve Dutch listeners showed a surprisingly similar confusion structure for the six emotions. For example, both native Chinese and Dutch naïve listeners strongly confused 'happiness' with 'surprise' (22.2% and 24.1%, respectively). Moreover, native Chinese and Dutch naïve listeners showed the same tendency of confusing 'sarcasm' with 'neutrality'.

In the perception of native-produced Chinese emotional prosody, advanced Dutch learners of Chinese performed significantly better than the other two listener groups. However, there was no significantly better listener group in the perception of L2 produced Chinese emotional prosody. Moreover, in both of the perceptual experiments, the confusion categories which advanced Dutch learners of Chinese fell into are quite similar to those of naïve Dutch listeners. For example, in the perception of native-produced emotional prosody, naïve Dutch listeners and advanced Dutch learners of Chinese both mistook 'anger' for 'neutrality' by 15.1% and 16.5%, respectively, and for 'surprise' by 10.9% and 12.5%, respectively. They both misrecognized 'happiness' as 'anger' by 35.7% and 30.2%. Furthermore, in perceiving non-native-produced emotional prosody, advanced Dutch learners of Chinese showed the exact same tendency as naïve Dutch listeners for 'sarcasm': they often confused 'sarcasm' with 'neutrality' (19.7%) and 'surprise' (19.7%); and naïve Dutch listeners confused it with 'neutrality' (18.4%) and 'surprise' (18.4%). In addition, in the second perception experiment the two Dutch listener groups both dramatically misidentified 'happiness' as 'neutrality' and 'surprise'; and they also confused 'neutrality' with 'sadness'. These findings support Chen's (2005) conclusion that L1-transfer is an important strategy in interpreting paralinguistic intonational meaning (e.g. emotional prosody) in L2.

Figure 4.1A-B shows the percent correct identification of six intended emotions by native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese in the two perception experiments. Figure 4.1A presents the results of the three listener groups perceiving emotional prosody produced by native Chinese speakers. Figure 4.1B shows the results of the three listener groups recognizing emotional prosody encoded by Dutch L2 speakers of Chinese.

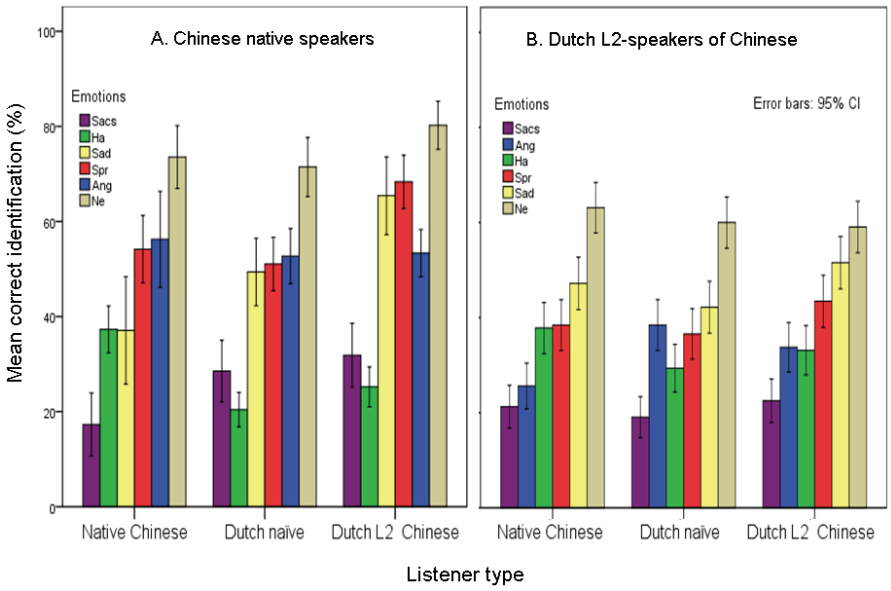

Figure 4.1A-B. *Percent correct identification of six intended emotions by native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese. Panel A presents the perceptual results of emotional prosody produced by native Chinese speakers (exp. 1). Panel B presents the perceptual results of emotional prosody encoded by advanced Dutch L2 speakers of Chinese (exp. 2). In each cluster of bars the intended emotions are ordered from left to right in ascending overall correct recognition rate. The order is indicated from top to bottom in the legends of the panels; note that the order differs between panels. The 95% confidence limits were computed for each bar on 20 listener means. Panel A repeats Figure 3.1B.*

Table 4.2, Table 4.3 and Figure 4.1A-B together indicate that native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese were able to recognize discrete Chinese emotional prosody, whether produced by native or by non-native Chinese speakers, above chance level (mean recognition rates: 48.6% and 39.0%, chance level: 16.7%). Moreover, emotions were identified (much) better than chance by each of the three listener groups in the two perception tests. In the perception of native-produced emotional prosodies, even the Dutch naïve listeners obtained a score of 45.6% correct, closely followed by the native Chinese listeners (45.9% correct), and with the best performance obtained by the advanced Dutch learners of Mandarin (54.1% correct). Furthermore, the difference between the three listener groups is statistically significant by a one-way Analysis of Variance, $F(2, 57) = 5.8$, $p = .005$. A Bonferroni post-hoc test ($\alpha = .05$) showed that the advanced Dutch learner group performed better than the other two groups in perception of native-produced emotional prosody. The other two listener groups did not differ from each other. In the perception of non-native-produced Chinese emotional prosody, there was no statistical significance between the three listener groups, even though advanced Dutch learners of Chinese performed slightly better than the other two groups (2% or 3% higher). This indicates that native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese performed

equally well/poorly in perceiving Chinese emotional prosody encoded by L2 speakers of Chinese.

Somewhat surprisingly, the success with which native Chinese listeners and Dutch naïve listeners identified vocal emotions in each of the perception experiments was correlated. Emotions that native listeners found difficult (or easy) to identify were also found difficult (or easy) for naïve listeners. For example, both groups identified 'anger', 'surprise', and 'sadness' more successfully than 'happiness' and 'sarcasm' in the first perceptual experiment. However, they found 'sadness' and 'surprise' less difficult to recognize than 'anger' in the second perceptual experiment. In Figure 4.1A, the order of difficulty among the six emotions was somewhat different for the advanced Dutch learners of Chinese than for the other listener groups. Specifically, the advanced Dutch learners of Chinese showed much higher identification rates for the emotions of 'sadness', 'surprise' and 'neutrality' portrayed by the native speakers. This finding supports Chen's study (2005) that L2 learners at different levels seem to have acquired different degrees of understanding of the differences between their L1 and L2, and adjust their interpretation of pitch variation in L2 accordingly.

4.3.2 Confidence rating

In the second part of this results section, I will analyze the confidence ratings. Although, as mentioned earlier in this section, there was no effect of weighting on the results and only unweighted identification results were presented, I would like to make use of the confidence ratings all the same to investigate the social behaviour of the listener groups. In this case, I just present means and observe unexpected differences between the groups.

Figure 4.2A-B shows the confidence rating of six intended emotions by native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese in the two perceptual experiments. Figure 4.2A shows that Chinese native listeners were less confident than the non-native listeners (mean = 1.49) in their identifications of native-produced emotions. Within the Dutch listeners, the advanced learners of Mandarin were more confident (mean = 2.29) than the naive listeners (mean = 1.96). The effect of listener group is significant by a one-way ANOVA, $F(2, 57) = 45.4$, ($p < .001$). Bonferroni post-hoc tests revealed that all differences between the three groups were significant ($\alpha = .05$). Figure 4.2B shows the opposite result that Chinese native listeners were as confident as advanced Dutch learners of Chinese (mean = 2.30), but Dutch naïve listeners were the least confident (mean = 1.92). Therefore, it can be concluded that native Chinese listeners are more confident in identifying emotional prosody produced by (Dutch) non-native speakers. The reason for this behavior is not clear since one would expect listeners to be more confident when having to make decisions based on materials produced by speakers who share the same linguistic code.

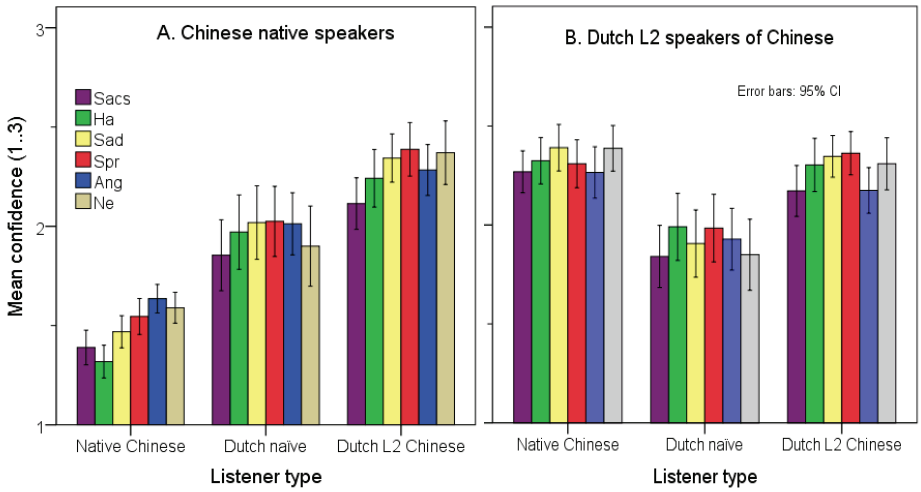

Figure 4.2A-B: *Confidence rating (3 = most) of six intended emotions by native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese. Panel A presents the perceptual results of emotional prosody produced by native Chinese speakers. Panel B presents the perceptual results of emotional prosody encoded by advanced Dutch L2 speakers of Chinese. In each cluster in both panels the intended emotions are ordered as in Figure 4.1A. Confidence limits are based on 20 listener means.*

4.3.3 Peripheral findings of the production of emotional prosody by L2 speakers[13]

In the last part of this section, I would like to address briefly the production of the Dutch L2 speakers who produced the stimuli for the present study. The results show that even advanced Dutch L2 speakers of Chinese are generally not as good as native speakers at vocally expressing emotions in Chinese. This could possibly be explained along the lines of Ross et al. (1986), who found that less use was made of short-term changes in F0 to express emotion in tone languages (in which short-term F0 contours are used to carry lexical information) than in Indo-European languages (in which F0 plays no lexical role). It would appear, then, that in some cases at least, linguistic use of a particular acoustic feature in spoken language limits its use for the communication of emotion. Obviously, native Chinese speakers can produce emotional prosody in Chinese without having problems of getting the lexical tones right. Therefore, we may assume that native speakers of a tonal language automatically encode lexical information when producing emotions in their native language. However, L2 speakers of Chinese may not know how to pronounce Chinese lexical tones correctly while at the same time expressing emotional prosody on top of the lexical tones. Perhaps, that is why the three listener groups did not perceive the non-native-produced emotional prosodies as well as those encoded by natives. It is unclear whether Dutch L2 speakers

[13] A more comprehensive account of the production of emotional prosody in L2 and L1 is presented in Chapters 5 and 6.

of Chinese used an L1-transfer strategy during their production of Chinese emotional prosody, meaning that they would use Dutch vocal cues to express emotions in Chinese. If they did, then the naïve Dutch listeners should have picked up the Dutch features straightaway. However, the Dutch naïve listeners did not show any better scores in this case. Or it might be that L2 speakers have used the L1-transfer strategy in producing emotion in their second language. But it may not have been easy for Dutch naïve listeners to pick up the cues, if they could not distinguish between Chinese lexical tones and emotional prosody. In other words, L1-transfer strategy in terms of production of emotional prosody might not work on L2 speakers of a tonal language. The viability of this speculation can only be checked when an acoustic analysis is applied. I will come back to this point in Chapter 6.

4.4 Conclusions and discussion

The results of this investigation indicate that native-produced emotional prosodies were recognized better by all the listener groups than those expressed by L2 speakers. In other words, emotional prosodies produced by L2 speakers of Chinese were less recognizable overall than those encoded by natives. Nevertheless, the three listener groups could recognize emotions well above chance level, regardless the speaker type. Moreover, the results showed that the three listener groups could recognize some negative emotions equally well, regardless the speaker type, for example, 'anger' and 'sadness'. More specifically, the native-produced 'anger' was recognized by the three listeners groups equally well in first perception experiment. These findings support previous studies which claimed that recognizing negative emotions from vocal cues, independently of language, is compatible with evolutionary views that vocal signals associated with threat that much must be highly salient to ensure human survival (Ohman et al. 2001, Tooby & Cosmides 1990). According to Darwin's evolution theory, 'anger' signals aggression towards the offended and warns the offender to expect an aggressive counter-reaction. In other words, 'anger' symbolizes danger to both the offender and the offended. Therefore, this emotion should be recognized equally well by all human beings irrespective of their linguistic and cultural backgrounds, as people/ animals have an instinct to sense, and protect themselves from, danger. Furthermore, 'neutrality' is identified most accurately by all the listener groups in this investigation, which finding is in line with previous literature (Cornew et al. 2010). However, one might claim that neutrality is the default response category so that its correct identification rate is higher than that of other emotions as a result of response uncertainty. Therefore, we can infer that emotion perception could be universal to some extent.

The recognition rates of 'happiness' and 'sarcasm' were relatively low in the two perception experiments. This finding is compatible with the previous finding that 'happiness' was recognized relatively poorly in Mandarin where the emotion must be recognized through audio channel only (e.g. Banse & Scherer 1996, Castro & Lima 2010, Liu & Pell 2012). And the recognition rates of 'happiness' by the three listener groups also varied regardless of speaker type. This implies that perception of positive emotions depends more on listener's linguistic and cultural background. Moreover,

even some primary emotions (e.g. 'anger') can be expressed in variable ways by different speaker groups. For instance, 'anger' encoded by the L2 speakers was identified poorly by all the listeners. It implies that, if an emotional prosody was not able to trigger the language-specific cues borne in L1/L2 listeners' linguistic knowledge, the recognition rate of the prosody should decrease to some extent. Altogether, we can conclude that perception of emotional prosody can be partly universal and partly language-or-culture specific. It means that, on one hand, native and non-native listeners (including L2 listeners) might have drawn on the universal resources embedded in human beings to decode some primary emotional prosodies, e.g. 'anger', 'sadness' or 'neutrality'. On the other hand, they might have also resorted to their own language-or-culture specific cues, which vary very much from culture to culture, when perceiving emotional prosody produced by different speaker groups. According to the results of this chapter, it seems that the non-native produced emotional prosodies neither properly triggered the universal cues nor the language-or-culture-specific cues of all the listener groups. This is why they were not recognized as well as those produced by the native speakers.

There are also some other findings in this chapter. In the perception of native-produced emotional prosodies, Chinese native listeners were not able to identify emotions more accurately and confidently than naïve Dutch listeners and advanced Dutch learners of Chinese. Surprisingly, advanced Dutch learners of Chinese recognized emotional prosody in Chinese significantly better than Chinese natives did themselves. This finding contradicts the conclusion of Graham's study (2001) that the ability to accurately identify emotions being portrayed through vocal cues in a second language may not be acquired by L2 learners without extensive exposure to such emotions in a native context or without special attention to developing these skills in an instructional context. Moreover, advanced Dutch learners of Chinese can identify Chinese emotional prosody substantially (and significantly) better than naïve Dutch listeners. Possible explanations for this finding can be found below. This finding confirms the result of Shoshi and Gagnié's study (2010) that trained second language learners may recognize emotional prosody in the target language better than listeners with no experience in the target language.

Moreover, naïve non-native listeners can recognize unknown emotional prosody as well/poorly as natives, regardless of speaker type. The in-group advantage found by other researchers therefore does not apply universally to all cultural groups (e.g. Chinese), which means that native listeners would perform significantly better than non-native listeners in perceiving emotional prosody in their L1. Natives and naïve non-natives may have drawn on very similar cognitive resources when identifying emotional prosody; even the incorrectly recognized emotional prosodies of natives and non-natives may fall into similar confusion categories. However, the detailed cognitive resources are still not known at the present stage. It might be that there are some universal cognitive resources shared by the two listener groups. Advanced learners of Chinese followed a slightly different order of success in the perception of native-produced emotional prosody. It indicates that emotions which are difficult for native and naïve non-native listeners to identify, are not necessarily difficult for them to recognize, for example: sadness and surprise. These findings support the conclusions of

Chen's study (2005) to some extent: L1-transfer is an important strategy in interpreting pitch variation in L2. L2 learners at different levels of proficiency seem to have acquired different degrees of understanding of the differences between their L1 and L2 and adjust their interpretation of pitch variation in L2 accordingly.

Finally, I will briefly summarize some additional findings that relate to the performance of the two speaker groups in the present study.

Firstly, L2 speakers are not able to vocally produce emotions in their L2 as well as natives, even though their Chinese proficiency is high. This finding supports previous studies (Gorelick & Ross 1987, Lieberman & Michaels 1962, Ross et al. 1986, Scherer et al. 1984): although spoken language constrains emotional expression to some extent, linguistic and emotional expression can be dissociated and typically function independently of one another. From this observation we can possibly conclude that a second language might constrain emotional expression more than a first language does, especially when the second language is a tonal language (e.g. Chinese).

Secondly, we do not know at this stage whether this L1-transfer strategy is also used by L2 speakers in production of emotional prosody in their L2, since the Dutch non-native listeners did not pick up any Dutch vocal cues from the Chinese emotional prosodies encoded by Dutch L2 speakers of Chinese. Otherwise, they would have scored better than the native listeners in the perceptual test.

There may be several possible explanations for the findings that (i) L2-produced emotional prosodies were overall less recognizable than those produced by natives, and (ii) Chinese emotions were identified more successfully by (advanced) Dutch learners of Mandarin than by native Chinese listeners themselves.

First of all, Ross et al. (1986) have shown there is less use of short-term changes in F0 to express emotion in tone languages (in which short-term F0 contours are used to carry lexical information) than in Indo-European languages (in which F0 plays no lexical role). Thus it seems that, in some cases at least, use of a particular acoustic feature in spoken language limits its use for the communication of emotion. This insight is incorporated into a functional view which claims that the prosodic space which languages may use is finite. The parameters (or dimensions) of the phonetic space (and of the prosodic space within it) can be used to express linguistic as well as paralinguistic contrasts. The functional principle holds that one can use a particular parameter in the phonetic space only once. It follows from the functional principle that if a language uses a prosodic parameter for linguistic purposes, it can no longer use the same parameter for non-/paralinguistic uses – or, in a less extreme version of the theory – cannot use the same parameter as effectively for the expression of paralinguistic or extralinguistic meanings. The prediction follows that speakers of a lexical tone language (such as Mandarin) have less room to express emotion through prosody (specifically through paralinguistic use of speech melody) than speakers of a non-tone language (such as Dutch or English). Apparently, native Chinese speakers can filter out Mandarin lexical tones automatically during the production of emotional

prosody in their native language, but L2 speakers of Chinese cannot. In this case, L2 speakers of Mandarin cannot easily separate emotional prosody from lexical tones during their production of Chinese emotional prosody, so that they cannot express it as well as natives. As a consequence of a functional view, listeners of a tonal language will be less intent on (and well in fact be less experienced in) decoding the paralinguistic use of prosody than listeners of a non-tonal language. In other words, listeners of a non-tonal language are generally better at recognizing emotional prosody than listeners of a tonal language. This would explain why naïve Dutch listeners can recognize Chinese emotional prosody as well as natives, and why advanced Dutch L2 learners of Chinese can identify the same emotions even better.

It is worthwhile rerunning this experiment with different linguistic groups to see if the results are similar, for example: British naïve listeners and British L2 learners of Chinese; or German naïve listeners and German learners of Chinese. The ultimate test of this explanation would be to examine emotion recognition by speakers of another tone language, e.g. Vietnamese or Thai. The prediction, obviously, would be that such listeners should recognize emotions in Mandarin more poorly than native Chinese listeners do – since (i) they are relatively insensitive to emotional prosody because their mother tongue is a tone language, and (ii) because being non-native listeners they are not familiar with the expression of emotion in the target language.

Secondly, the unexpected results might be caused by the absence of particles in the Chinese stimuli. In everyday Chinese speech particles often appear at the end of a sentence, carrying considerable emotional information that is alternatively expressed by intonation in other languages. Since this kind of lexical markers were deliberately left out in the present study due to the research purpose, their absence might have affected the perception of the emotional prosodies by L1 listeners but not by the Dutch listeners. Moreover, Dutch listeners might generally be more intent on the message in the sentence prosody, according to the functional view. That is possibly why the Chinese L1 listeners did not perform better than the non-native listeners in the perception experiments in which the stimuli with no final particles attached were presented through audio channel only. Testing this hypothesis is beyond the scope of the present study. Part of the endeavor would be to determine how much use Mandarin and Dutch make of particles expressing emotions on the part of the speaker and what the division of work would be between the use of such particles and emotional prosodies.

Thirdly, Chinese society is quite reserved when it comes to the overt expression of emotion, either in speech or in other modes of communication (Klineberg 1938). Showing emotion in public is interpreted as a sign of weakness in China (Wu & Tseng 1985). If this is indeed the case, then native speakers of Chinese will have had little exposure to clear instances of vocally expressed emotions. This would explain why the native Chinese listeners did relatively poorly when instructed to identify vocally expressed emotions in Chinese. It would also explain why Dutch listeners obtained equal or better identification rates for the Chinese emotions than the native listeners themselves. Especially the advanced Dutch learners of Chinese can pick up some

Chinese subtle emotions produced by native speakers (e.g. sadness and sarcasm) more successfully than natives.

Further studies could be carried out in the areas of second-language acquisition and cognitive psychology to find out more about the perception and production of emotional prosody by natives and non-natives/L2 learners of the target language.