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

Zhu, Y.

Chapter Five

Production of Emotional Prosody in L2 and in L1

Abstract

This chapter investigated how well Dutch L2 speakers of Chinese produced the six emotional prosodies (neutrality, happiness, anger, surprise, sadness and sarcasm) in their L2 (Chinese) and how well they produced the same emotional prosodies in their L1 (Dutch).[14] Two recognition studies were carried out in this chapter. The first recognition study was designed to test how well Dutch L2 speakers of Chinese produced the six emotional prosodies in Chinese. Native Chinese speakers participated in the first recognition study as the control group. The second recognition study aimed to find out how well the same Dutch L2 speakers of Chinese expressed the same vocal emotions in their native language – Dutch. Twenty Chinese native listeners, 20 naïve listeners (Dutch), and 20 advanced Dutch L2 learners of Chinese participated in the first recognition study and another 20 Dutch native listeners participated in the second recognition study as listeners/judges. The results showed that emotional prosodies produced by L2 speakers of Chinese in their L2 were overall less recognizable than those encoded by Chinese natives. Dutch L2 speakers of Chinese are better at vocally producing emotions in their L1 than in the L2. The prediction made in the beginning of this chapter is confirmed, which claims that second language limits L2 speakers' communication of emotion. A detailed acoustic analysis of selected stimuli is deferred to Chapter 6. The results also show that the naïve Dutch listeners could recognize the emotions in the unknown language (Mandarin Chinese) as well as the natives did. Moreover, naïve Dutch native listeners showed an in-group advantage in that they identified the same emotions in Dutch more accurately than in Mandarin Chinese.

[14] This chapter is the first part of Y. Zhu (2013). Production of emotional prosody in L2 and in L1 (submitted).

5.1 Introduction

Perception and production of emotion is an essential part of human/animal communication (Darwin 1872). The research question 'can listeners infer emotion from vocal cues?' has been studied by many researchers (e.g. Frick 1985, Scherer 1986, Standke 1992, Van Bezooijen 1984). These studies all show that listeners are rather good at inferring affective state and speaker attitude from vocal expression. Furthermore, the previous studies also claim that the vocal expression of emotions is differentially patterned (Scherer 1996). There is considerable evidence that emotion produces changes in respiration, phonation and articulation. A large number of different emotional and motivational states are indexed and communicated by specific acoustic characteristics of the concurrent vocalizations (Scherer 1989). The acoustic variables that are strongly involved in the production of vocally expressed emotion are summarized in Scherer's (1991, 1996) studies. However, previous studies mainly touched on the vocal production of emotion by native speakers from one particular linguistic group but not on speakers' L2.

Another finding is from Ross et al. (1986). They have shown that 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 typically 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. Inspired by Ross et al. I would like to predict that, if a language uses a prosodic parameter for linguistic purposes, it will have less space for non-/para-linguistic uses of the same cue. If this prediction were true, it would effectively mean 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).

Therefore, I carried out the present study to:

(1) Investigate: (i) how L2 speakers of Chinese vocally produce emotions in Chinese; and how they portray the same emotions in their L1; and what would be the differences; (ii) what would be the differences between Chinese native and L2 speakers of Chinese vocally producing emotion in Chinese; (iii) will a tonal language limit the vocal production of emotion?

(2) As a secondary aim, 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 to find out how these listener groups perceive Chinese emotions vocally produced by native speakers.

There is little literature which studied the first research question properly, especially when the target language is a tonal language, such as Mandarin. Anolli et al. (2008) conducted research on vocal production of emotion by Chinese and Italian young adults. They confirm that different emotions may be expressed through variations in the modulation of vocal cues, in both cultures; on the other hand, differences in the specific patterns of vocal cues in expressing emotions were identified between Chinese and Italian participants. Fortunately, there are a few studies which touched on perception of emotional prosody by both native and non-native listeners. To some

extent, previous findings all claimed that perception of emotion by different culture groups is partly universal and partly language/culture-specific. For instance, 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, who did not have any knowledge of Dutch, perceived Dutch emotional prosodies at the sentence level. Perceptual experiments showed that Dutch native listeners got the highest correct identification rate and Japanese listeners performed poorest. But both of the listener groups performed well above chance level. 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. Thompson and Balkwill (2006) conducted a similar 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. This would indicate that, again, emotional prosody is decoded by a combination of universal and culture-specific cues. Shoshi and Gagnié (2010) investigated 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 did; 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.

This chapter will focus on the vocal production of emotion in speakers' L2 and L1. 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, anger, surprise, sadness and sarcasm. [15] Sentences expressed in different emotions are used as stimuli for the perception experiment. There is no semantic link between the sentences.

[15] Discrete emotion theory assumes that humans universally express and recognize a small number (six to eight) of basic cross-culturally shared 'core' emotions, which are communicated through innate mechanisms (for a survey of positions see Ekman & Friesen 1971, Colombetti 2009; see also footnote 8).

5.2 Methods

Two recognition studies were conducted: the first recognition study aimed to test how well Dutch L2 Chinese speakers vocally expressed emotions in their second language, compared to Chinese native speakers (the control group). Actually, this recognition study is the combination of the first and the second judgment study presented in Chapters 3 and 4, respectively. It is now reviewed from the production perspective; the second recognition study was designed to test how well the same Dutch L2 speakers of Chinese vocally produced emotions in their mother tongue i.e. Dutch. Three groups of listeners who were used as judges voluntarily participated in the first recognition study; 20 native Dutch listeners were used as judges in the second recognition study.

5.2.1 Speakers

Four Dutch L2 speakers of Chinese (2 males, 2 females, mean age = 33 years) voluntarily participated in the recording of the stimuli for the two recognition studies. These four Dutch L2 speakers of Chinese, whose mother tongue was Dutch, were teachers from 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 at the time the recordings were made. All had spent at least one year living or studying in mainland China or Taiwan. In order to set up a control group for the first recognition study, 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 perception experiment. The four Chinese speakers were amateur actors/actresses who all had stage performance experience.

5.2.2 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 the first recognition study. They were asked to decide which emotion was intended by the speaker and how confident they were of their choice. The results can tell us how well the speakers had produced the six emotions in their L1 and L2. 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 at 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 eighteen. There was no special course in the curriculum designed for training these students to recognize emotions in Chinese.

Twenty Dutch native listeners who did not have any Chinese knowledge voluntarily participated in the second recognition study as listeners. They were bachelor or master students at Leiden University, majoring in linguistics. Although they were different subjects from those naïve Dutch listeners who took part in the first recognition study, both groups represent the same population statistically.

5.2.3 Materials and procedures

5.2.3.1 First recognition study

The first recognition study includes two perception experiments: the first perception experiment was set up to test how well the Chinese control group vocally expressed the six emotions in Chinese. In this experiment, I used six Mandarin statements (e.g. *She is three months pregnant; He has been to Xiao Ge's place once*). The reasons that I selected these six sentences particularly are: (1) these six sentences contain all the tones in Mandarin, i.e. 'high-level tone', 'rising tone', 'falling-rising tone', 'falling tone' and 'neutral tone' (e.g. Howie 1976); (2) according to the consensus of the speakers, these sentences are semantically neutral but can easily be expressed with different emotions; (3) both short and longer sentences were included, in case utterance length might play a role in the perception of emotional prosody. Each of the six statements was expressed in six different emotions (neutrality, happiness, anger, surprise, sadness and sarcasm).

The second perception experiment was carried out to find out how well the Dutch L2 speakers of Chinese encoded the six emotions in their second language, compared to the control group. In this experiment, two Mandarin statements were discarded from the original six Mandarin statements, as they were not very well perceived by the three groups of listeners in the first perception experiment. The list of stimulus sentences for the first recognition study is shown in Table 5.1.

In the first perception experiment, each of the six Mandarin statements was vocally expressed in six different emotions (neutrality, happiness, 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.

In 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 5.1), 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.

Table 5.1. *Stimulus list in Chinese (Pinyin orthography) 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 sentence was 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.

In both perception experiments, all the participants (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 of 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 mainly by guessing.' This confidence scale was introduced as a potential weighting factor. It would enable us to see which emotional utterances were identified by the listeners with more confidence and which were not. Therefore, we would later be able to compute the recognition rates based on the weighting. The first experiment lasted 25 minutes and the second one lasted 15 minutes, including the time for the listeners to read the instructions (in their native language) before they started the experiment and a 6-second pause in between the emotional utterances for the listeners 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).

5.2.3.2 Second recognition study

The second recognition study only included one perception experiment, in which 20 native Dutch listeners perceived the six emotions produced by the same four Dutch L2 speakers of Chinese, but in their mother tongue, i.e. Dutch. In this experiment, the four Mandarin statements used in the first recognition study were translated into Dutch by

the four Dutch L2 speakers of Chinese where sentence length, syntactic structure, syllables and sentence meaning were well controlled. Therefore, the final stimulus set for the perception experiment consisted of 4 Dutch statements × 4 Dutch L2 speakers × 6 emotions = 96 discrete emotional utterances. The list of stimulus sentences for the second recognition study is shown in Table 5.2. The same procedure as in the first recognition study was used to obtain the judgments.

Table 5.2. *Stimulus list of Dutch sentences with broad IPA transcription and English glosses.*

1.	*Dank je wel.* dɑŋk jə ʋɛl 'Thank you.'
2.	*Xiaowang weet dat helemaal niet.* ʃau ʋɑŋ ʋetɑt heləmal nit 'Xiao Wang does not know about this matter.'
3.	*Vanmiddag kan hij niet naar de vergadering.* ʋɑmɪdɑχ kɑni nit nar də ʋərɣadərɪŋ 'He cannot attend the meeting this afternoon.'
4.	*Zij is drie maanden zwanger.* zɛɪ ɪs dri mandə zʋɑŋər 'She is three months pregnant. '

5.3 Results

At the beginning of the present study, I explained the reason of introducing a confidence rating scale, which would be used as a potential weighting factor. However, it turned out that there was no effect of weighting on the results, according to the statistical analysis. Therefore, I report unweighted identification results in this chapter only.

5.3.1 Results of production

5.3.1.1 Production of emotional prosody in speakers' L2

Tables 5.3 and 5.4 (repeated and extended versions of Tables 3.2 and 4.2, respectively) 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. 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 (overall recognition rate: 48.7%) substantially better than those encoded by Dutch L2 speakers of Chinese (overall recognition rate: 39.3%). Figure 5.1A (which repeats

Figures 3.1B and 4.1A) presents the results of the three listener groups perceiving emotional prosody produced by native Chinese speakers (the control group). Figure 5.1B (which repeats Figure 4.1B) shows the results of the three listener groups recognizing emotional prosody encoded by Dutch L2 speakers of Chinese.

Table 5.3. *Perception of Chinese 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). This table repeats and extends Tables 3.2 and 4.2).**

Intended	Responded emotion by Chinese native listeners							Grand mean
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	46.0	
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**		
Total	20.4	13.5	26.6	6.3	13.4	19.6	100	
	Responded emotion by Naïve Dutch listeners							
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	46.0	
Sarcastic	13.3	6.5	15.7	**28.8**	16.1	19.6		48.7
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**		
Total	19.1	7.4	25.8	10.4	17.8	18.6	100	
	Responded emotion by advanced Dutch learners of Chinese							
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	54.0	
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**		
Total	16.5	7.8	24.8	9.8	20.2	21.1	100	

*Note: 'Mean' = mean correct identification rate of each listener group; 'Grand mean' = mean correct identification rate of the three listener groups.

Table 5.4. *Perception of Chinese emotional prosody produced by 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). This table repeats and extends Table 4.3).**

Intended	Responded emotion by Chinese native listeners							Grand mean
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	39.0	
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**		
Total	10.9	16.2	27.0	9.1	20.2	16.4	100	
	Responded emotion by Naïve Dutch listeners							
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	38.0	
Sarcastic	13.8	15.6	18.4	**19.1**	14.7	18.4		39.3
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**		
Total	16.8	12.9	26.3	8.4	18.5	16.1	100	
	Responded emotion by advanced Dutch learners of Chinese							
	Ang	Hap	Neu	Sar	Sad	Spr	Mean	
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	41.0	
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**		
Total	13.7	13.5	24.0	9.1	21.0	17.6	100	

*Note: 'Mean' = mean correct identification rate of each listener group; 'Grand mean' = mean correct identification rate of the three listener groups.

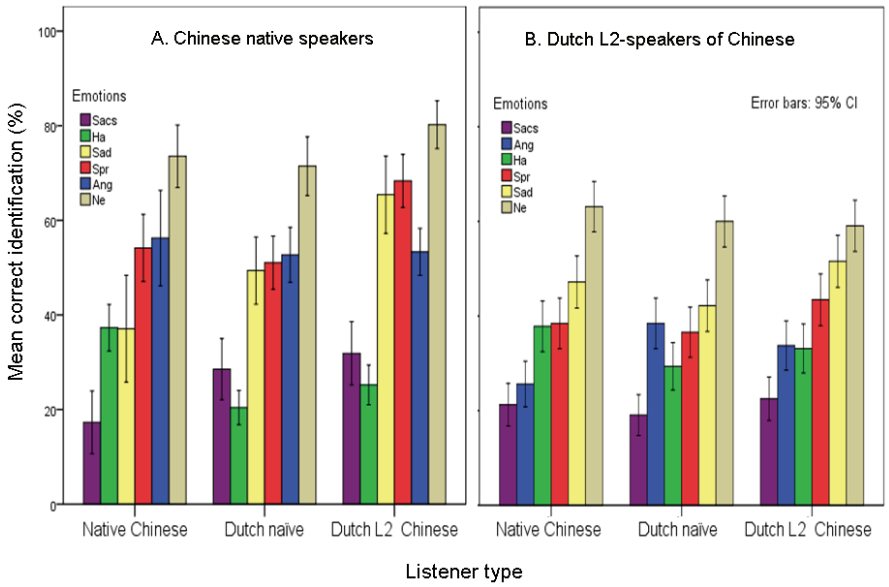

Figure 5.1A-B (= Figure 4.1A-B). *Percent correct identification of six intended Chinese emotions by native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese in the two perception experiments. Figure A presents the perceptual results of Chinese emotional prosody produced by native Chinese speakers. Figure B presents the perceptual results of Chinese emotional prosody encoded by Dutch L2 speakers of Chinese. The correct recognition rate of native Chinese speakers is 10 percentage points higher than that of Dutch L2 speakers of Chinese.*

Tables 5.3-4 and Figure 5.1A-B together indicate that both native and non-native produced Chinese emotional prosodies were recognized by the three listener groups above chance level (chance level: 16.7%). It means that Dutch L2 speakers of Chinese were able to vocally produce emotions in the L2. However, given the significant difference between the mean recognition rates of the three listener groups in the two perception experiments, we can conclude that Dutch L2 speakers of Chinese are generally not as good as native Chinese speakers are at vocally expressing emotions in Chinese. Specifically, the native-produced Chinese emotional prosodies (mean = 48.7%) were significantly better recognized overall than those produced by the L2 speakers (mean = 39.3%) by the three listener groups. This would mean that, in some cases at least, L2 speakers might not have the same ability as natives of expressing emotional prosody in the L2, even though their language proficiency of the L2 is very high. In the present study, native Chinese speakers can produce emotional prosody in Chinese without having problems of getting the lexical tones right. Therefore, we can assume that native speakers of a tonal language could automatically work out 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. In other words, even though Chinese lexical tones might limit the production of Chinese emotional

prosody to both L1 and L2 speakers, they might limit L2 speakers more. Perhaps, that is why the three listener groups did not perceive non-native-produced Chinese emotional prosodies as well as those encoded by Chinese natives.

5.3.1.2 Production of emotional prosody in speakers' L1

The second recognition study only included one perception experiment in which 20 Dutch native listeners were used as judges to test how well the same four Dutch L2 speakers of Chinese had produced the six emotional prosodies in their L1. Table 5.5 is the confusion matrix of intended versus perceived emotions in the perception experiment by the Dutch native listeners. It shows that the emotional prosodies produced in Dutch by the same Dutch L2 speakers of Chinese were recognized by the native Dutch listeners above chance level (chance level = 16.7%). Moreover, the overall correct recognition rate of the Dutch native listeners increased dramatically from 39% when the emotional prosodies expressed in speakers' L2-Chinese to 57% when the emotional prosodies were produced in the speakers' mother tongue, i.e. Dutch. This indicates that the four Dutch L2 speakers of Chinese are able to express emotional prosodies both in the L2 and in their native language. However, they are better at producing emotional prosody in their L1 than producing it in their L2. It further supports the claim that L2 limits an L2 speaker's production of emotional prosody.

Table 5.5. *Perception of Dutch emotional prosody produced by Dutch L2 speakers of Chinese: Confusion matrix of intended and perceived emotions by native Dutch listeners. Correct responses are located on the main diagonal (shaded and bolded).*

Intended	Responded emotion by Dutch native listeners						
	Ang	Hap	Neu	Sar	Sad	Spr	Mean
Angry	**55.6**	9.5	9.0	12.5	7.9	3.7	
Happy	2.3	**44.9**	4.6	17.4	1.2	12.5	
Neutral	15.9	13.0	**68.3**	17.6	19.7	5.3	57.0
Sarcastic	13.4	7.9	3.9	**34.3**	6.3	6.0	
Sad	8.1	9.7	13.4	4.4	**64.6**	.0	
Surprised	5.6	15.0	.7	13.9	.5	**72.5**	
Total	16.3	13.1	20.0	10.0	18.9	21.2	100

5.3.2 Perception results

Although the focus of this chapter is the vocal production of emotion in speakers' L2 and L1, I would like to analyse the perceptual performance of the listener groups in the two recognition studies, as production can never be separated from perception in the study of speech, especially not in the study of emotional prosody. I believe that investigating the perception of the emotional prosodies can tell us more about the production of the emotions.

5.3.2.1 Perception of native and non-native Chinese emotional prosodies

As can be seen from Table 5.3, for the perception of native-produced emotional prosody, native Chinese listeners and naïve Dutch listeners show quite different confusion patterns. For instance, Chinese listeners tended to mistake 'happiness' mainly for 'neutrality' (34.8%) while naïve Dutch listeners massively confused 'happiness' and 'anger' (35.7%). In the perception of non-native produced 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 mistaking 'sarcasm' for 'neutrality'.

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). 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 were no statistically significant differences between the three listener groups, even though advanced Dutch learners of Chinese performed slightly better than the other two groups (2 or 3 percentage points 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.

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 perceiving non-native-produced emotional prosody, advanced Dutch learners of Chinese showed the exact same tendency as naïve Dutch listeners for 'sarcasm': they often misidentified 'sarcasm' as 'neutrality' (19.7%) and 'surprise' (19.7%); and naïve Dutch listeners mistook it for 'neutrality' (18.4%) and 'surprise' (18.4%). These observations suggest that L1-transfer is an important strategy in interpreting paralinguistic meaning (e.g. emotional prosody) in L2.

5.3.2.2 Perception of the Dutch emotional prosodies by the Dutch listeners

Some confusion tendencies shown in Table 5.4 can be also seen in Table 5.5. For instance, the Dutch native listeners tended to mistake 'anger' mainly for 'sarcasm' (12.5%) when the emotion was produced in their L1; the naïve Dutch listeners also strongly misperceived 'anger' as 'sarcasm' (17.2%) when the emotion was expressed in Chinese (L2). Moreover, the same thing happened with perceiving 'sadness' and 'surprise': the naïve Dutch native listeners in both recognition studies confused 'sadness' mainly with 'neutrality', and confused 'surprise' mainly with 'happiness' regardless

the language in which the emotional prosodies were produced. From this, we can infer that both listeners and speakers use L1-transfer as a strategy in perception and production of emotional prosody in L2. Furthermore, we can also infer that knowing the meaning of the utterances does not influence the perception of emotional prosody very much. In other words, perception of emotional prosody is universal to some extent.

5.3.3 Combining the two recognition studies

In this section, I will report a summary analysis of the two recognition studies. Figure 5.2 presents the percentage of correctly identified emotions by seven combinations of speaker and listener type. Braces define speaker-listener combinations that do not differ significantly from each other (Bonferroni post-hoc test with $\alpha = .05$).

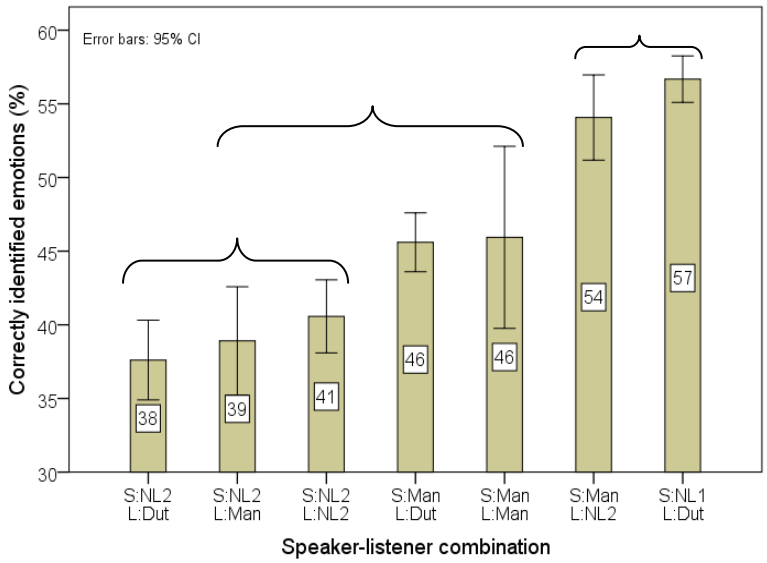

Figure 5.2. *Percent correct identification (%) of combined speaker and listener types.* Conditions under the same brace do not differ significantly from each other by a Bonferroni post hoc test ($\alpha = .05$).*

* Note: 'S'= speaker type; 'L' = listener type; 'Man'= native Mandarin speaker; 'Dut'= native Dutch listener (naïve); 'NL2' = emotional prosodies produced by Dutch L2 speakers of Chinese in their L2-Mandarin; 'NL1'= emotional prosodies produced by Dutch L2 speakers of Chinese in their L1, i.e. Dutch.

As can be seen from Figure 5.2, emotions, whether expressed in one's L1 or L2, are overall recognized above chance level, regardless listener type. It shows that perception and production of emotional prosody is universal to some extent. However, emotions vocally produced in the speaker's L1 are much more recognizable than those expressed in the speaker's L2. It indicates that, although the speaker is able to produce emotional prosody both in his L1 and L2, he produces emotional prosody in his L1 more

recognizably than in his L2, especially when the L2 is a tonal language. Also, native Dutch listeners identified emotions expressed in their L1 more successfully than those produced in Chinese. This finding supports the finding of Thompson and Balkwill's (2006) study that L1 listeners show an in-group advantage by decoding emotional prosody in their L1 more successfully than in other, non-native languages.

Surprisingly, advanced Dutch learners of Chinese recognized emotional prosodies produced by native Chinese speakers significantly better than native Chinese listeners and naïve Dutch listeners did. It indicates that advanced learners of a second language who have acquired high language proficiency of the L2 are better at interpreting paralinguistic meanings in the L2 than naïve listeners do. This finding is also compatible with Shoshi and Gagnié's (2010) finding that trained Japanese learners of French recognized the French emotions better than the naïve Japanese listeners.

5.4 Conclusions and discussion

The results of this chapter indicate that emotional prosodies produced by L2 speakers of Chinese were less recognizable overall than those encoded by natives. In other words, L2 speakers are not able to vocally produce emotions in their L2 as well as natives are, even though their Chinese proficiency is high. Furthermore, neither are they able to vocally portray emotions in their L2 as well as they doing in their L1. These finding confirm previous studies (Gorelick & Ross 1987, Lieberman & Michaels 1962, Ross et al. 1986, Scherer et al. 1984): spoken language constrains emotional expression to some extent; and these two systems can be dissociated and function independently of one another. From this point, we can possibly conclude that speaking in second language might constrain emotional expression more than first language does. However, the three listener groups could recognize emotion well above chance level, regardless the speaker type. Moreover, 'neutrality' is identified most accurately by all the listener groups in the present study, which finding is in line with previous literature (Cornew et al. 2010). Therefore, we can infer that emotion production is universal to some extent.

Native-produced 'anger' is recognized reasonably well in the two recognition studies, but 'anger' encoded in Chinese by the L2 speakers of Chinese is identified poorly by all the three listener groups. It seems that the L2 speakers might have their own interpretation of how to express basic emotions in the L2. Therefore, we could assume that L2 speakers might not be able to produce basic emotions in their L2 as effectively as they do in their L1, although basic emotions should have contained more universal acoustic cues than non-basic emotions, according to Darwin's evolution theory. Therefore, we could further assume that emotion production in an L2 depends very much on the L2 speaker's understanding of the L2, even for some basic emotions, such as 'anger'. Furthermore, the results suggest that production of emotional prosody in one's L1 is the combination of universal acoustic cues and culture-or-language-specific variables.

In the perception of native-produced Chinese emotional prosodies, Chinese native listeners are not able to identify emotions more accurately than naïve Dutch listeners

and advanced Dutch learners of Chinese did. Surprisingly, advanced Dutch learners of Chinese recognize emotional prosody in Chinese significantly better than the natives do themselves. This finding contradicts the conclusion of Graham's (2001) study 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 significantly (and substantially) better than naïve Dutch listeners. This finding confirms the result of Shoshi and Gagnié's (2010) study that trained second language learners recognize emotional prosody in the target language better than listeners with no experience in the target language.

There may be several possible explanations for the findings that Dutch L2 speakers of Chinese were not able to produce emotional prosody in their L2 (i) as well as Chinese natives and (ii) as successfully as in their L1.

First of all, Ross et al. (1986) have shown that 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 typically 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. Inspired by Ross et al., one might predict that the prosodic space which languages may use is finite. The parameters (or dimensions) of the phonetic space (and the prosodic space within it) can be used to express linguistic as well as paralinguistic contrasts. In other words, 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 pronounce Mandarin lexical tones correctly without thinking during the production of emotional prosody in their native language, but L2 speakers of Chinese cannot. In this case, L2 speakers of Mandarin are not able to easily separate emotional prosody from lexical tones during their production of Chinese emotional prosody, so that they cannot express it as well as natives. It can also explain why Dutch L2 speakers cannot vocally produce emotions as well as they do in their L1. As a consequence of the prediction, listeners of a tonal language will be less intent on (and in fact 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 worth 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.

Secondly, 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 Dutch L2 speakers could not produce emotional prosody in their L2 as well as natives, which is simply due to the same reason that they lack clear input of exemplars of emotional prosody produced in Chinese.

In order to better understand the production of emotional prosody in the speakers' L2 and L1, I carried out an acoustic analysis, which is presented in the next chapter. Three groups of speakers, i.e., L1 Dutch speakers, L2 Mandarin speakers and L1 Mandarin speakers (the former two are the same individuals), will be studied in this chapter.