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

Zhu, Y.

Chapter Seven
Perception of Emotional Prosody in a Listener's L1 and in an Unknown Language

Abstract

This chapter investigates the perception of emotional prosody by native and novice listeners in a reciprocal way. Twenty Chinese and 20 Dutch native listeners without any knowledge of Dutch and Chinese, respectively, identified emotional prosodies in these two languages. The results show that novice Dutch listeners could recognize emotional prosody in the unknown language (Chinese) as well as natives; and they performed significantly better in identifying emotional prosody expressed in their native language (Dutch). Chinese novice listeners, on the other hand, were able to recognize emotional prosody in their L1 only reasonably well but failed to identify vocal emotion in the unknown language (Dutch) 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 cross-cultural perception of vocal emotion is not symmetrical, meaning that some cultural group might be generally more sensitive than some other cultural group in the perception of emotional prosody. This finding lends support to the functional view that predicts that listeners of a tonal language should be generally less intent on the perception of vocal emotion than listeners of a non-tonal language. If this view is accepted, the asymmetry in emotion perception may not only be explained from a difference in culture but also as the result of a difference in linguistic structure.

7.1 Introduction

Studies on perception of vocal emotions cross-culturally have been widely carried out since Darwin claimed that affective expressions, including those produced via the vocal channel, are veridical (Darwin 1872). Earlier findings obtained in cross-cultural and/or cross-linguistic studies have borne out that the perception of emotion is partly universal and partly language/culture-specific. Perception of some vocal emotions, for instance, 'anger', 'sadness' or 'neutrality', depends on general biological and physiological mechanisms shared by all humans, meaning that listeners will be able to recognize these emotions even if they are expressed in an unknown language. However, 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. For instance, Albas et al. (1976) asked male Caucasian speakers of English and Amerindian Cree speakers to express the basic emotions of happiness, sadness, anger, and love, in their respective native languages using any words that came to their minds. These speech samples were electronically low-pass filtered and presented to Caucasian and Cree listeners. The results showed that the emotions were recognized above chance level in all cases but better in the listener's native language than in the unknown language. Therefore, perception of emotion is partly universal and partly culture or language specific. The authors suggest that language and culture are crucial factors in the transmission of emotion, even on the nonverbal level, but admit that the data are difficult to interpret because the content of the speech materials used for encoding was not controlled. 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 listeners and Japanese, respectively. The Asian listeners' identifications correlated at $r = .6$ with the Dutch identification percentages but correlated somewhat more strongly between Japanese and Taiwanese ($r = .7$). 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 the first group of vocal emotions relies very much on a universal code whereas the latter two depend largely on language-specific cues. Moreover, Scherer et al. (2001) report results from a study conducted in nine countries in Europe, the United States, and Asia on vocal emotion portrayals of anger, sadness, fear, joy, and neutral voice as produced by professional German actors. Data show an overall accuracy of 66% across all emotions and countries. Although accuracy was substantially better than chance, there were sizable differences ranging from 74% in Germany to 52% in Indonesia. Generally, accuracy decreased with increasing language dissimilarity from German in spite of the use of language-free speech samples. It is concluded that culture- and language-specific paralinguistic patterns may influence the decoding process. The results of these studies also showed a tendency for vocal emotion to be generally better recognized within the same cultural group. This tendency is called *ethnic bias* by some theorists (e.g. Kilbride & Yarczower 1983, Markham & Wang 1996), claiming that it is possible that recognition

accuracy is higher when emotions are both expressed and perceived by members of the same cultural group. A preferred term for the same tendency is called 'in-group advantage' by other researchers, further claiming that listeners generally better recognize emotional prosody produced in their L1 than in an unknown language. For example, Elfenbein and Ambaby (2002) conducted a meta-analysis which examined emotion recognition within and across cultures based on literature search. They concluded that emotions were universally recognized at better-than-chance levels. Accuracy was higher when emotions were both expressed and recognized by members of the same national, ethnic, or regional group, suggesting an in-group advantage. They also found that recognition of emotion is partly universal and partly cultural-specific. More recently, Thompson and Balkwill (2006) conducted an experiment in which twenty 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, and 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). On the basis of their findings the authors argued that the ability to understand vocally expressed emotions in speech is partly independent of linguistic ability and involves universal principles, although this ability is additionally shaped by linguistic and cultural variables.[18]

Previous studies typically investigated perception of vocal emotion cross-culturally one-way, i.e., vocal emotion encoded in language A was perceived by native listeners and other culture groups B, C, D, etc. (e.g. Scherer et al. 2001). Or, in some cases, culture group A perceived emotional prosody expressed in its L1 and in several other (and often unknown) languages (e.g. Thompson & Balkwill 2006). However, studies which investigated perception of vocal emotion by different cultural groups in a reciprocal manner are rare. In the reciprocal approach both culture groups A and B perceive emotional prosody not only expressed in their own, native language (A > A, B > B) but also emotions expressed in the other language (A > B, B > A). As an example of the latter situation, English listeners may be asked to recognize emotional prosody in Japanese, and vice versa. Although some studies (e.g. Albas et al. 1976, Dennis 1982, Gitter et al. 1972) used this reciprocal approach, the two cultural groups involved were

[18] 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.

also ethnically different rather than just culturally-or-linguistically dissimilar. Other studies also adopted the reciprocal method (e.g. Ekman 1972) but only investigated the perception of facially expressed emotions between two culture groups. Even though previous studies have clearly indicated an in-group advantage in the cross-cultural perception of emotion, the reciprocal method was not used so that those studies present an incomplete picture, especially when it comes to the perception of vocal emotion. Therefore, I conducted the present study applying the reciprocal method to two cultural groups whose cultures and languages are distant from each other. In the present study, I chose Mandarin and Dutch listener groups for this research purpose. This is because Mandarin is a Sino-Tibetan tonal language which uses a wide pitch range (with pitch movements up to 12 semitones, Xu 1999), has monosyllabic words and a simple syllable structure (Duanmu 2007a, b), while Dutch is an Indo-European non-tone language, with a rather restricted pitch range (De Pijper 1983, 't Hart et al. 1990), with often long, polysyllabic (compound) words that may contain complex consonant clusters (Booij 1995). Moreover, Chinese and Dutch cultures are very much dissimilar, as one is Asian culture and the other is West-European culture. Therefore, the present study asks the following questions:

1) How well can Chinese and Dutch novice native listeners identify emotional prosody in Chinese and Dutch, and vice versa? In other words, what is the difference between the two listener groups in perceiving emotional prosody in their L1 and in an unknown language?
2) Is the cross-cultural perception of vocal emotion symmetrical between Chinese and Dutch, i.e., will Dutch and Mandarin listeners have similar abilities of identifying emotional prosody expressed in the other language?

My prediction for research question 2 is negative. 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. A functional principle holds that one can use a particular parameter in the phonetic space only once (e.g. Berinstein 1979, Potisuk et al. 1997, Remijsen 2002a, b). It follows from this 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. As a consequence of the functional view, 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. Therefore, I would like to further predict that Dutch listeners would be overall better than Chinese listeners in correctly identifying emotional prosody in the present study.

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 emotional prosodies have been studied: neutrality, happiness, (hot) anger, surprise, sadness and sarcasm.

7.2 Methods

Two perception experiments were conducted in the present study: the first perception experiment was designed to test how well Chinese native and novice Dutch listeners of Chinese perceive emotional prosody produced in Chinese. The second perception experiment aimed to find out how well the same two listener groups perceive the same emotional prosodies but portrayed in Dutch. This is the reciprocal situation: Chinese natives became the novice listeners of Dutch.

7.2.1 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 native Dutch speakers (2 males, 2 females, mean age = 33 years) voluntarily participated in the recording of the stimuli for the second perception experiment. These four native Dutch participants were also advanced L2 speakers of Chinese who learnt Chinese for 6 to 10 years and had been teaching Chinese as a foreign language for 2 to 10 years when the recordings were made. These speakers can easily switch between Dutch and Chinese. They translated the stimuli from Chinese to Dutch for the second perception experiment (see below). Most of the native Chinese and Dutch speakers had experience in stage performance in their mother tongue. Moreover, a mood induction technique in which different background stories were told to the speakers to express a stimulus in different emotions was applied.

7.2.2 Listeners

Twenty native Mandarin listeners (10 males, 10 females, mean age = 24 years) and 20 native Dutch listeners (10 males, 10 females, mean age = 33 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; all spoke Mandarin. The native 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 native Dutch listeners spoke any Mandarin; neither did the native Chinese listeners speak any Dutch. The two listener groups were novice listeners of each other's native language (Chinese and Dutch). There was neither a special course in the curriculum nor any pre-test training designed for training these listeners to recognize emotions in their L1 or in an unknown language.

7.2.3 Materials and procedure

The testing materials for the two perception experiments were designed in a particular order. The stimuli were first collected and produced in Chinese and then translated into Dutch and portrayed in Dutch afterwards. The rationale behind this is that Chinese, a tone language, made the collection of stimuli more difficult, as the stimuli in Chinese had to meet the certain requirements (details see below). Dutch is a non-tonal language in which the requirements can be met with less difficulty. Therefore, it was easier to collect the stimuli in Chinese first and translate them into Dutch afterwards.

I selected six Mandarin statements as vocal stimuli (e.g. *She is three months pregnant; He has been to Xiao Ge's place once*) for the first perception test. The requirements for the stimulus selection are: (1) stimuli 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) stimuli have to be semantically neutral but can easily be expressed with different emotions; (3) both short and longer sentences have to be included, in case utterance length might play a role in the perception of emotional prosody.

According to the consensus of the native Chinese speakers and the Dutch L2 speakers of Chinese, the ensemble of six selected sentences met the requirements adequately. The stimuli were then translated into Dutch by the four Dutch L2 speakers of Chinese where sentence length, syntactic structure, syllables and sentence meaning were well controlled. In fact, 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. Each of the six statements was expressed in six different emotions (neutrality, happiness, anger, surprise, sadness and sarcasm). The list of stimulus sentences in Chinese is shown in Table 7.1.

Table 7.1. *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: '*': 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.

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 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 7.1), as these two were less well perceived by the two listener groups in the first perception test. Therefore, the final stimulus set for the second perception experiment consisted of 4 Dutch statements × 4 Dutch L2 speakers of Chinese × 6 emotions = 96 discrete emotional utterances. The list of stimulus sentences in Dutch is shown in Table 7.2. It made the second experiment shorter than the first one. In the comparison between the two experiments, I only used the shared materials.

Table 7.2. *Dutch stimulus sentences with Broad IPA transcriptions 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.'

In both perception experiments, both of the listener groups including native Chinese and Dutch listeners 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 expressed the intended emotion moderately well. I am not so sure of 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 great 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 listeners to read the instructions in their native language before they started the test and a 6-second pause in between the emotional sentences for the participants to make a choice.

Each listener 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).

7.3 Results

The results proved insensitive to any weighting based on response confidence. Therefore, I report unweighted identification results only. Tables 7.3 and 7.4 are confusion matrices of intended versus perceived emotions in the two perception experiments by the two listener groups, i.e. native Chinese listeners and native Dutch listeners.

Table 7.3. *Perception of emotional prosody produced in Chinese by native Chinese speakers: Confusion matrix of intended and perceived emotions by Chinese (upper panel) and novice Dutch listeners (lower panel). Correct responses are located on the main diagonal (shaded).*

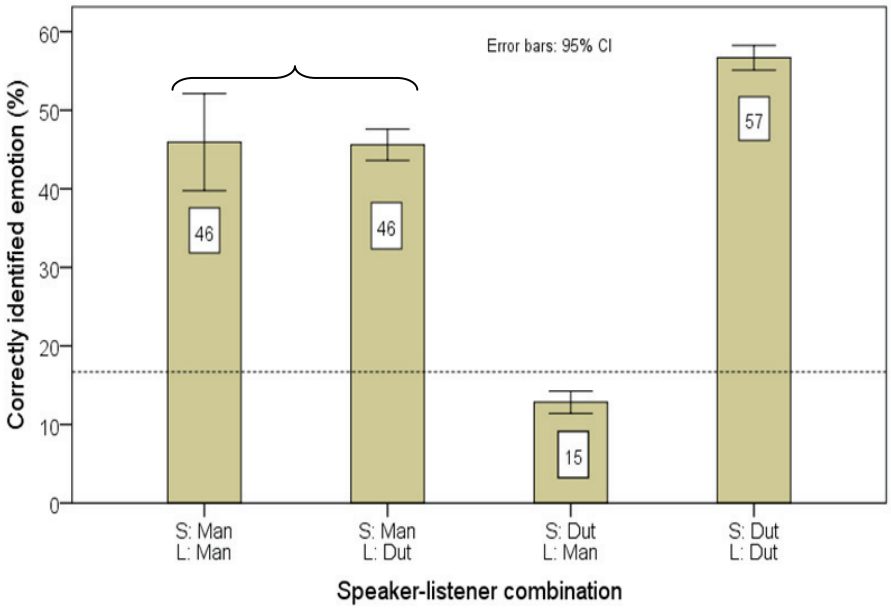
Intended	Ang	Hap	Neu	Sar	Sad	Spr	Overall correct
	Responded emotion by Chinese native listeners						
Angry	**56.3**	4.8	10.2	5.2	10.0	13.5	
Happy	12.1	**37.3**	34.8	1.7	.8	13.3	
Neutral	7.3	7.3	**73.5**	2.5	4.2	5.2	
Sarcastic	11.7	17.5	34.0	**17.3**	3.1	16.5	46.0
Sad	13.3	8.1	32.7	4.4	**37.1**	4.4	
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	
Responded emotion by Novice 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	46.0
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	

Table 7.4. *Perception of emotional prosody produced in Dutch by native Dutch speakers: Confusion matrix of intended and perceived emotions by native Dutch (upper panel) and novice Chinese listeners (lower panel). Correct responses are located on the main diagonal (shaded).*

Intended	Ang	Hap	Neu	Sar	Sad	Spr	Overall correct
	Responded emotion by Dutch native listeners						
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	
Sarcastic	13.4	7.9	3.9	**34.3**	6.3	6.0	57.0
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	
	Responded emotion by Novice Chinese listeners						
Angry	**8.0**	17.8	32.8	5.5	20.0	16.0	
Happy	13.0	**8.2**	36.5	16.0	13.5	12.8	
Neutral	31.0	7.2	**22.2**	11.0	14.0	14.5	
Sarcastic	18.0	8.8	33.2	**11.8**	9.2	19.0	15.0
Sad	17.5	12.0	26.8	13.2	**21.2**	18.8	
Surprised	16.7	10.2	30.4	11.2	14.3	**17.2**	
Total	8.9	9.1	24.7	13.1	23.6	19.1	

The confusion matrices and Figure 7.1 together show that Dutch native listeners were able to recognize the vocal emotions well above chance level (= 16.7%) regardless the speaker type; in contrast to this, Chinese native listeners were only able to identify the emotional prosodies produced in their native language reasonably well, but failed to recognize them above chance level when the emotions were vocally portrayed in the unknown language (Dutch). Table 7.3 and Figure 7.1 indicate that Dutch novice listeners perceived the six emotional prosodies produced by native Chinese speakers as well as Chinese native listeners, with a mean correct identification rate of 46%. Moreover, Figure 7.1 shows that Dutch native listeners recognized the same six emotional prosodies produced in their native language (Dutch) substantially better than in Chinese, with a mean correct recognition rate of 57%. It can be also seen from Figure 7.1 and Table 7.3 that Chinese native listeners identified the emotional prosodies portrayed in their L1 significantly better than in an unknown language too, with a mean correct recognition rate at 46% in the first perception experiment and at 15% in the second perception experiment. These findings support previous studies, e.g., Elfenbein & Ambaby (2002), Pell et al. (2009), Thompson & Balkwill (2006), claiming that listeners generally recognize emotional prosody produced in their L1 better than in an unknown language, which phenomenon is known as the 'in-group advantage'. In addition, Van Bezooijen's (1984) study shows that Dutch native listeners were significantly better than Asian listener groups (Chinese and Japanese) in recognizing emotional prosody in Dutch, which is along the line of the present finding. The present result also shows that Dutch native listeners could identify vocal emotion in an unknown language equally well as native listeners of that language could. These findings imply that some culture group (e.g., Dutch) might be generally better in recognizing vocal emotion than some

other culture group, for example, Chinese, no matter whether vocal emotion is produced in the listeners' L1 or in an unknown language. In other words, it might be the case that perception of emotional prosody cross-culturally is not symmetrical.



*Note: 'S' = speaker type; 'L' = listener type; 'Man' = native Mandarin speaker; 'Dut' = native Dutch listener. Chance level = 16.7%, which is marked by the dotted line in the graph.

Figure 7.1. *Percent correct identification (%) of emotions for four combinations of speaker (S) and listener (L) types.* Means and 95%-confidence limits are indicated. Conditions under the same brace do not differ significantly from each other by a Bonferroni post-hoc test* ($\alpha = .05$).

The finding that Chinese native listeners were not able to identify (six) emotional prosody produced in Dutch above chance (= 17%) is not in line with the finding of Van Bezooijen (1984). In her study, novice (Taiwanese) Chinese listeners were able to recognize (ten) Dutch emotional prosody well above chance level (= 10%). The mean correct identification rate for the (Taiwanese) Chinese was 37%, which however, was significantly lower than that of the Dutch native listeners (66%). The contradictory finding might result from the varieties of the native Chinese listeners' dialectical backgrounds, as the listeners hailed from all over mainland China which covers at least ten official dialects (Li 1987, a-2).

Surprisingly, Table 7.3 indicates that native Chinese and novice Dutch listeners followed a rather similar recognition order, such that they both found 'neutrality' the easiest emotion to identify, followed by 'anger', 'surprise', 'sadness', 'happiness' or 'sarcasm'. Table 7.4 shows the perception of the same emotional prosodies produced in

Dutch by the two listener groups. The detailed recognition order of the six emotional prosodies by the two listener groups in the two perception tests is shown in Table 7.5. In the second perception experiment, Dutch native and Chinese novice listeners did not follow any similar recognition pattern. It implies that the two culture groups resort to different cognitive resources when perceiving the vocal emotions. The concrete cognitive resources can hardly be known at this stage.

Table 7.5. *Recognition order of the six emotional prosodies by native Chinese and Dutch listeners. The recognition order of the six Chinese emotional prosodies is presented in the upper panel; the recognition order of the six Dutch emotional prosodies is presented in the lower panel.**

Listener group	Recognition order of the six Chinese emotional prosodies					
Native Chinese	neutrality >	anger >	surprise >	happiness >	sadness >	sarcasm
Dutch novice	neutrality >	anger >	surprise >	sadness >	sarcasm >	happiness
Listener group	Recognition order of the six Dutch emotional prosodies					
Native Dutch	surprise >	neutrality >	sadness >	anger >	happiness >	sarcasm
Chinese novice	neutrality >	sadness >	surprise >	sarcasm >	happiness >	anger

*: '>' means 'better identified than'.

Table 7.5 shows that both Chinese and Dutch native listeners perceive emotional prosody differently in their L1 and in the unknown language in terms of recognition order. It seems that negative emotions, such as sadness or anger, are relatively easy to recognize for both listener groups. This finding supports some previous studies, e.g., Ohman et al. (2001), Tooby & Cosmides (1990), claiming that negative emotions are generally better recognized by human beings. Even though anger is generally considered a basic emotion, the Chinese novice listeners massively confused 'anger' with 'neutrality' for Dutch emotional prosody. It implies that perception of primary emotions (e.g., anger or neutrality) only through the audio-channel is not universal all the time. It confirmed that perception of emotional prosody is partly universal and partly culture-or-language specific, even for some basic emotions, for example, anger, neutrality or sadness. In some extreme case, it is possible that listeners are not able to identify emotional prosody in an unknown language above chance, especially the unknown language is typologically remote from the listener's L1.

Finally, I will analyze the confidence ratings. Although it was mentioned earlier in this section that there was no effect of weighting on the results and therefore 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 7.2A-B shows that Chinese native listeners were less confident than the Dutch listeners in their identification of vocal emotions in their L1. Surprisingly, however, Chinese novice listeners were more confident than Dutch natives themselves when recognizing Dutch emotional prosody. The Chinese listeners showed less confidence when responding to emotions produced in their own language (mean = 1.5) than when responding to the Dutch emotions (mean = 2.3). The difference is highly significant by a paired-samples t-test, $t(19) = -39.1$ ($p < .001$). The same statistical method was applied to the Dutch listeners' recognition of emotional prosody in Chinese and Dutch. The result shows that there was also a significant effect for the Dutch listeners in the two perception experiments, $t(19) = -7.5$ ($p < .001$), indicating that Dutch listeners are more confident in identifying vocal emotion both in their L1 than in the unknown language.

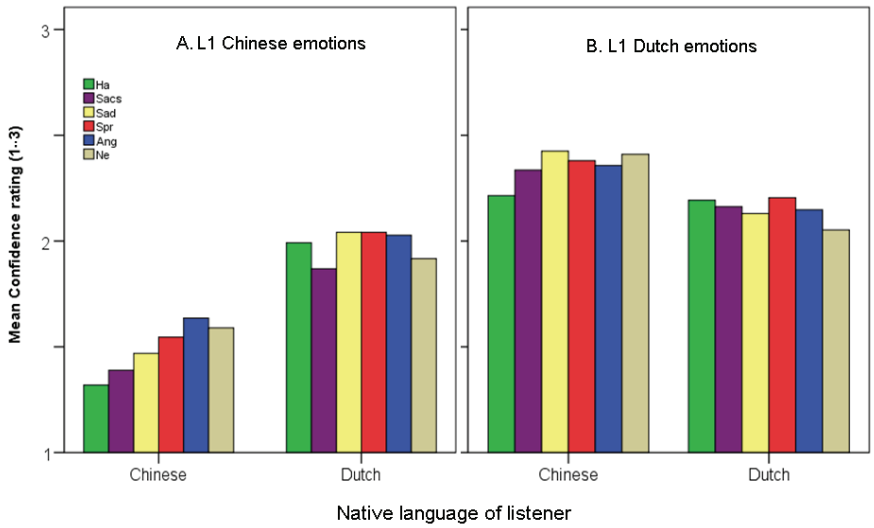


Figure 7.2. *Confidence rate (3 = most) of six intended emotions by native Chinese and Dutch listeners. Panel A presents the ratings of emotional prosody produced by native Chinese speakers. Panel B presents rating of emotional Dutch prosody encoded by advanced Dutch L2 speakers of Chinese.*

7.4 Conclusions and discussion

The results of this investigation indicate that Chinese native listeners are only able to recognize emotional prosody in their native language reasonably well, but cannot identify emotions in an unknown language above chance level, meaning that their identification in the unknown language is based on guessing. In contrast to this, Dutch native listeners are able to identify vocal emotion in an unknown language as accurately

as native listeners. And they recognize the same emotional prosodies expressed in their L1 even more correctly. Moreover, 'anger' was perceived relatively successfully by the two listener groups, apart from being identified by the Chinese listeners in Dutch. These findings show that perception of emotional prosody can be partly culture or language specific and partly universal. It confirms previous studies which claim that perception of vocal emotion is a combination of universal elements and cultural or linguistic variables (for example, Pell et al. 2009, Scherer 2000, Van Bezooijen (1984). The results show a significant difference between listeners' identifications of vocal emotions in their L1 and in the unknown language. Therefore, the existence of the in-group advantage found by other researchers (e.g. Elfenbein & Ambaby 2002, Pell et al. 2009, Thompson & Balkwill 2006) is confirmed, which claims that listeners generally better recognize emotional prosody produced in their L1 than in an unknown language. Since the Dutch novice listeners recognized the Chinese emotional prosodies as well as the Chinese natives did, therefore, the 'in-group advantage effect' might not be developed extensively, predicting that native listeners are generally better than novice listeners in perceiving vocal emotions in the native language. In fact, for some cultural groups (e.g. Dutch), novice listeners can identify vocal emotion in unknown native language as well (and as confidently) as native listeners can (or even better).

Chinese native and Dutch novice listeners followed a rather similar recognition order for the perception of Chinese vocal emotions. This finding is in line with findings by Scherer et al. (2001), who reported that patterns of confusion were very similar across all cultural groups: 'These data suggest the existence of similar inference rules from vocal expression across cultures' (Scherer et al. 2001). However, the confusion patterns of the two listener groups' perception of Dutch vocal emotions differed substantially in the present study, implying that there is an overlap between Dutch and Chinese inference rules from vocal expression; however, Dutch inference rules might cover a relatively wider range than those of Chinese. My results further suggest that perception of emotional prosody is not symmetrical cross-culturally, such that listeners of two barely related cultural or linguistic groups might have different perceptual abilities of inferring vocal emotions expressed in the other language.

A peripheral finding of the present study is that Chinese listeners are more confident in identifying vocal emotion in an unknown language than in their L1, even though their identifications of emotional prosody in the unknown language seemed mainly based on guessing. This unexpected social behavior of Chinese listener group cannot be explained at the current stage. In contrast to this, the mean confidence rating by Dutch listeners increased when the listeners were asked to perceive the vocal emotions in their L1. This finding can be seen as a spin-off of the in-group advantage, predicting that listeners are generally more confident at identifying vocal emotion in their L1 than in an unknown language. However, this prediction might not apply to all the cultural groups, for example, Chinese. It needs to be further tested by involving more cultural groups, especially cultural groups which culturally differ from each other very much.

The asymmetrical performance between Chinese and Dutch in the two perception experiments forces us to answer research question 2 negatively. In the asymmetrical approach one cultural group may be generally more sensitive to correctly identifying

emotional prosody than some other cultural group, e.g. Chinese. This view needs to be further tested by involving more and more diverse cultural groups. Moreover, Scherer et al. (2001) found that, generally, accuracy of identification of vocal emotions by different cultural groups decreased with increasing language dissimilarity between the different cultural groups. It is concluded that culture- and language-specific paralinguistic patterns may influence the decoding process. Up to this point, it can be true that Chinese emotional prosody generally contains more universal cues for novice listeners to detect. In contrast to this, Dutch emotion prosody may bear more culture-or-language-specific variables which are not easy for novice listeners to decode. This would explain why Chinese native listeners did poorly when instructed to identify emotional prosody in Dutch. But this does not explain the asymmetry in the success with which Dutch and Mandarin listeners identified the emotional prosodies in the other language. It has been suggested that such asymmetries may arise as a consequence of either cultural differences or differences between the phonologies of the two languages involved. McCluskey et al. (1975) consider, on the basis of the asymmetry they found, that Mexican-Spanish children, in contradistictions to their (Canadian) English peers, are brought up in a cultural setting that emphasizes the importance of attending to emotional prosody. This would then explain why Mexican children identify emotional prosody better than Canadian-English children, not only in their own language (Mexican Spanish) but also in Canadian English. But we may also think of strictly linguistic reasons that could explain asymmetries in affect perception, especially in the comparison of Dutch and Mandarin.

Apparently, our prediction made earlier that Dutch listeners would be overall more sensitive than Chinese listeners in correctly identifying emotional prosody was supported by the results. Therefore, it implies that the functional view might be true. If a language uses a prosodic parameter for linguistic purposes, it can no longer use the same parameter for non-/paralinguistic uses; or cannot use the same parameter as effectively for the expression of paralinguistic or extralinguistic meanings. In addition, 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 can be incorporated into the functional view, indicating that lexical tone might suppress or inhibit the expression of emotional prosody. In the present case, Mandarin is a tonal language that uses a wider pitch range than Dutch, and does so for linguistic purposes. Accordingly, its prosodic space for emotion is smaller than that of Dutch. As a consequence of the functional principle, listeners of Chinese are, in fact, less experienced in decoding the paralinguistic use of prosody than listeners of Dutch. This would explain why the Chinese listeners performed no better than the Dutch novices for the Chinese emotional prosodies and why they did so poorly in identifying Dutch vocal emotion, compared to the natives.

The unexpected performance difference between the Chinese and Dutch listeners in the two perception tests might be enhanced by the absence of particles in the Chinese stimuli used in the experiment. 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.[19] 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 better equipped for extracting emotional meaning from prosody, according to the functional view. That is possibly why the Chinese L1 listeners did not outperform the Dutch listeners in the perceptual study in which the stimuli with no final particles attached were presented through the 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.

[19] Lexical markers here refer to final particles in Chinese which can carry emotional information. Examples would be *ya* (friendly) or *a* (enthusiastic). Syntactic markers may be used imply negative emotions such as the 'annoyance' marking construction *nán dào … (ma)?*, which is a rhetorical question confronting the listener with his/her ignorance (less negative with *ma* than without).