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

Zhu, Y.

# Chapter Three

# Perception of Chinese Emotional Prosody by Chinese Natives, Naïve Dutch Listeners and Dutch L2 Learners of Chinese

**Abstract**

This chapter investigated the perception of six Chinese emotional prosodies (neutrality, happiness, anger, surprise, sadness and sarcasm) by 20 Chinese native listeners, 20 naïve Dutch listeners and 20 advanced Dutch L2 learners of Chinese.[7] The results showed that advanced Dutch L2 learners of Chinese recognized Chinese emotional prosody significantly better than Chinese native listeners and Dutch naïve listeners. The results also indicated that naïve non-native listeners could recognize emotions in an unknown language as well as the natives did. Chinese native listeners did not show an in-group advantage for identifying emotions in Chinese more accurately and confidently. 'Neutrality' was the easiest emotion for all the three listener groups to identify and 'anger' was recognized equally well by all the listener groups. The prediction made in the beginning of the study is confirmed, which claims that listeners of a tonal language will be less intent on paralinguistic use of prosody than listeners of a non-tonal language. The results in this chapter will be used as the control group for Chapter 4.

[7] This chapter has appeared as Y. Zhu (2013a). Which is the best listener group? Perception of Chinese emotional prosody by Chinese natives, naïve Dutch listeners and Dutch L2 learners of Chinese, *Dutch Journal of Applied Linguistics*, 2, 170–183.

## 3.1 Introduction

It is customary to distinguish between segmental and suprasegmental (also called prosodic) aspects of speech. Segmental properties are inherent properties of individual vowels and consonants or co-intrinsic properties that are predictable from the sequence of segments. Prosody then refers to the ensemble of properties of speech that cannot be derived from the mere sequence of segments (e.g. Lehiste 1970, Nooteboom 1997). It follows from this definition that prosody includes the temporal and melodic effects of lexical tone, word-stress, phrasing, accentuation and intonation, as well as the articulatory setting and voice quality of longer stretches of speech. Prosody fulfils both linguistic and paralinguistic functions in speech communication. Linguistic functions of prosody would be to divide the stream of speech into chunks of information that should be processed as meaningful units, highlight specific words as communicatively important, or differentiate between sentence types such as question and statement (Grandjean et al. 2006). A paralinguistic function of prosody is to signal the emotional state of the speaker (e.g. happiness, sadness, anger, fear, disgust) and/or the attitude of the speaker – either towards an addressee (e.g. dominance, submissiveness) or towards the verbal contents of the message (e.g. sincerity, irony, sarcasm). These paralinguistic functions of prosody are often subsumed under the term 'affect'. However, I only use the term 'emotional prosody' in this article to refer to the vocally expressed emotions and other affects. Moreover, although it has been shown that changes in articulatory setting, such as lip spreading (smiling) by a happy speaker (Tartter & Braun 1994), and in voice quality, such as a rough voice during anger (Grandjean et al. 2006), also contribute to the expression of emotion, I will concentrate on the use of speech timing and melody as the primary correlates of emotion.

A lot of attention has been drawn to vocal expressions of emotion and their acoustic accounts since Darwin concluded that affective expressions, including those produced via the vocal channel, are veridical (Darwin 1872). However, most of the previous studies within psycholinguistic and phonetics focused on human perception and use of different emotional prosodies of one particular language (e.g. Albas et al. 1976, Ladd et al. 1985). There were only few researchers who studied the perception of emotional prosody by both native and non-native listeners. For instance, Van Bezooijen (1984) studied ten emotional prosodies: neutral, disgust, surprise, shame, interest, joy, fear, contempt, sad, and angry. Basically, 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 the three listener groups were able to recognize the emotional prosodies well above chance level. Dutch native listeners got the highest correct identification rate and Japanese listeners performed poorest. Moreover, Dutch listeners correctly recognized 'joy' by 76% while Taiwanese and Japanese listeners only identified it by 24% and 20%. These results show that perception of vocal emotion in an unknown language is both universal and language or culture specific. Thompson and Balkwill (2006) had twenty English-speaking listeners judge the emotive intent of utterances produced by English, German, Chinese, Japanese, and Tagalog speakers. 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 emotional prosody is decoded by a combination of universal and culture-specific cues.

In addition, a few studies investigated the perception of vocal emotion by L2 learners. Some of them also studied the correlation between the learner's ability of recognizing 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 the advanced L2 learners being more successful than the beginning and the intermediate ones. Shochi, Gagnié, Rilliard, Erickson and Aubergé (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 advanced Japanese learners of French. They found out that facial information cues seem to be more salient than auditory cues. The advanced 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 advanced learners of French). This finding is in line with the conclusions of Van Bezooijen (1984) and Thompson and Balkwill (2006).

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 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. At this point, I would like to propose 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. This hypothesis holds that one can use a particular parameter in the phonetic space only once. 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 hypothesis – 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). As a consequence, 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. Therefore, I conducted the present study (1) to test whether the prediction made here will apply to perception of Chinese emotional prosody by natives and non-natives. Chinese native listeners, then, should have more difficulty identifying emotion from speech melody, so that the in-group advantage may be reduced or even absent for them as compared with Dutch listeners; (2) to investigate to what extent (i) native, (ii) naïve non-native and (iii) advanced L2 learners of Chinese can perceive Mandarin emotions expressed vocally and also to find out what would be the differences between native and non-native listeners of Chinese in perceiving Chinese emotional prosody; (3) to find out whether advanced L2 learners of Chinese will perform better than naïve listeners in perceiving Chinese emotional prosody.

In this study six Chinese emotional prosodies have been studied by using the discrete-emotion approach: 'neutrality', 'happiness', 'anger', 'surprise', 'sadness' and 'sarcasm'.[8]

## 3.2 Methods

### 3.2.1 Participants

Twenty native Mandarin listeners (10 males, 10 females, mean age = 24 years), twenty naïve Dutch listeners (10 males, 10 females, mean age = 33 years) and twenty advanced Dutch learners of Chinese (10 males, 10 females, mean age = 20 years) voluntarily participated in the perception test. 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 of 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 one MA student and three outstanding second-year BA students. Early bilinguals were excluded; none of the students started learning Mandarin before they enrolled at Leiden University, i.e. after the age of eighteen. There was no special course designed for training these students to recognize emotions in Chinese in the curriculum.

### 3.2.2 Materials and procedure

In this study 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 of 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

[8] The discrete-approach here means that listeners have to choose emotions from a limited set of discrete emotion labels in a perception experiment. This is different from the so-called emotion tracking technique, which asks (trained) listeners to adjust a pointing device (e.g. the Feeltrace device, cf. Cowie et al. 2000) in a two-dimensional space with continuously variable axes in accordance with the perceived strength of emotional parameters.

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).

Four native Mandarin speakers (2 males, 2 females, mean age = 45 years) whose mother tongue was standard Mandarin, voluntarily participated in the recording of the stimuli for the perception experiment. The stimuli were digitally recorded (44.1 KHz, 16 bits) in a sound-proofed booth through a Logitech external microphone. This procedure resulted in a stimulus set that consisted of 6 Chinese statements × 4 Mandarin speakers × 6 emotions = 144 discrete emotional utterances. The complete stimulus list can be found in Table 3.1.

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.' 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 be able to later correct the recognition rates based on the weighting. The entire experiment lasted 25 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 sentences 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).

Table 3.1. *Stimulus list in Chinese (in Chinese characters), Pinyin Chinese-phonetic notation (including tone) and English glosses.* 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.

| | |
|---|---|
| 1. | 是你。<br>shì nǐ<br>It is you. |
| 2. | 谢谢你。<br>xièxiè nǐ<br>Thank you. |
| 3. | 小王完全不知道这件事。<br>xiǎo wáng wánquán bù zhīdào zhè jiàn shì<br>Xiao Wang did not know about this matter. |
| 4. | 今天下午他不能来参加这个会。<br>jīntiān xiàwǔ tā bùnéng lái cānjiā zhège huì<br>He cannot attend the meeting this afternoon. |
| 5. | 她怀孕 3 个月了。<br>tā huáiyùn sān gè yuèle<br>She is three months pregnant. |
| 6. | 他去过小葛家一次。<br>tā qùguò xiǎo gé jiā yì cì<br>He has been to Xiao Ge's place once. |

### 3.3 Results

At the beginning of the present study, I explained the reason for introducing a confidence ratings 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 article only. The confidence ratings will be presented and analysed separately from the identification results. Table 3.2 shows the mean identification rates (in %) broken down by intended emotion. We can see from the table that native Chinese listeners and naïve Dutch listeners fell into very different confusion categories of emotions. For instance, Chinese listeners tended to mistake 'neutrality' for 'happiness' by 34.8%, while naïve Dutch listeners misidentified 'happiness' as 'anger' by 35.7%. Interestingly, the confusion categories that advanced Dutch learners of Chinese fell into are quite similar to those of naïve Dutch listeners. For example, Dutch learners of Chinese showed the same tendency as naïve Dutch listeners for 'happiness', which they often misidentified as 'anger' (30.2%). Moreover, both advanced Dutch learners of Chinese and naïve Dutch listeners misidentified 'sarcasm' as 'surprise' by 20%. This finding supports Chen's (2005) conclusion that L1-transfer is an important strategy in interpreting pitch variation in L2. 'Anger' was identified by the three listeners groups equally well, from which we conclude that anger could be the real basic and universal emotion to all human beings.

Table 3.2. *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 (bold and 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 | 0.8 | 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** |

Figure 3.1A shows the mean percentage of correct identification by the six intended emotions along the X-axis and broken down further by the three listener groups (in the legend). Figure 3.1B displays the same information, broken down first by listener group and second by emotion. Table 3.2 and Figure 3.1A-B together indicate that native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese were able to recognize discrete Chinese emotional prosodies above chance level (mean recognition rate: 48.6%, chance level: 16.7%). Moreover, emotions were identified much better than chance by each of the three listener groups. 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 data were analyzed by a repeated measures Analysis of Variance, with speaker, sentence and intended emotion as within-subjects factors and with listener type as a between-subjects factor. The dependent variable was the percentage of correctly identified emotions. Huynh-Feldt corrected degrees of freedom were used when the

assumption of sphericity was unreasonable. The difference between the three listener groups is statistically significant $F(2, 57) = 5.8$, $p = .005$, $p\eta^2 = .17$. A Bonferroni post-hoc test ($\alpha = .05$) showed that the advanced Dutch learner group performed better than the other two groups, which did not differ from each other. All main effects and all possible interactions were significant.

Surprisingly, native Chinese and naïve 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'. The detailed recognition order of the six emotional prosodies by the three listener groups is shown in Table 3.3.

Table 3.3. *Recognition order of the six emotional prosodies by native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese.**

| Listener group | Recognition order of the six emotional prosodies | | | | | |
|---|---|---|---|---|---|---|
| Native Chinese | neutrality > | anger > | surprise > | happiness > | sadness > | sarcasm |
| Dutch naïve | neutrality > | anger > | surprise > | sadness > | sarcasm > | happiness |
| Advanced learners | neutrality > | surprise > | sadness > | anger > | sarcasm > | happiness |

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

The three listener groups all found 'neutrality' the easiest emotion to identify. This result confirms earlier findings that neutral prosody is identified more accurately than emotional prosody (Cornew et al. 2010). Moreover, correct identification rates of native Chinese and naïve Dutch listeners are strongly correlated ($r = .837$, $N = 6$, $p < .001$), showing again that native and non-native listeners of Mandarin display very similar cognitive behavior in identifying Chinese emotional prosody. In other words, the emotions that native listeners found easier to identify are also considered easier by naïve non-native listeners, and vice versa.

In Figure 3.1B, though advanced Dutch learners of Chinese identified neutrality most successfully in the six intended emotions, they actually followed a slightly different recognition order, indicating that the emotions which native Chinese and naïve Dutch listeners found difficult to identify are not necessarily difficult for advanced Dutch learners of Chinese to recognize (e.g. sadness and surprise). This finding supports Chen's (2005) study 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 the L2 accordingly. Specifically, the advanced Dutch learners of Chinese showed higher identification rates for the emotions of 'sadness', 'surprise' and 'neutrality'. However, there was no significant difference between native Chinese listeners and naïve Dutch listeners, meaning that Chinese native listeners are not able to recognize emotions in their native language more successfully than naïve Dutch listeners. Interestingly, both naïve Dutch listeners and advanced Dutch learners

of Chinese showed better identification rates for 'sarcasm', which were 11.5% and 14.6% higher than that obtained by the Chinese listeners, respectively.

The findings of the perception experiment contradict the results demonstrated in previous studies that native listeners should recognize emotional prosody more accurately in their own language than non-native listeners do. (e.g. Dromey et al. 2004, Graham et al. 2001, Van Bezooijen 1984).

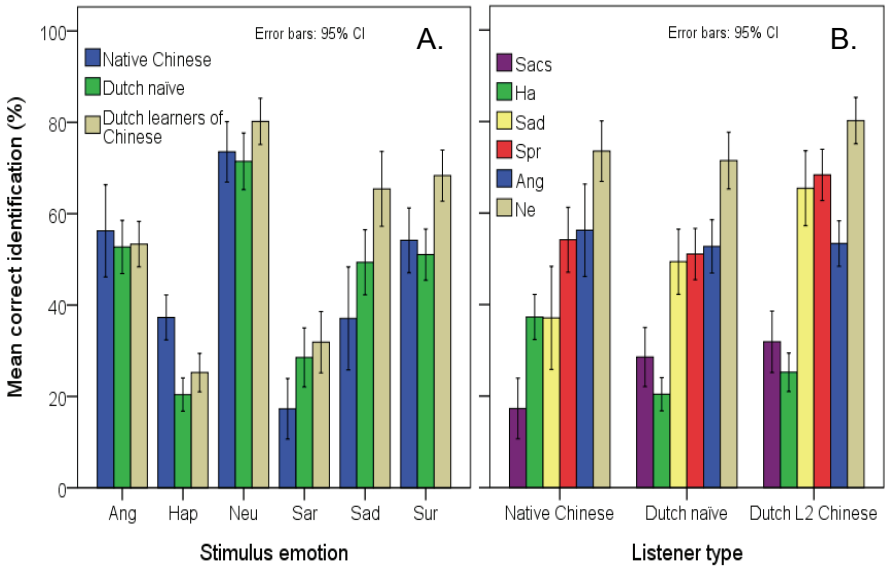


Figure 3.1A-B. *Percent correct identification of six intended emotions by native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese. Intended emotions in panel B are in ascending order of correct overall recognition. Confidence limits were computed for each bar on the basis of 20 listener means.*

Although there was no effect of weighting on the results for confidence, 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 mean confidence ratings and observe unexpected differences between the groups.

Figure 3.2 shows that Chinese native listeners were less confident (mean = 1.49) in their identifications than the Dutch listeners. Within the Dutch listeners the advanced learners of Mandarin were more confident (mean = 2.29) than the naïve listeners (mean = 1.96). The effect of listener group is significant by an ANOVA (same design as above), $F(2, 57) = 45.4$ ($p < .001$, $p\eta^2 = .614$). Bonferroni post-hoc tests revealed that the differences between all three groups were significant ($\alpha = .05$).

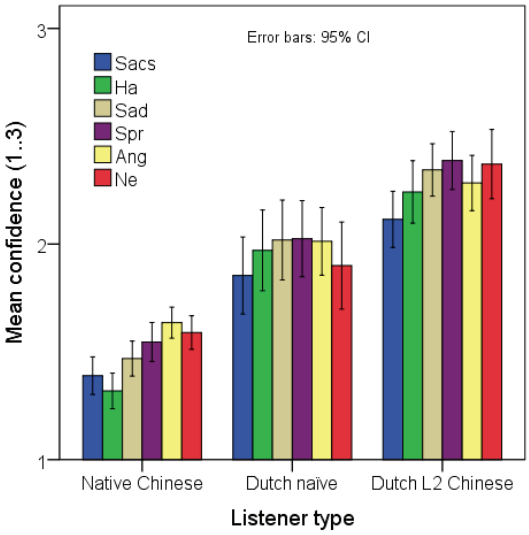


Figure 3.2. *Confidence rating (3 = most) of six intended emotions by native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese. Intended emotions are ordered as in Figure 3.1B. Confidence limits were computed for each bar on the basis of 20 listener means.*

### 3.4 Conclusions and discussion

The results of this investigation indicate that Chinese native listeners are not able to identify emotions in their native language more accurately and confidently than naïve Dutch listeners and advanced Dutch learners of Chinese. Surprisingly, advanced Dutch learners of Chinese can recognize emotional prosody in Chinese significantly better than natives. This finding contradicts the conclusion of Graham et al.'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 substantially better than naïve Dutch listeners. This finding is in line with the result of Shochi et al.'s (2010) study that trained second language learners may recognize emotional prosody in the target language better than naïve ones.

Naïve non-native listeners can recognize unknown emotional prosody as well as natives do. Both non-native listener types (including the advanced learners of Chinese) can even identify some emotions (e.g. sadness, sarcasm) more successfully than native listeners. The in-group advantage found by other researchers, therefore, does not apply universally to all cultural groups (e.g. Chinese). Natives and naïve non-natives may have drawn on very similar cognitive resources when identifying emotional prosody, but the

incorrectly recognized emotional prosodies of natives and naïve non-natives may fall into different confusion categories. Advanced learners of Chinese followed a slightly different recognition order, indicating that emotions which are difficult for native and naïve non-native listeners to identify are not necessarily difficulty for them to recognize, for example: sadness and surprise. These findings are in line with the conclusions of Chen's (2005) study: L1-transfer is an important strategy in interpreting pitch variation in L2. 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. Therefore, the advanced learners of Chinese followed a slightly different recognition order of identifying emotions in Chinese while at the same time they fell into very similar confusion categories to the ones which the naïve Dutch listeners also fell into.

'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 predictably higher than that of other emotions. 'Anger' is recognized by the three listeners groups equally well, which means that it could possibly be a truly basic and universal emotion for all human beings. According to Darwin's evolution theory, 'anger' is supposed to signal aggression of the offended and to warn the offender to expect an aggressive 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 regardless linguistic and cultural backgrounds, as people/animals have an instinct to sense, and protect themselves from, danger. From this point, we could further assume that emotion perception can be both universal and culture-specific depending on the particular emotions.

There may be several possible explanations for the finding that Chinese emotions were identified more successfully by (advanced) Dutch learners of Mandarin than by native Chinese listeners themselves.

1) Although voice quality and temporary changes in articulatory setting may also contribute to the expression of emotion (see introduction), these prosodic effects will not likely play a role in the comparison of Dutch and Mandarin.[9] Therefore, we would like to conclude that the prediction made at the beginning of the article is supported by our results. If a language – such as Mandarin – uses prosody for linguistic purposes, it can hardly use prosody for non-/paralinguistic uses. Since Mandarin uses prosody for linguistic purposes, specifically for expressing lexical tone contrast, Chinese listeners are less intent on the paralinguistic use of prosody than listeners of a non-tonal language (e.g. Dutch). In other words, listeners of a non-tonal language are generally better at recognizing emotional prosody than listeners of a tonal language. This explains 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

[9] The four Chinese native speakers portrayed the emotions in a normal voice. There was no special voice quality or phonation type, such as, hoarseness, roughness or smiling.

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 test emotion recognition by speakers of another tone language, e.g. Vietnamese or Thai.

2) It may be the case that Chinese listeners normally recognize emotions by lexical or syntactic markers and contextual connotations (Xing 1999). If so, they are not experienced at identifying emotion through the audio channel only.[10]

3) Chinese society is quite reserved when it comes to the overt expression of emotions, 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). Possibly, native listeners did not perform as well as advanced Dutch learners of Chinese because the latter have not been exposed to an emotion-free culture extensively, but the former are. Therefore, advanced Dutch learners of Chinese may be able to pick up some Chinese subtle emotions (e.g. sadness and sarcasm) more successfully than natives.

A basic problem with explanations of the type described above is that one can always reverse the argument. All human beings express emotions and are able to recognize emotions and respond to them. Now, if a cultural – or a linguistic code – prevents the speaker from expressing emotions plainly and overtly, the receiver (listener) is forced to attend to subtle expression of emotions. So on the one hand, such persons may be less sensitive to emotions as they have been exposed less to clear exemplars of the various affects, but on the other hand they may have learnt to be more attentive to subtle expression of emotions. In order to know what the effect is of growing up in an emotion-suppressive culture or linguistic environment, one can only turn to empirical observation.

[10] 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 to 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).