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

Zhu, Y.

Chapter Two

Background and Methodology

2.1 Introduction

In this chapter I will go through a literature review of what previous studies have done in terms of perception and production of emotional prosody by native and non-native listeners/speakers. Apart from that, I will also provide detailed information on research methods, including how I conducted the present study to answer my research questions and how I carried out the perception experiments in the present study, as well as how I analyzed the production of the six emotional prosodies portrayed in my speakers' L1 and L2.

2.2 Background

2.2.1 General introduction

Darwin (1872) was the first to claim that affective expressions, including those produced via the vocal channel, are veridical. He generally showed possible veridical associations between vocal acoustics and the vocalizer's emotional state. He also pointed out two aspects which very much influenced later studies in the field: (1) vocal signals can trigger emotional responses in listeners; (2) these signals can elicit learned emotional responses. These two observations are fundamental to any research on perception of vocally portrayed emotion, as they formally claim that there is a link between acoustic signals and emotions so that researchers since then were able to deepen the knowledge of how vocalizer (speaker) and recipient (listener) produce and perceive emotion via the audio-channel (Bachorowski & Owren 2008). Moreover, many previous studies have confirmed the two observations by Darwin, claiming that information concerning emotional state is encoded in vocal acoustics and subsequently decoded by listeners in order to respond to the speaker's emotional state (e.g. Juslin & Laukka 2001, Scherer 1986, 2003). They also found that specific patterns or configurations of vocal cues are reliably associated with different emotional/affective states (e.g. Borden & Harris 1994, Deller et al. 1993, Scherer 1986, Spackman et al. 2009). Furthermore, Bachorowski and Owren's (2008: 196) review states that there is now a substantial body of work focused on how emotion is conveyed by and perceived from vocal acoustics. Although this research has arguably not enjoyed the same degree of cumulative success as has work on the communication of emotion through the facial channel, there is nonetheless a solid body of evidence showing that specific vocal acoustic features are reliably associated with affect-related arousal (or activation) on the part of vocalizers, and that listeners in turn can reliably perceive arousal from vocal

acoustics. These claims together show that the audio channel plays an important role in the perception and production of emotion.

It is good to know some general background of perception of vocally produced emotion in listeners' native language (L1) before reviewing the perception of emotional prosody produced in listeners' non-native languages (L2). According to previous studies, humans can accurately decode discrete emotions from speech-embedded prosody at levels well above chance (e.g. Banse & Scherer 1996, Biehl et al. 1997, Haidt & Keltner 1999, Juslin & Laukka 2001, 2003, Rosenberg & Ekman 1995). Moreover, Johnstone and Scherer's (2000) review mentions that when listeners are asked to identify the intended emotion in utterances produced by actors, accuracy is significantly better than chance – although at a moderate level overall, typically about 55%. According to Bachorowski and Owren's (2008) review, identification rates are usually best for anger, fear, and sadness. Results for positive emotions have varied, but in an informative way. Accuracy is typically high when listeners are given only one positive response option (e.g., Johnson et al. 1986, Scherer et al. 1991). However, correct responses drop significantly when other positively toned options are tested, such as 'elation', 'contentment' or 'interest'. A similar effect may contribute to the identification of 'sadness', which is sometimes the only low-arousal option offered among the negative emotions. Another factor which might affect listener recognition accuracy is the experimental design, such as whether one should use forced-choice procedures or free-choice tests. It is concluded by previous researchers (e.g., Johnson et al. 1986, Pakosz 1983) that forced-choice procedures generate better recognition results than free-choice ones. In real human communication, there are other possible factors that can influence listener identification accuracy, for example, the speaker's age, sex and style, the listener's personal interpretation of the emotion label, the testing method, etc.

2.2.2 Perception of emotional prosody in L1, L2 or an unknown language

There are a few studies which touch on the 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, happy, sad, anxious, angry, afraid, surprised, disgusted, annoyed and embarrassed. 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 Asian listener groups performed well above chance level. Studies by Tickle (2000) and by Scherer et al. (2001) both further claim that listeners can recognize emotional prosody in an unknown language better than chance level but the misidentification rates increase as speaker and listener languages become more dissimilar. 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 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. 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; however, culture-specific attitudes (i.e. suspicious irony and obviousness) were confused by Japanese listeners (including trained listeners). Facial information cues seem to be more salient here than auditory cues.

2.2.3 Production of emotional prosody in speakers' L1 and L2

In Bachorowski and Owren's (2008) review, from the production point of view, Darwin's two observations (1872) form the core of an 'affect induction' view of vocal signaling, which began as a functional account of nonhuman primate calling (Owren & Rendall 1997, 2001, Rendall & Owren 2002), but may also apply to affect-related vocal signaling in humans (Owren et al. 2003). In sum, the affect-induction perspective argues that vocal expressions of emotion are not displays of vocalizer states as much as they are tools of social influence. This view shows that a listener's personal interpretation of the incoming signals from the speaker determines how successful the communication will be. Mismatches between a speaker's intended emotion and a listener's perceived emotion can cause a failure in the paralinguistic communication. The success of a communicative encounter requires not only the ability to convey one's own affect but also the ability to accurately gauge that of the other person. This process does not always work well, particularly when engaging in telephone communication, where face-to-face contact is not possible and the reading of affect is solely dependent on the auditory channel of communication (Mitchell & Ross 2013). Therefore, investigating how speakers encode emotional prosody will guide us to find out what vocal cues speakers use to vocally express emotion (or to trigger a listener's own affective response) and why these cues sometimes can be misinterpreted or mismatched to another affective state by listener.

Since Darwin (1872) claimed that there is a direct correspondence between particular signaler states and the communicative display produced, there were researchers who further developed the view, claiming that each emotion is associated with distinctive acoustic cues (e.g. Banse & Scherer 1996, Leinonen et al. 1997, Scherer 1986, 1989). Specifically, Banse and Scherer (1996) analyzed the vocalizations generated by 12

professional actors who each portrayed 14 emotions. In this study, 29 acoustic features were measured. They found that fundamental frequency (F0) and mean amplitude clearly showed the strongest connections to the emotions being portrayed. According to Scherer's (1996) review, there were other acoustic cues which also contributed to the production of emotional prosody: (a) the vocal energy (or intensity, perceived as loudness of the voice); (b) the distribution of the energy in the frequency spectrum (particularly the relative energy in the high vs. the low-frequency region, affecting the perception of voice quality or timbre); (c) the location of the formants (F1, F2…F*n*, related to the perception of articulation); and (d) a variety of temporal phenomena, including tempo and pausing; (e) F0 variability (including both overall range and moment-to moment perturbations, e.g. jitter). However, recently researchers not only studied jitter (or related measures) but also HNR (Harmonics to Noise Ratio) to better understand some particular emotions, for example, sadness.

Although there were plenty of studies on production of emotional prosody, previous studies mainly focused on the vocal production of emotion by native speakers from one particular linguistic group. There was little research on production of emotional prosody in an L2, especially when the L2 is a tonal language, e.g., Mandarin, Thai, or Vietnamese. The only related literature that can be found so far is Anolli et al.'s (2008) study, which directly studied the topic of vocal production of emotion cross-culturally. They 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. However, there is no earlier study which directly dealt with production of emotional prosody in speakers' L1 and in their L2. Therefore, the production part in the present study is a pioneering investigation on how well native speakers of a non-tonal language (e.g., English or Dutch native speakers) can express emotional prosody in a tonal L2.

2.3 Methods

2.3.1 Overview

In order to answer the research questions, I designed three judgment studies using an experimental approach in this dissertation. The first judgment study includes one perception experiment in which native Chinese listeners, naïve Dutch listeners and advanced Dutch learners of Chinese identify six Chinese emotional prosodies portrayed by native Chinese speakers. It aims to answer the first research question: how well do native and non-native listeners, including naïve listeners and L2 learners of the target language, perceive the emotional prosodies portrayed by native speakers? In this judgment study, apart from looking at the correct recognition rates of the three listener groups, I will also present confusion matrices and confidence ratings of the three listener groups. This will help us to obtain a clearer picture of what the differences are between native and non-native listeners in perceiving emotional prosody and whether having high language proficiency in the target language will help L2 learners of the target language perceive emotional prosody portrayed better in that language. The

results will serve as the control group in the entire research. The second judgment study consists of two perception experiments: in the first, the same three listener groups perceive the same six Chinese emotional prosodies but now produced by Dutch L2 speakers of Chinese. It is designed to find out: (i-a) compared to the native Chinese speakers (the control group), how well can the Dutch L2 speakers of Chinese vocally produce the Chinese emotional prosodies (the judgment will be based on the correct identification rates of the three listener groups); (i-b) what are the differences between listeners perceiving emotional prosody portrayed by native and L2 speakers of the target language.

In the second perception experiment, Dutch native listeners who do not have any knowledge of Chinese, will identify the same six emotional prosodies portrayed in their native language (Dutch) by the same Dutch L2 speakers of Chinese. This experiment is carried out to test how well the Dutch L2 speakers of Chinese can vocally produce the same emotional prosody in their native language. Moreover, it can tell us how Dutch native listeners perceive emotional prosodies generated in their L1. The results of this perception experiment will be compared with the results obtained in the first perception experiment. There will be an acoustic analysis based on selected stimuli after I run the two judgment studies. The results will reveal the vocal correlates that L1 and L2 speakers use to produce emotional prosody in their L1 and L2.

In sum, the second judgment study and the acoustic analysis will answer the following research questions: (ii) How well do native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese perceive the six Chinese emotional prosodies vocally portrayed by L2 speakers of Chinese? (iii) Can L2 speakers of Chinese vocally produce emotional prosodies in their L2 as well as they do in their L1? What will be the similarities and differences between the two productions? (iv) Does L2 limit the expression of emotional prosody, especially when the native language of L2 speakers of the tonal language is a non-tonal language? (vi) What acoustic parameters contribute to differentiate between emotional prosodies in general? What acoustic correlates do speakers and listeners use to produce and perceive the vocal emotions in their L1 and in an L2? Do Dutch L2 speakers of Chinese use L1-transfer to produce emotional prosody in Chinese? To what extent does automatic recognition reflect the perception of the emotional prosodies by the human listeners?

After the first and the second judgment study, research question (v) will be answered, i.e., is there any support for the functional view, predicting that listeners of a tonal language might be less intent on (and in fact less experienced in) decoding the paralinguistic use of prosody than listeners of a non-tonal language?

The third judgment study includes one perception experiment in which Chinese and Dutch novice listeners of each other's language perceive the six emotional prosodies portrayed in Chinese and Dutch by native speakers of the respective languages. This experiment will be conducted in a reciprocal way and it is designed to test whether the in-group advantage claimed by other researchers is universal, and to investigate whether or not the ability of Chinese and Dutch native listeners to perceive emotional prosody in the other (unknown) language is symmetrical. The third judgment study will answer research question (vii).

2.3.2 Speakers

2.3.2.1 Acted stimuli vs. natural stimuli

According to previous studies, it is suggested that using acted stimuli is better than natural stimuli (Scherer 2003). According to the review of Bachorowski and Owren (2008), the obvious problem with relying on acted samples is that these may not necessarily correspond to naturally produced vocalizations. One counterargument is that much of our verbal communication involves making impressions on others, and so having vocalizers act 'as if' they were experiencing a particular state is not markedly different from natural communicative circumstances. However, when the issue taken together with evidence from natural emotion-inducing circumstances showing that individual variability in vocalizer acoustics can be quite substantial (e.g., Streeter et al. 1983), it may be the case that the careful analysis of acoustic cues to acted emotion provides more information about emblematic portrayals of affective states than about naturally occurring cueing (Scherer 2003).

In addition, it has been seen quite often in previous studies that some individual actors (especially females) were much more convincing in their portrayals than others (e.g., Leinonen et al. 1997, Pell 2001, Scherer et al. 1991, Schröder 2000, Sobin & Alpert 1999, Walbott & Scherer 1986). Therefore, I decided to use four native speakers who are amateur actors/actresses and four non-native speakers to vocally produce the six intended emotional prosodies in their L1/L2 for all the perception studies, as it would help to balance off this kind of sex and talker differences in the production of stimuli. And I will use carefully selected and acted samples produced by the (non-)native speakers for acoustic analysis.

2.3.2.2 Native speakers of Chinese

In the present study, there were four native Mandarin Chinese speakers (2 males and 2 females, mean age = 45), who voluntarily produced the six emotional prosodies in spoken Chinese for the perception experiments. The four native Chinese speakers were amateur actors/actress who had acting training and stage performance experience.

2.3.2.3 Dutch L2 speakers of Chinese

Four Dutch L2 speakers of Chinese (2 males, 2 females, mean age = 33 years) voluntarily participated in the recording of the stimuli for the second judgment study. These four Dutch L2 speakers of Chinese 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; they had been teaching Chinese for 2 to 10 years when the recordings were made. All spent at least one year living or studying in mainland China or Taiwan.

2.3.3 Listeners

In the first and second judgment studies, 20 native Mandarin listeners (10 males, 10 females, mean age = 24 years), 20 naïve Dutch listeners (10 males, 10 females, mean age = 33 years) and 20 advanced Dutch learners of Chinese (10 males, 10 females, mean age = 20 years) voluntarily participated in each of the perception experiments, where they were required to perceive the emotional prosodies portrayed by L1 and L2 speakers of Chinese, respectively. 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 Chinese as their mother tongue. The naïve Dutch listeners were mainly bachelor students at the Humanities Faculty of Leiden University in the Netherlands and volunteers with variable education backgrounds. None of the naïve Dutch listeners spoke any Mandarin. The advanced Dutch learners of Chinese were mainly third-year BA students in the Chinese Program of Leiden University; the others were MA students and some outstanding second-year BA students. Early bilinguals were excluded; therefore, all students had learnt Mandarin after the age of eighteen. There was no special course in the curriculum designed for training these students to recognize emotions in Chinese.

In the third judgment experiment, 20 native Chinese listeners (10 males and 10 females) and 20 native Dutch listeners (10 males and 10 females) who did not know each other's language (novice listeners) voluntarily participated in the perception experiment in which they were required to identify the six emotional prosodies ('neutrality', 'happiness', 'anger', 'surprise', 'sadness' and 'sarcasm') portrayed in Dutch by native Dutch speakers, who were the same individuals as the Dutch L2 speakers of Chinese in the second judgment study.[4] This experiment was designed to test how well novice Chinese and Dutch native listeners perceive emotional prosody portrayed in the other language. The results would tell us whether the in-group advantage claimed by other researchers applies to these two cultural groups and whether the Chinese and Dutch native listeners possess similar ability of identifying emotional prosody correctly in an unknown language. For the aim of the third judgment study, the advanced Dutch L2 leaners of Chinese were excluded, even though they were also native Dutch listeners.

[4] These 20 native Chinese and Dutch (novice) listeners were drawn from the same population group as the 20 Chinese native listeners and the 20 naïve Dutch listeners in the first judgment study. The third judgment study was conducted three months later after the first two judgments were carried out.

2.3.4 Materials and procedure

2.3.4.1 The first judgment study

The first judgment study only included one perception experiment, which was set up to test how well the Chinese control group vocally expressed the six emotions in Chinese. In the first judgment study, I used six Mandarin statements (e.g. *She is three months pregnant; He has been to Xiao Ge's place once*). The requirements for the stimulus selection were: (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. The list of stimulus sentences for the first judgment study is shown in Table 2.1. The results of the first judgment study will be used as the control group.

Table 2.1. *Stimulus list in Chinese (Pinyin) with English glosses.*

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

Note: '*' means sentences were excluded in the second perception experiment of the first judgment study. 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 the 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.

It is concluded by previous researchers that forced-choice procedures generate better recognition results than free-choice tasks (e.g. Johnson et al. 1986, Pakosz 1983). Therefore, all the participants (including native Chinese listeners, naïve Dutch listeners and advanced learners of Chinese) were asked to make a forced choice of the speaker's intended emotion, from the six given emotions, immediately after they heard a stimulus. They also gave a confidence rating to each choice they made. A three-level confidence

rating scale was used, with the following interpretation: 3 = 'The speaker expressed the intended emotion well. I am very confident in my answer', 2 = 'The speaker reasonably expressed the intended emotion. But I am not so sure about my answer' and 1 = 'The speaker did not express the intended emotion well. I made the choice only by guessing.' The confidence scale was introduced in order to obtain a potential weighting factor such that responses given with great confidence would be weighted more heavily than responses that were largely based on guessing.

2.3.4.2 The second judgment study

The second judgment study included two perception experiments: the first perception experiment was carried out to test how well the three listener groups, i.e. native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese, perceived Chinese emotional prosody expressed by the Dutch L2 speakers of Chinese. It would thus reveal how well the Dutch L2 speakers of Chinese encoded the six emotions in Chinese, compared to the control group. The second experiment was conducted to determine how well Dutch L2 speakers of Chinese produced the same emotional prosodies in their mother tongue – Dutch.

In the first perception experiment, the four Dutch L2 speakers of Chinese were asked to express the six emotional prosodies in Chinese. The stimuli were digitally recorded under the same conditions as in the first judgment study. Two sentences were discarded from the stimulus set (see Table 2.1), 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 this experiment shorter than the perception experiment in the first judgment study. I only processed and showed the shared data in later comparisons. The three listener groups, including native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese, repeated the same experiment procedure of the first judgment study in this test round.

In the second perception experiment, 20 native Dutch listeners perceived the six emotions produced by the same four Dutch L2 speakers of Chinese, but in their mother tongue – Dutch.[5] This experiment was designed to find out how well the Dutch L2 speakers of Chinese produced the same emotional prosodies in their L1. The four Mandarin statements used in the first perception experiment 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. 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 for the second judgment study is shown in Table 2.2. The same procedure as in the first judgment study was used to obtain the judgments. In fact, some of the sentences may be associated more readily with some emotions than with

[5] These 20 native Dutch listeners were drawn from the same population as the 20 naïve Dutch listeners in the first judgment study. This second perception experiment was conducted three months after the first perception experiment was carried out.

others but on aggregate the lexico-syntactic materials will not be biased towards specific emotions.

Table 2.2. *Stimulus list in Dutch 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.'

2.3.4.3 The third judgment study

The third judgment study was designed to test whether the in-group advantage found by other researchers (e.g. Kilbride & Yarczower 1983, Markham & Wang 1996) is universal, claiming that native listeners are generally better at identifying emotional prosody in their L1 than in an unknown language. The third judgment study also aims to find out whether the ability to identify emotional prosody in an unknown language cross-culturally is symmetrical.

The third judgment study was conducted in a reciprocal way. 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 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 ethnically different rather than culturally-or-linguistically dis-similar. Other studies also adopted this method (e.g. Ekman 1972); however, they 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 perception of emotion cross-culturally, 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 would like to conduct the third judgment test applying the reciprocal method to the two cultural groups, i.e. Chinese and Dutch.

In this study, 20 Chinese and 20 Dutch native listeners who did not know each other's language (novice listeners) perceived the six emotional prosodies ('neutrality',

'happiness', 'anger', 'surprise', 'sadness' and 'sarcasm') portrayed in their L1 and in the other language by Chinese and Dutch native speakers.[6] The same experimental materials and procedures of the first two judgment studies were adopted in the reciprocal manner in this study (details in Chapter 7). The results and conclusions are also presented in Chapter 7.

2.3.5 Acoustic analysis

An acoustic analysis was conducted after the first and the second judgment studies. Two sentences were discarded from the stimulus list (see Table 2.1), as they were not well perceived by the listener groups in the first judgment study (the control group). The optimized stimulus list resulted in three sets of stimuli for the acoustic analysis in which each set consisted of 4 (non-)native speakers × 4 sentences × 6 emotional prosodies = 96 stimuli. Therefore, there were 288 stimuli in total (96 × 3 = 288) for the acoustic analysis: emotional prosody portrayed in L1 Chinese, emotional prosody portrayed in L2 Chinese by Dutch L2 speakers of Chinese, emotional prosody portrayed in L1 Dutch. First I extracted pitch contours of all the utterances, and then extracted selected other variables (see below). Finally, I compared the individual acoustic parameters between sets. In this manner, one can see the influence of each parameter in the vocal production of emotion. It also allows us to see what vocal correlates L1 and L2 speakers use in portraying emotional prosody in their L1/L2. Moreover, it will show us which parameters are relatively more important than others in the production and perception of emotional prosody, and which parameters can adequately differentiate between emotions.

In order to analyze the acoustic parameters which might contribute to the production of the emotional prosodies, I took Scherer's (1996) review as a point of reference. In addition, I investigated some parameters which were not mentioned in his review. Therefore, I studied the following acoustic variables obtained from the computer analyses of the speech signals: (a) tempo (duration/time); (b) mean fundamental frequency (F0) and F0 in the first and last quarters of the utterance duration, which is named 'F0 slope' in this dissertation, as well as standard deviation of F0; (c) distribution of the energy in four contiguous frequency bands (d) vocal energy (mean intensity standard deviation); (e) mean jitter; (f) mean HNR (Harmonics to Noise Ratio).

Automatic recognition was attempted on the basis of the acoustic analysis. The eight above-mentioned variables were entered into a Linear Discriminant Analysis (LDA) to identify the six emotional prosodies portrayed by the three speaker groups, i.e. L1 Dutch speaker, L2 Mandarin speakers and L1 Mandarin speakers (where the former two groups are the same individuals). This automatic recognition may reflect the variables used by the human listeners. The acoustic analysis and the automatic recognition will be presented in Chapter 6.

[6] Some data obtained in the first and second judgment study were re-used in the third judgment study.