Universiteit Leiden
The Netherlands

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

Zhu, Y.

Cover Page

The handle http://hdl.handle.net/1887/22850 holds various files of this Leiden University dissertation.

Author: Zhu, Yinyin
Title: Expression and recognition of emotion in native and foreign speech : the case of Mandarin and Dutch
Issue Date: 2013-12-12

Chapter Eight

Conclusion and Discussion

8.1 Introduction

In this dissertation, I have investigated 1) the perception of L1 and L2 produced Chinese emotional prosody by natives, naïve Dutch listeners and advanced learners of Chinese; 2) how well Dutch L2 speakers of Chinese vocally produced emotions in the L2 and how well they expressed the same emotions in their L1 (Dutch); 3) the in-group advantage and different cultural groups' abilities of identifying vocal emotions in an unknown language. In order to answer all the research questions listed in Chapter 1, three judgment studies were conducted in this dissertation. The first judgment study was designed to test how well native Chinese listeners, naïve Dutch listeners and advanced learners of Chinese perceive Chinese emotional prosodies produced by native Chinese speakers. The results were used as the baseline for comparisons. The second judgment study aimed to find out how well the same three listener groups perceive the same Chinese emotional prosodies produced by the Dutch L2 speakers of Chinese. The production of the Chinese emotional prosody by the Dutch L2 speakers was also investigated in this judgment study. The third judgment study was carried out in a reciprocal way to test the in-group advantage and the Chinese and Dutch native listeners' abilities of identifying emotional prosody in their L1 and in the other language. In this chapter, I will summarize the main findings of the three judgment studies and give the possible explanations of some unexpected results, as well as to show which direction future studies can go.

8.2 Answers to research questions

In this section, I will recapitulate the research questions formulated in Chapter 1 and provide integrated answers to them.

8.2.1 Perception of native-Chinese emotional prosody by three listener groups

Three listener groups participated in the first judgment study, i.e. native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese. The first judgment study was designed to test how well the three listener groups perceive Chinese emotional prosodies vocally expressed by Chinese native speakers, which is research question one. This study also aimed to map out the confusion patterns of the listener groups, which

addresses a sub-question of the research question one. The results were used as the control condition for the second judgment study.

The results show that the three listener groups were able to identify the six Chinese emotional prosodies above chance level (chance level = 16.7%). The Dutch L2 learners of Chinese performed best in the perception experiment (with the mean recognition rate 54%) and native Chinese and Dutch naïve listeners performed equally well (with the mean recognition rate 46% in both cases). These results show that Dutch naïve listeners could recognize emotional prosody in an unknown language as well as native listeners of that language (Chinese); Dutch L2 learners of Chinese identified the Chinese emotion prosodies better than naïve Dutch listeners and therefore also better than the native Chinese listeners themselves. This finding is compatible with earlier results by Shoshi and Gagnié (2010) showing that experienced L2 learners of a language are better than listeners with no experience of the L2 in perceiving vocal emotions in the L2. Detailed confusion patterns can be seen from Table 3.2. Basically, 'neutrality' was identified correctly most often among all the emotions by the three listener groups, followed by 'anger' and 'surprised'. 'Sarcasm' was not well identified by all the listener groups, implying that some non-primary emotions, e.g. 'sarcasm', are more individual-specific. Chinese native and Dutch naïve listeners shared some similar confusion patterns; however, the confusion patterns of Dutch L2 learners of Chinese were in between of those of the Chinese native and Dutch naïve listeners. This shows that L2 learners of a target language have a hybrid system when perceiving emotional prosody produced in the target language. This hybrid system is partly influenced by speakers' L1 and partly affected by the target language itself. This would in return explain why the confusion matrix of the Dutch advanced learners of Chinese was in-between of the matrices of the native Chinese and the Dutch naïve listeners in the first judgment study.

8.2.2 Perception of L2 Chinese emotional prosody by three listener groups

The second group of research questions is: how well can native Chinese, Dutch naïve listeners and advanced Dutch learners of Chinese perceive the six Chinese emotional prosodies vocally portrayed by the Dutch L2 speakers of Chinese? What will be the confusion patterns of the three listener groups? Will the confusion patterns be similar to those in the perception experiment where the emotional prosodies were vocally portrayed by native Chinese speakers? In order to answer these questions, the same listener groups, i.e. native Chinese, naïve Dutch listeners and advanced Dutch learners of Chinese, participated in the first perception experiment of the second judgment study, in which they perceived Chinese emotional prosodies produced by the L2 speakers of Chinese.

The results show that the three listener groups could not perceive the same six Chinese emotional prosodies vocally portrayed by L2 speakers of Chinese as well as they did with those produced by native speakers. The mean recognition rates of the three listener groups are 39% with Chinese native listeners, 38% with Dutch naïve listeners and 41% with advanced Dutch learners of Chinese. Even though the advanced Dutch learners of Chinese performed slightly better than the other two listener groups, there was no significantly better group in the perception test, which means that the three

listener groups did equally well/poor. It also implies that L2 speakers produced emotional prosodies that are less recognizable than those expressed by natives. The detailed confusion patterns of the three listener groups can be seen in Table 4.3. Admittedly, there are similarities between the confusion patterns obtained from the Chinese native and the Dutch naïve listeners in the perception of the non-native produced emotional prosodies. However, the confusion patterns are not very similar to those that were obtained in the perception experiment where the emotional prosodies were vocally portrayed by native Chinese speakers. It implies that the native Chinese speakers and the L2 speakers of Chinese may have used different vocal correlates to express the emotions. In the perception of the L2-produced Chinese emotional prosodies, the confusion categories which advanced Dutch learners of Chinese fell into are quite similar to those of naïve Dutch listeners. For example, they showed the exact same tendency as naïve Dutch listeners for 'sarcasm': they often misidentified 'sarcasm' as 'neutrality' (19.7%) and 'surprise' (19.7%); and naïve Dutch listeners misrecognized it with 'neutrality' (18.4%) and 'surprise' (18.4%). At this point, one can assume that L1-transfer is an important strategy in interpreting paralinguistic meaning (e.g. emotional prosody) in L2. However, the hybrid system seen in the first judgment study did not apply to the advanced Dutch learners of Chinese in the perception of the L2-produced Chinese emotional prosodies. It implies that advanced learners of a target language may mainly resort to their native language to perceive emotional prosody expressed by L2 speakers from the same linguistic group.

Overall speaking, the difference between the confusion matrices of the three listener groups in the perception of native and non-native produced Chinese emotional prosodies is big. It implies that native and non-native speakers may have used very different acoustic correlates to produce emotional prosody in the target language. It also indicates that perception of non-native produced emotional prosody is more language-or-culture specific.

The confidence scale was originally introduced 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. It turned out that the results of the first and second judgment studies proved insensitive to any weighting based on response confidence. However, I would like to make a use of the confidence rating all the same to look tangentially at the social behavior of the three listener groups in the two judgment studies. The results show that Chinese native listeners were more confident in perceiving emotional prosodies produced by L2 speakers of Chinese. The reason for this behavior is not clear since one would expect listeners to be more confident when having to make decisions based on materials produced by speakers who share the same linguistic code. In contrast to the above, both of the naïve Dutch listeners and the advanced Dutch learners of Chinese were confident in the perception of Chinese emotional prosody produced by both natives and non-native speakers.

8.2.3 Production of emotional prosody in Dutch L2 speakers' L2 and L1

The second judgment study had two aims: one is to test whether L2 speakers of Chinese are able to produce emotional prosodies in the L2 as well as they do in their L1; the other is to find out the similarities and differences between the two productions. Twenty native Dutch listeners identified the same six emotional prosodies produced by the same L2 speakers of Chinese in their native language – Dutch. The results show that the mean correct recognition rate of the native Dutch listeners increased significantly when the emotions were vocally produced in speakers' L1. It means that the Dutch L2 speakers of Chinese are better at vocally expressing emotions in their native language. According to the combination study of the first and second judgment studies in Chapter 5, we can see that the Dutch L2 speakers of Chinese were not able to vocally produce emotion very successfully as native Chinese did, even though their proficiency of Chinese is very high. Neither were they able to express emotional prosody in their L2 as well as in their L1. Therefore, we can conclude that L2 limits L2 speakers' vocal expression of emotions, especially when L2 speakers' L1 is not a tonal language but their L2 is.

The confusion matrices suggest that the advanced L2 speakers of Chinese have developed an in-between manner of producing emotional prosody in the L2. This in-between manner is neither very much similar to the production of Chinese native speakers nor completely like the manner they use to produce emotional prosody in their L1 (Dutch). There are more explanations about this in-between manner in § 8.2.6 in this chapter, which is named as 'the hybrid system'. Moreover, the results also suggest that L2 speakers' ability of successfully producing emotional prosody in the L2 cannot be obtained automatically during the learning process of the L2 in general. It seems that this ability does not go along with the increasing of one's language proficiency in the L2. It is not clear at this stage whether this ability can be trained by designed curriculums. This needs to be tested in practice.

8.2.4 Lexical tone and expression of emotional prosody

At the beginning of the present study, I predicted that if a language uses a prosodic parameter for linguistic purposes, it will have less space for non-/paralinguistic use of the same parameter. If this prediction were true, speakers of a lexical tone language (such as Mandarin) should 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). In a more extreme version, it can be interpreted such that listeners of a non-tonal language are generally better in perceiving emotional prosody than listeners of a tonal language. Interestingly, the results of the first and second judgment study showed that naïve Dutch listeners, whose L1 is a non-tonal language, performed equally well as Chinese native listeners, whose L1 is a tonal language. Moreover, the L2 learners of Chinese performed even better than native Chinese listeners did.

In addition, a perception experiment designed in a reciprocal way was carried out later, in which Chinese and Dutch novice listeners perceived the six vocal emotions

expressed in their native language and in the other language. The results showed that Chinese listeners were only able to identify emotional prosody expressed in their L1 (Chinese) reasonably well but failed to recognize the same emotional prosody portrayed in an unknown language (Dutch) above chance level (= 16.7%). This means that Chinese novice listeners of Dutch are not able to identify emotional prosody in an unknown language (Dutch); they can only recognize emotional prosody in their L1 well. On the contrary, Dutch novice listeners of Chinese could recognize emotional prosody expressed in an unknown language (Chinese) reasonably well (mean recognition rate = 46%); and they identified emotional prosody portrayed in their native language (Dutch) significantly better than they did with the unknown language – Chinese. These findings together support the prediction made earlier – that listeners of a non-tonal language are generally better at perceiving emotional prosody than listeners of a tonal language. Furthermore, this prediction can also explain why L2 speakers of a tonal language are not able to express emotional prosody as well as they do in their L1. If a tonal language limits both L1 and L2 speakers of the language in the production of emotional prosody, it would limit L2 speakers more. Because L1 speakers of the tonal language can automatically get the lexical tones correctly while producing emotional prosody, but L2 speakers of the tonal language cannot do it automatically. In other words, when L2 speakers of the tonal language try to vocally portray emotions in the tonal language, they may have to pay much attention to getting the lexical tones right while portraying the emotion at the same time. It will result in that they have even lesser room than native speakers for paralinguistic uses, as they have to make lexical tones right first consciously or unconsciously and then put emotional prosody on top of the lexical tone afterwards. It is very difficult for L2 speakers of a tonal language to do these two things at the same time, even though they have very high proficiency of the L2. Therefore, the Dutch L2 speakers did not portray the emotional prosodies as well as they did in their L1. In summary, the prediction is a sensible explanation for the results. But it needs to be tested more extensively, for example, it can be tested with British naïve listeners and British L2 learners of Chinese; or German naïve listeners and German learners of Chinese perceiving Chinese emotional prosody and vice versa.

8.2.5 Acoustic correlates of emotions: recognition by humans and machine

In Chapter 6, I examined the value of eight parameters as correlates of the six emotions studied. The eight parameters were the same for each of the three groups of speakers, i.e. Mandarin L1, Mandarin L2 and Dutch L1 (the latter two were the same individuals). The acoustic analysis shows that fundamental frequency, including mean F0, SD_F0 and 'slope of the F0', is an influential variable in the production of vocal emotions by the three groups of speakers. This finding confirms the study of Scherer (1996), who claimed that F0 plays a crucial role in the production of emotional prosody. However, jitter and standard deviation of the intensity did not contribute much to differentiating between emotions in the present study. Never were more than two subgroups of the emotional prosodies differentiated for any of the three speaker groups. This finding contradicts to the traditional claims, indicating that jitter and intensity play an important role in the production and perception of vocal emotions (e.g. Bachorowski 1999, Biersack & Kempe 2005, Scherer 1996). The acoustic analysis also shows that 'tempo' and 'compactness' were only sensitive to Mandarin L1 speakers, for whom three

subgroups of the emotional prosodies were found. Slope of the F0 indicates that Chinese uses rising intonation to express surprise, which confirms the previous studies (e.g. Yip 2006), claiming that many tonal languages use rising intonation to express surprise. Moreover, HNR was only relatively sensitive to Mandarin L2 speakers but not to Dutch and Mandarin L1 speakers. These altogether further show that fundamental frequency is a very influential variable in the production and perception of vocal emotion in general; however, other acoustic factors are not universally important; they are more language-specific or emotion-specific.

The acoustic analysis indicates that some basic emotions such as 'happy' and 'angry' can be clearly discriminated from each other by mean F0 and SD_F0, regardless the speaker type. 'Happy' is characterized by high values for mean and SD of F0 (z-values close to 1) while 'angry' has z-values close to 0. 'Neutral' is also universally differentiated from 'happy' and 'angry', viz. by low values for mean and SD of F0 (z-values close to –1). However, more controlled emotions, e.g. 'surprised' and 'sarcastic', are not well classified by any of the eight parameters examined in Chapter 6. HNR can clearly distinguish 'sad' from 'neutral' with Mandarin L2 and Dutch L1, who were actually the same individuals, but not in the case of L1 Mandarin speakers.

Furthermore, some other factors can also influence the perception of vocal emotion, for instance, personal interpretation of the emotional label. Since 'surprised' includes both positive and negative surprise, the human listeners sometimes misinterpreted this emotion as 'happy' or 'angry', respectively. In addition, some emotions were expressed differently by male and female speakers within one particular language. Moreover, the acoustic analysis shows that vocal emotions cross-culturally were produced diversely too. Therefore, we can conclude that production and perception of vocal emotion by humans is a much more complex and integrated procedure. It involves not only acoustic correlations but also other factors, such as, sex, language or personal interpretation of the emotional label. This conclusion is in the line with the previous studies, claiming that other factors including talker sex, talker linguistic background, talker identity and personal interpretation of emotional label may well also prove to be important in the production of vocal emotions (Bachorowski 1999, Scherer 2003).

In Chapter 6, I also conducted the automatic recognition (Linear Discriminant Analysis – LDA) of the six emotional prosodies. The results of LDA show that the automatic recognition in the present study can recognize human-produced emotional prosody well above chance level (50% overall correct). There was significant correlation between confusions obtained by the automatic recognition and by the human listeners in the present study. Moreover, the overall recognition rate of LDA is slightly better than that of the human perception. This indicates that automatic recognition can reflect human perception of emotional prosody to some extent; however, the human perception is still different from the computer perception. There are still acoustic correlates which used by the algorithm to discriminate between emotions but not used by L1 and L2 listeners in reality. In addition, the Stepwise LDA shows that there are four parameters which significantly contribute to the production and perception of emotional prosody: 'utterance duration', 'fundamental frequency', 'compactness' and 'HNR'. There may also be some other acoustic parameters contributing to the

production and the perception of emotional prosody in general, which have been missed in this dissertation.

8.2.6 L1-transfer and the hybrid system

The acoustic analysis indicates that Dutch L2 speakers use some acoustic parameters in the production of emotional prosodies in the L2 (Chinese) the same way they do in their L1 (Dutch), e.g. standard deviation of F0 and standard deviation of intensity. Therefore, we may conclude that L1-transfer is a strategy for L2 speakers to vocally produce emotions in the L2. However, this strategy may not work for all the emotions, e.g. not for 'surprise' and 'sarcasm'. Moreover, the results of the acoustic analysis show that the advanced L2 speakers of Chinese have developed a hybrid system of producing emotional prosody in the L2. This (L2) hybrid system approximates to some extent the Chinese native manner of portraying vocal emotion (the way it involves utterance duration, mean F0, slope of the F0, compactness and jitter), but exploits the variability in F0 and intensity that the L2 speakers use to produce emotional prosody in their L1. Emotional prosodies produced in this in-between manner were identified above chance level by both the native and non-native listeners in the present study. However, these emotional prosodies are less recognizable overall (41% correct within-group identification) than those produced in the Chinese native manner (46% correct). This would indicate that the expression of emotion through prosody is limited in an interlanguage. It further supports that L2 limits the expression of emotional prosody. The results in the present study also suggest that the L2 speakers did not automatically acquire the native approach to vocally produce emotional prosody in the target language during their L2 learning process. Therefore, it seems that, in a situation where there is no particular training for how to produce emotional prosody in L2, L2 speakers will create their own hybrid system/manner to express vocal emotions in the L2. This hybrid solution works for some emotions but not for all. Furthermore, this hybrid solution was also seen in the perception of native-produced Chinese emotional prosody by Dutch advanced learners of Chinese (see § 8.2.2). It implies that this hybrid system/solution may be applied in both of the production and the perception of emotional prosody in an L2.

In fact, the hybrid system can be seen as an extension of the interlanguage theory in terms of paralinguistic communication in L2. The interlanguage theory claims that L2 learners of a target language who have not become fully proficient yet but are approximating the target language, will preserve some features of their first language (or L1) or overgeneralize target language rules in speaking or writing the target language and creating innovations (Selinker 1972). In the present study, we have seen the same phenomena happen in the production and the perception of emotional prosody in Chinese by Dutch L2 advanced learners of Chinese. It remains to be seen to what extent the learning of emotional prosody can be modelled with rules and to what extent the concepts of the interlanguage theory apply to paralinguistics.

8.2.7 In-group advantage and cross-cultural perceptual ability of vocal emotion

The third judgment study conducted in a reciprocal way was designed to find the answer to the sixth research question, which is whether the in-group advantage found by other researchers is universal, claiming that listeners are better at recognizing emotional prosody produced in their native language than in their L2 or an unknown language. Twenty Chinese and 20 Dutch native listeners who did not know any Dutch and Chinese (respectively) perceived the emotional prosodies in their L1 and in the other language. In this case, Dutch was the unknown language for the Chinese native listeners; conversely, Mandarin Chinese was the unknown language for the Dutch native listeners. In the third judgment study, all the emotional prosodies were expressed by native speakers of Chinese and Dutch. The results show that both Chinese and Dutch native listeners recognized emotional prosodies better in their native language than in the unknown language. This finding suggests that the in-group advantage might be universal – but only in this weaker sense; it does not necessarily mean that vocal emotions are always recognized best by native listeners of the language the emotions are produced in. As a case in point, Dutch novice listeners identified the Chinese emotional prosodies as well as the natives did – and recognized the emotional prosody in their L1 much better. In contrast to this, Chinese native listeners identified the emotional prosody in their L1 reasonably well but failed to recognize the vocal emotion expressed in the unknown language (Dutch). This indicates that Dutch and Chinese listeners have very different abilities of identifying emotional prosody in an unknown language. It means that some culture groups might be generally better than some other culture groups in identifying emotional prosody. This further supports the view that cross-cultural perception of emotional prosody is not necessarily symmetrical.

8.2.8 Universal vs. culture-or-language specific

The ultimate goal of this dissertation is to find the answer to the last research question: whether perception and production of vocal emotion are universal or rather more language-specific and/or culture specific. The first two judgment studies tested how well native Chinese, naïve listeners and advanced learners of Chinese perceived Chinese emotional prosodies produced by native and L2 speakers; and how well L2 speakers of Chinese produced emotional prosody in the L2 and in their native language. The third judgment study investigated how well Chinese and Dutch novice listeners perceive emotional prosodies expressed in their L1 and in the other language. The three judgment studies all show that both perception and production of emotional prosody are partly universal and partly language-or-culture specific.

Perception of some basic emotions, e.g. 'anger' or 'neutrality', involves more universal acoustic cues. However, for some non-basic emotions, for instance, 'sarcasm' or 'happiness', perception is more culture-or-language specific. According to the confusion matrices of the naïve Dutch listeners and the advanced Dutch L2 learners of Chinese, we can infer that L1 influences L2 learners' perception of emotional prosody in the L2 to some extent. It further supports that perception of emotional prosody either in one's L1 or in one's L2 is partly universal and partly language-or-culture specific. Moreover, the third judgment study showed a strong in-group effect. It

confirms that emotion is generally better recognized between native listeners and native speakers from the same cultural group (e.g. Kilbride & Yarczower 1983, Markham & Wang 1996). However, the Chinese novice listeners in the third judgment study were not able to identify the emotional prosodies in an unknown language (Dutch) above chance level. Therefore, we can conclude that perception of emotional prosody in an unknown language is only universal to some extent – and that it is rather more cultural-or-language specific than universal.

From production point of view, L1 still plays a role in the production of emotional prosody, which means that L2 speakers of a target language might use L1-transfer to vocally express emotions in the target language. It seems that the advanced L2 speakers have already created their own approach (the hybrid system) to portray recognizable emotional prosody in the L2 after being learning the target language for several years. However, when L2 speakers were asked to produce emotional prosody in their L1, they expressed the emotional prosody much better than they did in the L2. It shows that production of emotional prosody is more language specific. However, for some basic emotions, such as 'anger', 'sadness' or 'neutrality', the acoustic correlates that L1 and L2 speakers used were very similar, implying that vocal production of emotion is universal to some extent. In summary, production of emotional prosody is more language-or-culture specific, although it is universal to some extent for some basic emotions. Moreover, different cultural groups have their own manner to express emotional prosody, which cannot be acquired by L2 speakers of that language along the increase of their proficiency of the target language.

8.3 General discussion

I carried out three judgment studies in this dissertation to answer the research questions. The eight research questions including the sub-questions of each research question listed in Chapter 1 have been carefully answered one by one. However, there are still ambiguities which need to be clarified in the future:

1)   The functional view needs to be further tested

In the beginning of this dissertation, I introduced a functional view which claims that the prosodic space which languages may use is finite. Therefore, if a language uses vocal pitch for lexical purposes (i.e. for the marking of lexical tone), pitch will be used to a lesser extent for the expression of paralinguistic contrasts, such as signaling emotion. It could also be interpreted as 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 advanced Dutch L2 learners of Chinese performed better than the native Chinese listeners when perceiving the emotions in Chinese. Moreover, novice Dutch listeners recognized the Chinese emotional prosodies as well as the native listeners did. These results altogether suggest that this functional view might be true, indicating that listeners of a non-tonal language are generally better than listeners of a tonal language at recognizing vocal emotions. However, this dissertation only investigated one tonal language (Mandarin Chinese) in terms of the production and the perception of emotional prosody. Therefore, it is difficult to know

at this stage whether the functional view is universal or not. It may be universal to some extent; for instance, it explained the unexpected results that the Dutch listeners outperformed the Chinese listeners or performed at least as well as the native listeners did. However, in order to get a complete picture of whether the functional view applies universally, future studies should investigate more linguistic groups. For instance, future studies can test how well listeners of a western non-tonal language (e.g. German, English or Spanish) perceive vocal emotion expressed in an Asian tonal language (Thai, Lao or Vietnamese) or in a sub-Saharan African language (Wolof, Koyra or Fulani) and vice versa.

2)   'L2 limits the expression of vocal emotion.' should be further investigated

The results of the first two judgment studies suggest that L2 limits the expression of vocal emotion. However, one may argue that it was not L2 which limited the expression of vocal emotion in the present study; actually, it might be the lexical tones which constrained the Dutch L2 speakers of Chinese to vocally produce emotions in the L2. This argument may be true, but it is not possible to purely separate lexical tone from emotional prosody in practice, since tone and (emotional) prosody both are phonetic suprasegments and they interact and entangle with each other in the daily communication. Therefore, it is suggested that future studies should involve some other L2 speaker groups whose L2 is a non-tonal language (e.g. Dutch L2 speakers of French) to further investigate whether it is the L2 or the lexical tone or both which limit the expression of vocal emotion in the L2.

3)   More acoustic correlates need to be further studied

In this dissertation I only examined eight acoustic parameters, i.e. tempo, mean fundamental frequency, standard deviation of fundamental frequency, slope of the fundamental frequency, compactness, standard deviation of intensity, jitter and Harmonics to Noise Ratio in which tempo, F0, intensity and jitter were proved to be important variables in the production of vocal emotions in the previous studies. I also ran the automatic recognition of the six emotional prosodies to test whether the Linear Discriminant Analysis reflects the perception of emotional prosody by human listeners. The results showed that the LDA reflected human perception of vocal emotion to some extent. However, the human perception of the six emotional prosodies was still different from the perception of the LDA. It shows that some acoustic parameters used by LDA might not be used by the human listeners or vice versa. There may also be some other acoustic parameters contributing to the production and the perception of emotional prosody in general, which have been missed in this dissertation. Therefore, it is necessary for future studies to investigate more acoustic variables in the production of vocal emotion in speakers' (non)-native language; or adopt other methodologies to further study what factors influence the production of vocal emotion in an L2.

4) The hybrid system needs to be further tested

The results of the acoustic analysis showed a hybrid system developed by the L2 speakers of Chinese. This hybrid system needs to be further investigated, since the present study only studied one L2 speaker group, i.e. Dutch L2 speakers of Chinese. Therefore, it is not clear whether this hybrid solution is adopted universally by L2 speakers of other languages when they vocally produce emotions in the L2. Therefore, it is worth re-running the production experiment with some other L2 speaker groups, including both speakers of tonal and non-tonal language as a second language. For instance, one can further test how well English L2 speakers of French or English L2 speakers of Thai produce emotional prosody in French or Thai. The more L2 speaker groups are investigated, the clearer the picture is. However, I suspect that it is not easy for researchers to find L2 speakers of a tonal language whose native language is a non-tonal language to participate in a production experiment as subjects, e.g. English L2 speakers of Thai or Vietnamese, as these tonal languages are not commonly taught as foreign languages in the west.