JAIST Repository
https://dspace.jaist.ac.jp/

Title	Comparison of emotional speech perception among multiple languages on emotion dimensions by different native language groups
Author(s)	韓，笑
Citation	
Issue Date	2015-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12651
Rights	
Description	Supervisor:赤木　正人，情報科学研究科，修士

JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Comparison of emotional speech perception among multiple languages on emotion dimensions by different native language groups

By Xiao HAN

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Professor Masato Akagi

March, 2015

Comparison of emotional speech perception among multiple languages on emotion dimensions by different native language groups

By Xiao HAN (1210205)

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Professor Masato Akagi

and approved by
Professor Masato Akagi
Professor Jianwu Dang
Associate Professor Masashi Unoki

February, 2015 (Submitted)

Copyright © 2015 by Xiao HAN

Abstract

From the experience in our daily life, it is found that human beings can judge the emotional states of a voice only by listening. This kind of situation occurred not only when they listened to their native languages, but also when they listened to non-native languages they do not familiar with. It suggests that there is another way to communicate each other without common language. Since the investigation of the fundamental knowledge on how human beings perceive emotinonal states of different languages can help us to understand human perception of emotional states, can guide us to build a speech emotion recognition system independent to languages and can provide us fundamental knowledge to sythesis expressive speech regardless of language, it is considered to have important meanings.

Considering to the emotion perception of different languages, researchers have already concerned with comparison of speech emotion perception among multiple languages and carried out several experiments among subjects in different native speakers. There are mainly two classifying approaches have been used to capture and describe the emotional content in speech in these previous studies: categorical approach and dimensional approach. In this research, since by using the dimensional approach, the degree of intensity in emotional state in real-life can be clearly presented, the dimensional approach are selected.

To compare the commonalities and differences among multiple languages on Valence-Activation approach, a listening test was carried out. In the listening test, thirty listeners from three different countries, China, Japan and Vietnam, were invited to evaluate emotional contents included in five different languages, Chinese, Japanese, Vietnameses, American English and German. Four common emotional states: neutral, happy, angry and sad are selected from the five databases for subjects to evaluate.

The results of the listening test reveal that the position of neutral states on Valence-Activation approach are not significantly different, the directions of emotional states are similar and the distance are significantly different among the three different native language groups. From the results of the listening test, we understand that the commonalities are the evaluation resulsts by different native language groups among multiple languages have similar position of neutral voice and similar direction of emotional states on Valence-Activation approach. On the other hand, the difference among multiple languages evaluated by different native language grouops are they have different evaluations of the degree of emotional states. Moreover, the commonalities and differences we found out in this research can help us to understand that when human beings perceive the emotional states in speech, different listeners would have same feeling when they perceive neutral voice, but different feeling when they perceive other emotinal states. It also reveals that, only by using the same position of neutral and controling the degree of each emotional states, researchers can construct a speech emotional recognition system and synthesis expressive voices independent one languages. *Keywords*: Human perception, multiple languages, native language group, emotional state, dimensional approach

Acknowledgements

I take this opportunity to express thanks to those who made a difference in my life in Japan. First and foremost, I would like to express most sincere gratitude to my parents, who give me my life, give me great love and fully understanding through all these years, which is my strongest motivation to go on my study and pursue my life goals.

Then, I am deeply indebted to my supervisor, Professor Masato Akagi, who has always supported me with his knowledge and enthusiasm. Although he is very busy, he is always willing to discuss with me if there are problems with my research and give me advices. Without his consistent help, this thesis would have not been completed or even written.

Thirdly, I would like to thank Associate Professor Masashi Unoki, my sub supervisor for his guidance and support throughout my study. Thanks to his comments, not only my presentation slides but also my thesis become better and better. I also would like to express my thanks to Professor Jianwu Dang for taking time to be at my mid-term defense as well as the final dissertation and give helpful comments for my research.

Moreover, I would like to thank Doctor Reda Elbarougy, assistant professor Daisuke Morikawa and assistant professor Ryota Miyauchi, who give me insightful comments on my research. Also thank to all of the members in our laboratory, who give me useful knowledge that help me to finish my experiment. And thank to the subjects who participate in me experiment, without their kind helping, the research could not be finished.

Last but not least, thank to my friends and my families, their helps and encouragements make me through many difficult periods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Communication is a essential part of human beings' social life, and speech communication is one of the most common way for us to communicate each other. When human beings communicate with others, there are three kind of information included in speech signal, linguistic information, paralinguistic information and nonlinguistic information, though there are no clear boundaries to distiguish them. In previous study[1], the linguistic information have been defined as the symbolic information that is represented by a set of discrete symbols and rules for their combination, and paralinguistic information is difined as the information that is not inferable from the written counterpart but is deliberately added by the speaker to modify or supplement the linguistic information. On the other hand, nonlinguistic information concerns such factors as the age, gender, idiosyncracy, physical and emotional states of the speaker, etc. To make listener understand speaker's meaning, not only the linguistic information, but also the paralinguistic information and nonlinguistic information, especially the emotion of speaker's, play important roles in human communication on speech. Understanding the information of speaker's emotion in his or her speech can help listeners to undertand speaker's meaning more accurately.

Moreover, in the growing internationalization of modern society, international communication is considered more and more important. Therefore, how to communicate with foreigners more convenient has become a popular topic. From the experience in our daily life, it is found that human beings can judge the emotional states of a voice only by listening. This kind of situation occur not only when they listen to their native language, but also when they listen to non-native languages they do not familiar with. It suggests that there is another way to communicate each other without common language.

In order to make communication more grobal, and do not be influented by language, nationality and culture, investigation of the fundamental knowledge on how human beings perceive emotinonal states of different languages is considered to have important meanings. The investigation can help us to understand human perception of emotional states, can guide us to build speech emotion recognition system independent to the languages. Moreover, it can provide us fundamental knowledge to sythesis expressive speech.

1.2 Backgroud

Considering to the emotion perception of different languages, some previous studies[2, 3, 4] have already concerned with comparison of speech emotion perception among multiple languages. Researchers have carried out experiments among subjects in different native languages. Huang[5, 6] have compared Japanese expressive speech perception between Japanese and Taiwenese listeners. The comparison of the resulting connections showed that 60% of the adjectives are used commonly by both Taiwanese and Japanese listeners. This commonality suggests how expressive speech communication is improved by nonlinguistic information. On the other hand, Polzehl[7] comparied recognition of anger voices among German and English. As a result of his research, he obtain a single bi-lingual anger recognition system which performs just as well as two separate mono-lingual systems on the test data. The feneral results of his experiments indicated a reasonalble and reliable multi-lingual anger recognition. Moreover, Abelin[8] presented an experiment in cross-cultural multimodal interpretation of emotional expressions among Spanish, Swedish, English and Finnish listeners. The research is aimed to see if the interpretation is dependent on the listeners cultural and linguistic background. Furthermore, in order to develop a generalized emotion recognition system, Elbarougy[9] analysis his system in cross-linguage as well as mono-language. The results of his research suggest that the cross-lingual system he proposed performs just as well as the two separate mono-lingual systems for estimating emotions. In these previous studies, the commonalities among different languages on speech emotion perception, which are considered as the fundamental knowledge on how human beings perceive emotional states among multiple languages, are concerned on in acoustice feature domain.

1.3 Purpose of this study

The purpose of this research is to investigate commonalities and differences among multiple languages of human perception for emotional states from speech signal. We will compare the commonalities and differences among multiple languages, which are evaluated by different native language groups. The results of this research can help us to understand how human beings perceive emotional states in speech and can guide us to construct a human perception model regardless of languages. Moreover, by useing the fundamental knowledge we collected in this research, a grobal speech emotion recognition system can be expected in the futrue. Futhermore, these results can provide some basic data to synthesis expressive speech.

1.4 Outline of the thesis

The thsis is organized as follow:

Chapter 1 is the introduction about this thesis. We present the research motivation, research background and the purpose of this research in this chapter.

Chapter 2 introduces a very important theory of emotion representation, which is used in my research. Two types of emotion representation will be described in the chapter. By discussing the lack points and the merits of these two emotion representation, which representation will be selected in the research and the reason why we select it will be presented.

Chapter 3 presents the detail information of the experiment. It includes databases, subjects, equipment and two parts of the listening test. we will introduce the procedure of these two parts, and then concerns on the results of the listening test.

Chapter 4 discusses results of my research. Not only discuss the results of the listening test, but also discuss the purpose of the research in general meaning.

Chapter 5 is the conlusion chapter. We will summary the research, draw contributions and mention the future work of the research in the chapter.

Chapter 2

Emotion Representation

To investigate the commonalities and differences among multiple languages on human perception of emotional states, the first question we faced is how to represent emotional states. In the previous studies[10, 11], there are mainly two classifying approaches have been used to capture and describe the emotional content in speech: categorical approach and dimensional approach. Categorical approach is based on the concept of basic emotions such as anger, happyness and sadness, which are the most intense form of emotion, and all other emotions are considerd as the variations or combinations of them. On the other hand, dimensional approach represents emotional states using continuous multi-dimensional space. Both of the two emotion approaches provide complementary information about the emotional expressions observed in individual.

2.1 Categorical approach

2.1.1 Theory of catigorical approach

The categorical theory proposes the presence of basic emotional states. The simplest description of emotions is the use of emotion category labels. Several previous researchers[12, 13] have used the discrete emotion categories such as neutral, happyness, anger, sadness and so on to represent emotional states. Discrete categorization allows a more particularized representation of emotions in applications where it is needed to recognize a predefined set of emotions. However, this approach ignores most of the spectrum of human emotional expressions. Some studies concentrate on only one or two selected categories

Figure 2.1 shows the concept of catigorical approach. The emotion in input speech signal is classified into discrete categories.

2.1.2 Weak point of categorical approach

Although the categories of emotional states can be clearly represented by using categorical approach, there is some weak points exist in the theory of catigorical approach. Since the choice of categories for a study varies and usually depend on an application that the

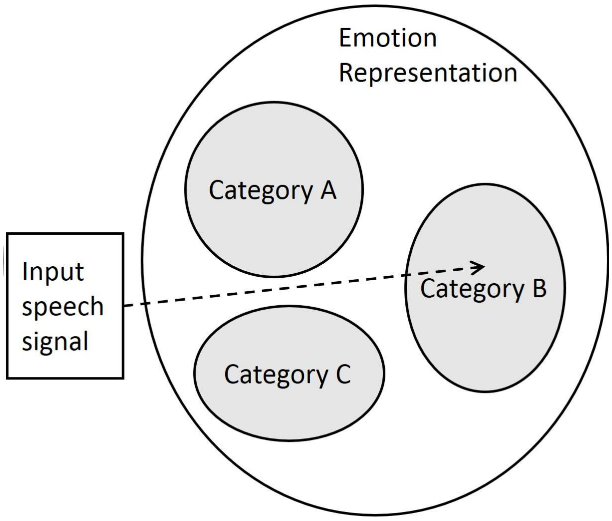


Figure 2.1: Concept figure of catigorical approach

researcher has in mind, the most frequently used categories may not be the most suitable ones for all research. Moreover, the problem that how many catigories they should use to describe the emotions in real-life is facing the researcher who using the catigorical approach. In the real-life case, emotions are not just a small number of discrete categories. When human beings perceive the emotional states in speech, not only the classification of emotional states as discrete categories is need to be focus on, the variability within a certain emotion is also important to be detected. This is supported by the fact that human soften or emphasize their emotional expressions flexibly depending on the situation in actual human speech communication. However, by using the catigorical approach, where each affective display is classified into a single category, a complex mental or affective state or blended emotions perhaps is too difficult to be handled[14].

2.2 Dimensional approach

2.2.1 Theory of dimensional approach

The emotions in real-life have different degree of intensity, and may change over time depending on the situation from low degree to high degree, for example, human beings may detect or describe the emotional states as little happy or very happy. In the categorical approach, since the emotions are classified into several discrete categories, the degree of intensity in emotions cannot be represented clearly. In order to overcome the weak point of categorical approach, the dimensional approach have been put forward. In 1954, psychologist Schossberg presented a thesis named *Three dimensions of emotions*[15]. In his thesis, he put forward the three dimension are Sleep-Tension dimension, Pleasantness-Unpleasantness dimension and Attention-Rejection dimension. Here, for Sleep-Tension dimension, the larger the value of Tension is, the stronger the emotion be percepted. For the other two dimensions, they are used to describe the emtions. Based on this three dimension model, previou researchers put forward the dimensional approach. On the dimensional approach[16], every emotional states are represented as different points on this multi-dimension space.

Figure 2.2 shows the concept of dimensional approach. The emotion in input speech signal is considered as a emotion vector on the multi-dimension space.

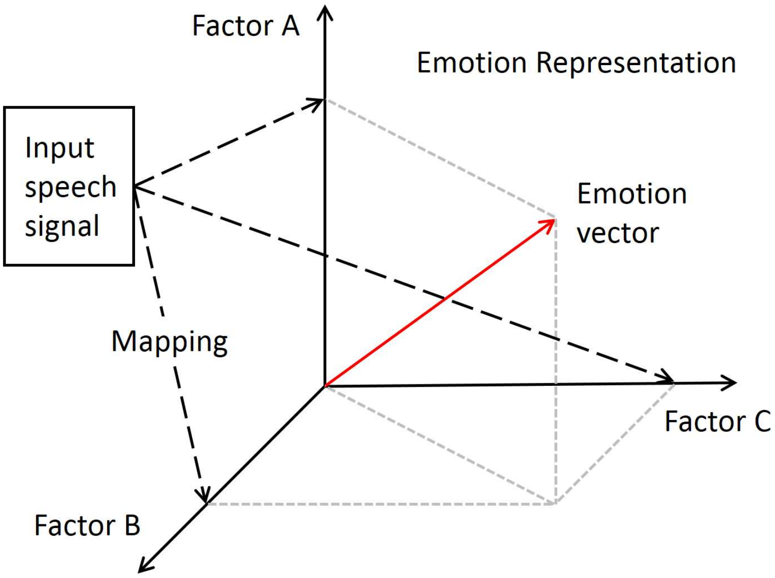


Figure 2.2: Concept figure of dimensional approach

2.2.2 Merit of dimensional approach

Emotion dimensions used in dimensional approach is a representation of emotional states which can fulfills the requirement of capability to express a large variety of emotional states from low-intensity to high-intensity. In the dimensional approach, the degree of intensity in emotional states can be represented clearly, as well as the category of emotional states. Hence, the continuous change of emotional states, which cannot be represented on categorical approach, can be easily represented by using the dimensional approach. Furthermore, the dimensional approach can represent not only single emotions in sppech signal, but also when more than one emotions are included in the speech signal, or when the emotion included in the speech signal are amphibolous. Therefore, the dimensional approach is more suitable than categorical approach to represented the emotional states in real-life.

2.2.3 Types of dimensional approach

Generally, there are several ways to represent emotions in a multi-dimension space. In previous studies, researchers mainly use two-dimensional approach and three-dimensional approach to represent emotions.

2.2.3.1 Two-dimensional approach

According to the three dimension model of psychologist Schlossberg, Cowie put forward a two-dimensional approach called Activation-Evaluation approach[17, 18]. On this emotion approach, the two dimensions are called Arousal (or Activation) and Valence (or Evaluation). Then Vogt[19] summaried the works of Cowie. He described valence or evaluation as values from positive to negative, arousal or activation as values from high to low.

Figure 2.3 are the figure which shows the concept of Arousal (Activation)-Valence (Evaluation) approach. In the figure, the positions of typical emotional states are shown on the space as examples.

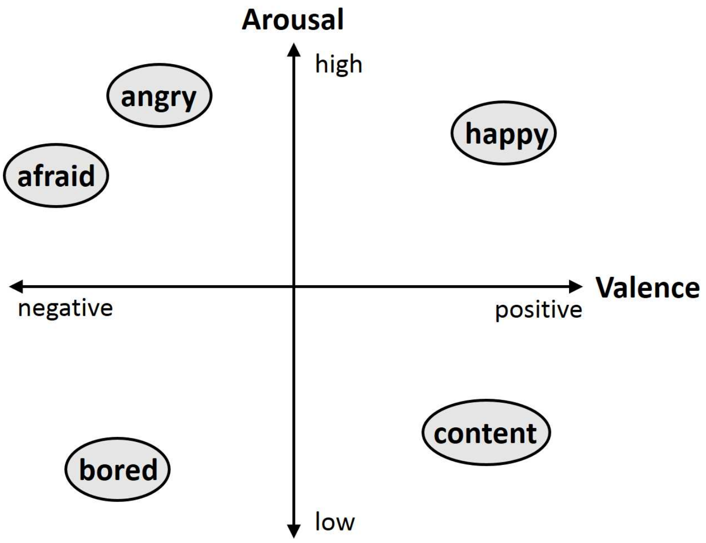


Figure 2.3: Concept figure of a two-dimensiona approach: Arousal (Activation)-Valence (Evaluation) approach.

2.2.3.2 Three-dimensional approach

Three-dimensional approach additionally include a third dimension defining the apparent strength of the person, which is referred to as dominance (or power).

Since using the Activation-Evaluation approach, which put forward by Cowie, the emotion anger and fear cannot be distinguished clearly, Grimm add the dominance dimension and discribe it as the values from strong to weak[20]. The three dimension in this approach are orthogonal. Thus, the approach can be represented as a cube just like Figure 2.4 shows us.

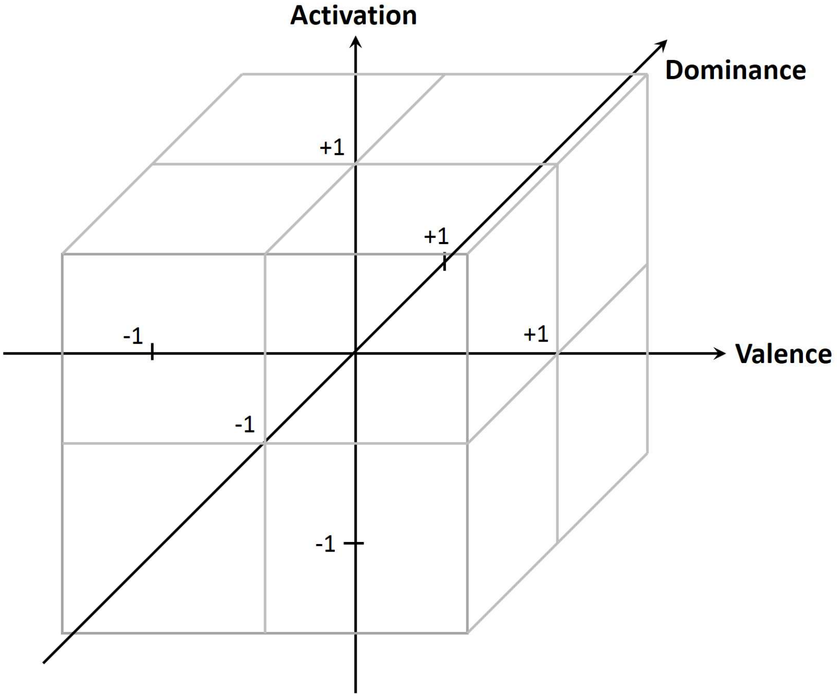


Figure 2.4: Three-dimensional approach: Valence-Activation-Dominance approach (or Activation-Evaluation-Dominance approach).

Moreover, Schroder[21, 22] also put forward a three-dimensional approach according to Activation-Evaluation approach. The third dimension in his works called power dimension, which is more important than activation dimension[23]. Especislly in the case of high activation[24], because that taking the level of control and social power of an individual into account is useful in distinguishing certain emotion. We can know from previous study[23] that anger and fearboth consist of similarly very negative and high actibation values and only can be distinguished due to their different values on the dominance dimension. Therefore, the third dimension is necessary to distinguish anger from fear.

2.3 Emotion representation used in the research

In this research, the commonalities and differences among multiple languages on human perception of emotional states are proposed to be investigated. According to the weak points of categorical approach and the merit of dimensional approach, which are discussed in the previous sections, the dimensional approach are selected to represented the emotional states because of its outstanding representation ability of the degree of itensity in emotios. Considering to the emotional states inclued in the databases we used are only four basic emotions, neutral, happyness, anger and sadness, a two-dimensional approach called Valence-Activation approach are selected.

Figure 2.5 shows the Valence-Activation approach used in the research. On the space, the four basic emtions, which are used in this research, are represented as examples, in order to make the theory of Valence-Activation approach easy to be understand.

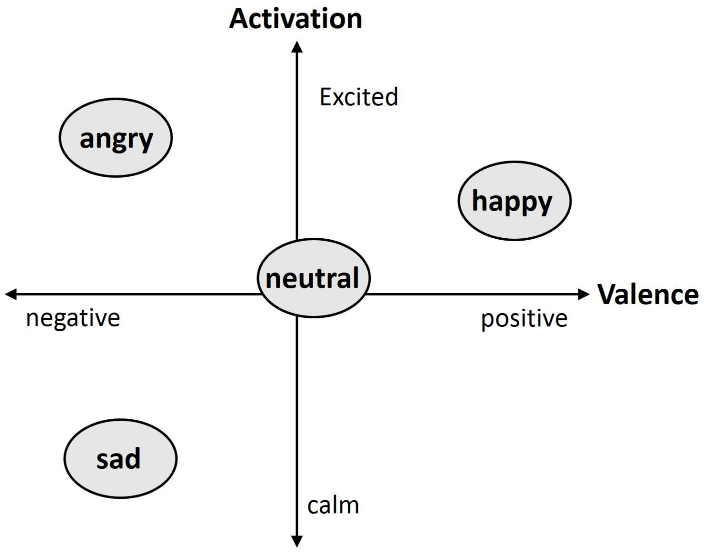


Figure 2.5: Two dimensional approach we used in this study: Valence-Activation approach

2.4 Emotion vector

On Valence-Activation approach, emotional states are presented as points distributing on the space and emotion vectors are used to discribe the emotional states. The emotion vector, which mentioned here, is the vector from the center of neutral states to the center of other emotional states. For one emotion vector, three factors of the emotion vector are have big meaning to present emotional states on Valence-Activation approach. The are the starting point, the angle and the length of emotion vectors, and these three factors present the position of neutral states, the direction of emotional states and the degree of emotional states on dimensional approach, respectively.

Chapter 3

Experiment

In the study, commonalities and differences among multiple languages for human perception of emotional states in speech are investigated on Valence-Activation approach. To compare the commonalities and differences among multiple languages, a listening test was carried out. In the listening test, thirty listeners from three different countries, China, Japan and Vietnam, were invited to evaluate emotional contents included in five different languages, Chinese, Japanese, Vietnameses, American English and German. Four common emotional states: neutral, happy, angry and sad are selected from the five databases. For each emotion category, the average value of valence and activation are used to calculate the central position of valence and activation of these emotional states. These central positions of all emotional states are compared among three listener groups for the five databases individually. The results of comparison can help us to find out the commonalities and differences among multiple languages.

3.1 Stimuli

In order to compare the emotional states in multiple languages, five emotional speech databases consisted of acted emotions in five different languages are selected in the listening test. The five databases, CASIA database, Fujitsu database, VNU database, IEMOCAP database and Berlin database are selected to be used as the databases of Chinese, Japanese, Vietnamese, Amarican English and German, respectively. The utterances we selected from the five databases are covered the four basic emotions, neutral, happiness, anger and sadness.

The Chinese database is produced and recorded by CASIA by 2 professional actresses and 2 professional actors. There are six different emotional states in this database: neutral, happiness, anger, sadness, fear and surprise. For each emotional state, there are 400 different sentences spoken by four speakers respectively. In these 400 sentences, 300 sentences are recorded by four subjects in four emotional states. The rest 100 sentences are different among four subjects. The total number of utterances in Chinese database are 1600.

For Japanese database, the Fujitsu database is selected. The Fujitsu database is pro-

Table 3.1: Numbers of selected utterances

Language Emotion	Chinese	Japanese	Vietnamese	English	German	Total
Neutal	28	20	28	28	28	132
Happy	28	30	28	26	28	140
Angry	28	30	28	28	28	142
Sad	28	30	28	28	28	142
Total	112	110	112	110	112	556

duced and recorded by Fujitsu Laboratory, and all utterances are produced by a professional actress. There are five emotion categories, neutral, happy, cold angry, sad and hot angry, are included in this database. Twenty different sentences are uttered in all of these emotional states, and each sentence has one utterance in neutral and two utterances in each of the other categories. Thus, there are nine utterances for each sentence and totally 180 utterances in this database.

The Vietnamese database is a database recorded by Vietnam National University - Hanoi[25]. Two famous artists, one is actor and the other is actress, are take part in the recording. When recording, the speakers spoke 19 different sentences in five different emotions, neutral, happy, cold angry, sad and hot angry. Hence, there are totally 190 utterances are included in this database.

Considering the English database, we selected an American English database named IEMOCAP database[26]. The full name of this database is interactive emotional dyadic motion capture database, and collected by the Speech Analysis and Interpretation Laboratory at the University of Southern California. The database was recorded from ten actors in dyadic sessions (5 ssions with 2 subjcts each). They wrer asked to perform three selected scripts with clear emotinalcontent. In addition to the scripts, the suvjects were also asked to improvise dialogs in hypothetical scenarios, designed to elicit specific emotions, happiness, anger, sadness, frustration and neutral state. In total, the database contains approximately twelve hours of data.

The German database is selected from the Berlin database, which comprises seven emotional states, neutral, happiness, anger, sadness, boredom, disgust and anxiety[27]. Each emotional category has ten different sentences, and all of these utterances are spoken by 5 professional German actors and 5 professional German actress. These sentences are not equally distributed among the various emotional states. There are 79 utterances in neutral, 71 utterances in happy, 127 utterances in anger, 61 utterances in san, 81 utterances in bored, 46 utterances in disgusted and 69 utterances in frightened. Finally, there are 534 utterances in Berlin database.

The numbers of utterances we selected from each database and every emotional states are listed in Table 3.1. The table shows that the numbers of utterances selected from each database are around 110 and the total number of utterances are 556.

Table 3.2: Numbers of listeners who take part in the listening test

Listener group	Female	Male	Total
Chinese subjects	5	5	10
Japanese subjects	5	5	10
Vietnamese subjects	5	5	10

Table 3.3: Subjects' knowledge of the five languages, where ⊙ means native speaker, ◯ means well understanding, △ means undertand a little and × means do not understand at all.

Language National	Chinese	Japanese	Vietnamese	English	German
Chinese subjects	⊙	△	×	◯	×
Japanese subjects	×	⊙	×	◯	×
Vietnamese subjects	×	△	⊙	◯	×

3.2 Subjects

In this research, the five databsses are evaluated by thirty subjects from three different native languages, in terms of valence and activation using a listening test. In the listening test, all utterances are evaluated by thirty listeners from three different countries and speak three different native languages. Since the languages of databases we used in the experiment are Chinese, Japanese, Vietnese, American English and German, It is better to select subjects from the five countries, China, Japan, Vietnam, American and Germany. Considering to the numbers of foreigners in this area, finally we selected ten Chinese, ten japanese and ten Vietnamese to participate in the listening test. In addition, all of them are graduate students from 20 to 35 years old. No listeners have hearing impairment.

Table 3.2 shows the numbers of subjects selected from each country. The number of female subjects are same to the number of males in the listening test.

For every subjects, their knowledge of the five languages are listed in Table 3.3, where ⊙ means native speaker, ◯ means well understanding, △ means undertand a little and × means do not understand at all. From this table, we can know that for the thirty subjects, their native language Chinese, Japanese and Vietnamese are all selected in the experiment. Chinese and Vietnamese subjects do not understand the language of each other, and they all undertand Japanese a little because now they are living in Japan. Japanese subjects do not understand Chinese and Vietnamese at all. For all subjects, they understan English very well and do not understand German at all.

3.3 Equipment

In the listening test, in order to achieve more exact values evaluated by subjects, all 556 stimuli were presented through a binaural headphones (STAX SPM-a/MK-2) at a comfortable sound pressure level in a soundproof room. A M-AUDIO Fast Track Pro

Table 3.4: Numbers of listeners who take part in the listening test

Dimension Score	Valence	Activation
-3	very negative	very calm
-2	mid negative	mid calm
-1	low negative	low calm
0	neutral	neutral
+1	low positive	excited
+2	mid positive	mid excited
+3	very positive	very excited

and Asio sterio driver are used as D/A device and driver to ensure the subjects can hear steady speech signal we selected. When subjects evaluate the values of valence and activation they are asked to used a notebook computer which have been well preparation. A MATLAB GUI was used to input given scores for evaluation. All subjects use same computer, same D/A device, same headphone and same program to evaluate the values in same soundproof.

3.4 Procedure

Most of the existing emotional speech databases are annotated using the categorical approach, including the selected databases. However, according to the advantages of dimensional approach, listening test in this research are required to annotate each utterance in the used databases using the dimensional representation. In order to obtain the positions of each utterance in the Valence-Activation approach, the five selected emotional databases are evaluated using a listening test in two terms, valence and activation, bythree listener groups in the listening test. Every listener is required to evaluate the emotion in the voices by his/her own perceived impression based on the way of speaking, but not on the content itself. A seven-point-scale are used for their evaluation. The seven point are -3, -2, -1, 0, 1, 2, 3 for both valence dimension and activation dimension. Table 3.4 shows values of the two dimensions and the meanings of every points in theses two dimensions.

The listening test was divided into two sessions, the evaluation of valence dimension and the evaluation of activation dimension. In the two sessions, listeners are asked to evaluate the scores of valence axis and activation axis independently. In each session, before start the evaluation, subjects are asked to read a document about the experiment and the attention point. Then, the purpose of the experiment, the basic theory of dimensional representation are introduced to them[28]. After they understand all of these, they need to sign on an agreement file.

After finished these preparation, subjects can start the listening test. One session of the listening test is consisted of three sections, training, pre-test and main-test. The training section and pre-test section can be considered as the preparation section of main-test. The results of training and pre-test do not be used in the analysis. Only the results of main-test are used in the analysis.

3.5 Training and pre-test

Before subjects take part in the main-test, they need to join the training and pre-test in order to make them been accustomed to the listening test and more suitable for the experiment.

3.5.1 Training

Training section is the first section of the listening test. In order to help subjects to understand the 7-point-scale of valence and activation, all subjects are asked to listen to all the examples in the training section. The utterances used in this section are covered all the seven points and all these utterances can be listened many times as subjects like.

3.5.2 Pre-test

After subjects listened all utterances in the training section, they can start the pre-test section. This section means to check subjects' understanding of valence and activation dimensions. The procedure of pre-test is same to the main-test. The only differences between the pre-test and the main-test are the numbers of utterances and whether there is accuracy and correlation presented after finished all the evalution. In the training section, the utterances covered all the 7-point-scale of valence and activation dimensions are presented to subjects. Subjects need to listened to the utterances and then selected one value from the seven points by their own feeling. After subjects finished evaluation of all utterances in the pre-test, there are two scores presented on screen. One score is the accuracy, which calculated by the equation:

$$Accuracy = num_s/num * 100\% \tag{3.1}$$

where num_s are the numbers of utterance which subjects give the same value as we expected and num are the number of utterance in this section. Another score presented on screen are the correlation between the subject's evaluation and the values of utterance we expected. These two scores are only the reference for me to check the understanding of subjects, and will not influence the results of main-test.

3.5.3 Agreement

If subjects could not get a high scores in pre-test, they would asked to back to the training section and then do the pre-test one more time. Finally, all subjects get over 45% accuracy and over 0.80 correlations by repeat these two section in at most 3 times.

3.6 Main-test

The aim of the main-test is to extract the data for the comparision of commonalities and differences among multiple languages on emotion perception. In order to achieve the data, thirty subjects are asked to evaluated 556 utterances in valence session and activation session, respectively. In the main-test, Every utteran can be repeated when subjects make their choices. Once the dicision be made, it cannot be modified and back to the previous utterance. The average of the subjects' rating for each utterances on valence and activation was calculated.

3.6.1 Agreement among subjects

In order to reduce the influence of the misunderstanding by subjects, the correlations among subjects in one native language group are calculated. As consequence, the correlations among the ten subjects in one native language group of activation dimension are all higher then 0.6 for the three native language groups. However, for valence dimension, the correlations among Chinese subjects are all higher than 0.6, and the correlations among Japanese subjects and Vietnamese subjects are not all higher than 0.6. There is one Japanese and two Vietnamese have very low correlations with other subjects in their groups. It is suggest that, these three subjects maybe do not understand the experiment and the theory of Valence-Activation approach. Therefore, in order to reduce the influence of the misunderstanding by subjects, when we analyse the results of the experiment, the evaluation of these three subjects are not be used.

Moreover, after calculated the correlations among subjects in one native language group, we calculate the deviation among the ten subjects in one native language group for all utterances. For one utterance, there are four deviations achieved. The three are for the three native language groups and the other is for all subjects. If the deviation values of two native language groups are higher than 1.5 and the deviation for all subjects is also higher then 1.5, the evaluatin results of this utterances are considered as not good results. These results are not used when we analyse the results of the experiment.

3.7 Results of the listening test

To compare the results of listening test on Valence-Activation approach among multiple languages, subjects were required to evaluated valence and activation values of each utterance, and position of each utterance on Valence-Activation approach is investigated. Figure 3.1 to Figure 3.5 show the position of each utterance on Valence-Activation approach for the five databases. The figures from left to right are the results evaluated by Chinese subjects, Japanese subjects and Vietanmese subjects, respectively. The marke in blue circle, purple square, red triangle and black hexagonal star represent the four basic emotions used in the experiment, neutral, happiess, anger and sad, respectively. Each mark represents one utterance, plotted by mean value of valence from all subjects on the the horizontal axis and mean value of activation on the vertical axis, and the evaluation of different utterances might overlap each each by the reason of same evaluation results.

In order to show the mean values and the standard deviations of valence dimension and activation dimension on Valence-Activation approach, every emotional states are presented as ellipse distributed on the space. Figure 3.6 shows the concept of the ellipse of emotional states. Coordinate (x_E, y_E) presents the center of the ellipse, in which x_E and y_E are the averages of valence and activation of the emotional states (E). Moreover, the standard deviations of valence and activation are presented by the horizontal and vertical radii of the ellipse.

Figures 3.7 (a)-(e) show the positions of the four emotional states for the five databases, Chinese, Japanese, Vietnamese, American English and German, on Valence-Activation approach, respectively. For each database, the marks circle, square, triangle and hexagonal star represent the four emotions neutral, happiness, anger and sadness, respectively. Then the color red, blue and green represente the evaluation results of Chinese subjects, Japanese subjects and Vietnamese subjects, respectively.

3.7.1 Position of neutral states

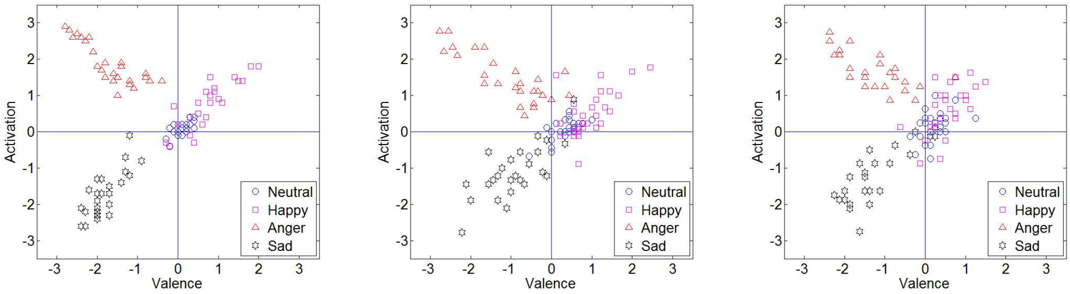
The first thing that should be investigated is that whether there are commonalities and differences or not, when subjects from different native languages perceive neutral voices. In order to discuss the starting points of emotion vectors on Valece-Activation approach, we extract the ellipse of neutral states for each databases. Figure 3.8 (a)-(e) show the position of neutral states on Valecen-Activation approach. For each database, colors red, blue and green represent the evaluated results of three native language groups, Chinese, Japanese and Vietnamese, respectively. From these figures, we can find out that the starting points of emotion vectors are not significantly different among three listener groups. It means that the position of neutral voice are not significantly different among three listener groups for all databases.

3.7.2 Direction of emotional states

When human beings perceive emotional states, are the directions from neutral state to other emotional states similar or not among different native language groups? To answer the question, the angle of emotion vectors which from the center of neutral states to the center of other emotional states are needed to be calculated. Figure 3.9 shows what is the angle of emotion vectors. In the figure, the emotion vector is from the center of neutral state (x_N, y_N) to the center of emotional state E (x_E, y_E).The angle is calculated by the following equation:

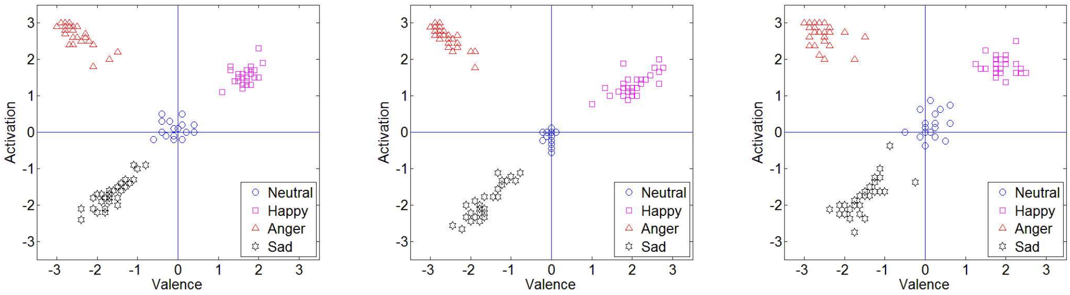
$$angle = arctan(\frac{y_E - y_N}{x_E - x_N}), \tag{3.2}$$

where (x_E, y_E) is the center of the emotional state (E), and (x_N, y_N) is the center of the neutral state.



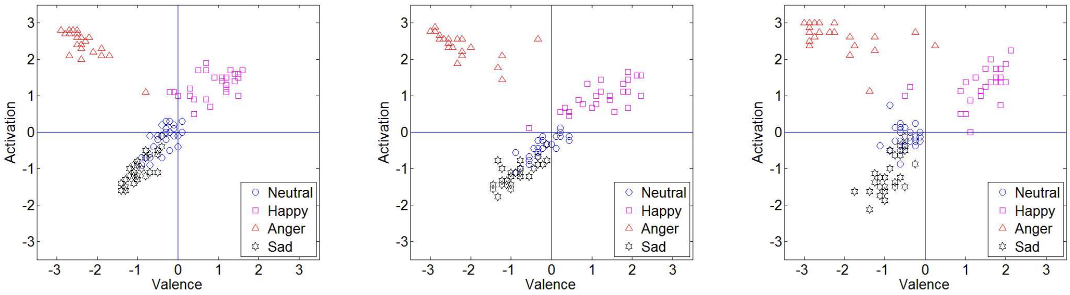
(a) Chinese subjects (b) Japanese subjects (c) Vietnamese subjects

Figure 3.1: Scatter figure of Chinese database.



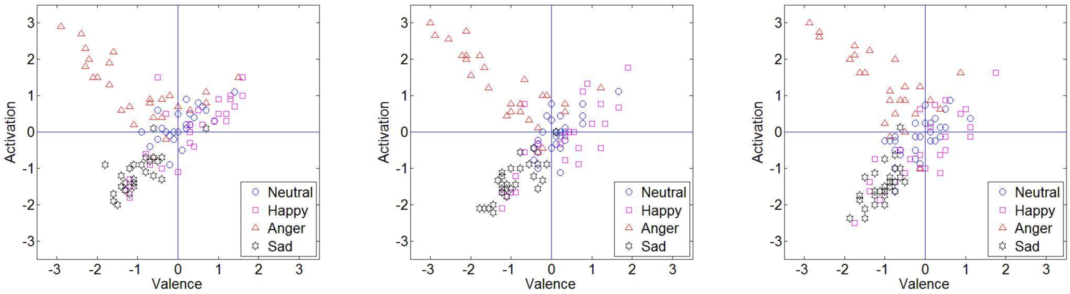
(a) Chinese subjects (b) Japanese subjects (c) Vietnamese subjects

Figure 3.2: Scatter figure of Japanese database.



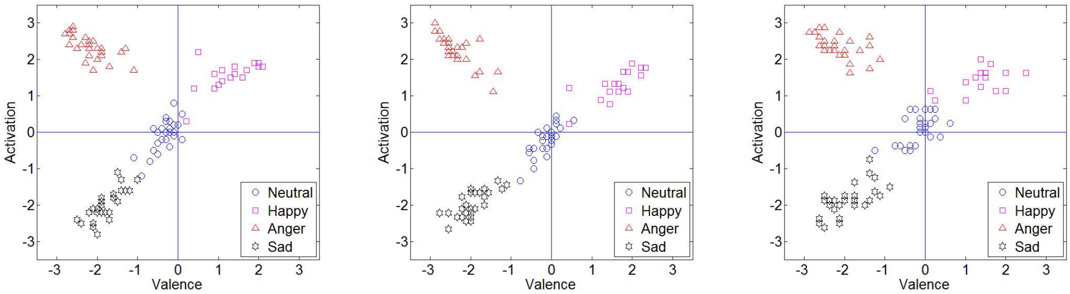
(a) Chinese subjects (b) Japanese subjects (c) Vietnamese subjects

Figure 3.3: Scatter figure of Vietnamese database.



(a) Chinese subjects (b) Japanese subjects (c) Vietnamese subjects

Figure 3.4: Scatter figure of American English database.



(a) Chinese subjects (b) Japanese subjects (c) Vietnamese subjects

Figure 3.5: Scatter figure of German database.

The calculated results are listed in Table 3.5. In the table, most of the differences between two different native language groups are smaller than 10 degree, and only the results of happy voice in English database are larger than 20 degree. It reveals that the angle of emotion vector are similar among three different native language groups. It means that the direction of emotion states on Valence-Activation approach are similar among three differen native language groups for the five databases.

3.7.3 Degree of emotional states

The third thing that to be investigated is that whether there are any commonalities and differences when human beings perceive the degree of one emotional states. To calculate degrees of emotional states, the length of emotion vectors are used as a metric. The length of the emotion vectors is equal to the distance from neutral to other emotional states on Valence-Activation approach. Figure 3.10 shows the concept of the degree of one emotional states. In another word, it shows that how long the distances are from the center of neutral state to the center of other emotional states. Large distance means strong emotional states and vice-versa [29]. The Euclidean-distances between neutral to other emotional states on Valence-Activation approach are calculated by the following

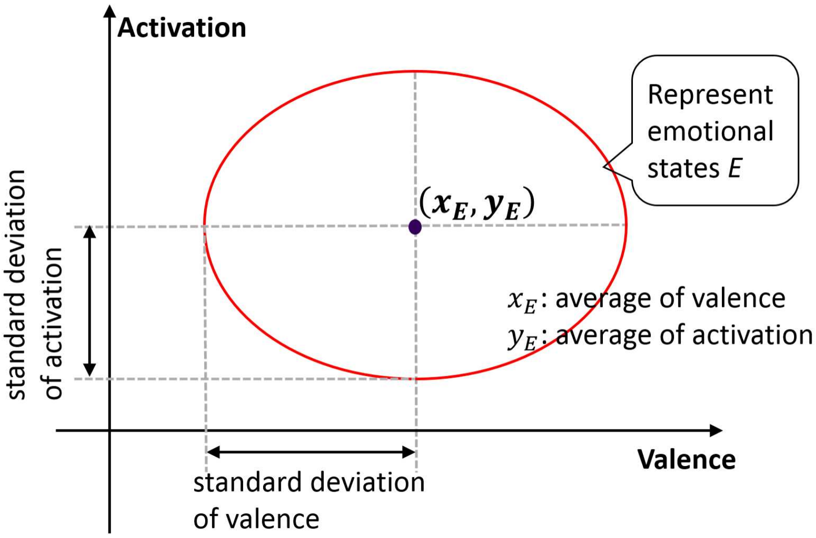

Activation
Represent emotional states E
(x_E, y_E)
standard deviation of activation
x_E: average of valence
y_E: average of activation
Valence
standard deviation of valence

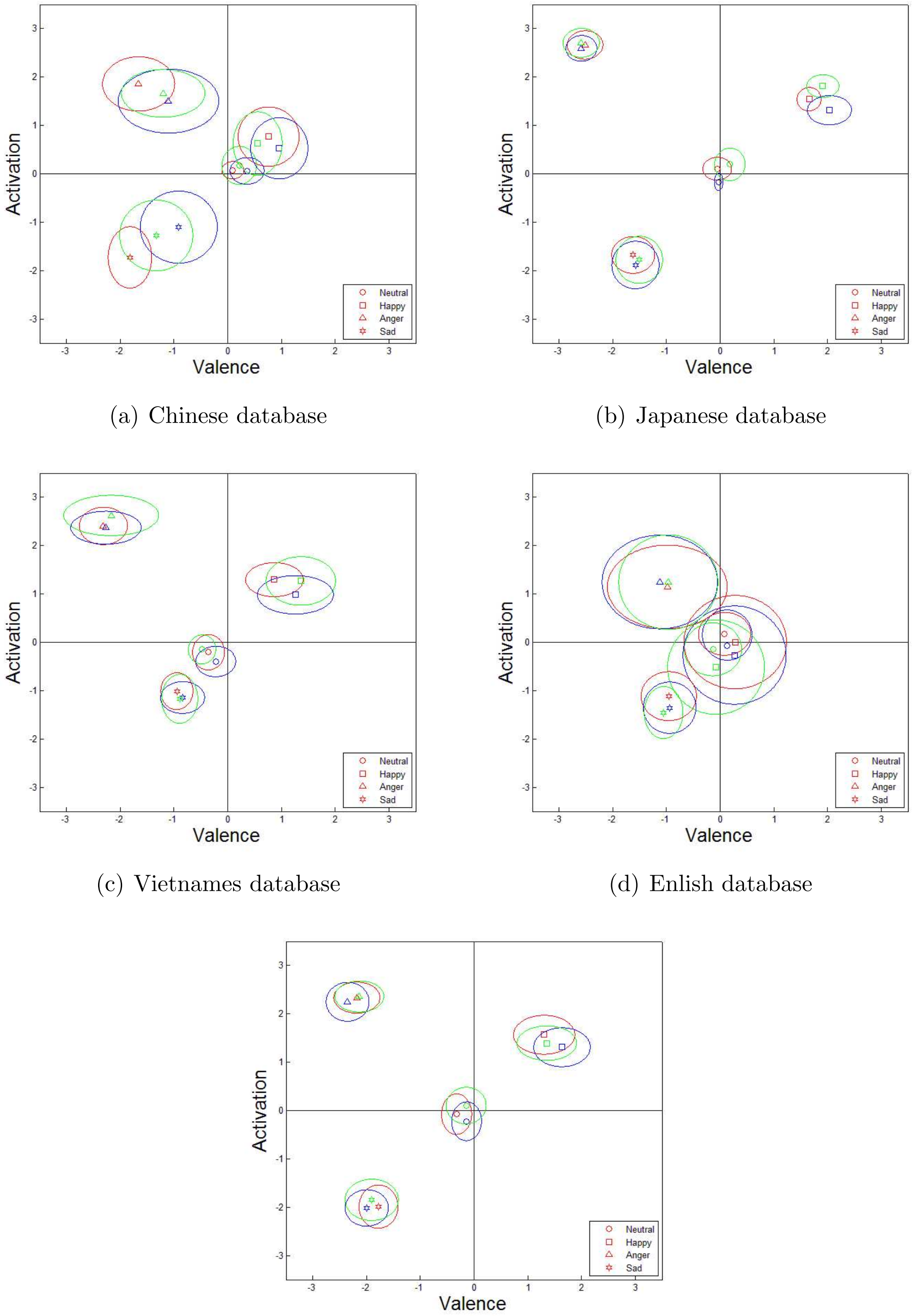
Figure 3.6: The ellipse which used to represente emotional states.

equation:

$$d(E,N) = \sqrt{(x_E - x_N)^2 + (y_E - y_N)^2} \tag{3.3}$$

where $(x_E,\, y_E)$ is the center of the emotional state (E), and $(x_N,\, y_N)$ is the center of the neutral state.

Table 3.6 listed the results. The table shows that the length of emotion vectors of native speakers are almost the longest. It means that the response of native speakers are almost the strongest. It is found that the degree of emotional states are significantly different three differen native language groups.



(a) Chinese database

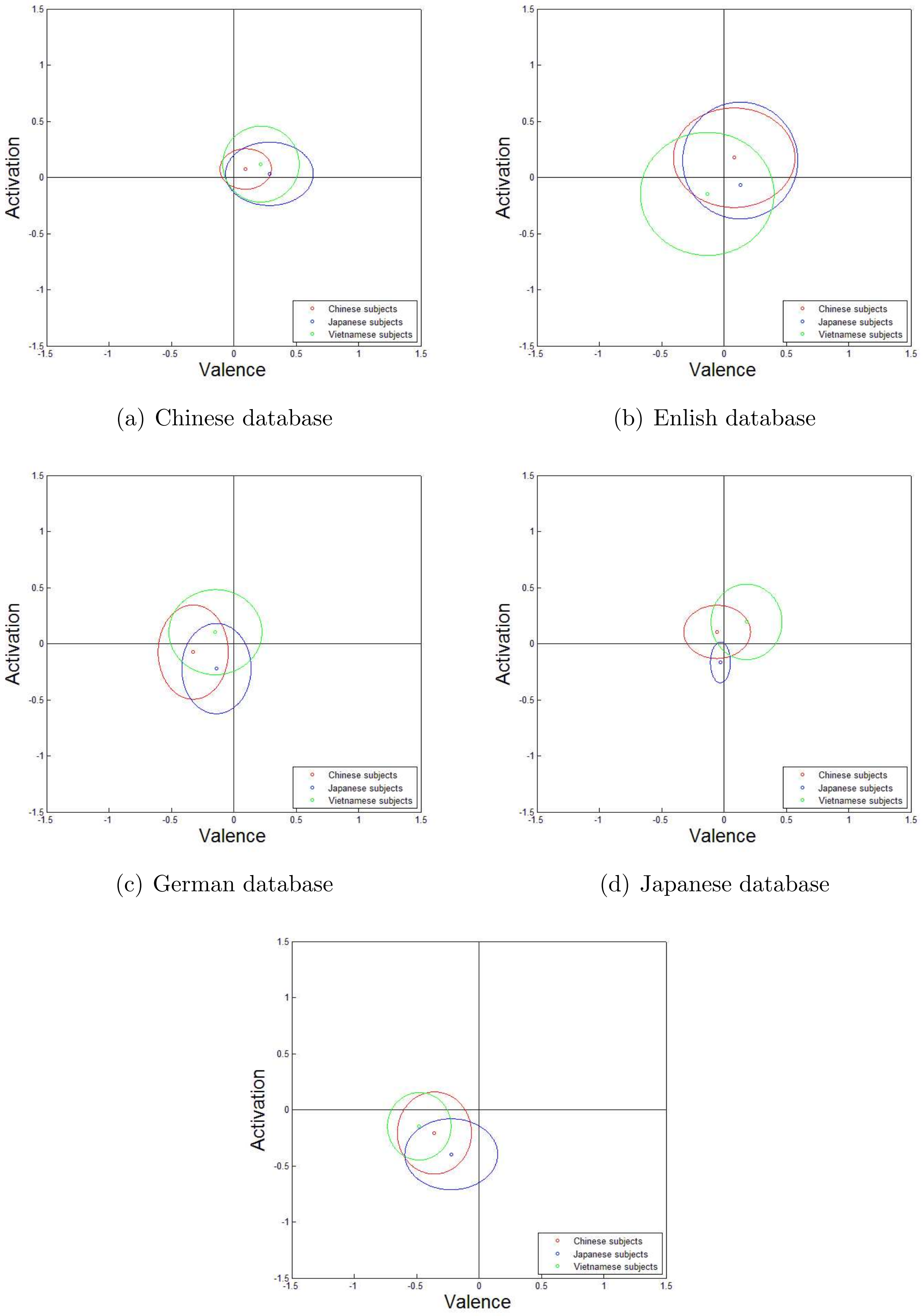
(b) Japanese database

(c) Vietnames database

(d) Enlish database

(e) German database

Figure 3.7: Emotional states' position on Valence-Activation approach.



(a) Chinese database

(b) Enlish database

(c) German database

(d) Japanese database

(e) Vietnames database

Figure 3.8: Position of neutral states on Valence-Activation approach.

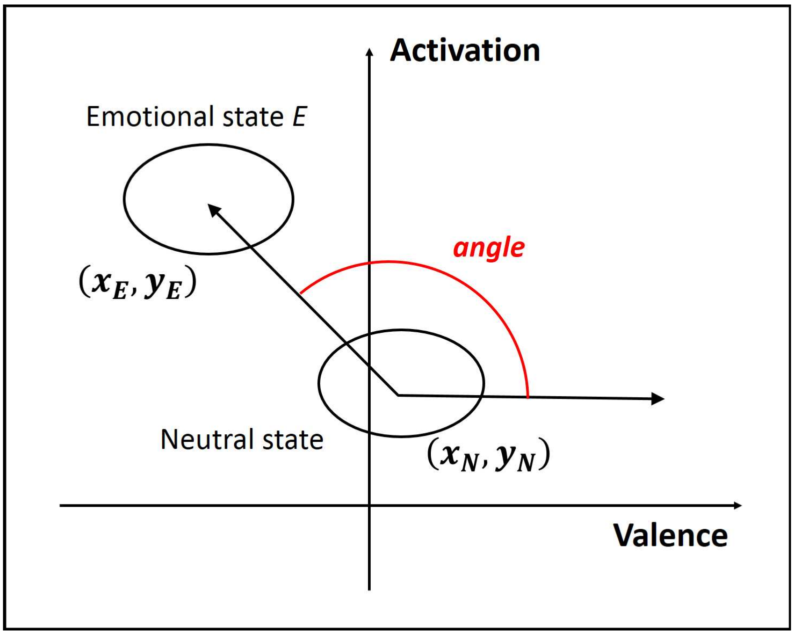


Figure 3.9: Example of direction of emotional states

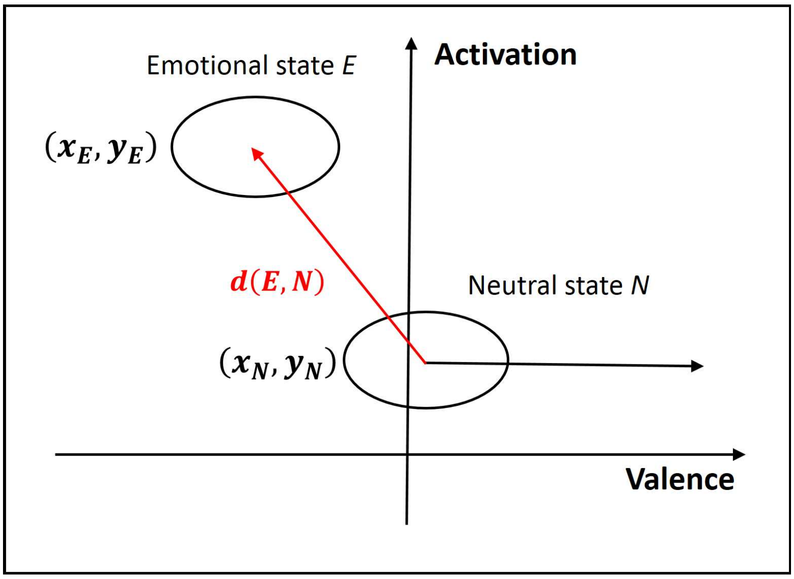


Figure 3.10: Example of distance between neutral states to other emotional states

Table 3.5: Angles of the vector from neutral state to other emotional states.

(a) Chinese Database

Subject	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	46.2°	41.9°	48.9°
Neutral-Angry	136.0°	138.6°	141.6°
Neutral-Sad	222.2°	220.5°	225.2°

(b) Japanese Database

Subject	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	40.0°	35.6°	44.3°
Neutral-Angry	136.1°	137.6°	138.1°
Neutral-Sad	228.7°	228.2°	227.4°

(c) Vietnamese Database

Subject	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	53.5°	46.1°	41.7°
Neutral-Angry	145.9°	149.3°	154.6°
Neutral-Sad	233.8°	229.5°	244.3°

(d) English Database

Subject	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	40.7°	49.4°	59.3°
Neutral-Angry	134.3°	141.1°	155.9°
Neutral-Sad	231.1°	231.9°	234.4°

(e) German Database

Subject	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	50.2°	48.4°	54.4°
Neutral-Angry	142.5°	142.4°	140.3°
Neutral-Sad	232.8°	225.8°	227.5°

Table 3.6: Distances between neutral state and other emotional states.

(a) Chinese Database

Distance	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	0.96	0.70	0.52
Neutral-Angry	2.52	1.96	1.92
Neutral-Sad	2.57	1.71	1.97

(b) Japanese Database

Distance	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	2.24	2.45	2.32
Neutral-Angry	3.55	3.64	3.38
Neutral-Sad	2.37	2.27	2.35

(c) Vietnamese Database

Distance	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	1.88	1.93	2.21
Neutral-Angry	3.15	3.15	3.12
Neutral-Sad	1.00	0.91	0.96

(d) English Database

Distance	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	0.26	0.23	0.37
Neutral-Angry	1.43	1.72	1.53
Neutral-Sad	1.66	1.53	1.19

(e) German Database

Distance	Chinese subjects	Japanese subjects	Vietnamese subjects
Neutral-Happy	2.24	2.25	1.79
Neutral-Angry	3.03	3.08	2.78
Neutral-Sad	2.40	2.49	2.54

Chapter 4

Discussion

The purpose of these research is to investigate the commonalities and differences among multiple languages on human perception of emotional states. In order to compare the commonalities and differences among multiple languages on Valence-Activation approach, an listening test which invite thirty subjects in three different native languages to evaluate 556 utterances in five databases on Valence-Activation approach. In the previous section, the results of the listening test was analysed in three points of views: the position of neutral voice, the direction of emotinal states and the degree of emotional states. In this section, the comparison of commonalities and differences among multiple languages evaluated by different native language groups are discussed from these three points of views in detail. Then a section of general discussion are presented to discuss the results of the listening test in general view of point.

4.1 Position of neutral states

To investigate the commonalites and differences among multiple languages evaluated by different native language groups, the first thing should be investigated is whether the positions of neutral voices are same or not among the evaluation by different native language groups.

The results of listening test about the position of neutral voice are shown in Figure 3.8 (a)-(e). The figures show that the diveation between the evaluation of two different native language groups are from 0.135 to 0.388. It suggests that the positions of neutral voice are not significantly different among the evaluations of three different native language groups for all databases. Moreover, the positions of neutral voice evaluated by the three listener groups are not located at the center of the Valence-Activation approach. Considering the reason why the positions of neutral voice are not located at the center of the approach occurred in all of the five databases, there are three hypotheses, which have high possibility, need to be discussed in detail.

1. The first hypothesis is the positions of neutral states in these languages are not at the center of approach. In another word, the characteristics of these languages might influent the positions of neutral voice evaluated by the three listener groups, and

lead to the shift between the positions of neutral voice and the center of the Valence-Activation approach. For Chinese, Japanese and Vietnamese databases, since the responces of native speakers, Chinese subjects, Japanese subject and Vietnamese subjects, are not at the center of the approach either, the probablity that the shift between the positions of neutral state and the center of the approach caused by the languages is very low in these three databases. On the other hand, because there are no native speakers of English and German, we cannot know whether the hypothesis influent the results or not.

2. The second hypothesis is that the personality of the speaker influent the results. There are a lot of nonlinguistic information included in speech signal inluent the position of neutral states on Valence-Activation approach, the personality of speaker is one the these nonlinguistic information. When human beings are speaking, every speakers have different personalities. In this research, the differences of personalities might influent the position of neutral states. Since for Chinese, English and German databases, the number of speakers in one database are more than four, it reduce the influence of speaker's personality. But for Japanese and Vietnamese databases, there are only 1 or 2 speakers. Thus, in these two databases, the influence of the speakers' personality should be stronger than it in Chinese, English and German databases.

3. The third hypothesis is that when speakers recording neutral voices, they spoke with other emotional states. When speakers recording the databases, there might have other emotions included in their voice and the emotions of one speaker might affect other speakers unconsciously. As a consequence, all speakers might speak with some other emotions when they recording the utterances of neutral voice. Since this hypothesis is every common when human beings communicate with others, and it cannot be controlled by speakers, this hypothesis would give strong influence to the results of positions of neutral voice in all of these five databases.

4.2 Direction of emotional states

After discussed the positions of neutral voices, the commonalities and differences on other emotional states are need to be discussed next. Thus, the next two discussions focus on the position of other emoitonal states and the relationship between neutral state and other emotional states.

In order to discuss the position of other emotional states, the direction of the emotonal states is very important. The direction of one emotional state is presented by the angle of the vector from the center of neutral state to the center of other emotional state. As Table 3.5 shows, the differeces between the angles of two different native language groups in Chinese, Japanese and German database are all smaller than 10 degree. It suggests that the angles from neutral state to other emotional states evaluated by the three different native languages groups are similar with each other. On the other hand, though the differences between the angles of two different native language groups in Vietmanese and

English databases are not so small as them in Chinese, Japanese and German databases, they are still not so large. Only for the happy voice in English database, the difference is 26.4 degree (larger than 20 degree). From Table 3.5 and Figure 3.7, we can understand that the directions of emotinal states on Valence-Activation approach are almost similar among the three different native languages groups. This result supposes human beings can perceive emotional state using these angles no matter they understand the languages or not.

4.3 Degree of emotional states

When discuss the position of other emotional states and the relationship between neutral state and other emotional state, the discussion of the degree of emotional state is very necessary. In human beings' real-life, emotions are not only several discrete categories. In the real-life, the emotions are continuous and have different degrees of intensity. To investigate the commonalities and difference on human perception of emotional states, the discussions of the degree of emotional state is required. From Figure 3.7 , in Chinese database, the responces of Chinese subjects are stronger than the responces of Japanese and Vietanmese subjects. Similarly with Chinese database, the responces of Japanese subjects are the stronggest in Japanese database, and the responces of Vietnamese subjects are the stronggest in Vietnamese database. Moreover, Table 3.6 shows the calculated results of the distance between neutral states and other emotional states. For Chinese, Japanese and Vietnamese database, the distances evaluated by native speakers of one language are almost the largest one in the three different native language groups. It suggests that native speaker always gives stronger responce than other people. On the other hand, for English and German database, since there are no native speakers of these two languages in the three different native language groups , it is difficult to distinguish which listener group gave the stronggest response.

4.4 General discussion

The purpose of this research is to investigate the commonalities and differences among multiple languages on human perception of emotinal states evaluated by different native language groups on Valence-Activation approach. In order to achieve the purpose of this research, subjects from three different native language groups are asked to evaluated the values of valence and activation in five different languages. From the results of the listening test, we understand that the commonalities are the evaluation resulsts by different native language groups among multiple languages have similar position of neutral voice on Valence-Activation approach and similar direction of emotional states. On the other hand, the difference among multiple languages evaluated by different native language grouops are they have different evaluation of the degree of emotional states. Moreover, the commonalities and differences we found out in this research can help us to understand how human beings perceive emotional states among multiple languages. From the similar position of neutral states we can know that when human beings perceive the emotional states in speech, different listeners would have same feeling when they perceive neutral voice, and from the similar direction and different degree of emotional states we can know that when human beings perceive emotional states among multiple languages, different native language groups have different feeling when they percept other emotinal states. According to these two results, we can conclude that human beings can perceive emotional states regardless of language. The results of the research also reveals that, only by using the same position of neutral and controling the degree of each emotional states, researchers can construct a speech emotional recognition system and synthesis expressive voices independent one languages.

Chapter 5

Conclusion

In the previous study, we discussed the results of the listening test in three views of point on Valence-Activation approach and in general viewpoint. In this chapter, we will summarize the this reseach we have done and then list the contribution according to the results of this research. Finally, we will expound what are need to be done in the future.

5.1 Summary

In this research, we attempted to investigate the commonalites and differences among multiple languages on Valence-Activation approach for human perception of emotional states in speech. In order to find the commonalies and the differences, an experiment is carried out. We invite thirty subjects from China, Japan and Vietnamese, and divide them into three different native language groups to partacipate in the experiment. In the listening test, 556 utterance covered four basic emotional states, neutral, happiness, anger and sadness, selected from five languages, Chinese, Japanese, Vietnamese, American English and German. The, we compared the results of experiment in three views of points,

- the position of neutral state
- the direction of emotional states
- the degree of emotional statesand

among the evaluation by different native language group on Valence-Activation approach for all databases. According to the analysis, we achieved that the commonalities are that the evaluation by different native language groups have same position of neutral state and same direction from neutral states to other emotional state, and the difference is the degree of emotional state different among native language groups. Moreover, the results suggest that when human beings perceive emotional states among multiple language, although they different feeling in other emotional states, they have same feeling in neutral voice and can perceive emotional states regardless of languages.

5.2 Contribution

According to the results of this research, the comparison among different native language groups on Valence-Activation approach in multiple languages can help us:

- to understand how human beings perceive emotional states among multiple languages in speech, which is one of the most part of human perception.
- to build a human perception system, in order to simulate the way of human perception.
- to construct a grobal speech emotion recognition system, which can recognize emotional state from speech automatically and the recognition can regardless of the input languages.
- to improve the traditional synthesis system, in order to synthesize the expressive voices which are more similar with human beings' voice.

5.3 Future work

In the future, my focus is on comparison of the commonalities and differences not only on Valence-Activation approach, but also on the acoustic features. The commonalities and differences among multiple languages on acoustic feature is the basis of speech in acoustic domain and important. In order to provided fundamental knowledge of human perception to construct a grobal speech emotion recognition system and to improve the traditioanl synthesis system, the commonalities and differences for acoustic features among multiple language is necessary.

Bibliography

[1] H. Fujisaki, “Prosody, information, and modeling - with Emphasis on Tonal Features of Speech - ”, *Proc. Speech Prosody 2004 Nara*, pp. 1-10(2004).

[2] H. R. Pfitzinger “Cross-language Perception of Hebrew and German Authentic Emotional Speech”, ICPhS, Hong Kong, 17-21 August 2001.

[3] A. Abelin, “Cross-Cultural Multimodal Interpretation of Emotional Expressions - An Experimental Study of Spanish and Swedish”, *Proc. Speech Prosody 2004 Nara*, pp. 23-26(2004).

[4] X. Zuo, T. Li and P. Fung, “A Multilingual Database of Natural Stress Emotion”

[5] C. F. Huang, D. Erickson and M. Akagi, “A Study on Expressive Epeech and Perception of Semantic Primitives: Comparison between Taiwanese and Japanese”, *IEICE Technical Report*, SP2007-32.

[6] C. F. Huang, D. Erickson and M. Akagi, “Comparison of Japanese expressive speech perception by Japanese and Taiwanese listeners”, *Acoustics2008 Paris*, Paris, pp. 2317–2322, 2008.

[7] T. Polzehl, A. Schmitt and F. Metze, “Approaching Multi-Lingual Emotion Recognition from Speech - On Language Dependency of Acousti Prosodic Features for Anger Recognition”, *Speech-Prosody*, Chicago, USA (2010).

[8] A. Abelin and J. Allwood, “Cross Linguistic Interpretation of Emotional Prosody” *Speech and Emotion*, Newcastle, Northern Ireland, UK, September 5-7, 2000.

[9] R. Elbarougy and M. Akagi, “Cross-lingual Speech Emotion Recognition System Based on a Three-Layer Model for Human Perception”, *Proceeding of International Conference (APSIPA2013 ASC)*, Kaohsiung, Taiwan, November 2013.

[10] R. Elbarougy and M. Akagi, “Speech Emotion Recognition System Based on a Dimensional Approach Using a Three-layered Model”, *Proceeding of International Conference (APSIPA2013 ASC)*, Hollywood, USA, December 2012.

[11] C. F. Huang and M. Akagi, “A three-layered model for expressive speech perception”, *Speech Communication*, 2008, 50 (10).

[12] J. W. Dang, A. J. Li, D. Erickson, A. Suemitsu, M. Akagi, K. Sakuraba, N. Minematsu and K. Hirose, "Comparison of Emotion Perception among Different Cultures", *Acoust. Sci. & Tech.*, **31**, 6, 2010.

[13] K. sawamura, J. Dang, M. Akagi, D. Erikson, A. Li, K. Sakuraba, N. Minematsu and K. Hirose, "Common factors in emotion perception among different cultures", *Proc. ICPhS 2007*, 2113-2116 (2007).

[14] C. Yu, P. M. Aoki and A. Woodruff, "Detecting user engagement in everyday conversation", *arXiv preprint*, cs/0410027, 2004.

[15] H. Schlosberg, "Three dimensions of emotion", *Psychol. Rev.*, 61, 81-88 (1954).

[16] M. Grimm and K. Kroschel, "Emotion Estimation in Speech Using a 3D Emotion Space Concept", in *Robust Speech Recognition and Understanding*, M. Grimm and K. Kroschel, Eds. (I-Tech Education and Publishing, Vienna, 2007), Chap.16.

[17] R. Cowie, *et al.* E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. Taylor, "Emotion recognition in human-computer interaction", *IEEE Signal Process. Mag.*, 18, 32-80, 2001.

[18] R. Cowie and R. Cornelius, "Describing the emotional states that are expressed in speech", *Speech Commun.*, **40**, 5-32, 2003.

[19] T. Vogt, E. Andre and J. Wagner, "Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realization", *Affect and Emotion in HCI*, C. Peter and R. Beale, Eds. (Springer, Berling/Heidelberg, 2008), pp. 75-91.

[20] M. Grimm, K. Kroschel, E. Mower and S. Narayanan, "Primitives-based evaluation and estimation of emotions in speech", *Speech Commun.*, 49, 787-800 (2007).

[21] M. Schroder, "Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis", *Doct. thesis, Phonus 7. Rep. Inst. Phonet.*, Saarland Univ. (2004).

[22] M. Schroder, "Dimensional emotion representation as a basis for speech synthesis with non-extreme emotion", in *ADS 2004*, E. Andre *et al.*, Eds. (Springer, Berlin/Heidelberg, 2004), pp.209-220.

[23] J. A. Russell and A. Mehrabian "Evidence for a Three-Factor Theory of Emotions", *Journal of research in personality*, 11, 273-294 (1977).

[24] T. L. Gehm and K. R. Scherer, "Factors Determining the Dimensions of Subjective Emotional Space", 1988.

[25] T. D. Ngo and M. Akagi, "Toward a Rule-Based Synthesis of Vietnamese Emotional Speech", *Knowledge and Systems Engineering*, Springer International Publishing, 2015. 129-143.

[26] C. Busso, M. Bulut, C. C. Lee *et al.*, *Language resources and evaluation*, 42. 4 (2008): 335-359.

[27] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, “A Database of German Emotional Speech”, *Proceedings of Interspeech*, Lissabon, Portugal, 2005.

[28] H. Mori, T. Satake, M. Nakamura and H. Kasuya, “Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics”, *Speech Communication*, **53**, 36-50, 2011.

[29] R. Plutchik, “Emotions and Life: Perspectives from psychology, biology, and evolution”, *American Psychology Association*, Washington, 2002.

[30] M. Akagi, “Analysis of production and perception characteristics of non-linguistic information in speech and its application to inter-language comunications”, *Proc. APSIPA 2009*, Sapporo, pp. 513-519 (2009).

List of Publications

[1] X. Han, R. Elbarougy, M. Akagi and J. F. Li, “Comparison of perceived results of emotional states on Valence-Activation space among multiple-languages”, *IEICE Technical Report*, SP2014-55, WIT2014-10 (2014-6).

[2] X. Han, R. Elbarougy, M. Akagi and J. F. Li, “Study on perceived emotional states in multiple languages on Valence-Activation space”, *Acoustic Society of Japan, 2014 autumn conference*, Sapporo, Japan, 2014.

[3] X. Han, R. Elbarougy, M. Akagi, J. F. Li, T. D. Ngo, and T. D. Bui, “A study on perception of emotional states in multiple languages on Valence-Activation approach”, *NCSP 2015*, Kuala Lumpur, Malaysia, 2015.