

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

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

Xiao Han (1210205)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 12, 2015

Keywords: human perception, multiple languages, native language group, emotional state, dimensional approach.

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. 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 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. In order to make communication more global, and do not be influenced by language, nationality and culture, investigation of the fundamental knowledge on how human beings perceive emotional 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 synthesis expressive speech. 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. 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. Therefore, The purpose of this research is to investigate commonalities and differences among multiple languages of human perception for emotional states from speech signal.

In previous studies, 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, happiness and sadness, which are the most intense form of emotion, and all other emotions are considered as the variations or combinations of them. By using the categorical approach, the category of emotion can be clearly presented on this approach. On the other hand, dimensional approach represents emotional states using continuous multi-dimensional space. This approach can be used to present the degree of intensity in emotional states in the real-life. Both of the two emotion approaches provide complementary information about the emotional expressions observed in individual. 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 is selected to represent the emotional states because of its outstanding representation ability of the degree of intensity in emotions. Considering the emotional states included in the databases we used are only four basic emotions, neutral, happiness, anger and sadness, a two-dimensional approach called Valence-Activation approach is 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, Vietnamese, 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.

These comparisons are carried out among different native language groups on three views of point. They are the position of the neutral voice, the direction of emotional states and the degree of emotional states. Firstly, the position of neutral voice are presented as ellipse distributed on the space. In the ellipse, the center of the ellipse are the averages of valence and activation of the emotional states and the standard deviations of valence and activation are presented by the horizontal and vertical radii of the ellipse. It is found that the position of neutral state on Valence-Activation approach for three different native language groups are not significantly different. Secondly, the direction of one emotional state is represented by angle of the vector from the center of neutral state to the center of the emotional state. The results reveal that directions of emotional states on Valence-Activation approach are similar among three different native language groups. Thirdly, the degrees of emotional states presented by the distances from the center of neutral state to the center of other emotional states on Valence-Activation approach. It is found that the distance are significantly different among three different native language groups.

From the results of the listening test, we understand that the commonalities are the evaluation results 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 groups 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 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 percept other emotional states. It also reveals that, only by using the same position

of neutral and controlling the degree of each emotional states, researchers can construct a speech emotional recognition system and synthesis expressive voices independent one languages.

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 global speech emotion recognition system, which can recognize emotional state from speech automatically and the recognition can regardless of the input languages and to improve the traditional synthesis system, in order to synthesize the expressive voices which are more similar with human beings' voice.

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 global speech emotion recognition system and to improve the traditional synthesis system, the commonalities and differences for acoustic features among multiple language is necessary.