PREGORESS REPORT FOR THREE-LAYER PERCEPTUAL MODEL OF EXPRESSIVE SPEECH PROJECT

Chun-Fang Huang

The aim of research

In speech communication, emotion plays an essential role. Many studies concerning with the vocal expression of emotion dealt with the relationship between expressive speech and acoustic features. They measured acoustic features in speech signal that reflect the emotional state of speakers. However, doing in this way still can not provide a complete and appropriate solution for explaining how human perceive emotion from speech, and further, synthesizing natural emotional speech. The goal of the research project is to develop a perceptual model that explains how human perceive emotion from speech from engineering, psychological and physical points of view.

The approach and idea

In order to achieve the goal, the project is divided into 4 tasks:

- ***The first task is to conceive a perceptual model.***

  We considered that emotional speech is not directly related to acoustic features. In fact, using a computer to deal with emotional speech should involve three fields of knowledge - engineering, psychology and physiology. Integrating these three fields, the approach proposed in the research is a three-layer model.

  The three-layer model consists of expressive speech layer, semantic primitive layer, and acoustic feature layer where the expressive speech layer is various categories of emotional speech. In this study, it includes 6 categories of emotional speech, natural, joy, cold anger, sadness, and hot anger. The semantic primitives layer is defined as a set of descriptions or adjectives that are used to describe voice quality. The acoustic feature layer is a set of physical values of speech signals.

- ***The second task is to build the perceptual model by a top-down approach.***

  The perceptual model was constructed by 4 steps:

  Step 1: To investigate what semantic primitives should be used

  Step 2: To build the relationship between the expressive speech layer and the semantic primitive layer

  Step 3: To analyze acoustic features

  Step 4: To build the relationship between the semantic primitive layer and the acoustic feature layer

- ***The third task is to verify the perceptual model by a bottom-up approach.***

  The bottom-up approach is to verify the perceptual model by resynthesis (morphing) and more perceptual experiments. Therefore, this task needs the following steps to achieve it.

  Step 1: According to the analyzing results of acoustic features, to establish prosody rules for resynthesis.

  Step 2: To develop a program, by which the original speech signal can be resynthesized following the prosody rules.

  Step 3: To verify the two relationships of the model by conducting perceptual experiments to examine resynthesized voice.

- ***The fourth task is the application of the perceptual model***

  The purpose of this step is to find the commonality/difference in expressive speech perception between people with different cultures/languages background. This task needs the following steps to achieve it.

  Step 1: Using different group of subjects than the second task by the same process of the second task to build the perceptual model.

  Step 2: Compare the model built in the second task and this task to find how people of different groups perceive expressive speech.

Progress of 2007

In 2007, we conducted the fourth task. We use the same Japanese utterances but with different groups of subjects, Taiwanese and Japanese. We found that even without the understanding of Japanese, Taiwanese clearly identified the intended expressive speech categories. We also found that these two groups of people tend to use the same set of primary semantic primitives to describe expressive voices but different in secondary ones. We expect to apply the model to create universal-applicable expressive voices synthesizer.

Publication in 2005

Huang, C-F, Erickson, D., and Agaki, M. (2007). A study of expressive speech and perception of semantic primitives: Comparison between Taiwanese and Japanese. Technical Report, IEICE-CE (July2007), SP2007-32(2007-7), pp. 49-54

Huang, C-F., Erickson, D., Akagi, M. (2007). Perception of Japanese Expressive Speech: Comparison between Japanese and Taiwanese Listeners. ASJ fall meeting.