Title	スピーチにおける感情的な知覚の多層ファジィ論理的なモデルの構築
Author(s)	黄, 純芳
Citation	
Issue Date	2004-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1884
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

A MULTI-LAYER FUZZY LOGICAL MODEL FOR EMOTIONAL SPEECH PERCEPTION

Chun-Fang Huang (210204)

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

August 13, 2004

Keywords: Emotional Speech Perception, Fuzzy Logic, Multidimensional Scaling, Acoustic Features.

1 Introduction

This thesis proposes a multi-layered perceptual model, which attempts to simulate human ability to perceive emotions from speech. Regarding emotional speech, previous research mainly concerned the direct relationship between emotions and acoustic features. However, they all overlooked one important point. Human ability to percept emotions from speech was not directly based on acoustic features. Differing from their approach, the proposed perceptual model includes 3 layers, emotion layer, primitive feature layer, and acoustic feature layer, where primitive feature is defined as an adjective that is used to describe emotional speech. And 5 emotions of Neutral, Joy, Cold Anger, Sadness and Hot Anger were investigated. To accomplish the perceptual model, two approaches should be considered. The top-down approach is adopted to construct a framework for the model, and the bottom-up approach is adopted to verify the existing framework. This thesis focuses on the top-down approach, mainly. The purpose of this thesis is to construct the proposed multi-layered perceptual model by the top-down approach. Two relationships, the emotion and the primitive feature, the primitive feature and the acoustic feature, were investigated sequentially.

2 Relationship between the emotion and the primitive feature

Three perceptual experiments were conducted and fuzzy logic was applying to deal with the in order to build the relationship between the emotion and the primitive feature.

2.1 Three experiments

The first experiment examined the utterances in terms of the emotions. Subjects were asked to rate 171utterances in terms of the 5 emotions. The rating results showed how the 171 utterances were related to the 5 emotions. The second and third experiments, although with different purposes, find suitable primitive features collectively. The second experiment was to construct a psychological distance model for 15 utterances chosen according to the results of the first experiment. The method of paired comparison was implemented. The subjects were asked to rate each utterance pair according to how similar they perceived the pair. By applying multidimensional scaling techniques (MDS) to deal with the rating results, a reasonable psychological distance model was constructed. Before performing the third experiment, a pre-experiment was conducted and 34 adjectives were chosen from 60 adjectives. Then, in the third experiment, the subjects were asked to rate the 34 adjectives. According to the combination of the rating results and how each of the 34 adjectives were related to a certain emotion, depending on what direction an adjective was superimposed onto the psychological distance model, 15 adjectives were selected to be the members of the primitive features for the proposed perceptual model.

2.2 Building the relationship by Fuzzy logic

By applying fuzzy logic to deal with the results of the three experiments, the relationship between the emotions and the primitive features was built. For each emotion, a fuzzy inference system (FIS) was constructed by using MATLAB Fuzzy Logic Toolbox. The construction process could be concluded in 3 main steps. First an initial model was constructed. Membership functions and rules of the initial model were selected. Second, the initial model was trained by adaptive neuro-fuzzy technique. Third, the membership functions were readjusted according to the checking errors and the training errors. By plotting the output (i.e. emotion degree) against input data (i.e. primitive features) and calculating the regression line to fit the curve, the slope of the regression line indicated the relationship between the emotions and primitive features.

3 Relationship between the primitive feature and the acoustic feature

In order to find the relationship between the primitive feature and the acoustic feature for the perceptual model, the acoustic features were analyzed in terms of the three attributes of sound, pitch, loudness and timbre. Their correlated physical value F0, power, and power spectrum of utterances were calculated by STRAIGHT.

Regarding F0 and power, several acoustic features in terms of both the accentual phrase and the overall utterance were measured. According to the correlation coefficients between the measured acoustic features and the primitive features, those acoustic features that were most related to primitive features were selected. For F0, they were highest pitch (HP), average pitch (AP), mean value of rising slope (RS), and rising slope of the first accentual phrase (RS1st). For power, they were mean value of power range in accentual phrase (PRAP), power range (PWR), rising slope of the first accentual phrase (RS1st), the ratio between the mean value of power in high frequency, and the mean value of power through the whole frequency domain (RHT). Regarding power spectrum, although the analysis is not completely done yet since the time limit, part of the analytic results provided very useful clues to find the acoustic features in terms of timbre that affect the primitive features. Two values were measured from power spectrum. They were average power spectrum and variance in vowel power spectrum.

In this thesis, the relationships between the primitive features and the acoustic features in terms of pitch and loudness have been constructed. But, the relationship between the primitive feature and the acoustic feature in terms of timbre is not completely constructed yet.

4 Conclusion

For the perception of emotional speech, this thesis reports a three-layer perceptual model. In order to construct the model builds, the relationship between the emotion and the primitive feature was built by conducting 3 experiments and using fuzzy logic to deal with the experimental results. As the results of each FIS showed, FIS well modeled the relationship. Moreover, the relationship between the primitive feature and the acoustic feature was built partially.