A Quantitative Study of the Broadcasting Keynote Based on Physiological and Acoustic Features

Jie YANG

Laboratory of Language Sciences, Peking University, Beijing, China
Department of Chinese Language and Literature, Peking University, Beijing, China

In the spread process of broadcast television language, the broadcasting keynote is the intuitive basis for judging whether it is accurate or not. It refers to the total emotional coloring and weight mastered by the announcer and host when broadcasting a manuscript, which reflects the host's understanding, feeling, and expression of the work. At present, most of the research on the broadcasting keynote focuses on specific cases, and few scholars summarize the expression rules of the broadcasting keynote. When it comes to the broadcasting keynote of a specific production, it is often necessary to use more than one adjective to describe the keynote, which also makes their cognition of the keynote of the broadcast not unified enough, and then not accurate enough in the presentation of the sound form. Based on the method of phonetic experiment, this paper conducts a physiological acoustic study on the common broadcast keynote of different program types in the context of radio and television, constructs the speech emotion evaluation scale according to the distinctive features of speech emotion, and empirically describes the broadcast keynote from a new perspective.

The corpus for this study was derived from four professional announcers. The corpus includes the same text with different emotions and the radio/TV program clips of different types and themes, a total of 680 corpora. Perceptual experiments showed that the emotional speech classification criteria used in this paper had cognitive consensus. In this study, five kinds of signals, including speech signal, voice signal, chest breathing signal, abdominal breathing signal and video signal, were collected, and these signals constructed a relatively complete multimodal database of speech emotion. This study focuses on the analysis of speech signals and voice signals. In this study, the Speech Breath Rhythm Analysis Platform with the national software copyright of the Linguistics Laboratory of Peking University and Praat were used to extract the fundamental frequency, amplitude, speaking rate, articulation rate, the first three formants, open quotient, speed quotient.

We conducted the research on the distinctive features of speech emotion based on the corpus of the same text with different emotions. Through non-parametric test, we found that the vocal area characteristics, speech speed & articulation speed characteristics, resonance characteristics and voice characteristics of speech with different emotions were significantly different among different emotions, and there was at least one group of significant distinctive features between different emotions. Based on the mean values of different parameters, the distinctive features matrix is constructed, and the sound expression forms of different speech emotions are described. Finally, a set of distinctive features system for effectively describing and defining speech emotions is established.

On the basis of the research on the above-mentioned distinctive features, we used the mean of fundamental frequency, the standard deviation of fundamental frequency, the mean of open quotient, the standard deviation of open quotient, the mean value of speed quotient,

the standard deviation of speed quotient, the mean value of amplitude, the standard deviation of amplitude, the mean and standard deviation of the first three formants, the mean of speech speed and articulation speed, a total of 16 parameters to construct the speech emotion evaluation scale model. The model showed good prediction accuracy in both closed-set and open-set tests. Comparing parameters of the model, we found that the combined use of prosody parameters, voice parameters and resonance parameters had higher prediction accuracy than the model constructed by using a certain type of parameters alone. In addition, with the increase of the time window width of the test data, the prediction accuracy of the eight kinds of speech emotions gradually improved at first and then stabilized, and it was found that a window width of 10 seconds could make the prediction accuracy of the eight speech emotions stabilize. The description and discussion of the time-domain change characteristics reveal that the expression of different speech emotions is not invariant and changes in a regular manner. The results show that the scale model can effectively describe and define speech emotion in real time.

Based on the above research, the physiological and acoustic research were conducted on the keynotes of four different program types: news broadcast, documentary commentary, host of variety show and host of social life program, and the analysis data are input into the scale model of speech emotion evaluation. The results showed that: 1) different keynote may have multiple emotional components at the same time, which is reflected as "compound phonetic emotion"; 2) in the same program type, the common components of different keynote expression reflect the common characteristics of sound expression given by the program type, and after ignoring the common emotional components, the shape direction in the radar chart is very distinct, and the emotional component of the direction is the main emotional feature presented in the broadcasting keynote; 3) in terms of gender, for the keynote expression of different program types, male voices are mainly based on "normal", and attached with the corresponding emotional characteristics of the keynote, while female voice form mainly through a significant change to reflect the difference of different keynote. This suggests that the broadcasting keynote has "regularity" and "measurability".

Based on the theory of speech production, this paper makes the empirical research on the broadcasting keynote by using the methods of modern phonetics. The speech emotion evaluation scale model is used to test the components of speech emotion in the keynote and quantify and describe the broadcasting keynote from a new perspective. Then, this study also provides intuitive expression skills for announcers, hosts and students majoring in broadcasting and anchoring to grasp the keynote of broadcasting. At the same time, as the virtual anchor is becoming increasingly prominent today, the study of broadcasting keynote provides new research ideas and methods for emotional speech recognition and synthesis technology. In the future, further research on the perception of speech emotion can be conducted to explore issues related to the perception of speech emotion.