# Affect Related Acoustic Features of Speech and Their Modification*

Dandan Cui, Fanbo Meng, Lianhong Cai, and Liuyi Sun

Department of Computer Science and Technology, Tsinghua University,
Beijing 100084, P.R. China
cuidd02@mails.tsinghua.edu.cn

**Abstract.** This paper presents our recent work on the investigation of affect related acoustic features of Mandarin speech and their modification. A syllabic F0 contour model so called "Gravity Model" is proposed, which can describe the detailed changes of the F0 contour on the syllable level between different emotions. Then, different representations of spectrum, together with different modification methods, are also tried and compared; a maximal-mean amplitude representation proves to be the best, with its corresponding modifying method.

## 1 Introduction

Acoustic features are very important for affective speech research. However, during our preliminary analyzing, converting, and modifying, we found the conventional features are facing many problems [1] [2]: utterances with the similar acoustic features may have different emotions; features may conflict during modification. This paper investigates acoustic features and their modification methods, especially for the purpose of affect conversion. The work is based on the affective speech corpus ACCorpus_SA developed in Tsinghua University, which includes 11 typical emotional categories that can cover all the octants of PAD space: exuberant, relaxed, docile, disdainful, disgusted, angry, fearful, anxious, surprised, sad and neutral [2].

## 2 Investigation on Prosodic Features: Gravity Model

To describe the detailed changes of F0 contour between different emotions, taking the principle of Pitch Target Model for reference but giving the F0 contour more variability and stability, we proposed a more detailed syllabic F0 model. First, referring to the SPiS [3], we use 6 parameters to build up the F0 framework of a syllable: B for MinF0, H for MaxF0, $N_1$ for the position of B, $N_2$ for the position of H, F for the start position of F0 contour, and E for the end. The kernel of

* The work in this paper is supported by National Science Foundation of China (No. 60433030, No. 60418012) and the Special Funds for Major State Basic Research Program of China (973 Program) (No. 2006CB303101).

A. Paiva, R. Prada, and R.W. Picard (Eds.): ACII 2007, LNCS 4738, pp. 779–780, 2007.

the whole model is to portray the difference between the shapes of F0 between emotions. We take 2 kinds of changes into account and model them by 3 kinds of Gravity Effects. In modification experiment, Gravity Model is superior to Pitch Target Model and conventional stat. features in both objective and perceptive measures. With the Gravity Affects, more details of F0 contour are concerned. At the same time, the framework insures our model wont produce unreasonable outputs like Pitch Target Model does, even if the emotions are really extreme, or the style is pretty free.

## 3 Characterization and Modification of Spectrum

After trying several approaches including Spectral Centroid, we suppose that: the affect related information may be embedded more in the distribution along amplitude-axis, rather than along the frequency-axis. So, the representation of spectrum is revised as follows: The maximal amplitude of spectrum (SpecMax), i.e. the range of spectrum; the average amplitude of spectrum (SpecMean), i.e. the major distribution of spectrum, which is modulated nonlinearly within the range of spectral amplitude. Perceptive test shows that introducing spectral features into conversion and taking the amplitude-axis oriented method in its characterization and modification, the perceptive distance between the modified and target speech is reduced. Besides, the two features we selected can be both easily calculated and modified.

## 4 Conclusion

As stated above, this paper presents our recent work on the investigation of affect related acoustic features of Mandarin speech and their modification. A syllabic F0 contour model so called Gravity Model is proposed, which can describe the detailed changes of the F0 contour on the syllable level between different emotions. Then, spectral features are also introduced: a maximal-mean amplitude representation proves to be the best, with its corresponding modifying method. The work is a basis of our succeeding affective speech modeling and system building.

## References

1. Cui, D., Cai, L., Wang, Y., Zhang, X.: Investigation on pleasure related acoustic features of affective speech. In: Huo, Q., Ma, B., Chng, E.-S., Li, H. (eds.) ISCSLP 2006. LNCS (LNAI), vol. 4274, pp. 67–78. Springer, Heidelberg (2006)
2. Cui, D., Cai, L.: Acoustic and physiological feature analysis of affective speech. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) ICIC 2006. LNCS, vol. 4113, pp. 912–917. Springer, Heidelberg (2006)
3. Jianhua Tao. The prosodic modeling in Mandarin TTS. [PhD dessertation]. Beijing: Tsinghua University (2001)