Analysis of glottal source waves for emotional speech using ARX-LF model

Yongwei Li, Masato Akagi
Japan Advanced Institute of Science and Technology

Abstract : It is crucial to analyze the glottal source waves and vocal tract shapes for emotional speech synthesis that plays an important role in many real applications. Such as speech-to-speech translation and story teller systems. This paper focuses on estimating glottal source waves of emotional speech estimated by an improved ARX-LF model. Specifically, the ARX-LF model that was originally introduced for neutral speech analysis is improved by further integrating the GOI and GCI cues of speech, resulting in an improved analysis performance for emotional speech. The improved ARX-LF model was then used to estimate the glottal source waves of three basic emotional states (joy, anger, sad) and one neutral vowel /a/. Estimated glottal source waves are evaluated by re-synthetized speech with fixed vocal tract shapes. The emotions of original vowel /a/ and re-synthetized one are compared by listening test. The result showed that the emotions of original vowel /a/ are same with re-synthetized one, except joy state.

Keywords : ARX-LF model, source-filter separation, emotional speech, glottal source wave.

1. Introduction

As the most natural means of human-human communication, speech consists of not only linguistic information but also non-linguistic information, such as emotion, gender and age. Although speech includes various non-linguistic information, emotion is the mean focus of this research. Emotional speech analysis provides an important basis for expressive speech synthesis and many other speech applications. Most previous studies on emotional speech analysis tried to find out the acoustic features that are closely related to the emotions in speech, on which emotional speech synthesis and recognition systems have been built up. It seems quite difficult for this research methodology to further improve the performance of emotional speech synthesis and recognition systems, which in still far from that of human being. It is well-known that glottal source and vocal tract shapes play important roles in speech production. Thus, source filter model are widely used for estimating glottal source parameters and synthesizing speeches.

One of source filter models is the ARX-LF model has been widely used, since it enables to separate the roles of glottal source and vocal tract shape in speech production. The ARX-LF model was originally suggested for neutral speech analysis and synthesis. However it is still challenging for the ARX-LF model to analyze the emotional speech. In this paper, we propose a new approach to analyze emotional speech using ARX-LF model.

2. Method

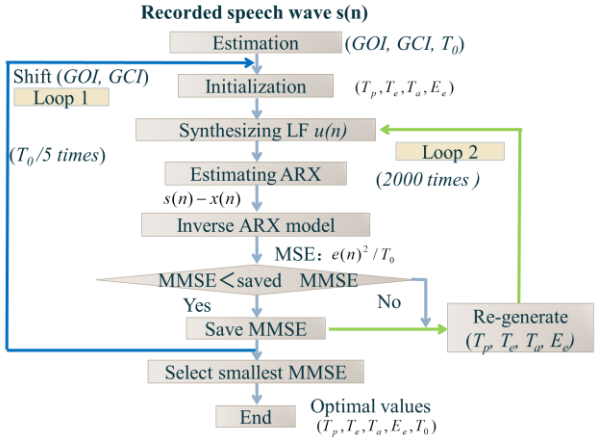

Figure 1: The analysis procedure of a period of glottal source wave.

3. Evaluation

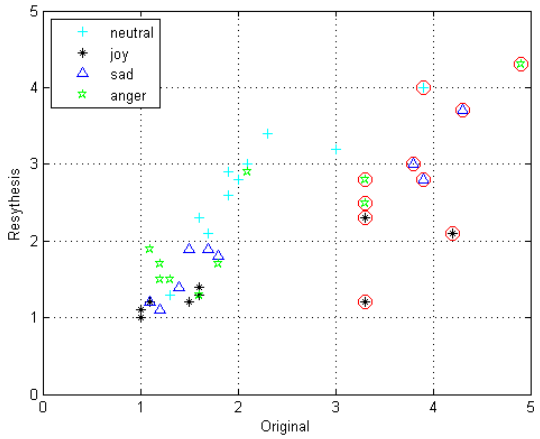

Figure 2: Comparison of original speech and resynthesized speech with fixed vocal tract shape.

4. Conclusion

The glottal source waves were estimated and analyzed for emotional speeches using the proposed estimation algorithms of the ARX-LF model. The listening test was used for evaluation, the results most are expected, except joy state. The future work is to consider solve problem of joy state.