**Realization of Personalized Speech Communication in a High Reliable Internet**

----Text-to-Speech System for Chinese using Physiological Articulatory Model

Jianguo Wei

Japan Advanced Institute of Science and Technology

1-1, Asahidai, Nomi-shi, Ishikawa, 923-1211, Japan

E-mail: jianguo@jaist.ac.jp

1. **Purpose of research:**

Speech synthesis is the technology that underlies the ability of computers to talk. Talking machines are currently being deployed in a variety of applications. It is believed that the use of speech synthesis technology will increase in the coming years in areas such as speech instruction, automated announcement systems, and telephone access to internet services and information for the majority of people without online access.

The ultimate goal for speech synthesis is to generate human-like speech. However, it is difficult to achieve such a goal by the current popular speech synthesis technology such as concatenative and formant methods which cannot implement the personalized information in the synthesizer. The ultimate objective of my research is to implement human mechanism in controlling a physiological speech production model to realize a TTS system for Chinese. This system is based on a physiological articulatory model constructed by Dang et al [1].

2. **Approach**

One of the advantages of physiological articulatory model based speech synthesizer is that it can realize coarticulation in an intuitive physical way, which is greatly correlated with naturalness of synthetic speech. So in first step, we focus on investigating and implementing the coarticulation in the physiological articulatory model.

In order to achieve this target, the information of phoneme target in articulatory space, articulatory movement and speech sound in acoustic space need to be investigated. In order to realize the coarticulation mechanism, we need to use the typical articulatory targets and parameters of coarticulation model which are unknown to us. We use optimization method to estimate them by using the observed articulatory data from electromagnetic midsagittal articulographic (EMMA) experiments. For evaluation, we test the model by combining the model with the physiological articulatory model on model simulation.

3. **Research progress of 2005**

A coarticulation model, namely ‘carrier model’, has been proposed by Dang et al [2] to improve the naturalness of synthesized speech using a physiological articulatory model. The form of the
carrier model offers a good framework to account for the coarticulation in the planning stage, while its parameters need to be refined in order to improve the performance of the model. For use of this model, we suppose that there is a typical spatial target for each phoneme, which is actually unknown. The objective of the study in this fiscal year is to refine the parameters of the carrier model and learn the typical articulatory targets by reducing the difference between model simulations and observations. These two tasks were combined in a model-based simulation using an optimization framework. In the optimization, a bilevel optimization strategy was employed in a high level, obtaining the planned targets from the typical target via the carrier model, to decompose the complicated problem into a set of subproblems. A direct search method was applied in low level that consists of the processes from the planned targets to articulatory movements. A general evaluation was carried out by combining the refined carrier model and the learned typical targets together with the physiological articulatory model. As the result, the refined parameters of carrier model and the learned typical articulatory targets that obtained by above method showed a good performance in the simulation.

4. Future direction

- Collecting more articulatory data using EMA. For good generalization of the model, more practical data needs to be collected which will be used to optimize the model parameters.

- Investigating the articulatory movement trajectory and its relationship with the speech sound in order to produce the natural sound by the physiological articulatory model.

5. Publication


6. Reference
