PROGRESS REPORT

Jianguo Wei
IIPL laboratory, Japan Advanced Institute of Science and Technology

1. Research Topic

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

2. Background:

Speech synthesis is the technology that underlies the ability of computers to talk. Talking machines are currently being deployed in a variety of applications. And there is main reason to believe that the use of speech synthesis technology will increase in the coming years, and expand into areas such as language instruction, automated announcement systems, and telephone access to internet services and information for the majority of people without online access. In short, speech synthesis technology is having an increasing impact on how we work and live in our high tech society.

Nowadays, great achievement has been accomplished in the speech synthesis system which has come out of lab and become commercial system. But despite the actual and future potential of speech synthesis, users are not generally happy with the quality of current systems. Concatenative synthesis is currently the most popular technique in commercial and research TTS system. There is no commercial and out-lab synthesizer based on physiological articulatory model, which is our final target. Yet authors of The Bell Labs Approach state: “We use Concatenative synthesis because that is currently the best available method to produce synthetic speech of consistently high quality . . . however; at the same time we also believe that in the long run concatenative synthesis is not the answer”. The synthesizer based on physiological articulatory model show the characteristics to solve the problems which could not be handled by concatenative method, because articulatory model pays attention to describe the system that produces the sound rather than the result of the process. For example, articulatory method should handle coarticulation correctly, since the articulators themselves are objects following the laws of physics.

3. The Purpose of My Research

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 based on a physiological articulatory model constructed by Dang et al.

To do this, the first task of this research is to implement a coarticulation model,”Carrier model” (Dang et al) in this physiological articulatory model in order to drive this it produces sound automatically and naturally. In this task, we need obtain the coefficients of”Carrier model” by means of analyze the electromagnetic articulographic system data.

The second task of this research is to realize the prosody model in this system, so as to this system can produce sentence level utterance naturally.

The third task of this research is to obtain the articulatory target of Chinese special phonemes. The step can result in produce natural Chinese sound by this physiological articulatory model. In this step, we need record the acoustic and articulatory data of Chinese by means of electromagnetic articulographic system. Then we need analysis this data set to obtain the

parameters in order to driving our system produce Chinese phonemes and syllables.

The final task of this research is release this system on internet so as to everyone can evaluates it online.

4. The Finished Work of Last Year

In the last year, we are keeping work in the study of coarticulation which is the key issue to drive the physiological articulatory model synthesis sound automatically. In this research, we quantify the coefficients in modeling the coarticulation mechanism, and verify it using simulation with the physiological articulatory model. In this research, we estimate the coefficients of 'Carrier Model' based on articulatory data analysis. The articulatory data used in this study were recorded using the electromagnetic articulographic system for reading sentences by three Japanese male subjects. We used the statistical methods to study the distribution of the articulator's position and movement tendency corresponding to every vowel and consonants. We also studied the correlation between central phonemes with bilateral phonemes to quantify the effect of bilateral vowel and consonant on the central phoneme, and then we can modulate the articulatory gesture of the central phoneme considered the coarticulation. We obtain the coefficient about α, β and degree of articulatory constrains of five Japanese vowels and consonances /g/,/k/,/d/,/t/,which be used in the 'Carrier Model'. DAC denote the capability of phoneme insist on its position when surrounding phonemes affect it. The Carrier Model is implemented in the physiological articulatory model to take the coarticulation into account in the target planning stage. VCV phoneme sequences were used in the simulation, in which the "abstract" target is out of context-independent typical target codebook. The coefficients obtained above were implemented in the model. The simulation showed reasonable results by comparing to the observation.

5.The List of Publication

1. Jianguo Wei, Takeharu Suzuki, Jianwu Dang, ``Modeling and Simulation of Coarticulation Using a Physiological Articulatory Model," Proc. of The 2005 Spring Meeting of The Acoustical Society of Japan, 1-1-6, pp.165-166, Mar 2005.

2. Jianwu Dang,Jianguo Wei,Takeharu Suzuki,Kiyoshi Honda,Pascal Perrier and Masaaki Honda, " Investigation and modeling of coarticulation in speech production", International Symposium of Chinese Spoken Language 2004, 25-28.