Construction and Control of a 3D physiological articulatory model for speech production

Qiang FANG
IIPL, Japan Advanced Institute of Science and Technology

1. Purpose of this study

For years, speech scientist attempted to make it clear what the nature of phonetic targets are and how the observed articulatory movement and corresponding speech signals are rendered from the phonetic targets by coarticulation via the observed data. However, from our point of view, the observed articulatory movements as well as speech signals are the consequence of the interaction between intended targets and the mechanical properties of speech apparatus. This brings difficulties to the understanding of the issues what we are interested in. Firstly, the phonetic targets can not be observed directly and need to be inferred from the observed articulatory movements. Secondly, the articulatory movements are rendered by the co-effects of anticipation and carry over effects involved in the coarticulation process. Thirdly, anticipation and carry over effects are difficult to be separated. Therefore, if there is an effective method to separate these effects, it would help people to understand what the phonetic targets are and how speech is controlled, and further, how the speech perception correlates with speech production.

Effects of the anticipation are hypothesized to be mainly concerned with the planning process, while effects of the carry over are considered to be mainly concerned with the mechanical properties of the articulatory systems. Hence, the carry over effects are the inherent properties embedded in the human physiological mechanism that is able to be realized using a faithful physiological articulatory model. Accordingly, we construct a 3D physiological articulatory model to account for the carry over effects, while the anticipation function is realized in the high level planning only. Thus, these two effects can be separated in the model simulation. In this study, we attempt to construct a faithful physiological articulatory model and control it to investigate the mechanism of speech production.

2. Progress

2.1 Model configuration

In the last year, a complete 3D model was developed based on the previous pseudo-3D physiological articulatory model. It includes a 3D tongue, jaw, and surrounding vocal-tract wall, which consists of the hard palate, soft palate, pharyngeal wall and the larynx tube. (So far, the soft palate moves with the nasopharyngeal port) The vocal-tract wall is treated as unmovable structures in the current model. Furthermore, there are 9 tongue muscles and 2 jaw muscle groups involved in the model to drive the tongue and jaw for articulation and other movements. Basic evaluation showed that the model behaved properly when the muscles were activated.

2.2 Model simulation

Differing from the psudo-3D model, in the 3D physiological articulatory model, the tongue deforms and moves in the transversal dimension as well as sagittal dimension. For this reason, the control strategy for the psudo-3D model cannot be applied to the full 3D model directly. Our first focus is to develop a new control strategy for the 3D model. Our control method is a static target-force mapping, which is derived from a number of target articulations and correspond muscle forces. Therefore, a large number of simulations are necessary. Based preliminary analysis, we hypothesize that 1) the tongue muscles could be roughly classified into 2 categories: one for the sagittal movements of the tongue, and the other for the transversal movement of the tongue; 2) some muscle pairs responsible for the sagittal movement act as antagonist pairs for some part of tongue while have cooperative effects for the other part. Based on the above assumption, the number of the simulations is decreased

to about 1/9 of the exhaustive searching method. And the results were reasonable from acoustic point of view and articulatory point of view as well.

2.3 Posture analysis

To simplify the task of model control, at the current stage, we consider the control for vowels only. The postures for the vowels were extracted from the simulation results according to the acoustic and articulatory constraints. The posture of the articulatory system is described by the coordinates of the nodes of the tongue and jaw, which is high dimensional and has great redundancies. To depict the posture of the vowels with efficient parameters, the postures of the vowels are analyzed by the Linear Component Analysis method. The results show that the 96% variance can be explained by 5 extracted components, which mainly concerned with jaw height component (JH), tongue body front-back component (TBA), tongue body width component (TBW), tongue dorsum arch component (TDA) and tongue tip front-back component (TTA). And the RMS between the reconstructed posture and the original posture fall in an acceptable range (0.15mm).

2.4 Vocal tract and area function

To generate speech sound by the proposed physiological articulatory model, a module is required to extract the vocal tract shape and estimate the associated area function. In most previous study, the area function of the VT was estimated by using the width of the midsagittal plane (2D model) or the width in both midsagittal and parasagittal plane (partial 3D model). Those methods heavily depend on the predefined the coefficients estimated from a small sample set, therefore readily introduce error. To overcome this problem, a 3D vocal tract model was proposed, in which the area of each cross-sectional slice is calculated by direct integration rather than by the parametric model based on sagittal dimensions. The result shows that the accuracy of the area function estimated by the 3D model is better than that by previous method, and is comparable with that measured by MR images.

3. Summary

To apply the physiological articulatory model to the research of speech production, the modules, articulatory model, control strategy, and speech sound generation are indispensable. So far, the 3D physiological articulatory model has been established. For the purpose of efficiently controlling the model to produce vowel, the postures and corresponding muscle forces are necessary for building the target-force mapping. At the current stage, the postures of vowels are extracted from results of model simulation, and efficiently described by 5 components which are assumed to associate with the freedom of the articulatory apparatus for speech. The mapping between the postures and the associated muscle forces would be constructed in the next step. As for generating speech sounds from articulatory movements, the module for estimate the area function of vocal tract and the acoustic model for produce sound have been accomplished.

4. Future work

- To analyze the muscle force patterns associated with the extract postures for vowels
- To establish the mapping between posture and muscle forces.

Publication

[1] Qiang FANG, Jianguo WEI, Xugang LU, Jianwu DANG, “A 3D physiological articulatory model for speech synthesis”, The Japan-China joint conference of Acoustics, June ,2007

[2] Qiang FANG, Satoru Fujita, Xugang Lu, Jianwu Dang, “Analysis of 3-D tongue shape in speech production,” The 2007 Autumn Conference of The Acoustical Society of Japan,, pp.335-336, Sep 2007

[3] Qiang FANG, Akikazu Nishikido, Satoru Fujita, Xugang LU, Jianwu DANG, “Investigation of 3D tongue shape for model control”, The 2008 Spring Conference of The Acoustical Society of Japan

[4] Qiang FANG, Satoru Fujita, Xugang LU, Jianwu DANG, “Investigation of functional relationships of the tongue muscles for model control”, submitted to PCC_2008