Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 111 (2017) 463–470

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Information Technology, IAIT2016, 19-22 December 2016, Macau, China

Speech Behavior Analysis by Articulatory Observations

Yan Ji[a], Jianguo Wei[b], Jingshu Zhang[a], Qiang Fang[c], Wenhuan Lu[b,*], Kiyoshi Honda[a], Xugang Lu[d]

[a]*School of Computer Science and Technology, Yaguan Road 135, Tianjin University, 300350 Tianjin, China*
[b]*School of Computer Software, Yaguan Road 135, Tianjin University, 300350 Tianjin, China*
[c]*Chinese Academy of Social Sciences, 5 Jianguomennei Dajie, 100732 Beijing, China*
[d]*National Institute of Information and Communications Technology, Kyoto, Japan*

Abstract

In articulatory phonetics, the place and manner of articulation indicate setting and action of the articulators during production of speech sounds. The place and manner are varied in pronouncing phonetic components such as vowel, bilabial, and alveolar. Moreover, the place and manner differ across individual speakers in realizing the same phoneme. Hence, our study attempts at analyzing those ways of articulation during speech production. To do so, the electromagnetic articulograph is used for recording the articulatory information so that the speaker-specific behavior of the articulators can be explored. In this study, the ways of articulation among different speakers was investigated by applying an articulatory recognition technique, and the contribution of each sensor's location to articulation was evaluated. In the results, two different production behaviors were observed, suggesting different behavioral rules for producing varied phonemes, and the contribution of the tongue surface regions were larger than other articulators among the subjects.

Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology

Keywords: Speech behavior; Articulatory recognition; EMA data;

* Corresponding author. Tel.: +86 18722011170;
E-mail address: wenhuan@tju.edu.cn

1877-0509
Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology
10.1016/j.procs.2017.06.048

Introduction

Speech behavior is a potential key factor in the study of personalized speech synthesis, speaker identification, and emotional analysis. While common acoustic analysis on speech signals is limited for this purpose, examination on motor behavior for speech elucidates the roles of articulation on producing varied speech sounds. Also, the motor behavior varies across speakers, and its analysis reveals causal factors of individual difference in speech sound. This study aims at exploring the speech behavior as individually and phonetically different settings and actions of speech organs using the kinematic data on the motor behavior for speech production. To do so, a novel method based on speech recognition techniques was employed to reveal articulatory-to-acoustic processes, which we call called articulatory recognition in this study. Speech recognition technology has progressively advanced in recent years. However, the work mainly focuses on the traditional speech recognition based on acoustic features. In this study, a recognizer is used for matching speech production and perception mechanisms instead for patterning the acoustic features. Using the kinematic articulatory data, this approach leads us to analyze the effect of settings and actions of speech organs on phonetic realization of speech sounds and its individual difference.

This study is based on the reports from our group. Dang (2001) developed a physiological articulatory model, which includes the lips, jaw, teeth, tongue, velum to contract the vocal tract[1]. Different articulators control the production of diverse phonemes. For instance, the velum plays an important role in distinguish nasal sounds, while movements of the tongue dorsum are important for distinguishing vowels. Based on such knowledge, the ways of articulation are studied by articulatory recognition in this study. On one purpose, we study on realization processes of diverse phonemes. On another, we focus on individual difference of speakers' pronunciation. In the analysis, the articulatory database composed by the electromagnetic articulograph (EMA) at the NTT Communication Science Laboratories was employed. As Dang (2003) described, the articulatory data were obtained from eight sensor points on the articulators[2]. In the EMA database, x- and y-coordinates were recorded for each sensor location. As shown in Fig. 1, the sensor location is the upper lip (UL), lower lip (LL), lower jaw (LJ), tongue tip (T1), tongue blade (T2), tongue dorsum (T3), tongue rear (T4) and velum (VM). The sampling frequency for recording each sensor location was set at 250 Hz.

This paper is organized as follows. In section 2, we analyzed the Japanese Phonology on phonemes and the ways of articulation among subjects. In section 3, we introduce the framework of our proposed method. In section 4, we describe the material and method for analysis and evaluate articulators' regional contribution to each phoneme, which revealed speech behavior during speech production. Finally, the conclusion is given in section 5.

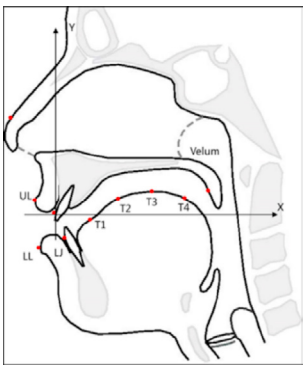

Fig. 1. Placement of the sensors for obtaining observation points from the articulatory data.