Morphological normalization of vowel images for articulatory speech recognition

Jianguo Wei¹,², Jingshu Zhang², Yan Ji², Wenhuan Lu¹ and Jianwu Dang³

¹School of Computer Software, Tianjin University,
²Tianjin Key Laboratory of Cognitive Computing and Application, School of Computer Science and Technology, Tianjin University,
³Chinese Academy of Social Sciences, Beijing,
{jianguo; jingshu; tjujiyan; wenhuan}@tju.edu.cn, fangqiang@cass.org.cn

ABSTRACT
Minimizing morphological variances of the vocal tract across speakers is a challenge for articulatory analysis and modeling. In order to reduce morphological differences in speech organs among speakers and retain speakers’ speech dynamics, our study proposes a method of normalizing the vocal-tract shapes of Mandarin and Japanese speakers by using a Thin-Plate Spline (TPS) method. We apply the properties of TPS in a two-dimensional space in order to normalize vocal-tract shapes. Furthermore, we also use DNN (Deep Neural Networks) based speech recognition for our evaluations. We obtained our template for normalization by measuring three speakers’ palates and tongue shapes. Our results show a reduction in variances among subjects. The similar vowel structure of pre/post-normalization data indicates that our framework retains speaker specific characteristics. Our results for the articulatory recognition of isolated phonemes show an improvement of 25%. Moreover, our phone error rate of continuous speech improved by 5.84%.

Index Terms— Vocal tract normalization, Articulatory data, Acoustic data, Thin-Plate Spline, DNN, articulatory recognition