

ACOUSTICS2008/852
Continuous estimation of VTL from vowels using a linearly
VTL-covariant speech feature

Christian Feldbauer, Jessica Monaghan and Roy Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience,
University of Cambridge, Downing Site, CB23EG Cambridge, UK

Vowel sounds encode information about the speaker's vocal tract length (VTL) as well as vowel type. This paper shows how VTL can be estimated on a frame-by-frame basis from vowels to facilitate speaker tracking and make automatic speech recognition (ASR) more robust in multi-speaker environments. The algorithm is based on a new, VTL-covariant speech feature that encodes the size information linearly. The paper shows that this new speech feature is better suited for VTL estimation than the more traditional cepstral coefficients. The VTL estimation is based on a Gaussian mixture model that was trained on the speech material of speakers with known body height but unknown VTL. Research supported by the Austrian Science Fund (FWF) [J2541-N15], EOARD [FA8655-05-1-3043], and the UK-MRC [G0500221].