Modelling and Segmentation of the Vocal Track during Speech Production by using Deformable Models in Magnetic Resonance Images

Maria João M. Vasconcelos, Sandra M. Rua Ventura, Diamantino Rui S. Freitas, João Manuel R. S. Tavares
Faculty of Engineering, University of Porto
Porto, PORTUGAL
maria.vasconcelos@fe.up.pt, smr@estsp.ipp.pt, dfreitas@fe.up.pt, tavares@fe.up.pt

ABSTRACT

Since ancient times, speech production has attracted particularly interest aiming at reaching a deeper understanding of the mechanisms involved by considering both morphological and speech acoustic aspects. The central anatomical aspects and the physiology of the human vocal tract are common to all individuals. However, speech production is an exceptionally complex and individualistic process. Therefore, the modelling of the mechanisms involved in speech production implies the enclosing of adequate flexibility in order to consider individual variations accurately. In this work, the shape of vocal tract in the articulation of some European Portuguese (EP) sounds is evaluated by using deformable models applied in Magnetic Resonance (MR) images. Additionally, the deformable models built are afterwards used to automatically segment the modelled vocal tract in MR images.

From the imaging modalities that have been taken into consideration in order to study the vocal tract shape and articulators, Magnetic Resonance Imaging (MRI) has been the most commonly accepted. Actually, the use of MRI allows the study of the entire human vocal tract and, in addition, the quality and resolution of soft-tissues and the use of non-ionizing radiation are key advantages presented by MRI.

The deformable model used, commonly known as Point Distribution Model (PDM), was built from a set of training images acquired during artificially sustained articulations of 21 EP sounds. In a brief review, one can assert that PDM's are obtained by a statistical analysis done on the co-ordinates of landmark points that represent the shape to be modelled: after aligning the training shapes, a Principal Component Analysis is performed in order to obtain the model mean shape and the modes of variation relatively to this mean shape.

Combining the geometrical information of the PDM with the grey levels of the landmark points used in its building one can build the Active Shape Models (ASM) and the Active Appearance Models (AAM). With these enhanced models is possible to segment the modelled shape in new images in a fully automated way.

From the experimental results obtained in this work, one may conclude that the PDM built could efficiently characterize the behaviour of the vocal tract shape during the production of the EP sounds studied with MRI. Furthermore, one can verify that the ASM and the AAM built could be used to segment the modelled vocal tract in MR images in a successful manner. Therefore, the deformable models built should be considered towards the efficient and automatic study of the vocal tract during speech production with MRI, in particular for enhanced speech production simulation and speech rehabilitation therapies.

Acknowledgments

The first and second authors would like to thank the support of the PhD grants with references SFRH/BD/28817/2006 and SFRH/PROTEC/49517/2009, respectively, from *Fundação para a Ciência e Tecnologia* (FCT).

This work was partially done in the scope of the project “Methodologies to Analyze Organs from Complex Medical Images – Applications to Female Pelvic Cavity”, with reference PTDC/EEA-CRO/103320/2008, financially supported by FCT.