Tongue Visualization For Specified Speech Task

Yin Yang Xiaohu Guo*

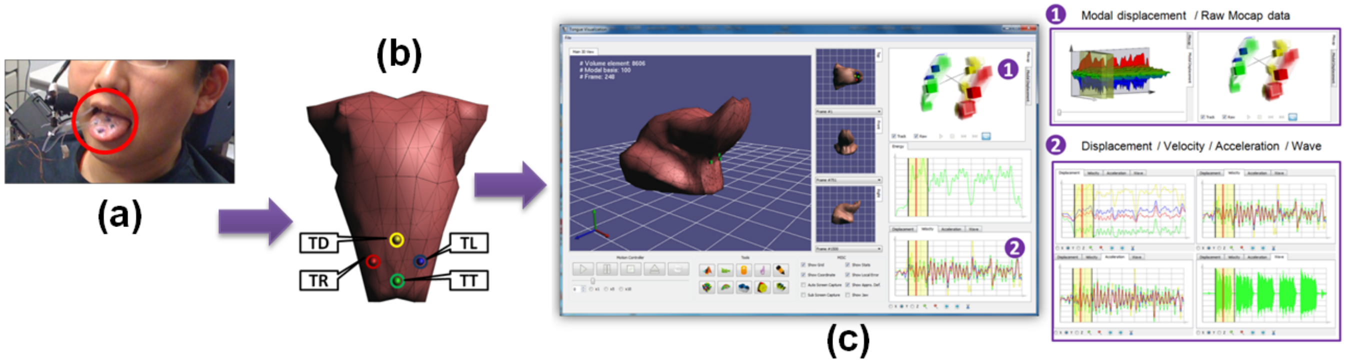

Figure 1: *An overview of the general work flow of the proposed framework.(a) 4 mocap sensors placed on the tongue. (b) The corresponding 4 nodes on the mesh. (c) A comprehensive visualization interface that provides enriched/abstract visual information to the user.*

1 Introduction

Tongue is a muscular organ which is particularly important for phonetic articulation during speech task. In many real cases, a visualized movement of the tongue could be of help, for example in the rehabilitation of speech disorders. The challenge lies in the fact that the tongue motion is subtle, swift and hardly visible during speech production. Many imaging modalities (*e.g.*, MRI/CT/ultrasound) have been adopted for the acquisition of tongue shape. Unfortunately, these methodologies are expensive and with fundamental limitations especially for real-time 3D imaging reconstruction of the tongue.

We propose a novel real-time visualization framework based on the motion capture (mocap) technique and constraint-driven elastic model that circumvents the aforementioned problems and produces the most physically reasonable estimate of the tongue shape. We also provide a deformation segmentation tool for the domain expert to fast identify the “tendency” of the deformation. Additional features including energy-based low dimension visualization and morphing are also available in this framework.

2 Method

Data gathering has 7 subjects including native/non-native English as well as an adult patient suffering with communication disorder involved. 5 mocap sensors were used in the experiment. 4 of them were directly attached to the tongue (Fig. 1 (a)) and 1 sensor was placed at mandibular symphysis so that the contribution of jaw movement to the tongue’s position can be removed. The elicited tasks were consonant-vowel (CV) syllable trains that were 5 syllables in length and the audio was also recorded by a head-mounted microphone. An optimized transformation that maps the gathered mocap signals to the finite element mesh was applied to minimize the geometrical deviation between the mesh and the tongue (Fig. 1 (b)).

Shape modeling is based on the theory of linear elasticity and modal warping [Choi and Ko 2005]. The constraint-driven dynamics utilizes the normalized mocap data and computes the appropriate constraint forces which, instead of muscle articulation [Stavness et al. 2011], forward the movement of the tongue (Fig. 2 (a)). Modal analysis significantly reduces the computation intensity and makes the visualization real-time.

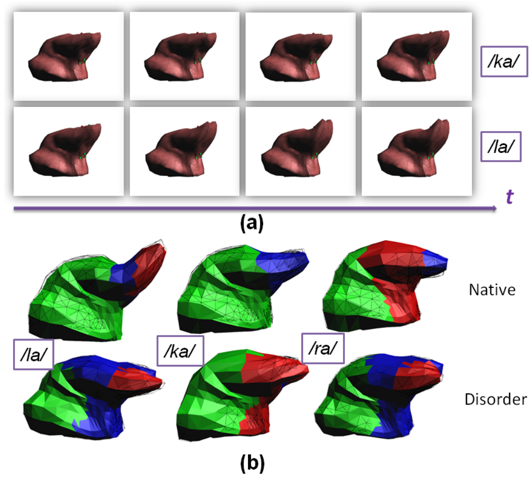

Figure 2: *(a) Sequence of snapshots of the tongue shapes during producing CV syllables /ka/ and /la/. (b) Segmentation results for a native English speaker and a patient with communication disorder.*

Deformation segmentation targets on providing the visualizaion of the general pattern of the 3D deformation. Like using a piecewise linear line segment to approximate a curve, we use a low-ranked local subspace deformation to fit the given displacement field. The hidden sematic information could be unveiled with this technique. For example, an asymmetrical tip deformation can be found at disordered speaker while the normal speaker has mostly symmetric shape (Fig. 2 (b)).

References

Choi, M. G., and Ko, H.-S. 2005. Modal warping: Real-time simulation of large rotational deformation and manipulation. *IEEE TVCG 11*, 1, 91–101.

Stavness, I., Lloyd, J. E., Payan, Y., and Fels, S. 2011. Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics. *Int. J. for Numerical Methods in Biomedical Engineering 27*, 3, 367–390.

*e-mail:{yinyang|xguo}@utdallas.edu

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008