A DYNAMIC THREE-DIMENSIONAL REPRESENTATION OF THE TONGUE SURFACE BASED ON ULTRASOUND SCANS FOR TIME-VARYING VOCALIZATIONS

M. Cordaro*, M. Stone**, M.H. Goldstein*, M. Unser**

*Sensory Communications Lab, The Johns Hopkins University, Baltimore, MD 21218
**National Institutes of Health, Bethesda, MD 20892

ABSTRACT

This paper describes a technique for producing a dynamic three-dimensional display of tongue surface motion based on independently collected two-dimensional ultrasound scans. Multiple coronal ultrasound scans of the tongue are recorded on video tape with a vocalization repeated for each scan angle. The tongue surface profile is extracted from each video field using customized image processing software. The tongue profiles are aligned in the time domain using a time-warping algorithm and spatially aligned based on the measured transducer positions. The dynamic tongue surface is presented as a movie in slow-motion or freeze-frame modes.

INTRODUCTION

Real-time ultrasound imaging of the tongue has been successfully used to characterize cross-sectional tongue shape during speech and is currently used as an imaging modality for the diagnosis of speech and swallowing disorders. [1,2] A pseudo-three-dimensional ultrasound reconstruction of the tongue surface was produced for steady-state vocalizations by sweeping the transducer below the mandible. [4] However, it would be useful for the speech researcher or pathologist to reconstruct the three-dimensional tongue surface for the case of time-varying vocalizations.

A personal computer-based method is presented to reconstruct the dynamic tongue surface from multiple independently collected two-dimensional ultrasound scans of the tongue.

METHODS

Multiple coronal views of the tongue are obtained using a mechanical sectoring real-time ultrasonic scanner (Advanced Technology Laboratory, Inc., Bellevue, WA) with a 3-MHz transducer which produces a complete scan 30 times per second as described by Stone, et al. [2] The images are recorded on a Sony U-Matic video tape recorder (VTR). A microphone is used to simultaneously record the acoustic waveform on one of the VTR's audio tracks. A real-time sequence of images (scan sequence) is collected as a vocalization is repeated for each transducer position.

A Macintosh IIci equipped with a QuickCapture frame grabber board (Data Translation 2255) is used to digitize the individual video fields of each scan sequence. A Region of Interest (ROI) is defined and saved on hard disk as a TIFF file.

For each image field, the upper surface of the tongue (surface profile) is detected using customized software developed by Unser and Stone. [3] The ultrasound image is first preprocessed by performing a 50% size reduction and low-pass filtering (blurring). A sector of interest is resampled in polar coordinates, which are better suited to the tongue's geometry. A matched filter (vertical edge detector) is applied to the polar sector to enhance the tissue-air interface. Finally, the optimal radial path or tongue surface profile is detected using a dynamic programming algorithm and this contour is transformed back to rectangular coordinates.

For each scan sequence, the acoustic signal is low-pass filtered to 5 KHz and digitized at a 10 KHz sampling rate from the VTR's audio track using an IBM PC AT equipped with a DT2821 I/O board (Data Translation) and Canadian Speech Research Environment (CSRE 3.0, University of Western Ontario) software. CSRE provides a Raw Parameter Tracking sub-system which calculates the frequency, bandwidth, and amplitude of formant frequency "candidates" as well as fundamental frequency for analysis windows of the acoustic waveform. A discrete 16 msec analysis window is hopped through the waveform to provide short-time spectral data corresponding to each video field of a scan sequence.

The three-dimensional display is produced by composing independently collected two-dimensional ultrasound scans which represent multiple repetitions of a vocalization. Since a human subject cannot reproduce a vocalization at exactly the same rate, the multiple scan views must be aligned in the time domain. In addition, repetitions of a speech sample do not differ in a linear fashion. They are not simply shortened or stretched versions of each other.

The multiple ultrasound scans are aligned in time or time-warped in a piecewise linear fashion. Each sequence is

broken down into the same number of temporal segments based on the content of the images and acoustic waveform. The shortest of each temporal segment is taken as a template. The remaining segments are linearly compressed to match the template by removing appropriate video fields.

Once time-aligned, the multiple extracted two-dimensional tongue surface profiles are composed into a three-dimensional representation of the tongue surface at each point in time based on measured transducer positions.

RESULTS

The result is a sequence of images such as that shown in Figure 1 which demonstrate the tongue surface over the time course of a dynamic vocalization. Each image of the sequence is effectively a video field in the resultant "movie" of speech. This "movie" can be displayed at several speeds and in freeze-frame mode.

Initial implementation of the dynamic three-dimensional reconstruction of the tongue surface indicates that it will be useful for qualitative analysis in the areas of speech research and pathology. Currently the technique lacks automation, but with further refinement, extensive data collection will be permitted.

The reconstruction technique allows the speech researcher or pathologist to view what previously had to be mentally abstracted from multiple ultrasound scans. This allows attention to be focused on more complex patterns of movement.

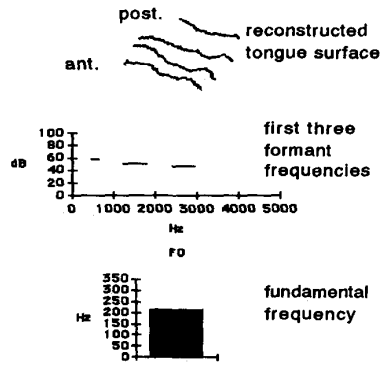

Figure 1. Single image field of resultant "movie" of the tongue surface.

REFERENCES

[1] B.C. Sonies, E.F. Ekman, et al., "Swallowing Dysfunction in Nephropathic Cystinosis", New England Journal of Medicine Vol. 323, pp. 565-570:1990.

[2] M. Stone, T.H. Shawker, et al., "Cross-Sectional Tongue Shape during the Production of Vowels", J. Acoust. Soc. Am. Vol. 83, pp. 1586-1596:1988.

[3] M. Unser, M. Stone, "Computerized Extraction of the Tongue Surface from Sequences of Ultrasound Images", J. Acoust. Soc. Am. Vol. 87 Suppl. 1, pp. S122:1990.

[4] K.L. Watkin, J.M. Rubin, "Pseudo-Three-Dimensional Reconstruction of Ultrasonic Images of the Tongue", J. Acoust. Soc. Am. Vol. 85, pp. 496-499:1989.

Marc A. Cordaro
c/o ACW Research, Inc.
1900 51st Street Suite 1-C
Brooklyn, NY 11204
(718) 252-1999