Automatic tongue surface extraction from 3D ultrasound using 3D SLURP

Steven M. Lulich[1], *Enamundram M. V. Naga Karthik*[2], *Catherine Laporte*[2]

[1] Indiana University (USA)
[2] École de technologie supérieure (Canada)

slulich@indiana.edu, muni-venkata-naga-karthik.enamundram.1@ens.etsmtl.ca, catherine.laporte@etsmtl.ca

Background: There are a number of published methods for extracting mid-sagittal tongue contours from two-dimensional ultrasound (2D-US) images (e.g. Li *et al*., 2005; Roussos *et al*., 2009; Fasel & Berry, 2010; Tang *et al*., 2012; Laporte & Ménard, 2018; Karimi *et al*., 2019). While these inherently 2D methods can be adapted to three-dimensional (3D-US) images (Laporte & Lulich, 2017), there may be advantages to developing a method that directly extracts tongue surfaces from 3D-US. Novel open-source software, called 3D SLURP, has been developed for this purpose and integrated as a plugin within the WASL application (Lulich, 2020) for MATLAB.

Method: The algorithm underlying the 3D SLURP software is an extension of Karimi *et al.* (2019)'s 2D frame-by-frame tongue contour extraction method to three dimensions (Naga Karthik *et al.*, 2020), and consists of 5 image processing steps:

- Computation of a denoised ultrasound volume using a phase symmetry map.
- Identification of tongue surface candidate voxels in the denoised volume, using measures of *sheetness* and *tongueness* based on eigenvalue decomposition of the local intensity Hessian matrix.
- Candidate voxel pruning.
- Surface growing using the Fast Marching Method.
- Surface merging and smoothing.

The property of *sheetness* refers to the degree to which candidate voxels are located in locally planar, sheet-like structures, while the property of *tongueness* refers to the degree in which the sheet-like structures are oriented orthogonally to the anatomical midsagittal plane. Note that 3D SLURP is not based on a tracking algorithm – it extracts a tongue surface from a single 3D-US image, without reference to previous 3D-US images. Furthermore, 3D SLURP is fully automatic, and does not rely on training data. 3D SLURP is integrated within WASL as an independent plugin.

Software integration: WASL is a MATLAB-based graphical user interface for visualization and analysis of ultrasound imaging data together with time-synchronized audio data. The architecture of WASL is designed to encourage rapid development of independent modules (i.e. plugins), and such a plugin was developed as a wrapper for the algorithm underlying the 3D SLURP software (cf. Figure 1).

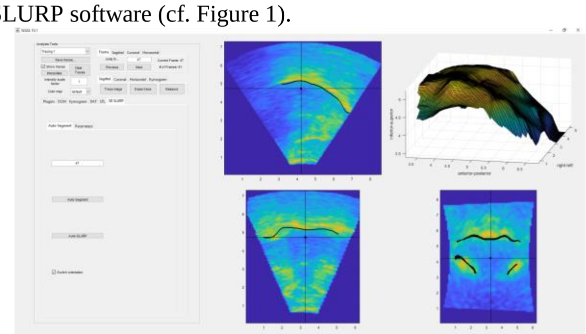

Figure 1: *Screen shot of 3D SLURP tongue surface in WASL.*

Results and discussion: The 3D SLURP software is robust, fast, and objective. Validation tests with a variety of speech stimuli (drawn from the *Goldman-Fristoe Test of Articulation, 3rd Edition*) produced by 8 speakers show median errors smaller than 2 millimeters, based on mean sum of distances (MSD) comparisons with manually segmented tongue surfaces. Midsagittal tongue contours extracted from the three-dimensional tongue surfaces had median MSD errors close to 1 millimeter, and greatly outperformed a previously developed 2.5D method (Laporte & Lulich, 2017).

In addition to providing robust 3D tongue surface measurements, automated 2D midsagittal tongue contour extraction from 3D-US images using 3D SLURP is also more robust than the 2D frame-by-frame method presented by Karimi et al. (2019), because 3D SLURP takes advantage of information in parasagittal planes through the inherently 3D sheetness measure. For example, midsagittal tongue contours extracted using 3D SLURP were superior, especially for speech sounds articulated with tongue tip retroflexion. In addition, 3D SLURP is robust to variation in speakers and speech content, with similar MSD errors across speakers (including children and adults) and speech sounds.

Tongue surface extraction using 3D SLURP is considerably faster than manual segmentation using WASL, even when the manually segmented surface is interpolated over a small set of manually segmented sagittal and coronal contours. While manual segmentation of a 3D-US image might require approximately 5 minutes of work, 3D SLURP extracts the tongue surface in approximately 15 seconds. Instances of segmentation failure when using 3D SLURP are due to familiar causes, such as poor image quality and ambiguous echoes.

Based on the experience of the first author's laboratory, using the 3D SLURP plugin for WASL, fully automatic tongue surface segmentation can be completed for a typical data set (approximately 100 recorded utterances) within 24 hours, with approximately 90% of 3D-US images judged to be segmented well enough as not to require manual corrections. This significantly streamlines a process that used to be the primary bottleneck in 3D-US data analysis.

References

Fasel, I., & Berry, J. (2010). Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. *International Conference on Pattern Recognition*, 1493-1496.

Karimi, E., Ménard, L., & Laporte, C. (2019). Fully-automated tongue detection in ultrasound images. *Computers in Biology and Medicine*, 111, 103335.

Naga Karthik, E. M. V., Karimi, E., Lulich, S. M., & Laporte, C. (2020). Automatic tongue surface extraction from 3D ultrasound vocal tract images. *Journal of the Acoustical Society of America*, 147, 1623-1633.

Laporte, C., & Lulich, S. M. (2017). Towards automated segmentation fo the moving tongue surface in 3D ultrasound. *UltraFest VIII*, 65-66.

Laporte, C., & Menard, L. (2018). Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech. *Medical Image Analysis*, 44, 98-114.

Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19, 545-554.

Lulich, S. M. (2020). Ultrasound visualization and analysis using WASL. *UltraFest IX*.

Tang, L., Bressmann, T., & Hamarneh, G. (2012). Tongue contour tracking in dynamic ultrasound via higher-order mrfs and efficient fusion moves. *Medical Image Analysis*, 16, 1503-1520.

Roussos, A., Katsamanis, A., & Maragos, P. (2009). Tongue tracking in ultrasound images with active appearance models. *IEEE International Conference on Image Processing*, 1733-1736.

Keywords: 3D/4D ultrasound, tongue surface, automatic segmentation