Error analysis of extracted tongue contours from 2D ultrasound images

Tamás Gábor Csapó[1], Steven M. Lulich[2], csapot@tmit.bme.hu, slulich@indiana.edu

[1]Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Hungary,
[2]Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN, USA

1. Introduction

Role of ultrasound in speech research

- Typical result of **2D ultrasound recordings**: a series of gray-scale images
  - The tongue surface contour has a greater brightness than the surrounding tissue and air
  - Image quality strongly depends on speaker, sex, age, hydration level of the tongue

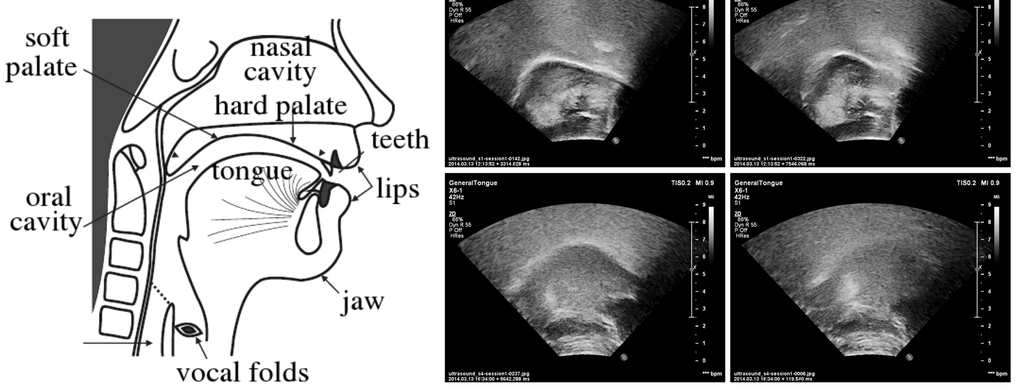

Fig. 1: Vocal tract (left) and sample ultrasound images (right) with the same orientation.

- **Extracting tongue contours** from these images is critical for later analyses
  - Comparison of tongue shapes of various languages
  - Measuring parameters related to tongue curvature
  - Addressing phonological questions related to articulation
- **Manual tracing is not a practical option** for a continuous ultrasound image sequence (5–10 sec long sentence, $30 - 100 fps$ recording, $2 - 5$ sec/tracing $\rightarrow$ $5 - 80$ min/sentence)

Previous studies of variability in manual tracing / automatic tongue tracking

- Variability in manual tongue contours traced by pairs of experts
  - Maximum errors: $0.49 - 0.7\ mm$, mean absolute errors: $0.73 - 2.04\ mm$
- Variability in tongue contour trackings generated by computer algorithms
  - EdgeTrak [1] : mean absolute errors between 0.54 mm and 1.06 mm
  - TongueTrack [2] : mean absolute errors between 2 and 4 mm
  - AutoTrace [3] : mean absolute error as 5.656 $pixels$ (estimation: around 1.67 mm)

Goals of the current paper

- Examine the **variability of tongue contours traced manually** by several individuals
- Characterize & quantify the **major errors of automatic tongue contour trackings**

2. General Methods

Subjects and recordings

- Recordings from **two female and two male adult subjects** (denoted F1, F2, M1, M2)
- Producing the sentence *'I owe you a yoyo'* twice
- **Philips EpiQ-7G ultrasound system with an xMatrix 6–1 MHz transducer**
- Image frame rates: $42 - 44 fps$; speaking rate: $2.53 - 4.65$ syllables per second
- Recorded in DICOM format (800×600 resolution) and converted to JPG images using Image-J
- Altogether 1,145 ultrasound images (389, 275, 241 and 240 for speakers F1, F2, M1, and M2)

3. Experiment 1: Manual tracing

Methods

- Manual tracings using a custom website, access to the images via a browser (see Fig. 2, left)
- Seven tracers: two authors + five undergraduate students
  - Training with feedback before starting to trace the images
  - Cumulative time required for each individual to trace the 1,145 images: $3 - 5$ hours
- **Radial coordinate system** to quantify variability (see Fig. 2, right)
  - Origin: the point of intersection between the lines defining the sides of the ultrasound wedge
  - Total of 41 radial lines, spanning -60 to 60 degrees (relative to vertical) in steps of 3 degrees
  - Contours up-sampled by linear interpolation, and then down-sampled to 41 points

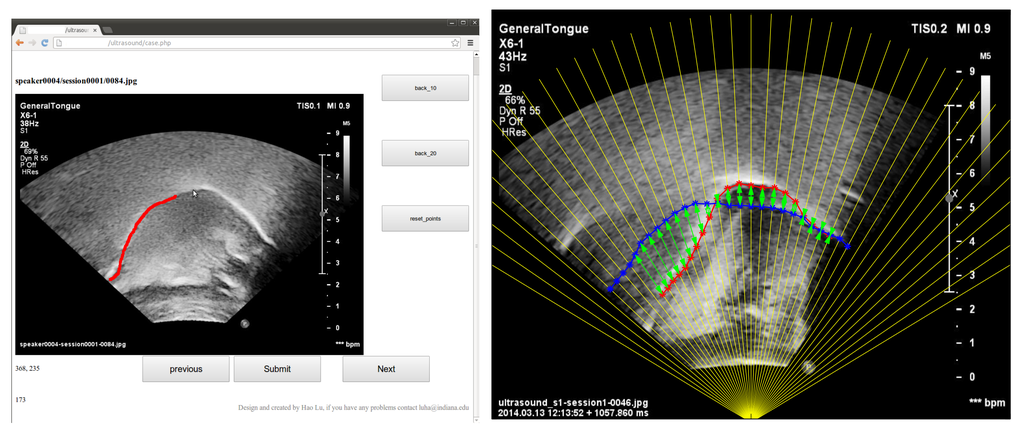

Fig. 2: Website for manual tracing (left) and sampled tongue tracings (right). Red: manual, blue: automatic, green: differences.

- Quantify the uncertainty associated with the manual tracings
  - Mean and unbiased standard deviation of values along each radial line for each frame
  - Grand mean and standard deviation calculated across all radial lines and across all frames

Results

- Distribution of standard deviations across radial lines and frames for each of the image sequences was skewed and roughly log-normal (data not shown)
- Grand mean and grand standard deviation were calculated from the log-transformed distributions and then transformed back to millimeter units
- Grand mean and standard deviation of the unbiased radial line standard deviations are:

Table 1: Grand mean and standard deviations of unbiased standard deviations (manual) and RMSEs (automatic) of tongue contour tracings (in mm).

tracer	F1	F2	M1	M2	avg
Manual	0.95 (0.29)	1.09 (0.32)	1.17 (0.31)	2.11 (0.32)	1.33 (0.31)
AutoTrace3.5	1.15 (0.35)	1.93 (0.31)	1.78 (0.29)	2.19 (0.28)	1.76 (0.31)
AutoTrace3	5.85 (0.33)	7.06 (0.43)	5.59 (0.32)	9.94 (0.28)	7.11 (0.34)
EdgeTrak	1.95 (0.45)	3.46 (0.37)	1.89 (0.41)	5.15 (0.40)	3.11 (0.41)
TongueTrack	1.96 (0.53)	3.15 (0.37)	2.76 (0.38)	3.60 (0.37)	2.87 (0.41)
Baseline	3.59 (0.40)	4.32 (0.33)	4.50 (0.33)	4.01 (0.37)	4.11 (0.36)

Fig. 3: Sample for the average (middle green line) and standard deviation (outside yellow lines) of manual tongue tracings.

- **Overall unbiased standard deviation:** 1.33 mm (speaker F1: 0.95 mm – M2: 2.11 mm)
- These standard deviations for seven tracers with varying levels of experience are very similar to those reported previously for pairs of expert tracers

4. Experiment 2: Automatic tracking

Methods and Analyses

- **EdgeTrak**: [1], uses a snakes-based algorithm; currently the *de facto* software for tongue tracking in ultrasound sequences
  - Initialized by providing the mean manual tracing of the first image, and the RoI was manually determined for each sequence
- **TongueTrack**: [2], uses a machine learning approach in combination with spatiotemporal constraint optimization
  - Anisotropic and despeckle filters were applied to the original 800×600 pixel JPG images and converted to MHD format
- **AutoTrace**: [3], uses deep belief networks (DBNs) that rely on prior tongue contour tracings for training
  - **'AutoTrace3.5'**, training set includes data from the test speaker $\rightarrow$ closely matched training and test data
  - **'AutoTrace3'**, training set does not include data from the test speaker $\rightarrow$ mismatched training and test matched data
- **Baseline**: by guessing that the tongue contours in all images of a sequence were identical to the mean manual tracing of the first image
- **Default 'out of the box' parameters** adopted for all programs; the results therefore do not necessarily represent optimal performances
- Resampled to radial lines (Fig. 2); absolute error (AE) along each radial; root mean square error (RMSE) across the log-transformed AEs

Results

- **AutoTrace3.5, EdgeTrak, and TongueTrack have mean RMSE values smaller than the baseline** (see Table 1)
- AEs were examined qualitatively in a series of graphs plotting AEs as a function of frame and radial line (see Fig. 4)
  - **Error type 1**: the periodic pattern in the baseline is caused by tongue movement; this is also reflected in AutoTrace3 and TongueTrack
  - **Error type 2**: extreme back/front of the tongue tracked poorly, other parts tracked well (AutoTrace3, speaker M2, angle < 20)
  - **Error type 3**: tracings frequently include only a smaller region of the tongue (typical of AutoTrace3.5 and AutoTrace3, speaker F2)

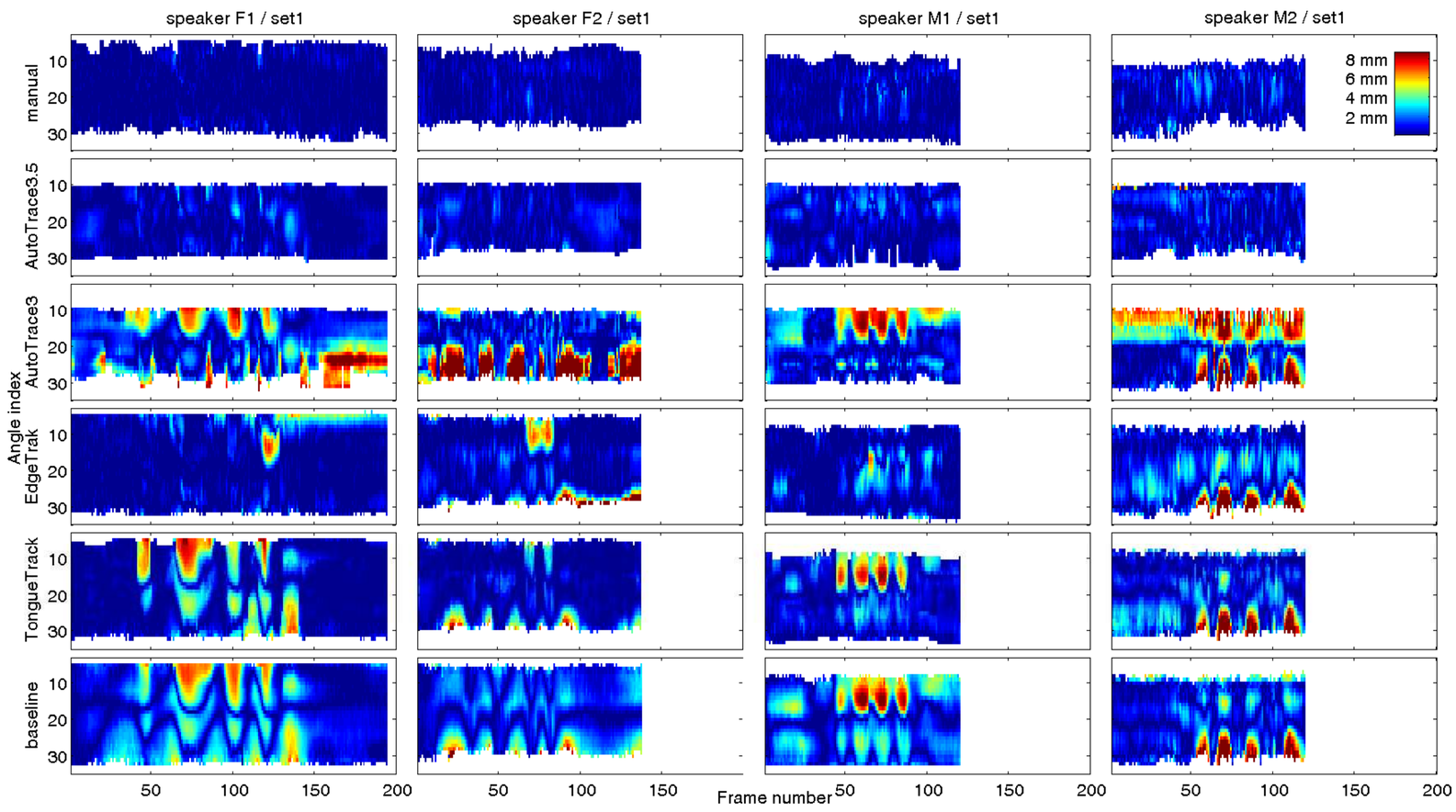

Fig. 4: Error maps of a manual tongue contour tracing (row 1) and of automatic tongue contour trackings (rows 2–5). Color scale represents Absolute Error (hotter color corresponds to larger AEs).

5. Discussion

- Manual tracings:
  - Previous errors (pairs of experts) 0.49 – 2.04 mm
  - Current mean errors (seven tracers) 0.95 – 2.11 mm
  - **Degree of variability: likely more sensitive to image quality than to expertise of tracer**
  - Mean tracings were accepted as the 'gold standard'
- Automatic trackings:
  - **Best performance: by AutoTrace3.5**, for which training and test sets were highly matched
  - Worst performance: by AutoTrace3, for which training and test sets were mismatched
  - EdgeTrak and TongueTrack sometimes returned errors approaching the magnitude of the baseline

6. Conclusions

- Expertise is likely to have a secondary influence on manual tracing accuracy, while image quality has a primary influence
- Manual tracings typically are in good agreement and close to the mean
- Automatic tracings can achieve very good accuracy under appropriate conditions
- Errors could be further decreased by optimizing the parameters (vs. 'out of the box' parameters)
- Results might be useful for articulatory-acoustic investigations and for analysis of 3D tongue shape

References

[1] M. Li, C. Kambhamettu, and M. Stone, "Automatic contour tracking in ultrasound images," *Clinical Linguistics & Phonetics*, vol. 19, no. 6-7, pp. 545–554, Jan. 2005.

[2] L. Tang, T. Bressmann, and G. Hamarneh, "Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves," *Medical Image Analysis*, vol. 16, no. 8, pp. 1503–1520, Dec. 2012.

[3] G. V. Hahn-powell, D. Archangeli, J. Berry, and I. Fasel, "AutoTrace: An automatic system for tracing tongue contours," *The Journal of the Acoustical Society of America*, vol. 136, no. 4, p. 2104, Oct. 2014.

Acknowledgements

The first author was supported by a Fulbright scholarship and by the travel grant of the Hungarian Academy of Engineering.