Investigating tongue movement during speech with ultrasound

Tamás Gábor CSAPÓ

csapot@tmit.bme.hu

Meet the Hungarian Fulbrighters of AY 2013-2014

April 28, 2015
1 Introduction
   - Indiana University Bloomington
   - Speech research with ultrasound
   - Goals of my scholarship

2 Research methods and results
   - Ultrasound recordings
   - Manual tongue contour tracing
   - Automatic tongue contour tracking

3 Personal experiences
   - Playgrounds and pictures

4 Summary and future plans
Bloomington, IN
Investigating tongue movement with ultrasound
Indiana University, Bloomington III

Investigating tongue movement with ultrasound
Indiana University

- Jacobs School of Music
- Kelley School of Business
- Dept. of Speech and Hearing Sciences
- you can even learn Hungarian!

Speech Production Laboratory

- professor: Dr. Steven M. Lulich
- brand new equipment for speech research
Ultrasound (US)
  - used in speech research since early ’80s
  - US transducer positioned below the chin during speech
  - record video of tongue movement
  - series of gray-scale images
  - tongue surface has a greater brightness than the surrounding tissue and air

[Stone et al., 1983, Stone, 2005]
Speech research with ultrasound II

Investigating tongue movement with ultrasound
Speech research with ultrasound III

Vocal tract

- soft palate
- nasal cavity
- hard palate
- teeth
- lips
- vocal folds

Ultrasound sample

- tongue

[Németh and Olaszy, 2010]

(听起来)
Speech research with ultrasound IV

Phonetic research examples

- reconstruct tongue shape during sustained vowels
- investigate speech sounds of under-researched languages
- compare articulatory characteristics of vowels
- analyze tongue shapes for clinical purposes

First step is always the tongue contour tracking!

My goals

This scholarship

- compare manual tongue tracings of several individuals
- compare automatic tongue contour extraction programs
- use 2D ultrasound at high frame rate

Long-term

- extend text-to-speech with tongue contour data based on ultrasound
- use real-time 3D ultrasound
Methods

Subjects

- two female and two male
- 3 speakers of American English
- 1 speaker of Hungarian

Speech material

- ‘I owe you a yo-yo.’ sentence two times
- 135 various English sentences
- 210 various Hungarian sentences
Recordings

Location
- Speech Production Lab, IU

Parallel recordings
- speech signal with a microphone
- video of the lips with a webcam
- video of the tongue with an ultrasound device (Philips EpiQ-7G, xMatrix 6-1 MHz)
Manual tracings

Ultrasound recordings
- JPG image sequence
- 800x600 pixels resolution

Tracers
- 7 individuals (2 professors and 5 students)
- drag a computer mouse cursor from the root of the tongue (left) to the tip of the tongue (right)
- about 150–200 points per image
- about 5–10 seconds per image
Manual tracing website

Investigating tongue movement with ultrasound
Automatic tongue contour tracking algorithms

5 freely available programs, baseline settings

- **AutoTrace** (University of Arizona, USA)
- **EdgeTrak** (University of Maryland, USA)
- **Palatoglossotron** (North Carolina State University, USA)
- **TongueTrack** (Simon Fraser University, Canada)
- **Ultra-CATS** (University of Toronto, Canada)

[Sung et al., 2013, Li et al., 2005, Baker et al., 2005, Tang et al., 2012, Bressmann et al., 2005]
Comparison of two tongue contours

Investigating tongue movement with ultrasound
Automatic trackings

RMSE (Root Mean Squared Error)

difference from mean of manual tracing

- Average for the best algorithm, AutoTrace: 9.66 pixel (1.93 mm)
- depending on the speaker, algorithm and image
- (compare with: 7.11 pixel inter-tracer variability)

US video samples

- speaker1 (click)
- speaker4 (click)
What can this be used for?

- investigate articulation during speech
- visual reconstruction of 3D tongue surface
- audiovisual speech synthesis
- language education: how to produce unfamiliar speech sounds?
- speech rehabilitation: learn to speak after a tongue surgery

[Stone et al., 2005]
Playgrounds I
Biking in Bloomington I
4th of July with the Lulich family
Roundtrip I

Investigating tongue movement with ultrasound
Roundtrip II
Roundtrip III

The sign reads:

Montgomery County Rotary Jail


(Continued on other side)
Summary I

This study

- ultrasound recordings with several speakers
- compared manual tongue tracings
- compared automatic tongue contour extraction programs

Future plans

- extend Hungarian / English Text-To-Speech with tongue contour data
- use 2D / real-time 3D ultrasound
Summary II

Presentations and papers during the scholarship

- T. G. Csapó, S. M. Lulich, „Comparison of tongue contour extraction methods from ultrasound images for use in TTS”, Conf. of HCA, Bloomington, IN, USA, April 6, 2014.

- TGCs, SML, „Comparison of tongue contour extraction methods”, virtual presentation at the lab meeting of University of Arizona, May 13, 2014.

- TGCs, SML, „Tongue contour tracings from 2D ultrasound image sequences: quantification of measurement error using manual and automatic tracing methods”, in preparation, 2014.
Presentations and papers after the scholarship


- TGCs, SML, "Error analysis of extracted tongue contours from 2D ultrasound images", submitted to Interspeech 2015.
Grants

- Bolyai post-doc grant, Modeling articulation using ultrasound, with special regard to text-to-speech synthesis (submitted).
- OTKA-NSF International collaboration grant (planned).
Acknowledgements

Support from
- Fulbright Hungary
- Hungarian Academy of Engineering

Thank you for your attention!

http://csapobloomington.blogspot.hu/
References I

Tracing the tongue with GLoSSatron.
In Ultrafest III, Tucson, AZ, USA.

Articulatory characteristics of Hungarian ‘transparent’ vowels.

Applications of 2D and 3D ultrasound imaging in speech-language pathology.

Automatic contour tracking in ultrasound images.

Articulation of the Kagayanen interdental approximant: An ultrasound study.

Németh, G. and Olaszy, G., editors (2010).
A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek.
Akadémiai Kiadó, Budapest.
A guide to analysing tongue motion from ultrasound images.

Predicting 3D Tongue Shapes from Midsagittal Contours.

Three-dimensional tongue surface shapes of English consonants and vowels.

Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system.

In *Ultrafest VI*, pages 9–10, Edinburgh, UK.

Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves.
*Medical Image Analysis*, 16(8):1503–1520.