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

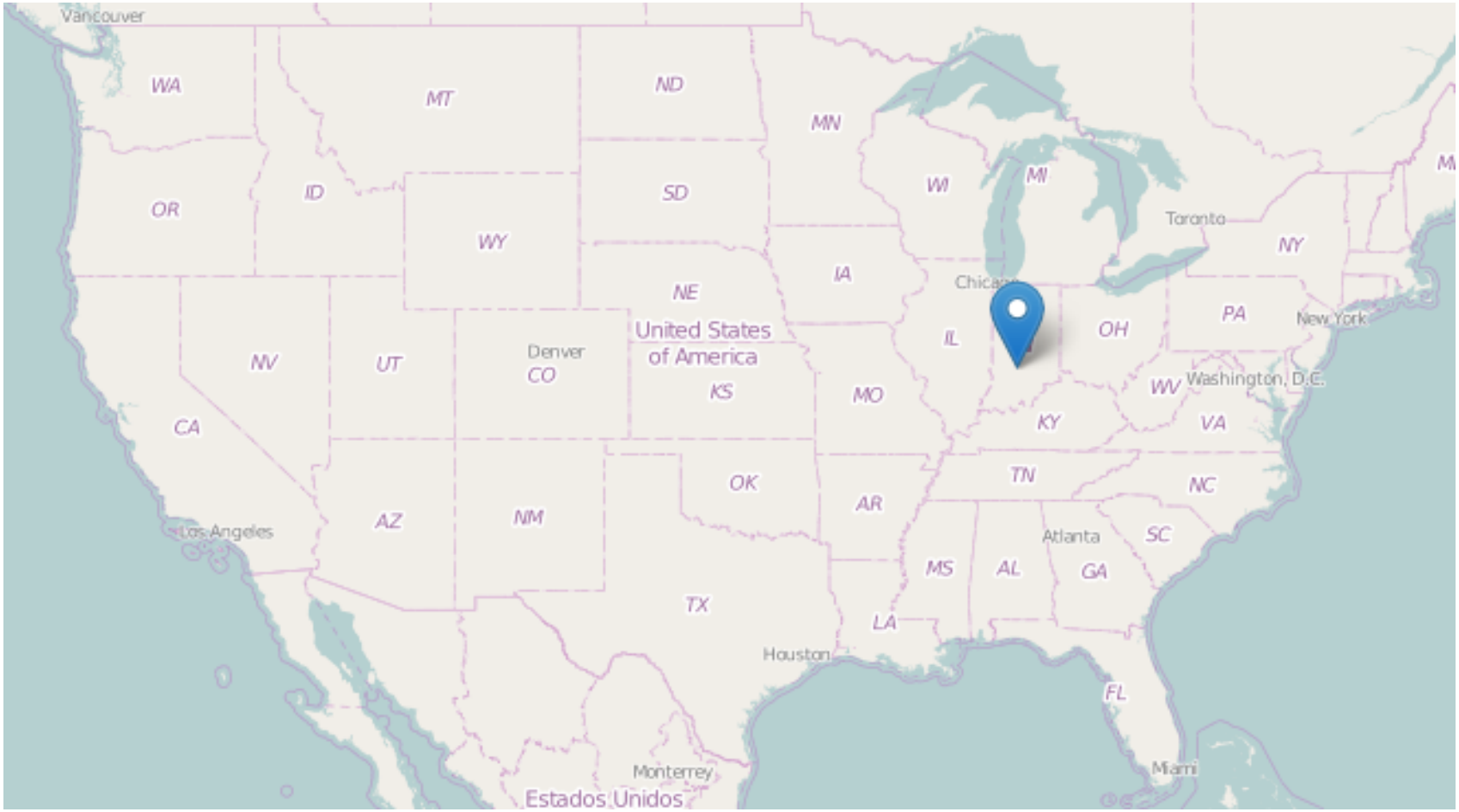
MŰEGYETEM 1782

INDIANA UNIVERSITY
BLOOMINGTON

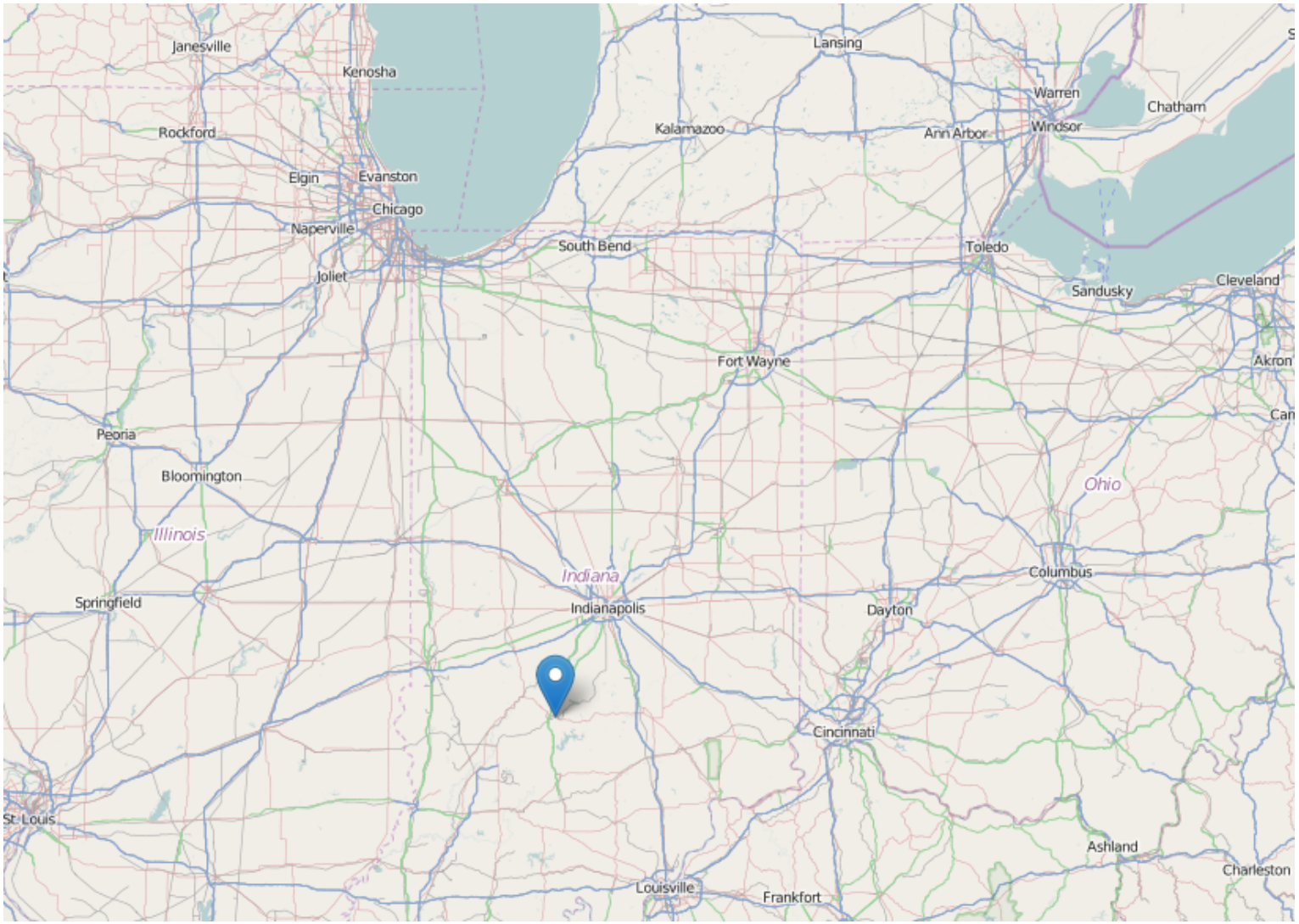
FULBRIGHT

Indiana University, Bloomington I

Bloomington, IN



Indiana University, Bloomington II



Indiana University, Bloomington III

Indiana University, Bloomington IV

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

Speech research with ultrasound I

Ultrasound (US)

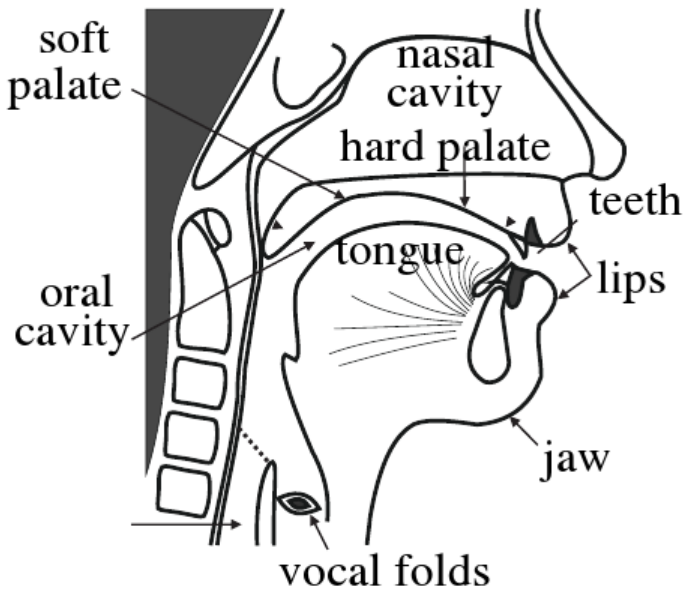
- used in speech research since early ’80s
- US transducer positioned below the chin during speech
- record video of tongue movement
- series of gray-scale images
- tongue surface has a greater brightness than the surrounding tissue and air

[Stone et al., 1983, Stone, 2005]

Speech research with ultrasound II

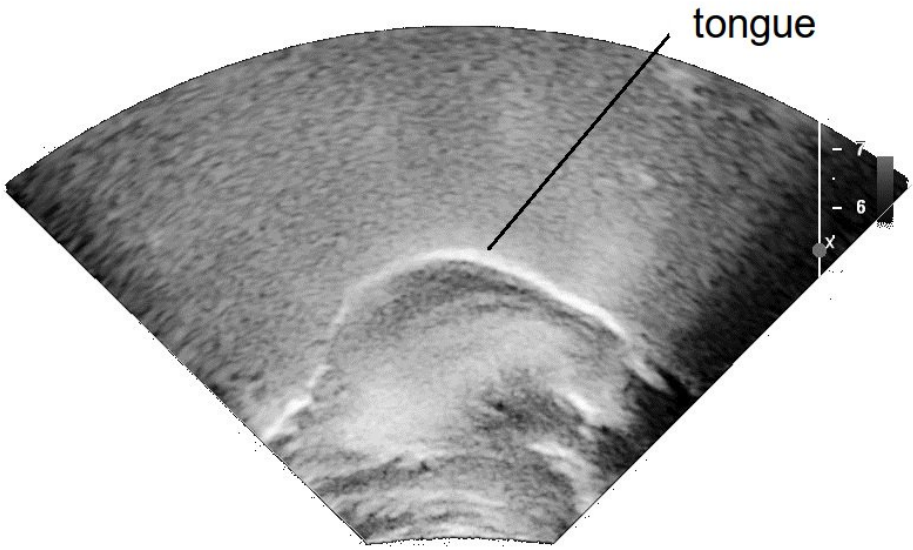
Speech research with ultrasound III

Vocal tract



[Németh and Olaszy, 2010]

Ultrasound sample



(click)

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!

[Stone and Lundberg, 1996, Mielke et al., 2011, Benus and Gafos, 2007, Zharkova, 2013]

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

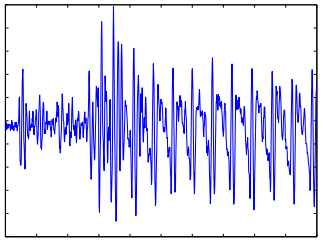
Recordings

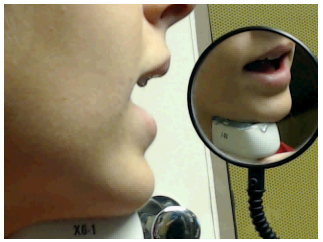
Location

- Speech Production Lab, IU

Parallel recordings

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





Recording setup

Manual tracings

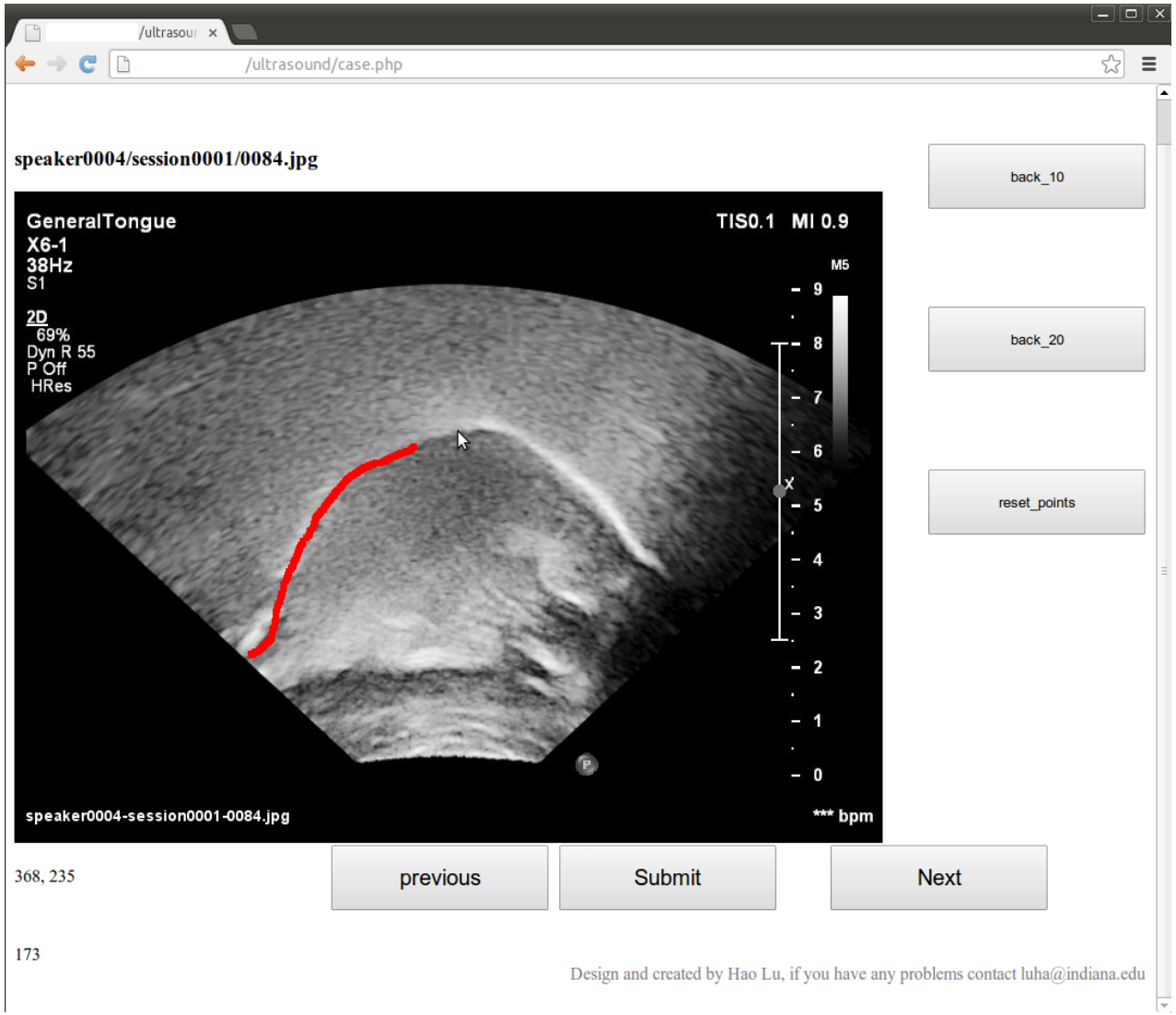
Ultrasound recordings

- 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



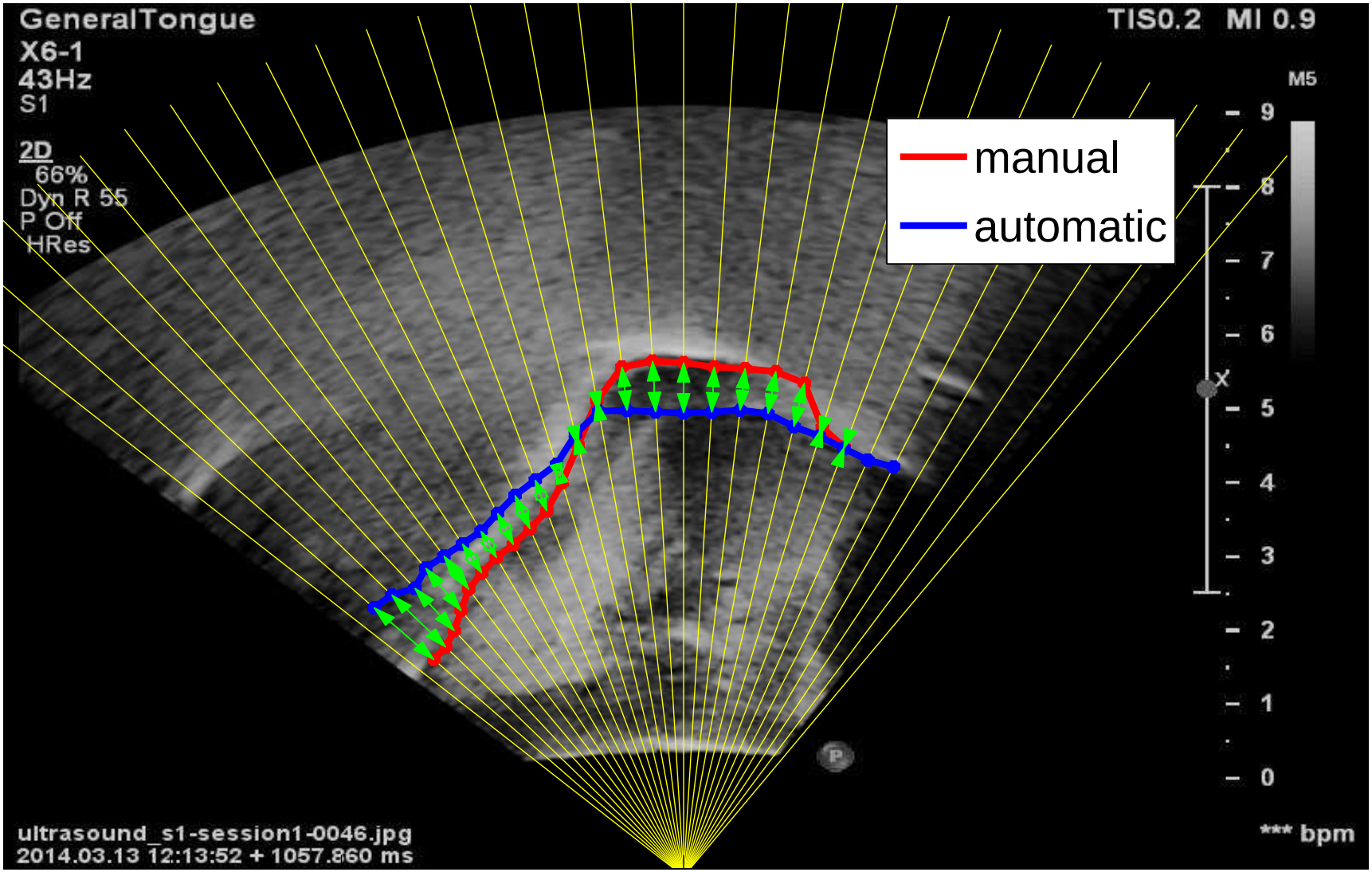
Automatic tongue contour tracking algorithms

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



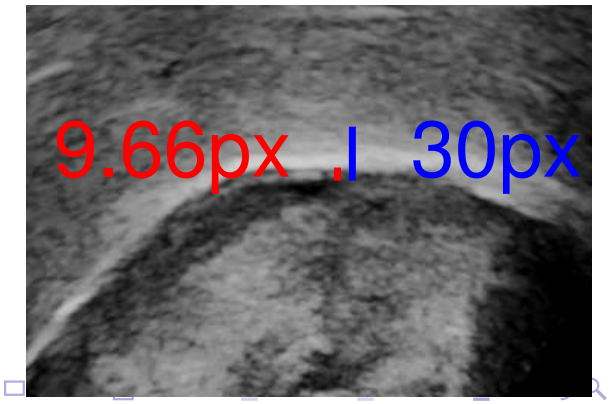
Automatic trackings

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)

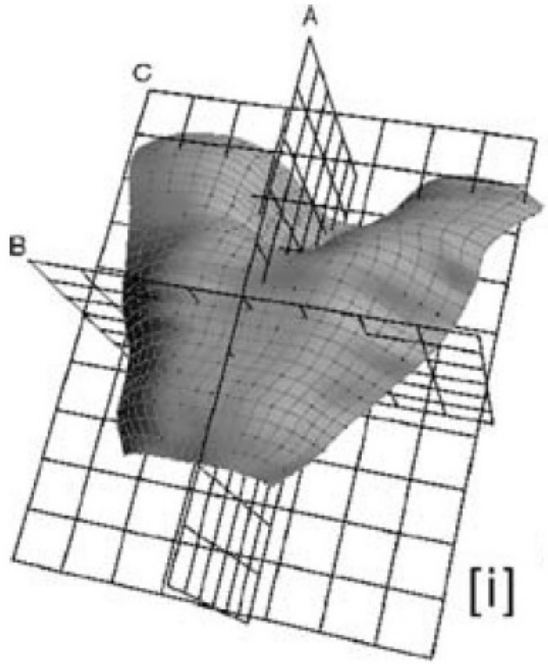
US video samples

- 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



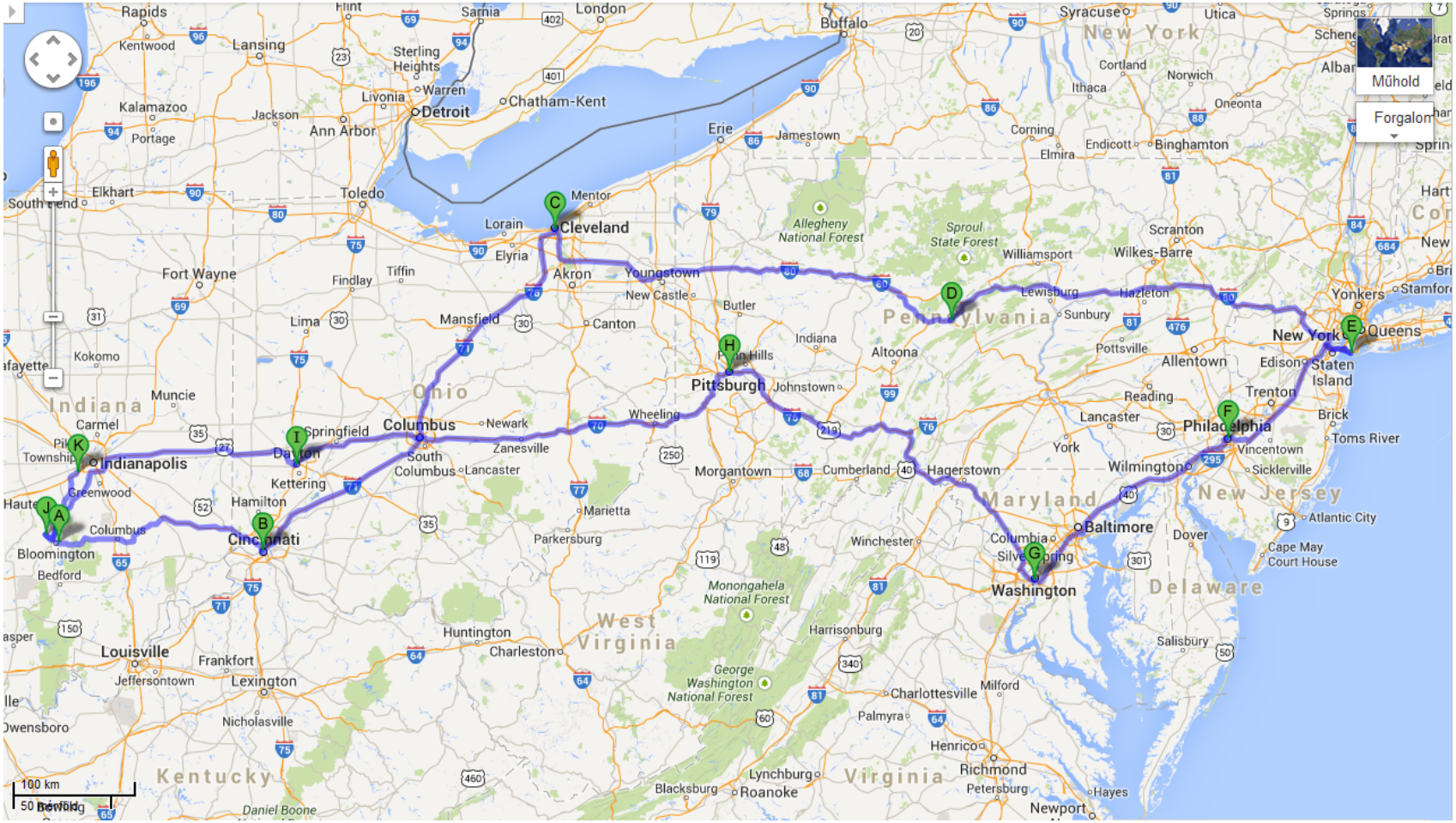
[Stone et al., 2005]

Playgrounds I

Biking in Bloomington I

4th of July with the Lulich family I

Roundtrip I


Cleveland
Pittsburgh
Columbus
Indianapolis
Cincinnati
Washington
Baltimore
Philadelphia
New York
Buffalo
Detroit
Toledo
Akron
Youngstown
Erie
Scranton
Williamsport
Harrisburg
Charleston
Louisville
Lexington
Richmond
Virginia
Pennsylvania
Ohio
Indiana
Kentucky
West Virginia
Maryland
Delaware
New Jersey
Műhold
Forgalom
100 km
50 mi

Roundtrip II

Roundtrip III

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.

Summary III

Presentations and papers after the scholarship

- R. Pedro, E. Mazzocco, TGCs, SML, „**Investigation of a tongue-internal coordinate system for two-dimensional ultrasound**”, 168th Meeting of the Acoustical Society of America, Indianapolis, IN, USA, Oct 27-31, 2014.
- D. Csopor, „**Mély neuronhálók alkalmazása ultrahangos nyelvkontúr követésre**”, supervised by TGCs, Scientific Students’ Association Annual Conference of BME VIK, Budapest, Nov 11, 2014.
- TGCs, D. Csopor, „**Ultrahangos nyelvkontúr követés automatikusan: a mély neuronhálókon alapuló AutoTrace eljárás vizsgálata**”, Beszédkutatás 2015, pp. 177-187, 2015.
- TGCs, SML, „**Error analysis of extracted tongue contours from 2D ultrasound images**”, submitted to Interspeech 2015.

Summary IV

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

Baker, A., Mielke, J., and Archangeli, D. (2005).
Tracing the tongue with GLoSsatron.
In *Ultrafest III*, Tucson, AZ, USA.

Benus, S. and Gafos, A. I. (2007).
Articulatory characteristics of Hungarian 'transparent' vowels.
Journal of Phonetics, 35(3):271–300.

Bressmann, T., Heng, C.-L., and Irish, J. C. (2005).
Applications of 2D and 3D ultrasound imaging in speech-language pathology.
Journal of Speech-Language Pathology and Audiology, 29(4):158–168.

Li, M., Kambhamettu, C., and Stone, M. (2005).
Automatic contour tracking in ultrasound images.
Clinical Linguistics & Phonetics, 19(6-7):545–554.

Mielke, J., Olson, K. S., Baker, A., and Archangeli, D. (2011).
Articulation of the Kagayanen interdental approximant: An ultrasound study.
Journal of Phonetics, 39(3):403–412.

Németh, G. and Olaszy, G., editors (2010).
A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek.
Akadémiai Kiadó, Budapest.

References II

Stone, M. (2005).
A guide to analysing tongue motion from ultrasound images.
Clinical Linguistics & Phonetics, 19(6-7):455–501.

Stone, M., Epstein, M. A., Li, M. I. N., and Kambhamettu, C. (2005).
Predicting 3D Tongue Shapes from Midsagittal Contours.
In Tabain, J. and M., editors, *Speech Production: Models, Phonetic Processes, and Techniques*, chapter 18, pages 315–331. Psychology Press.

Stone, M. and Lundberg, A. (1996).
Three-dimensional tongue surface shapes of English consonants and vowels.
The Journal of the Acoustical Society of America, 99(6):3728–37.

Stone, M., Sonies, B., Shawker, T., Weiss, G., and Nadel, L. (1983).
Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system.
Journal of Phonetics, 11:207–218.

Sung, J.-H., Berry, J., Cooper, M., Hahn-Powell, G., and Archangeli, D. (2013).
Testing AutoTrace: A Machine-learning Approach to Automated Tongue Contour Data Extraction.
In *Ultrafest VI*, pages 9–10, Edinburgh, UK.

Tang, L., Bressmann, T., and Hamarneh, G. (2012).
Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves.
Medical Image Analysis, 16(8):1503–1520.

References III

Zharkova, N. (2013).
A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes.
Clinical Linguistics & Phonetics, 27(6-7):484–96.