

# VoiceSauce:

A program

for voice analysis

# Yen-Liang Shue

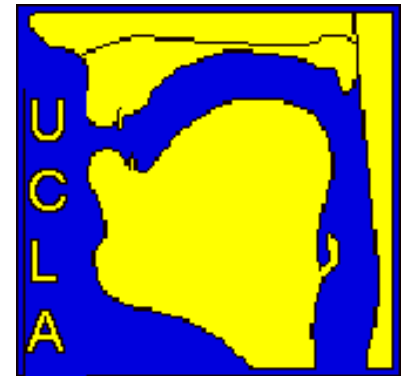
Speech Processing and Auditory Perception Laboratory  
Department of Electrical Engineering, UCLA

yshue@ee.ucla.edu



# Patricia Keating Chad Vicenik

Phonetics Lab,  
Linguistics, UCLA



# Overview

- **VoiceSauce** is a new application, currently implemented in Matlab, which provides **automated voice measurements over time** from audio recordings.
- VoiceSauce computes **many voice measures**, including **corrections** for formant frequencies and bandwidths.
- It outputs values as **text** or for **Emu** database.
- VoiceSauce is available **free** by downloading.

# VoiceSauce algorithms: F0 estimation

- First the **STRAIGHT** algorithm (Kawahara et al. 1998) is used to find F0 at 1 ms intervals.
- The **Snack Sound Toolkit** (Sjölander 2004) can also be used to estimate F0 at variable intervals.
- Future versions will allow integration with other F0 estimators.

# VoiceSauce algorithms: Harmonic magnitudes

- Harmonic spectra magnitudes are computed **pitch-synchronously**, over a 3-cycle window. This eliminates much of the variability in spectra computed over a fixed time window.
- Harmonics are found using standard **optimization** techniques to find the maximum of the spectrum around the peak locations as estimated by F0. This enables a much more accurate measure without relying on large FFT calculations.

# VoiceSauce algorithms: Formants estimation

- The [Snack Sound Toolkit](#) (Sjölander 2004) is used to find the frequencies and bandwidths of the first four formants, using as defaults the covariance method, pre-emphasis of .96, window length of 25 ms, and frame shift of 1 ms (to match STRAIGHT).
- Future versions will allow integration with other formant estimators.

# VoiceSauce algorithms: Formant corrections

- In previous work, Iseli & Alwan (2000, 2004) developed an algorithm that estimates the voice source parameters  $F_0$ ,  $H1^*-H2^*$ , and  $H1^*-A3^*$ , where the asterisk denotes that the corresponding spectral magnitudes ( $H1$ ,  $H2$  and  $A3$ ) are **corrected for the effect of formants** (frequencies and bandwidths). Further developments have been reported by Iseli et al. (2006, 2007).

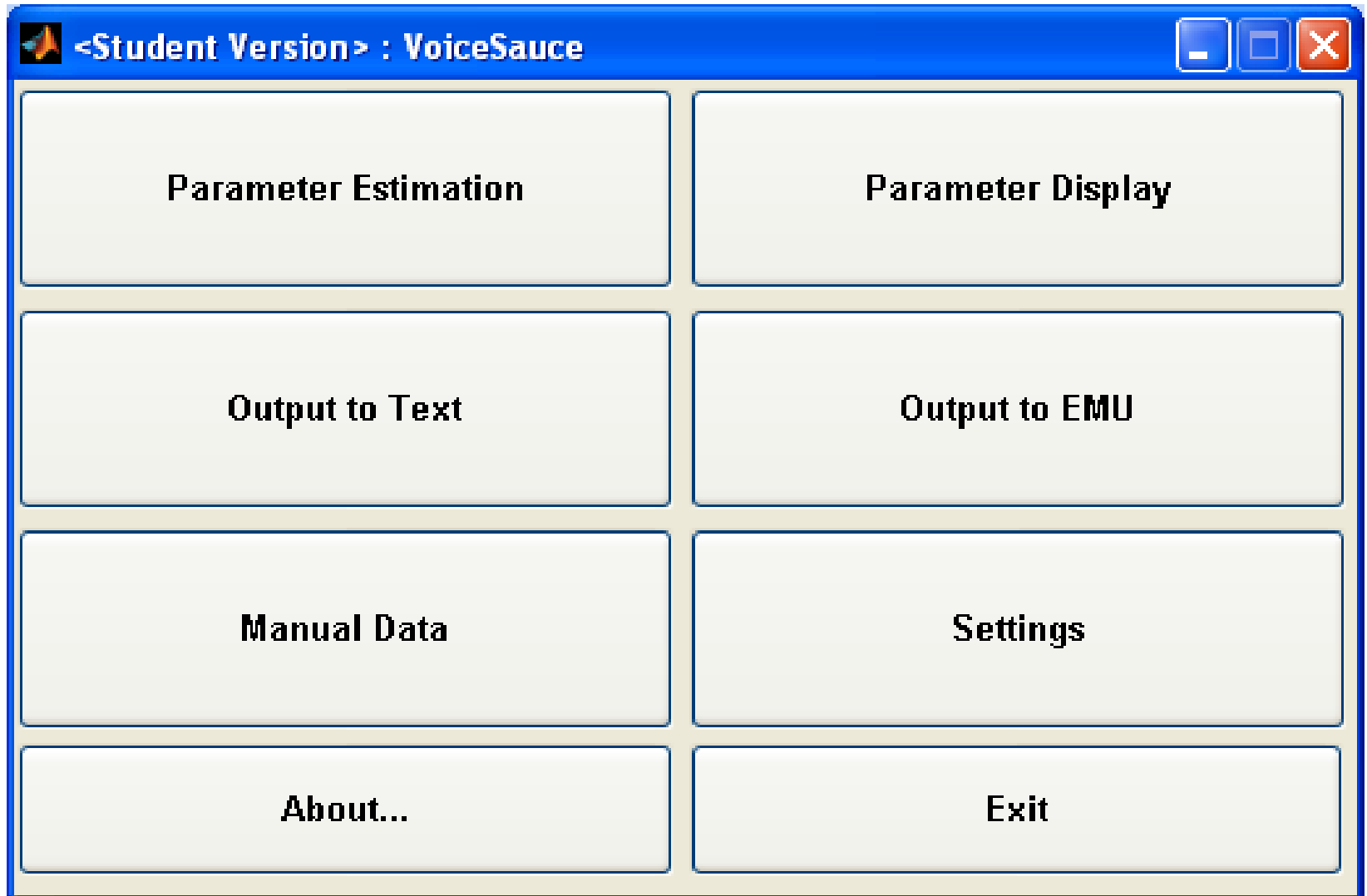


- In VoiceSauce, the harmonic amplitudes for all measures of spectral magnitude are corrected every frame using the Snack formant frequencies and bandwidths. (For H1\*-H2\*, only F1 and F2 are used in the correction; for e.g. H1\*-A3\*, F1 through F3 are used.)
- Finally, the measures are smoothed with a moving average filter with a default length of 20 samples.

# Variables computed

- F0 from Snack
- F0 from STRAIGHT
- F1-F4 and B1-B4 from Snack
- H1
- H2
- H4
- A1, A3, A3
- Cepstral Peak Prominence
- Energy
- H1-H2(\*)
- H1-A1(\*)
- H1-A2(\*)
- H1-A3(\*)
- H2-H4(\*)
- All harmonic measures come both corrected (\*) and uncorrected

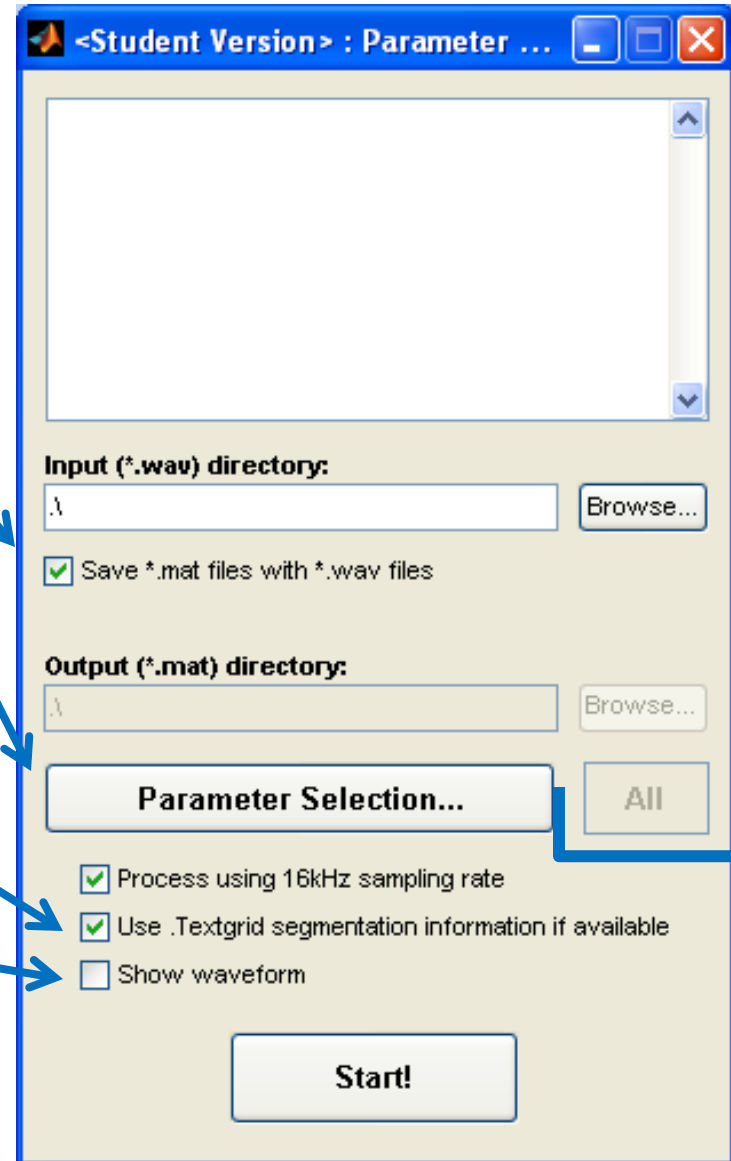
# VoiceSauce: User Interface



# Module: Parameter estimation

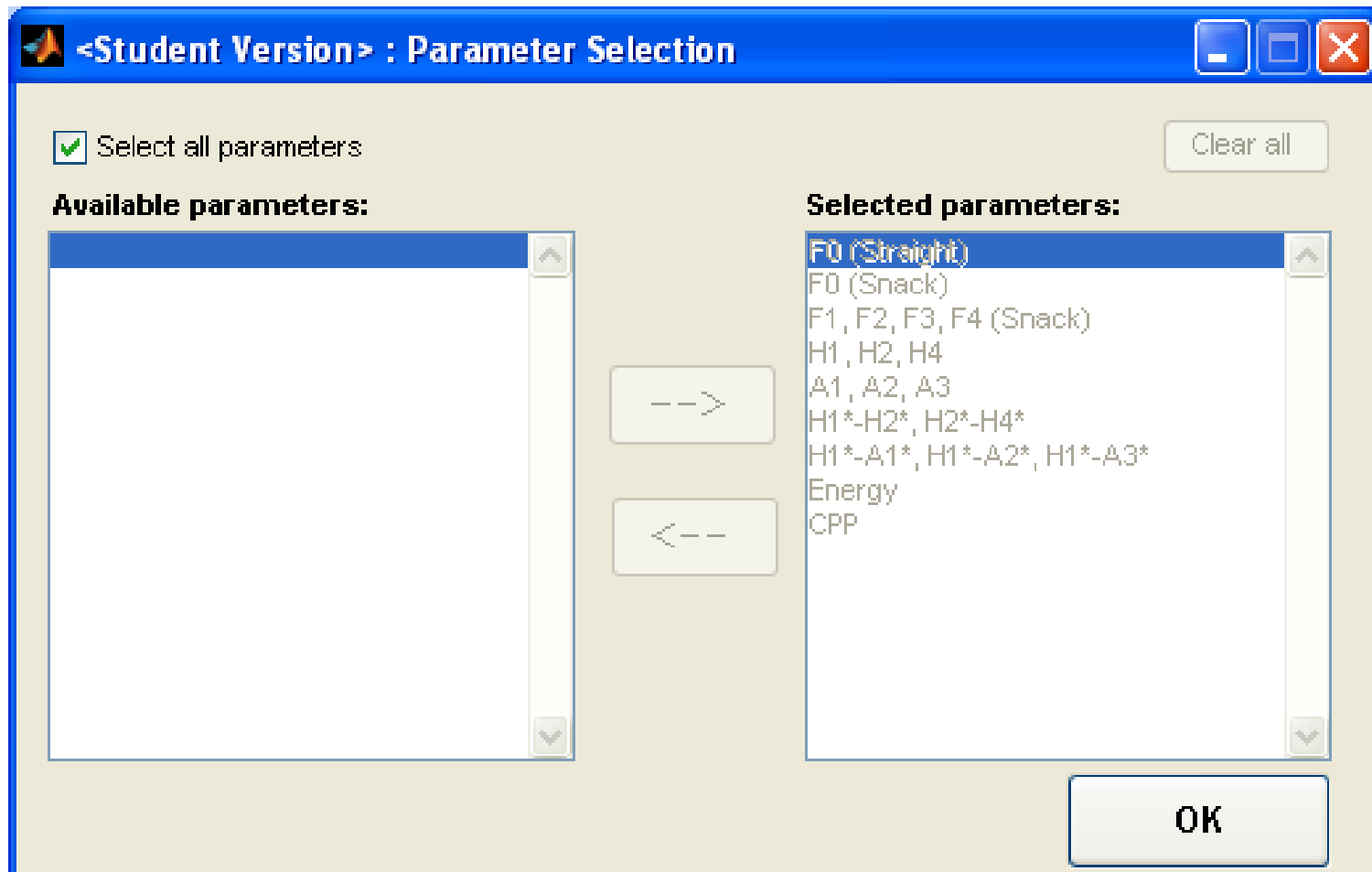
## User specifies:

- Where to find .wav files to analyze, and where to save .mat files of results
- Acoustic parameters to be calculated (next slide)
- whether to limit analysis to segments in Praat textgrid for each file (next slide)
- whether to display waveform of each file during its processing



# Parameter selection

Default setting is that all available parameters will be calculated:



# Using Praat textgrids in parameter estimation

- **Praat textgrids** (Boersma 2001) are used to delimit and label segments of interest.
- These can then be used to guide VoiceSauce acoustic **analysis** (only labeled segments are analyzed).
- They are also used to structure VoiceSauce **output**.

# Module: Settings

Change how parameters are calculated, how textgrids are used:

**<Student Version> : Settings**

Settings

**F0**

Used for parameter estimation:  Straight  Snack  Other

**Straight**

Max F0 (Hz): 500  
Min F0 (Hz): 40  
Max duration (s): 10

**Snack**

Max F0 (Hz): 500  
Min F0 (Hz): 40

**Other**

Enable  
Command:   
Offset (ms): 0

**Formants**

Used for parameter estimation:  Snack  Other

**Snack**

Pre-emphasis: 0.96

**Other**

Enable  
Command:   
Offset (ms): 0

**Common**

Window size (ms): 25  
Frame shift (ms): 1  
Not a number: 0

Recurse sub-directories  
 Link mat directories  
 Link wav directories

**Textgrid**

Ignore these labels: "", "", "SIL"  
Tier number: 1

**EGG Data**

Headers to search for: CQ, CQ\_H, CQ\_PM, CQ\_HT, peak\_Vel, peak\_  
Time label: Frame

OK

# Module: Output to text

- From Praat textgrids, VoiceSauce identifies all labeled segments.
- Writes out the results for those segments
  - all values (frame interval) or,
  - averages over N intervals
- User specifies which parameters to output.
- Output can be one giant text file with all parameters, or separate smaller text files with subsets of parameters.





Parameters and Settings

Parameters:

- F1 - Other (oF1)
- F2 - Other (oF2)
- F3 - Other (oF3)
- F4 - Other (oF4)
- B1 - Snack (sB1)
- B2 - Snack (sB2)
- B3 - Snack (sB3)
- B4 - Snack (sB4)
- B1 - Other (oB1)
- B2 - Other (oB2)
- B3 - Other (oB3)
- B4 - Other (oB4)

No. of parameters selected: 43

Input .mat directory:

Input .Textgrid directory:

Include EGG data

EGG data directory:

Output .txt directory:

Include subdirectories

mat files:

- 

Include Textgrid labels      Column delimiter:

Segments

No segments (write out all data)       Use segments

No. of segments:

Output Options

Single file       Multiple files

**Output file:**

**Output files:**

FO/CPP/E:

Formants:

Hx/Ax:

Hx-Hx:

Hx-Ax:

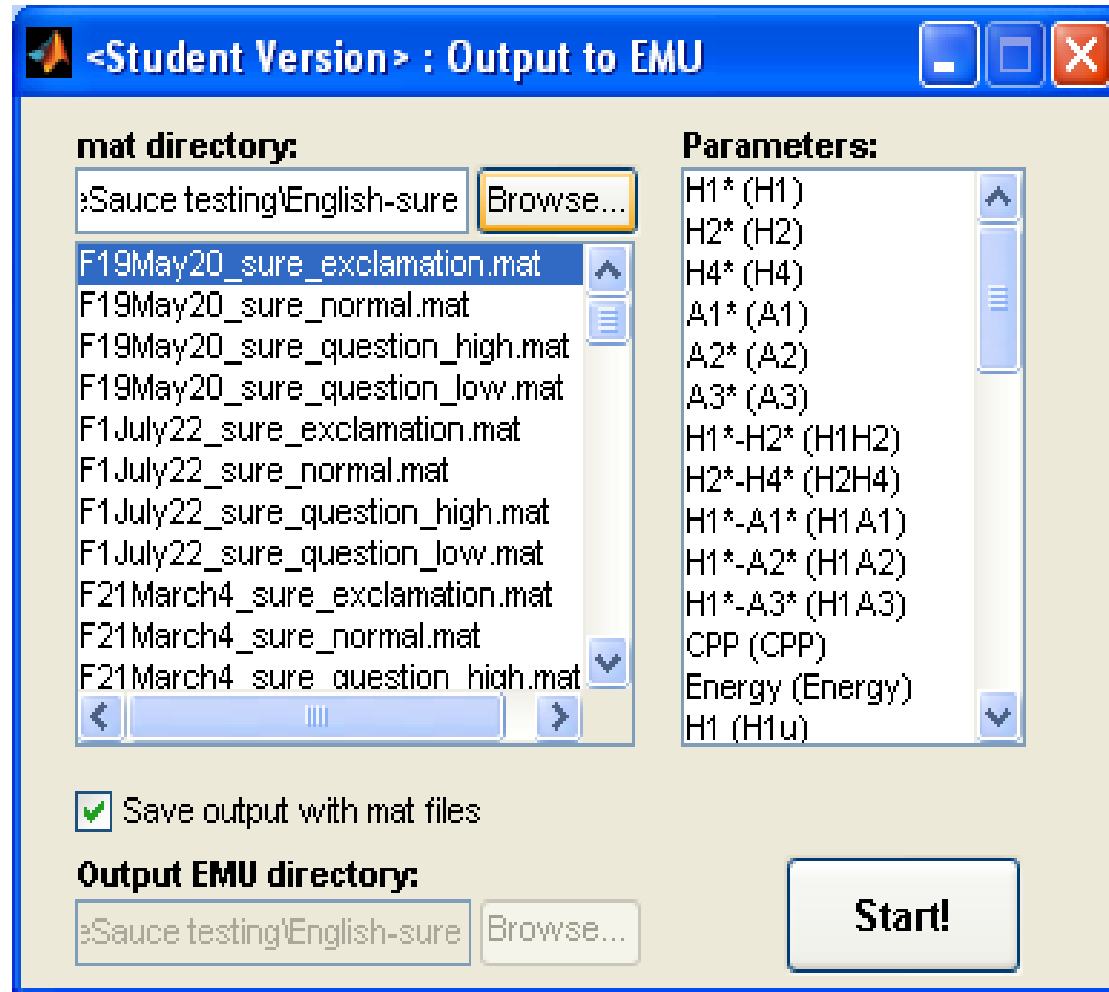
EGG:

# Including EGG data

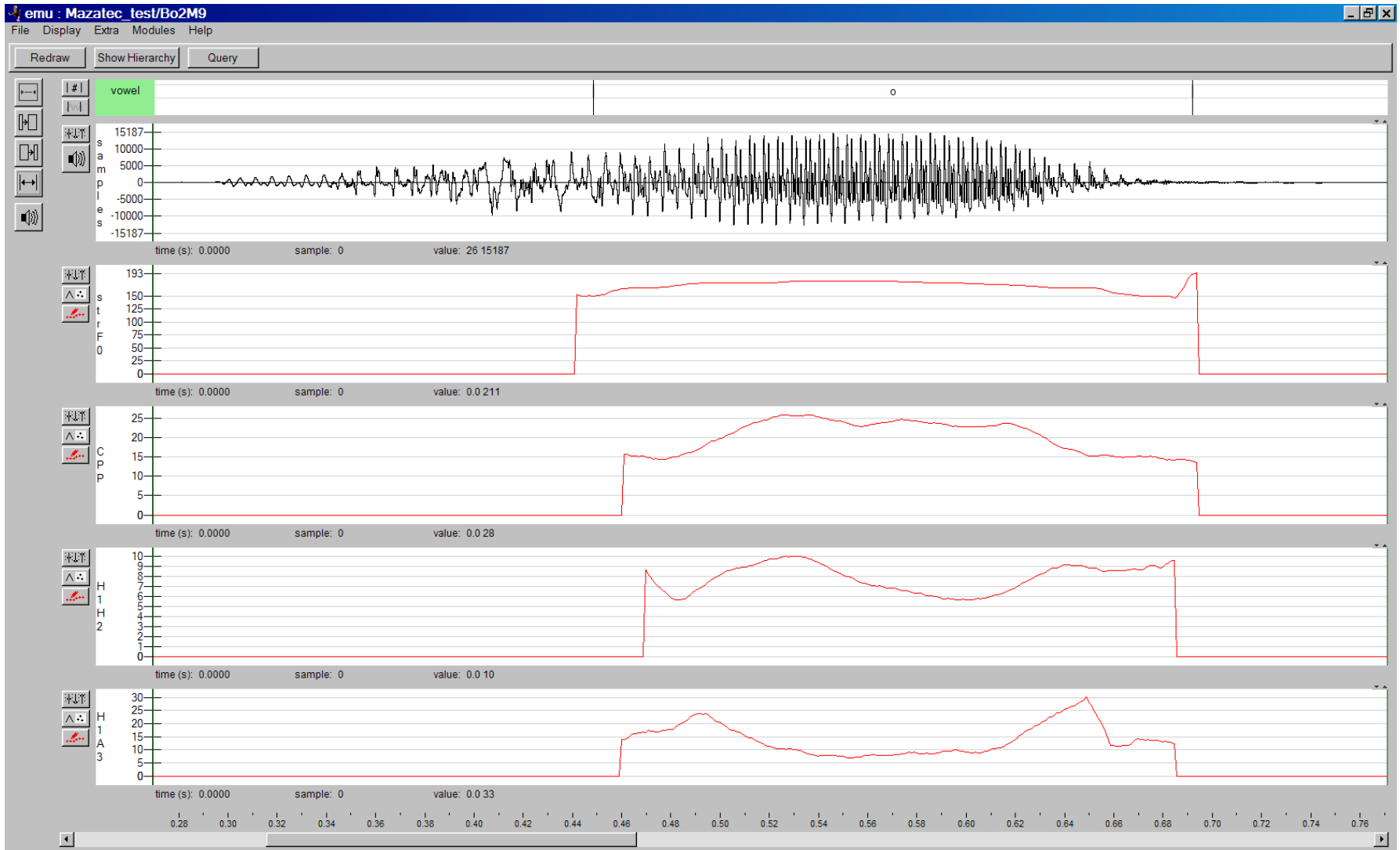
- **EGGWorks**, a free program by Henry Tehrani based on PCQuirerX, computes several **EGG** measures from .wav files of EGG channels, in batch mode.
- Its output file can be included as an input to VoiceSauce's output step.

# Module: Output to Emu

- For use in Emu speech databases (Harrington in press)
- Emu's trackdata files in SSFF format
- One track file per parameter per audio file as in Emu
- Can view, query, analyze in Emu, or in R using Emu library



# Sample display in Emu



# Comparing VoiceSauce to other methods

- Compare VoiceSauce's outputs to
  - By hand measurements, taken from FFT spectra (traditional method)
  - Praat (Boersma 2001)
- Same speech materials all three methods
  - Taken from Vicenik (2008)

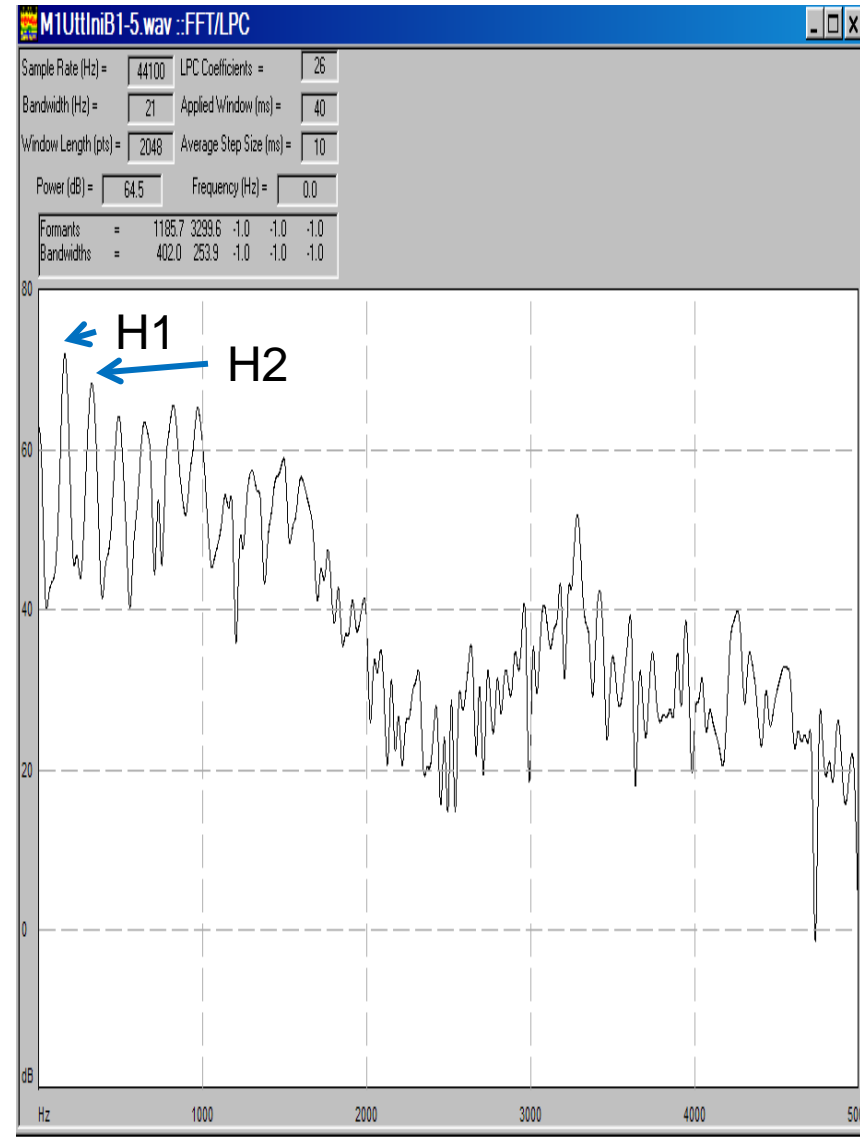
# Speech corpus

- Voice measures made for low vowel [a], after 9 Georgian stops
  - Three stop types - voiceless aspirated, voiced, ejective
  - Three places of articulation – bilabial, alveolar, velar
  - Measurements averaged over the first third of the vowel (Praat, VoiceSauce), or measured immediately after vowel onset (by hand)
- Five speakers – middle-aged women from Tblisi, Georgia

# H1-H2 by hand

## Measured in PCQuirer

- Created FFT spectrum with 21 Hz Bandwidth and 40 ms window
- Spectrum taken at vowel onset
- Manually marked and logged H1 and H2 using cursor
- Very slow
- Many files could not be analyzed



# H1-H2 by Praat

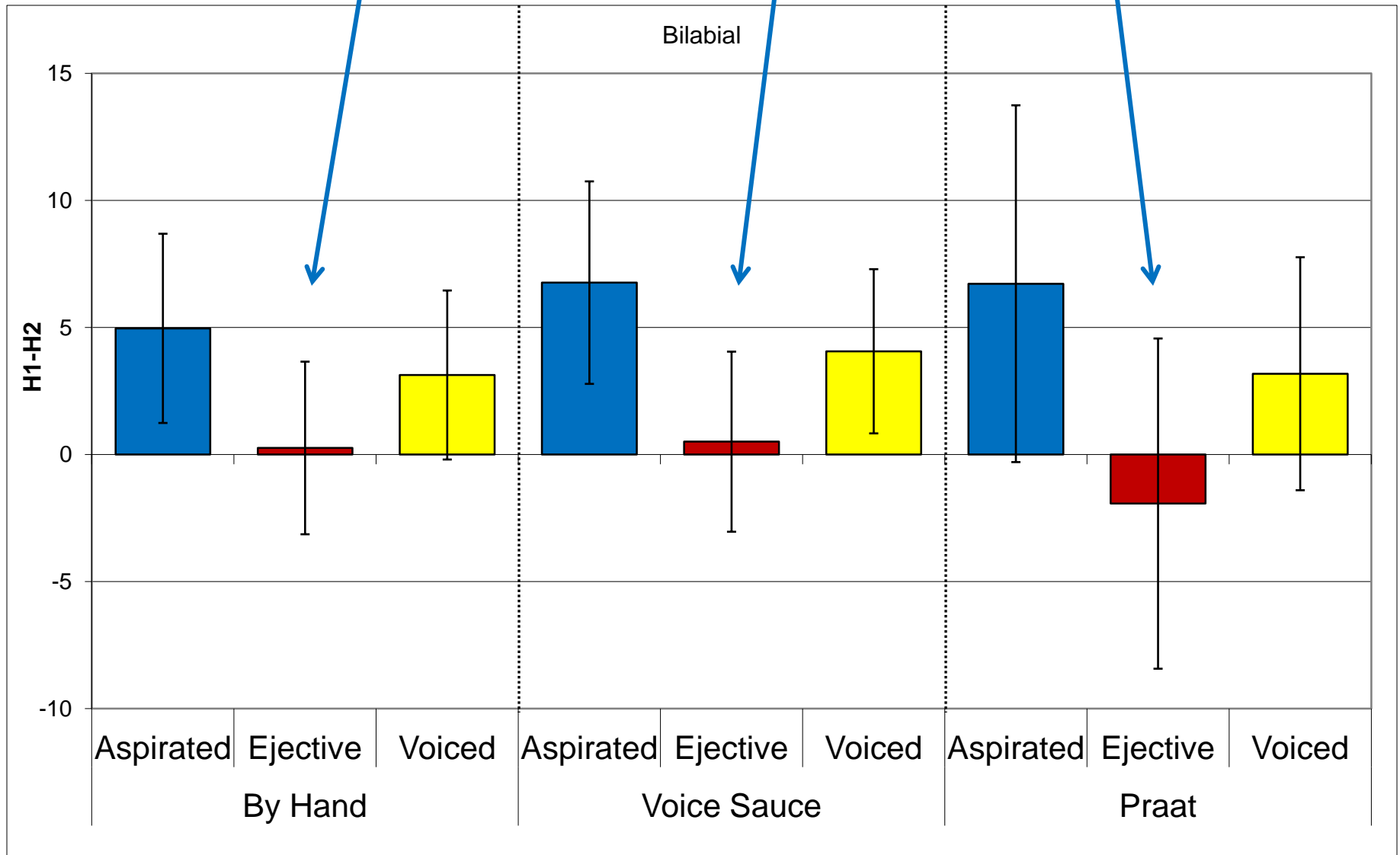
- Using a new **script** based on one by Bert Remijsen (his “msr&check\_f1f2\_indiv\_interv.psc”)
- Measures H1-H2, H1-A1, H1-A2 and H1-A3 for each labeled segment on a tier
- Not pitch-synchronous, not corrected
- If Praat cannot find F0 or all three formants, a file is skipped



# H1-H2 by VoiceSauce

- All VoiceSauce measures computed, but some are not available from other methods
- Here, uncorrected spectral magnitude measures were used

# H1-H2 for three consonant manners By-hand vs. VoiceSauce vs. Praat



# Differences

- Overall results from the 3 methods are similar.
- Measurements made by hand have the smallest H1-H2 range; Praat has largest H1-H2 range (larger category differences).
- BUT Praat measurements also have greater variation than from VoiceSauce or by hand – about twice as much for non-modal phonation.

# What makes Praat more variable?

- H1 and H2 measures both more variable
- Some possible reasons:
  - the STRAIGHT pitch-tracker used in VoiceSauce is very good
  - having F0 values every msec avoids discontinuities
  - harmonic amplitudes are found by optimization, which is equivalent to using a very long FFT window

# Conclusions

- We hope that VoiceSauce will be a useful and easy-to-use tool for researchers interested in multiple voice measures over running speech.
- Future versions will incorporate additional features.

# Download VoiceSauce

- VoiceSauce is free
- Currently requires Matlab to run
- Future versions will be compiled executables
- Available now from:  
<http://www.ee.ucla.edu/~spapl/voicesauce/>

# References

- Boersma, P. (2001) "Praat, a system for doing phonetics by computer", *Glott International* 5: 341-5.
- Harrington, J. (in press 2010) *Phonetic Analysis of Speech Corpora*, John Wiley and Sons
- Iseli, M. and A. Alwan (2000) "Inter- and Intra-speaker Variability of Glottal Flow Derivative using the LF Model," 6th International Conference on Spoken Language Processing, ICSLP 2000. Volume 1, pp. 477-480
- Iseli, M. and A. Alwan (2004) "An Improved Correction Formula for The Estimation of Harmonic Magnitudes and Its Application to Open Quotient Estimation," Proceedings of IEEE ICASSP, May 2004, pp. 669-672
- Iseli, M. Y. Shue and A. Alwan (2006) "Age- and Gender-Dependent Analysis of Voice Source Characteristics", Proceedings of IEEE ICASSP, vol. I, May 2006, pp. 389-392
- Iseli, M., Y.-L. Shue and A. Alwan (2007) "Age, sex, and vowel dependencies of acoustic measures related to the voice source", J. Acoust. Soc. Am. 121, 2283-2295
- Kawahara, H., A. de Cheveign, and R. D. Patterson (1998) "An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite," in Proceedings ICSLP'98, Sydney, Australia, December 1998
- Remijsen, B., Praat scripts: <http://www.ling.ed.ac.uk/~bert/praatscripts.html>
- Sjolander, K. (2004) "Snack sound toolkit," KTH Stockholm, Sweden.  
<http://www.speech.kth.se/snack>
- Tehrani, H, (2009) EGGWorks, a program for automated analysis of EGG signals.
- Vicenik, C. (2008). "An Acoustic Analysis of Georgian Stops." Unpublished UCLA Masters Thesis.

# Acknowledgments

- NSF grant BCS-0720304
- Code contributors: Henry Tehrani and Markus Iseli
- VoiceSauce beta users: Kristine Yu, Christina Esposito, Sameer Khan, Marc Garellek, Jianjing Kuang
- Co-PIs: Abeer Alwan and Jody Kreiman