Acoustic analysis software: A quantitative and qualitative comparison of four systems

ASHA Convention 2011 San Diego, California

Session 8855, Poster 373

18 November 2011

Carlyn Burris, Houri K. Vorperian, Marios Fourakis, and Raymond D. Kent

Waisman Center, **University of Wisconsin - Madison**

ABSTRACT:
Studies in speech acoustics often compare their findings with other published studies that use various acoustic analysis software. This study compares four acoustic analysis software (CSL-Praat-TF32-Wavesurfer) with respect to analysis features, and the accuracy and comparability of measurements of synthesized and natural vowels. Findings reveal differences of significance for clinicians and researchers.

INTRODUCTION:
Studies in speech acoustics often compare their findings with previously published studies that used various acoustic analysis software. Thus, the ***purpose*** of this study is to:

1. Test the assumption that all measurements generated by one acoustic analysis software package are accurate and comparable to measurements from other acoustic analysis software packages.
2. Assess how different software packages compare in terms of features and ease of use (*cf. Read, Buder & Kent 1990, 1992*).

METHODS:
Four commonly used acoustic analysis software packages were evaluated in this study:

- Praat (*version 5.1.31 by Boersma and Weenink*)
- Wavesurfer (*version 1.8.5 by the Centre for Speech Technology at KTH in Stockholm, Sweden*)
- TF32 (*formerly CSpeech by Milenkovic*)
- CSL (*Computerized Speech Laboratory model 4500, version 2.7.0 by Kay Elemetrics*).

Acoustic measurements compared across the four software packages, *using manufacturer recommended default settings*, included:

- Fundamental frequency (F0); First through fourth formant frequencies and bandwidths (F1-F4 & B1-B4)

Speech stimuli used to assess the accuracy and comparability of acoustic measurements included:

- ***Synthesized Speech/Vowels*** to assess accuracy of acoustic analysis software packages where software-generated measurements were compared against the input values used to synthesize them. Stimuli used were those synthesized by Fourakis, Preisel, and Hawks (1998) and consisted of:
  - Corner vowels /i/, /ae/, /u/, and /a/
  - Each vowel was synthesized with a male and female F0
- ***Natural Speech/Vowels*** to assess comparability of the four acoustic analysis software packages where stimuli collected by Hillenbrand, Getty, Clark, & Wheeler (1995) were reanalyzed, and measurements compared against their published results. Bandwidth values were compared against manual measurements. Stimuli used were the mid-vowel portion (150 ms) of words with corner vowels from speakers that were randomly selected from Hillenbrand's website (http://homepages.wmich.edu/~hillenbr/) and consisted of:
  - Corner vowels /i/, /ae/, /u/, and /a/
  - Five randomly selected speakers for each of the following four groups: Adult males, adult females, male children and female children

Analysis Procedures: Additional detail

- Analyses were blind, that is done without reference to the input values of the synthesized vowels or the published measurements of the Hillenbrand vowels.
- A protocol was established and followed for each software, which included use of a 150 ms midvowel analysis segment.
- Each software's default settings (e.g., analysis window type, window length) were used based on the understanding that software developers selected default settings that yield optimal results.
- Manual measurements of the first through fourth formant frequencies (F1-F4) and the associated bandwidths (B1-B4) were respectively taken from the LPC spectra formant peaks and 3dB below on either end of each formant peak using TF32.
- Edge phenomena' effects were eliminated by removing the first and last acoustic measurements from the list of measurements generated by each software package.

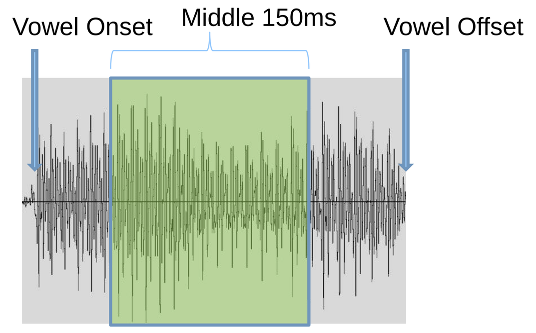

Figure 1: Illustration of how the 150ms midvowel analysis segment was determined.

RESULTS: *NOTE* Manufacturer's recommended default settings were used for all analyses.

Quantitative results: See **Figure 2** for Synthesized Vowels & **Figure 3** for Hillenbrand Vowels.

- **Fundamental frequency (f_0):**
  - Praat, Wavesurfer, and TF32 are accurate and comparable for both the synthesized and Hillenbrand vowels
  - CSL-generated values were more variable across speakers and vowels, especially for the synthesized vowels
- **Formant frequencies of synthesized vowels:**
  - All four software produce accurate and comparable results; however,
  - CSL had more variable measurements, particularly for female vowels /i/, /a/, and /ae/.
- **Formant frequencies of Hillenbrand vowels:**
  - All four software are accurate and comparable for F1 data
  - Praat, Wavesurfer, and TF32 produce F2-F4 measurements that are most comparable to Hillenbrand's findings, particularly for adults.
  - All four software generate more varied results in F3 and F4 analysis for children, with the degree of error varying by speaker and vowel type. However, Praat produced the most comparable and least variable results for both male and female child speakers. CSL produced the most varied results for most formants.
- **Bandwidth analyses:**
  - Synthesized vowels: Manual measurements of B1-B4 correlate closely to the input bandwidth values and are more accurate than the four software-generated values
  - Hillenbrand vowels: Praat, TF32, and Wavesurfer produced bandwidth measurements that were comparable to manual bandwidth measurements, but only for select vowels in select speakers, particularly males.

Synthesized Vowels

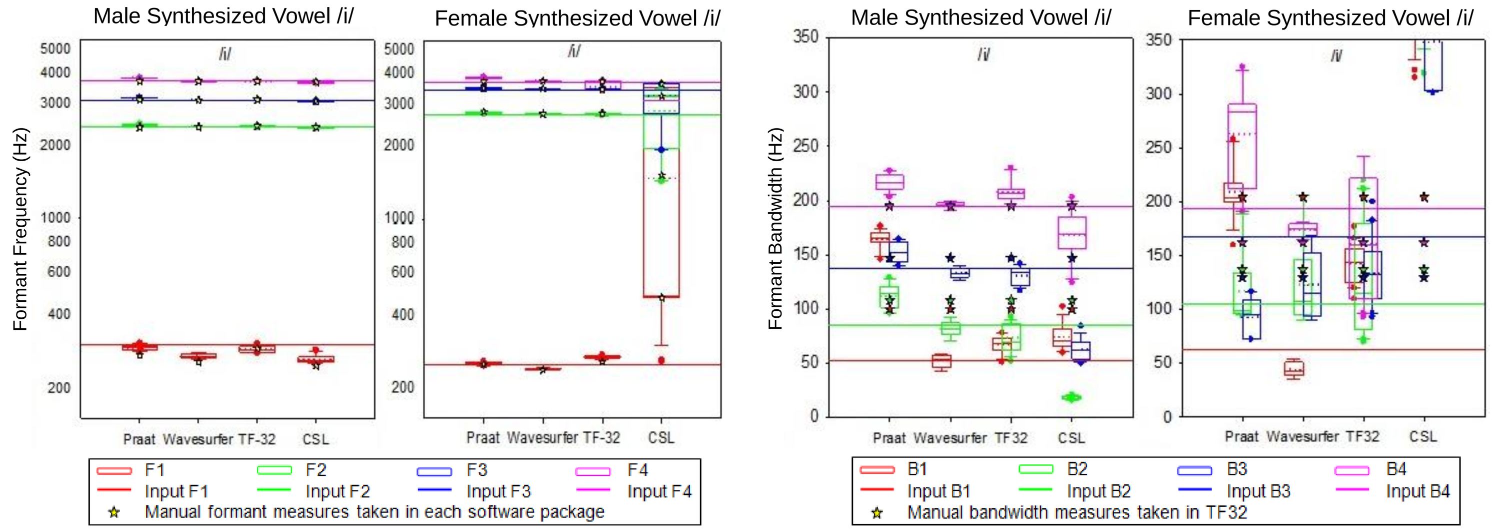

Figure 2: Synthesized vowels /i/ (with male and female F0) formant frequencies F1-F4 (left panel) and formant bandwidths B1-B4 (right panel) as a function of the four software packages. The software-generated measurements are displayed in box plots, while the synthesized input values are displayed as horizontal lines, and the manual measurements are displayed as gold stars.

Hillenbrand Vowels

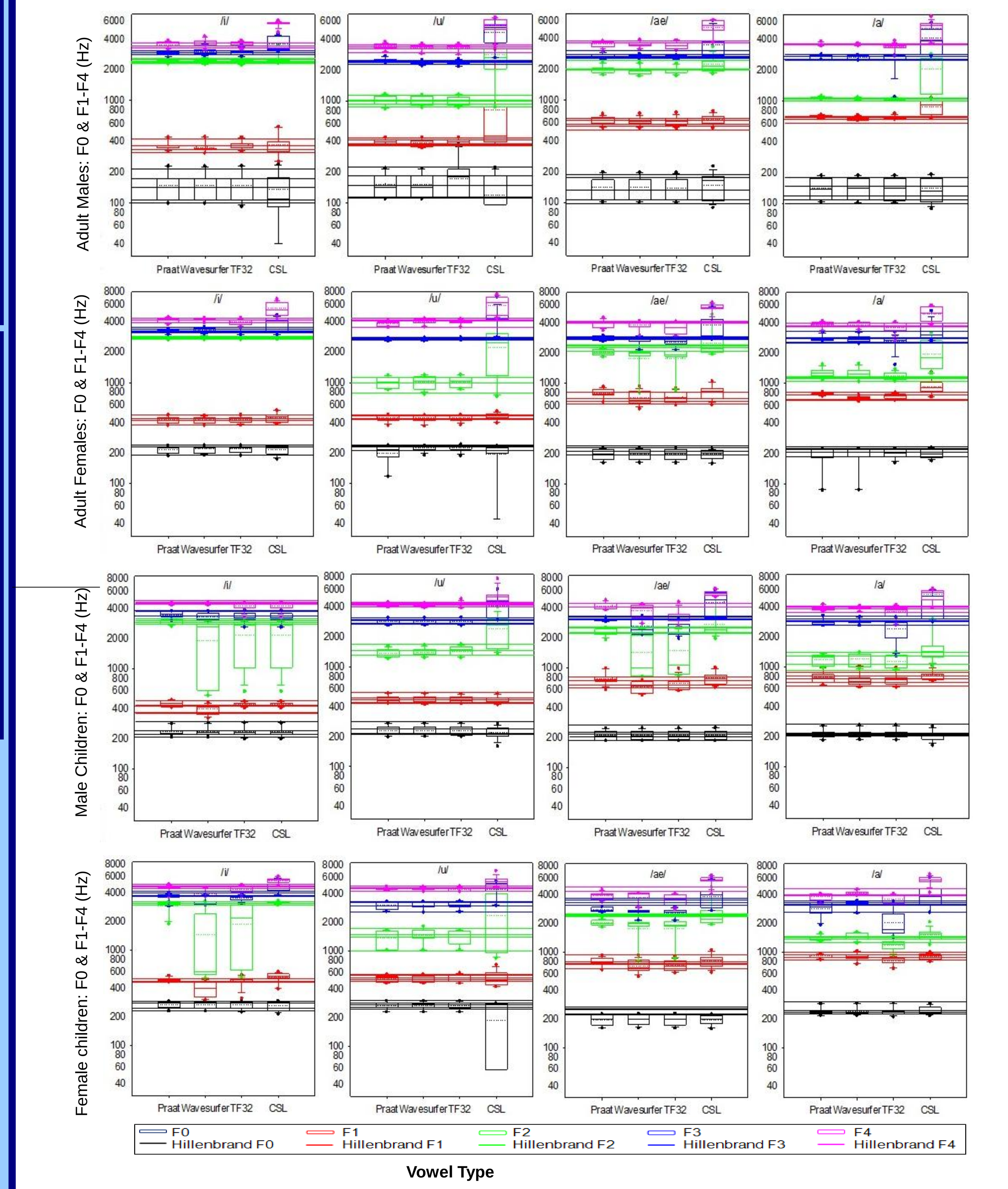

Figure 3: Average Hillenbrand vowels' fundamental and formant frequencies for adult males (1st panel), adult females (2nd panel) male children (3rd panel), and female children (4th panel) for each conrner vowel type. The box plots display the mean, 25th percentile, and 75th percentile values for F0 (black) and F1-F4 (red, green, blue & pink) for the five speakers per group/panel in each of the four software packages (i.e. each box represents 15 values). The F0 and F1-F4 values, as reported by Hillenbrand et al (1995), are displayed as a matching colored horizontal lines. One line for each of the five randomly selected speakers in each of the four speaker groups).

Qualitative findings: An assessment based on similarities and differences between the four software packages in terms of default settings, features, and ease of use:

- Praat, Wavesurfer, and CSL allow for a great deal of setting manipulation, such as window length and pre-emphasis, whereas TF32 only allows the user to manipulate a few settings, such as window type and analysis bandwidth.
- CSL is the only software package that allows for the purchase of additional modules, such as the commonly used Multi-Dimensional Voice Profile, which allows for a detailed analysis of a speaker's voice (e.g. jitter and shimmer).
- The table below outlines strengths and limitations of each software package and highlights the fact that each software package contains both positive and negative aspects. The identification of positive and negative aspects here represents the point of view of the authors and is a subjective judgment that may not be shared by all users of these systems.

TABLE 1: Features of each software package

Feature	TF32	Wavesurfer	Praat	CSL
Users Manual:	• No help menu • Online PDF of user's manual (116pgs)	• Manual in help menu • FAQ option in help menu	• Online PDF of tutorial (27pgs) • Somewhat outdated, for older version • Active user group to answer questions	• In-software tutorial
Zoom/scroll	• Zoom to selected, zoom out full • Scrollbar • Keyboard arrows to move along waveform	• Zoom to selection, zoom in/out, zoom out full • Scrollbar • Keyboard arrows to move along waveform	• Zoom to selection, zoom in/out, zoom out full • Scrollbar	• Select and zoom, zoom out full • Scrollbar
Time readout	• 0.001ms	• 1ms	• 0.001ms	• 0.01ms
Amplitude readout	• No	• Yes	• Yes	• Yes
View and play selected	• Yes	• Yes	• Yes	• Yes
Settings held when open multiple files	• Yes	• No	• Yes	• Yes
- FFT	• Yes	• Yes	• Yes	• Yes
- LPC	• Yes	• Yes	• Yes	• Yes
- Waterfall	• No	• No	• No	• Yes
- Freq and amp readout	• Yes	• Yes	• Yes	• Yes
Pitch	• Yes	• Yes	• Yes	• Yes
Simultaneous displays	• Yes	• Yes	• Yes	• Yes
Additional features and information	• Optional smoothing function • Available for free	• Automatic smoothing function • Available for free	• Labeling and segmenting • Available for free	• Multiple additional modules, including MDVP • Must purchase

TABLE 2: Strengths and limitations of each software package

Software	Strengths	Limitations
TF32	• Can move along waveform and spectrogram and make small time changes • Clearly displays formants on spectrogram • LPC spectrogram changes as arrow keys move to different sections of the waveform • Very user friendly and does not require a lot of set up prior to analysis • Can easily move from one sound file to another, if sound files are saved in the same folder • Settings are held from one file to another during continued use. Settings are reset when software is closed.	• No measures of F4 in the non-alpha version, but can get rough estimate from LPC spectrogram • No B1-B4 measures in the non-alpha version • Cannot set pre-emphasis value, can only turn on or turn off • Limited ability to manipulate settings
Wavesurfer	• Speech analysis option displays everything, no need to open each feature separately • LPC spectrogram changes as arrow keys move to different sections of the waveform • Easy and precise moving across waveform and spectrogram with computer mouse • Large number of manipulable settings	• Have to export data file to see values, not displayed in software • Settings are not held from one file to another during continued use • Large amount of settings may be overwhelming to a novice user
Praat	• Can label portions of a waveform and save data for future reference • Settings are held from one file to another during continued use (settings are reset when software is closed) • Shows listing of formant values and gives an average	• Limited ability to make small time adjustments along waveform • Adjustment of analysis bandwidth (eg. broad vs. narrow) only possible through adjusting window size • Can only extend selected area easily when using the segmenting feature
CSL	• Comprehensive in-software tutorial which shows the different features and how to use them • Clear display of formants and bandwidths • Option to purchase additional modules, such as MDVP for analyzing voice characteristic of speaker (e.g., jitter and shimmer)	• External proprietary hardware required • The most costly of the 4 systems with no option of a downloadable demo • Fine time adjustments can only be made with computer mouse; cannot use arrow keys for additional refinement • Tedious to make small zooming adjustments

DISCUSSION / MAJOR FINDINGS:

- This is the first study that has quantitatively compared software-generated measurements from four acoustic software analysis packages.
- Comparative evaluation of this kind is one step toward insuring the validity and reliability of published acoustic databases.
- Findings presented here are specific to synthesized and naturally produced vowels.
- Further research is warranted for consonants and disordered speech.
- Results indicate that users should not assume that all acoustic analysis software yield comparable results for all acoustic measures.
- Present findings indicate that results are speaker (age and sex) and signal (vowel type) dependent for both accuracy and comparability of measurements across software.
- Findings also show that manufacturer recommended default settings for analysis yield optimal results for adult speakers when using Praat, TF32 & Wavesurfer but not CSL. For child speakers, only Praat yields the most reliable and least variable results.
- Manual correction of formant measurements or applying a smoothing function (as is standard in Wavesurfer and available in TF32) is warranted in all software packages.
- Comparison of formant values automatically derived from the software to formants visible in FFT spectrograms is recommended to help detect erroneous formant values.
- In terms of measurement accuracy of synthesized vowels i.e. acoustic analysis software-generated measurements that were close to synthesized vowel input values:
  - All four software generate fairly accurate F0 values, however CSL measurements were highly variable.
  - All formant frequencies were fairly accurate but F1 was the least accurate, and CSL produced highly variable results particularly for vowels synthesized with a female F0.
  - All software-generated bandwidth measurements were highly variable and only manual BW measurements were closest to input values.
- In terms of comparability of software-generated measurements of natural vowels against published values as well as manual measurements:
  - Three of the four software packages (Praat, TF32 & Wavesurfer) yielded comparable F0 as well as F1-F4 values for adult speech. CSL measurements were more variable.
  - For children's speech, all four software generated fairly comparable F0 and F1 values, but F2-F4 values were more varied particularly CSL results. Praat produced the most reliable and and least variable measurements.
  - As for bandwidth measurements, software-generated values were not comparable and appear to be vowel-dependent as well as speaker-sex and speaker-age dependent. Manual BW measurements, though tedious, appear to be the most accurate approach to bandwidth analysis.

Qualitative:

- The decision of which acoustic analysis software package to use should be based on the user's needs and not just software availability.
- Each software package can benefit different users.
  - TF32 is a good choice for users who are new to acoustic analysis due to the user-friendly interface, fast generation of data, which would be helpful in a clinical setting, and small number of menus to navigate.
  - Praat has the advantage of labeling speech files
  - Wavesurfer allows for a large amount of setting manipulation
  - CSL supports the use of additional modules, such as the Multi-Dimensional Voice Profile (MDVP)

MAJOR CONCLUSIONS based on synthesized and natural vowels for:

Clinicians:

- Three of the four software (Praat, Wavesurfer, TF32) are available for free download and are fairly user-friendly.
- For adult speech, Praat, TF32, and Wavesurfer generate accurate and comparable measurements; whereas Praat performed the best for children's speech.
- Manuals are available online for TF32 and Praat. Praat also has an active online user group, which can be used for troubleshooting and learning more about the software.
- TF32 is a good choice for new users due to the user-friendly interface, fast generation of data, and small number of menus to navigate.
- Voice clinicians should be wary of F0 measurements generated by CSL.

Researchers:

- These results pertain only to selected synthesized and natural typical vowels. Further research is warranted for disordered speech and consonants.
- There are differences between the four software-generated measurements based on vowel type, speaker sex and age group.
- Knowledge of the strengths and limitations of each software is critical in selecting an acoustic analysis software that best meets research objectives.
- Changing software settings can improve reliability of software-generated measurements. (*For example, changing the filter order in CSL yields more accurate F1-F3 formant tracking. Similarly, the application of the smoothing function improves software-generated measurements for F3-F4. However, the general, rather than selective, application of the smoothing function may hinder accurate assessments of disordered speech*.)
- Very few studies report on formant bandwidths. (*For example, Robb, Chen, & Gilbert (1997) reported significant changes in B1 and B2 throughout infancy, despite very small changes in F1 and F2, highlighting the importance of investigating the accuracy and comparability of bandwidth measurements in addition to F0 and F1-F4 measurements.*)
- None of the software provide reliable bandwidth values, thus tedious manual measurements are the most reliable until software developers address this issue.
- Manual BW measurements require additional caution when:
  - measuring B1 and B2 because the very small values are often difficult to measure on the LPC spectrum.
  - measuring the bandwidths of closely positioned formants (e.g., F1 & F2 for vowels /a/ and /u/) because these formants are likely to influence one another and that interaction could lead to inaccurate manual measurements.

Software Developers:

- Software algorithms need to be revised to better detect bandwidth measures.
- The ability to label waveforms and save those labels for subsequent views, as is available in Praat, is a valuable software feature for researchers and clinicians.
- Including an in-software tutorial, as is available in CSL, is a valuable resource for new users.

ACKNOWLEDGMENTS:
This work is supported by NIH Research Grants R01DC 006282 & P-30 HD03352. This research was originally submitted as a Master's thesis by the first author to the Department of Communicative Disorders at the University of Wisconsin-Madison.

SELECT REFERENCES: This poster can be found at: http://www.waisman.wisc.edu/vocal/posters.html

(1) Fourakis, M., Preisel, C., and Hawks, J.W. (1998). Perception of vowel stimuli synthesized with different fundamental frequencies. *J. Acoust. Soc. Am., 104*, 1778. **(2)** Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5 Pt 1), 3099-3111. **(3)** Read, C., Buder, E. H., & Kent, R. D. (1990). Speech analysis systems: a survey. *Journal of Speech and Hearing Research, 33*(2), 363-374. **(4)** Read, C., Buder, E. H., & Kent, R. D. (1992). Speech analysis systems: an evaluation. *Journal of Speech and Hearing Research, 35*(2), 314-332. **(5)** Robb, M. P., Chen, Y., & Gilbert, H. R. (1997). Developmental aspects of formant frequency and bandwidth in infants and toddlers. *Folia Phoniatr Logop, 49*(2), 88-95.