AJSLP

Research Article

# Optimizing Vowel Formant Measurements in Four Acoustic Analysis Systems for Diverse Speaker Groups

Ekaterini Derdemezis,[a] Houri K. Vorperian,[a] Ray D. Kent,[a] Marios Fourakis,[b] Emily L. Reinicke,[a] and Daniel M. Bolt[b]

**Purpose:** This study systematically assessed the effects of select linear predictive coding (LPC) analysis parameter manipulations on vowel formant measurements for diverse speaker groups using 4 trademarked Speech Acoustic Analysis Software Packages (SAASPs): CSL, Praat, TF32, and WaveSurfer.
**Method:** Productions of 4 words containing the corner vowels were recorded from 4 speaker groups with typical development (male and female adults and male and female children) and 4 speaker groups with Down syndrome (male and female adults and male and female children). Formant frequencies were determined from manual measurements using a consensus analysis procedure to establish formant reference values, and from the 4 SAASPs (using both the default analysis parameters and with adjustments or manipulations to select parameters). Smaller differences between values obtained from the SAASPs and the consensus analysis implied more optimal analysis parameter settings.
**Results:** Manipulations of default analysis parameters in CSL, Praat, and TF32 yielded more accurate formant measurements, though the benefit was not uniform across speaker groups and formants. In WaveSurfer, manipulations did not improve formant measurements.
**Conclusions:** The effects of analysis parameter manipulations on accuracy of formant-frequency measurements varied by SAASP, speaker group, and formant. The information from this study helps to guide clinical and research applications of SAASPs.

In its clinical application, acoustic analysis can afford a greater sensitivity than commonly used perceptual methods and can provide the quantitative data needed for detailed assessment, documenting treatment outcome, or tracking disease progression. Advantages of acoustic analysis have been noted in studies of the speech disorder in individuals with cognitive impairment (Saz, Simon, Rodriguez, Lleida, & Vaquero, 2009), autism (Diehl & Paul, 2013), and Parkinson's disease (Chenausky, MacAuslan, & Goldhor, 2011), to name a few. But, as with any method, the desired clinical sensitivity obtains only if the analyses are accurate within the constraints of clinical practice. This article addresses the accuracy of one important set of acoustic measures, vowel formant frequencies.

Vowel formants have been central measures in numerous recent reports on speech disorders, especially in relation to calculations of vowel space area or similar metrics such as vowel distance. These metrics have been reported to be useful for several purposes, including assessment of speech motor function in children and adults with dysarthria (Higgins & Hodge, 2002; Hustad, Gorton, & Lee, 2010; Sapir, Polczynska, & Tobin, 2009), diagnosis and classification of dysarthria (Lansford & Liss, 2014), assessment of speech in adults with Down syndrome (DS; Bunton & Leddy, 2011), use as an early marker of Parkinson's disease (Rusz et al., 2013), monitoring disease progression in Parkinson's disease (Skodda, Gronheit, & Schlegel, 2012), evaluating dysarthria treatment in Parkinson's disease or stroke (Mahler & Ramig, 2012; Wenke, Cornwell, & Theodoros, 2010), and determining articulatory changes following manual circumlaryngeal therapy for muscle tension dysphonia (Roy, Nissen, Dromey, & Sapir, 2009). For these and similar clinical applications of formant pattern, accuracy of formant-frequency measurements is critical. The current study is part of a larger effort to evaluate the performance of Speech Acoustic Analysis Software Packages (SAASPs) in the

[a]Waisman Center, University of Wisconsin–Madison
[b]University of Wisconsin–Madison
Correspondence to Houri K. Vorperian: vorperian@waisman.wisc.edu

Editor: Krista Wilkinson
Associate Editor: Jack Ryalls
Received February 25, 2015
Revision received June 23, 2015
Accepted September 18, 2015
DOI: 10.1044/2015_AJSLP-15-0020

***Disclosure:*** *The authors have declared that no competing interests existed at the time of publication.*

study of disordered speech in male and female speakers across the life span. A primary objective is to determine the strengths and limitations of SAASPs for clinical purposes, including users who may not have an extensive background in the acoustic analysis of speech.

SAASPs typically allow for several types of analyses that are used by researchers, clinicians, forensic specialists, and others concerned with the acoustic properties of speech. Among the most widely used analyses are fast Fourier transform (FFT) analyses to generate spectrograms and power spectra and linear predictive coding (LPC) analyses to identify formants. These analyses are valuable tools in the study of vocal tract characteristics, speech motor control, or phonological and phonetic patterns. In general, SAASPs have default settings that are tailored to the speech characteristics of adult male speakers with typical speech, and general guidelines may be provided in their documentation (manuals and online user group exchanges) for the adjustment of analysis parameters to analyze the speech of female adults, children, or populations with speech disorders of either sex and/or different ages. The parameters that are most likely to be adjusted in SAASPs for vowel analysis are the pre-emphasis to shape the spectrum for further analyses, analyzing bandwidth used to generate FFT spectrograms, and LPC filter order (number of coefficients) to identify formants (Deng & Dang, 2007; Kent & Read, 2001; Vallabha & Tuller, 2002; Yao, Tilsen, Sprouse, & Johnson, 2010). Although research articles often mention that such manipulations were made in the acoustic analysis of speech to improve on formant measurements, they generally do not provide specific information, such as the criteria used to adjust LPC filter order.

Because LPC automatically determines formant frequencies and bandwidths, it offers convenience and efficiency in acoustic analysis, especially for nonnasalized vowels, for which the all-pole solution generally used in LPC is most appropriate. A particular advantage is that LPC analysis can output the data as a worksheet and/or a formant-tracking display that may be superimposed on a spectrogram. These forms of output facilitate user-made corrections to formant-tracking errors that may occur in the identification of closely spaced formants or formants of low intensity. In many implementations, LPC filter order is automatically adjusted on the basis of sampling rate, or it can be adjusted by the user if, for a given speaker and a given speech task, the default value is judged to be nonoptimal. The ideal filter order provides accurate analysis while conserving computation and memory requirements. The commonly recommended guidelines are to adjust the filter order to be equal to (a) the sampling frequency in kHz (e.g., 10 coefficients for a sampling rate of 10 kHz) or (b) 2 times the desired number of formants plus two (e.g., 10 coefficients to achieve an analysis of four formants; Deng & Dang, 2007; Vallabha & Tuller, 2002). The basis for the latter is to include a sufficient number of model poles (two poles per formant) for the signal bandwidth of interest, along with a small number of additional poles to compensate for windowing effects and limitations of the all-pole model. That is, the model can produce erroneous results because of nonformant effects associated with vocal tract excitation or lip radiation. Vallabha and Tuller (2002, 2004) proposed a method of adjusting the LPC filter order on the basis of the computation of reflection coefficients for a set of representative analysis frames. They recommended that LPC filter order be adjusted for individual speakers whenever possible, and this recommendation underscores the importance of analysis parameter adjustments. Clinical application of SAASPs places a premium on efficiency, given that time pressures in a busy practice often do not allow for multiple adjustments of analysis parameters and an inspection of the results of these adjustments. If acoustic analysis can be accomplished with few adjustments, and if clear guidelines are provided for these adjustments, then clinical application is facilitated.

The effects of commonly used analysis parameter manipulations on vowel formant measurements from SAASPs have rarely been assessed empirically. Yao et al. (2010) examined the effect of manipulating LPC order and pre-emphasis on the measurement of vowel formant frequencies and found that for different speakers, different combinations of the two parameters yielded optimal formant measurements. In a forensic application of formant measurements, it was shown that the results of independent analyses by three different laboratories came to closer agreement when consistent methodology was used (Duckworth, McDougall, de Jong, & Shockey, 2011). Burris, Vorperian, Fourakis, Kent, and Bolt (2014) examined the accuracy and comparability of vowel formant measurements made by four SAASPs for synthesized speech with known input values and natural speech samples from a published study (Hillenbrand, Getty, Clark, & Wheeler, 1995). The measurements were performed with each software's default settings. Burris et al. reported significant differences in measurements for both synthesized and natural speech. Substantial errors were noted for some measures. For example, Burris et al. observed that some bandwidths were so large as to render them nearly useless. They also reported that a trial assessment of analysis parameter manipulations yielded more accurate formant-frequency measurements but did not undertake a systematic evaluation of these manipulations for different SAASPs.

The natural speech samples used by Burris et al. (2014) were from typically developing (TD) children and healthy adults, and the speech samples used by Yao et al. (2010) were produced by healthy adults. To our knowledge, no study has examined how SAASPs perform when analyzing disordered speech, which can complicate acoustic analysis due to features such as atypical vocal quality (e.g., perturbations and noise), oral–nasal resonance imbalance (e.g., episodes of nasalization where it does not normally occur), large formant bandwidths (relating to characteristics of the vocal tract tissues or to nasalization), and overall instability in the acoustic pattern. Therefore, it is important to examine if analysis parameter manipulations can overcome some of these challenges to enhance the accuracy of SAASP vowel formant measurements for disordered speech. The samples of disordered speech used in this study were obtained from

children and adults with DS. Reduced intelligibility is a commonly reported feature of this syndrome and is likely related to a combination of disorders of voice, articulation of vowels and consonants, resonance, and prosody (Kent & Vorperian, 2013).

The purpose of this study was to systematically assess the effect of analysis parameter manipulations that are commonly reported and are of relevance to securing first to fourth vowel formant-frequency measurements derived from LPC analysis for each of four trademarked and commonly used SAASPs with a diverse group of speakers of both sexes, including TD children and adults as well as children and adults with DS. The manipulations are not equivalent across the SAASPs because of differences in software design including default settings. The aim of this evaluation was not to compare the performance of different SAASPs but rather to evaluate the effects of analysis parameter changes within each system that are most likely to be performed in clinical settings. The hypothesis was that parameter manipulations can improve the accuracy of vowel formant measurements for each SAASP but that such benefits would be software-specific and vary by speaker group and formant.

## Method

### *Source Materials for Acoustic Analysis*

This study was approved by the University of Wisconsin–Madison Health Sciences Institutional Review Board. The speech samples, obtained from the Acoustics Database of the Vocal Tract Development Lab, were monosyllabic words containing the vowels /i/, /u/, /æ/, and /ɑ/ as in *eat*, *hoot*, *hat*, and *hot*. Recordings were obtained from 42 individuals with DS (26 male speakers and 16 female speakers, ages 4–36 years) and 122 TD individuals (52 male speakers and 70 female speakers, ages 4–66 years). The speech samples were recorded in a quiet room, using the TOCS+ Platform Program (Hodge, Gotzke, & Daniels, 2009), which served as the interface for presenting the stimuli to be repeated by the participants. Speech was recorded using a cardioid Shure SM48 microphone (Shure Inc., Niles, IL), placed 15 cm from each participant's mouth and stabilized using a floor stand that was directly connected to a Marantz PMD660 digital audio recorder (Marantz Professional, Cumberland, RI) that digitizes speech at 48 kHz with 16-bit resolution on a SanDisk Ultra II flashcard (SanDisk Corp., Milpitas, CA). A laptop computer hosting the TOCS+ Platform Program was used to present the stimuli in random lists. The stimuli used to elicit productions from the participants consisted of 20 words (five words per vowel/set) that were presented as pictures on a computer monitor, with the orthographic word present on the bottom of the display and an audio recording (of a male adult) of each word and sentence presented over external speakers. Participants were instructed to repeat each word and sentence at a normal conversational loudness, and produced two practice words before beginning the word repetition task; this was done so that the sound-level meter could be adjusted accordingly. The targeted recording level was between 6 and 12 dB below the maximum level, and feedback regarding production volume was given to the participant throughout the recording session. All participants repeated each word at least once; the first two words of each vowel set were repeated twice, and words within each vowel set were randomized.

The vowel portion of the following four words spoken by each participant was acoustically analyzed: *eat*, *hoot*, *hat*, and *hot*. The participants, randomly selected from the database, included (a) 10 TD children (five boys; five girls) ages 5–10 years; (b) 10 children with DS (five boys; five girls) ages 5–10 years; (c) 10 adults with DS (five men; five women) ages 21–35 years; and (d) 20 TD adults (10 men; 10 women) ages 20–25 years, representing a total of eight speaker groups. As indicated previously, participants with DS were selected to provide samples of disordered speech because of the likelihood of combined speech and voice impairments. DS is also challenging in acoustic analysis because craniofacial anomalies may affect the acoustic properties of speech (Moura et al., 2008). The speech capability for the participants with DS was determined by a transcription task performed by five listeners. Recordings of single words were presented to the listeners, who were asked to transcribe what they heard using a computer keyboard. Percent scores were calculated for words and the vowels within the words. The mean results for each group of participants with DS were male children: 37% words correct, 67% vowels correct; female children: 40% words correct, 70% vowels correct; male adults: 65% words correct, 85% vowels correct; female adults: 72% words correct, 91% vowels correct.

### *Procedures of Acoustic Analysis*

The acoustic analyses were performed with four commonly used SAASPs (the same software versions as those evaluated by Burris et al., 2014): (a) Computerized Speech Laboratory (CSL; model 4500, version 3.4.1, KayPentax, 1996); (b) Praat (version 5.1.31, Boersma & Weenink, 2010); (c) TF32 (alpha test version 1.2 used by Burris et al., 2014; formerly known as CSpeech by Milenkovic, 2010); and (d) WaveSurfer (version 1.8.5; Sjolander & Beskow, 2005; a more recent version is Sjolander & Beskow, 2010).

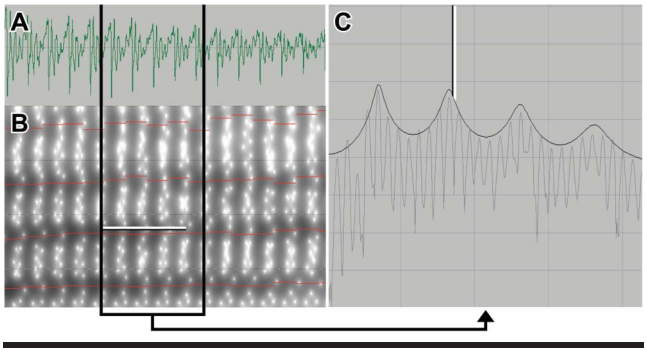
The waveforms of the recorded words were first opened in Praat to segment and save the vowel portion as a separate sound file. The onset and offset of the vowel were determined by visual inspection of the spectrogram and waveform and by listening to the sound segment. The main criteria were the visibility of the main vowel formants (F1 and F2) and the presence of periodic energy. For the purposes of this study, the primary objective of vowel segmentation was to provide a sample suitable for selection of an analysis interval representing the vowel steady state. Vowel duration was not further used in this study. Next, and as the first step of the consensus analysis procedure described below, the most stable vowel segment was determined by identifying the interval in the spectrogram for which there was overall stability in the formant tracks

and where all four formants were most visible. A segment duration between 25 and 50 ms consistently captured the vowel steady state as just defined. This range was shorter than the 150-ms duration used by Burris et al. (2014) but ensured representation of the vowel steady state for both typical and disordered speech. Two sets of acoustic measurements were made: first through fourth formant frequencies (F1–F4) and fundamental frequency (F0), described below. Measurements of F0 were made because F0 is known to affect formant measurements (Kent & Read, 2001). Formant bandwidth measurements were not made because Burris et al. (2014) reported such measurements by SAASPs to be highly variable and not reliable. Mehta and Wolfe (2015) explained that the accuracy of formant bandwidth estimation is due to the inherent statistical properties of LPC speech analysis. Two sets of acoustic measurements were made: reference measurements performed using a consensus analysis procedure (described below) and SAASP-generated measurements taken at both the software manufacturers default settings and after applying a select set of analysis parameter manipulations.

### *Reference Values (Consensus Analysis)*

To establish reference values, two individuals experienced in the acoustic analysis of speech participated in consensus analysis; a third individual also participated in most, but not all, analyses. This consensus analysis procedure involved a global visual inspection approach similar to the typical approach used in manual correction of LPC formant analyses (Fox & Jacewicz, 2009; Hillenbrand et al., 1995). These individuals viewed the waveform (Figure 1A) and spectrogram (Figure 1B) of the segmented vowel to: (a) select the steady-state segment of the vowel (i.e., the most stable 25–50 ms of the total vowel duration; marked by two vertical cursors in Figure 1A and 1B), and (b) determine the F1–F4 frequencies, henceforth referred to as consensus analysis reference values (CARVs). CARVs were measured in TF32 because this SAASP allows the overlay of the LPC on the FFT spectra (Figure 1C). Furthermore, TF32's spectrogram is synchronized with the LPC spectrum so that when the user moves the cursor across the LPC and FFT spectra (vertical/white cursor in Figure 1C), a slaved cursor moves vertically to the corresponding frequency on the spectrogram (horizontal/white cursor in Figure 1B). Manipulations to the spectrogram were made to help establish the most accurate CARV and included (a) changing the analysis bandwidth of the FFT on the basis of the speaker's sex, age, and F0 to display the formants in a wide-band spectrogram, (b) changing the dynamic range to assist in viewing and discriminating between the different formants, and (c) using a spectral slice from the 25–50 ms vowel analysis segment to aid in decision making. When consensus could not be reached, no measurements were taken. Such difficulty usually occurred when (a) two formants were close in frequency, (b) strong harmonics precluded certainty of formant tracking, (c) there were oral–nasal resonance imbalances, or (d) vocal quality was atypical. Of the total number

**Figure 1.** The left panel displays a 110-ms steady-state portion of the vowel /ae/ produced by a 21-year-old male speaker saying the word *bat*, with a 35-ms analysis segment of the vowel marked by two vertical cursors on the time domain (as displayed in TF32). The top portion (A) is an amplitude by time waveform, and the bottom portion (B) its frequency by time spectrogram. The right panel (C) displays the amplitude by frequency spectral slice (overlay of linear predictive coding spectrum on fast Fourier transform spectrum) of the marked analysis segment of the vowel, with vertical gridline markings in 1000-Hz intervals, and horizontal lines in 10-dB increments. The frequency of the white vertical cursor on the spectra (C), correspond to the frequency of the white horizontal cursor on the spectrogram (B).




of CARVs, 15% were abandoned because of uncertainty in formant detection. The majority of these were for the higher formants F3 and F4. In addition, manual F0 measurements were made from the steady-state portion of the waveform by computing the inverse of the total duration of the middle three cycles of the analysis segment divided by three.

The reliability of CARVs was confirmed by remeasuring 10% of the tokens 3 months later. Paired *t* test results were as follows: for TD speakers, $t = 1.643$; $p = .117$, and for speakers with DS, $t = 1.914$; $p = .061$. Approximately 80% of the remeasured values were within 50 Hz of the original values. The reliability of the consensus analysis procedure was further assessed using a subset of stimuli from an external source, (a subset of 10 Hillenbrand et al., 1995, tokens representative of the four corner vowels) that were downloaded from Hillenbrand's homepage (Hillenbrand, 1995) and used in the Burris et al. (2014) study. The middle 150 ms of the vowels were segmented and saved as separate sound files in Praat and were analyzed to make F1–F4 measurements using the consensus analysis procedure. These measurements were then compared against the published values in Hillenbrand et al. The larger analysis segment duration of 150 ms was used instead of the 25- to 50-ms consensus analysis duration because the published Hillenbrand et al. reference values were made for the 150-ms vowel segment. Paired *t* test comparison of F1–F4 measurements from the consensus analysis procedure and Hillenbrand et al. (1995) published values further confirmed the reliability of the consensus analysis procedure ($t = 1.068$; $p = .292$).

### *SAASP Measurements*

To assess the effect of analysis parameter manipulations on the accuracy of formant measurements in each

SAASP, measurements of F1–F4 and F0 were taken at the midpoint of the 25- to 50-ms steady-state vowel segment (or a point close to the midpoint when measurements could not be made for all four formants at the midpoint). F1–F4 values from each SAASP were based on LPC analysis values (i.e., from the LPC formant tracks superimposed on the spectrogram). Measurements of F1–F4 were recorded for (a) output measurements at the manufacturer's default settings (presumably optimal for adult male speech), with F0 also measured at the manufacturer's default settings; and (b) output after manipulating a select set of analysis parameters as permitted by each SAASP (to enhance the accuracy of formant-frequency measurements). Because the SAASPs are not uniform in the way analysis parameters are changed, it is necessary to describe the manipulations within the features of each system.

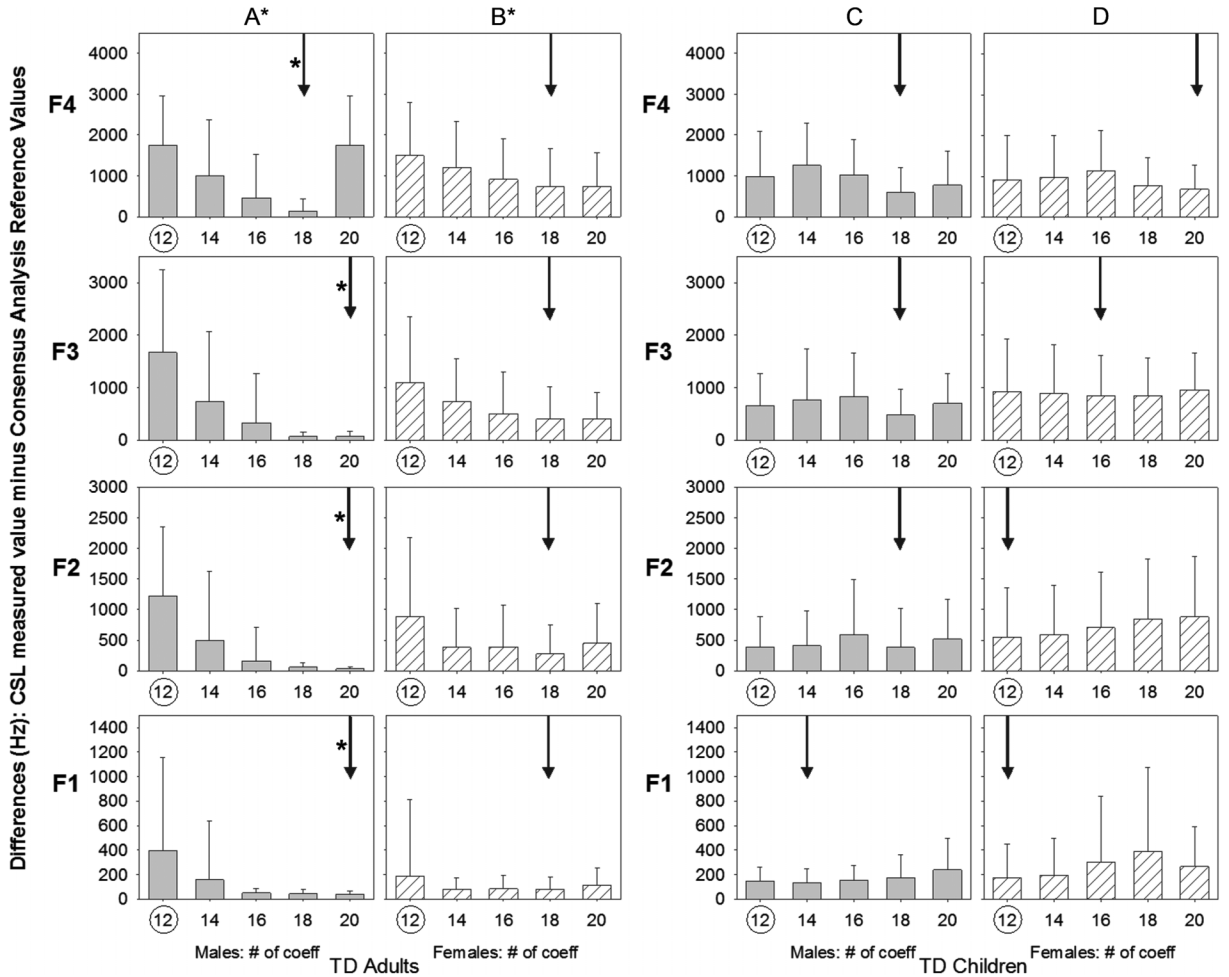
As summarized in Table A1, manipulations to analysis parameters included (a) changing the number of LPC coefficients and (b) applying the smoothing function, a manipulation that smoothes the computed LPC coefficients between pitch period frames and then applies a dynamic programming method to achieve continuity for formants when labeling the LPC poles. Such manipulations, described in more detail below, were made if the SAASP allowed the user to manipulate the analysis parameters. Table A1 outlines all permissible parameter manipulations for each SAASP, and the sequential order of parameter manipulations is denoted by column number (1 = first parameter manipulated; 2 = second, and 3 = last). In addition, for all four SAASPs, particularly the ones that downsample to 10 kHz, the maximum formant-frequency range was scrutinized to ensure that its upper limit was not below the F4 frequency range for female adults and children. This was done to ensure accuracy of formant-frequency estimation.

The sequence of parameter manipulations in each SAASP was as follows: In CSL, the first step was downsampling the analysis segment to 16 kHz and adding voice period marks, then changing the number of LPC coefficients in increments of two (across a range of 12–20), and finally applying smoothing. In TF32, downsampling was not needed because this SAASP automatically adjusts the default number of LPC coefficients on the basis of the sampling rate. Manipulations in TF32 were limited to changing the number of coefficients in increments of two (across a range of 48–56) and then applying smoothing. TF32 is unlike the other SAASP studied here in that it is designed to do analyses over a large frequency range, with an accordingly large number of LPC coefficients (the rationale for this design feature is explained in Appendix B). Praat automatically downsamples the analysis segment to 10 kHz and does not provide users the direct option to manipulate the number of coefficients. Users can, however, indirectly adjust the number of coefficients by changing the default number of five formants to four formants. This procedure changed the number of coefficients from 10 to eight, and was the only coefficient manipulation made given our interest in F1–F4 measurements. The time step was also changed to .00625, and for male adults, the maximum formant was changed to 5000 Hz, because the default 5500 Hz value is preferred for adult females and children. In WaveSurfer the system downsamples automatically to 10 kHz, so manipulations consisted of changing the number of LPC coefficients (across a range of 12–20), followed by automatic smoothing applied by the system. However, for children and female adults, the downsampling frequency was changed to 11 kHz.

## Results

Mean absolute differences of formant frequencies were determined between the SAASP-generated measurements (at both default and at different levels of parameter manipulations) and CARVs. These values (Hz) are displayed in Figures 2, 3, 4, 5, 6, 7, 8, and 9 as vertical bar charts with data pooled across vowels and speakers within a given speaker group, and the simple error bar represents the standard deviation from the mean. The smaller the difference(s) as compared to difference at default (denoted by a circle around the number of coefficients), the more optimal that specific analysis parameter was considered (denoted by an arrow above the vertical bar chart in Figures 2–9). Differences between the default and optimal analysis parameter (if different than default) were tested for statistical significance using the nonparametric Wilcoxon signed-ranks test across all four formants at a significance level of .05 and the individual formants at .0125 (to account for alpha inflation error; Dunn, 1961). The test was performed by attending to the sign (direction) of the difference in measurements for a given speaker and vowel under the default and optimal settings. Significant differences are denoted by an asterisk (*), where an asterisk to the left of the arrow indicates statistically significant differences between default and the optimal analysis parameter for the individual formant, and an asterisk next to the panel letter indicates statistically significant differences between the default setting and the optimal setting across the four formants. Optimal analysis setting across the four formants was determined when a particular analysis level was optimal for at least two individual formants. Although measurements from the same speaker across vowels tend to correlate at a given parameter setting, our analyses attend to the difference in these measurements for different parameter settings, an outcome we suspected would display little, if any, subject-level dependence. To investigate this issue more explicitly, we conducted some initial analyses that used Fisher exact tests looking for dependence between and the direction of difference (positive/negative) across parameter settings. Analyses conducted across 17 parameter settings (applied for the four different software packages and speaker groups) found only one instance of statistically detectable speaker effects related to the directional difference. Thus, our assumption to regard differences measurements from the same speaker as independent observations appeared reasonable. Next, we used the nonparametric Wilcoxon signed-ranks test (instead of a parametric paired *t* test), because the normality assumption

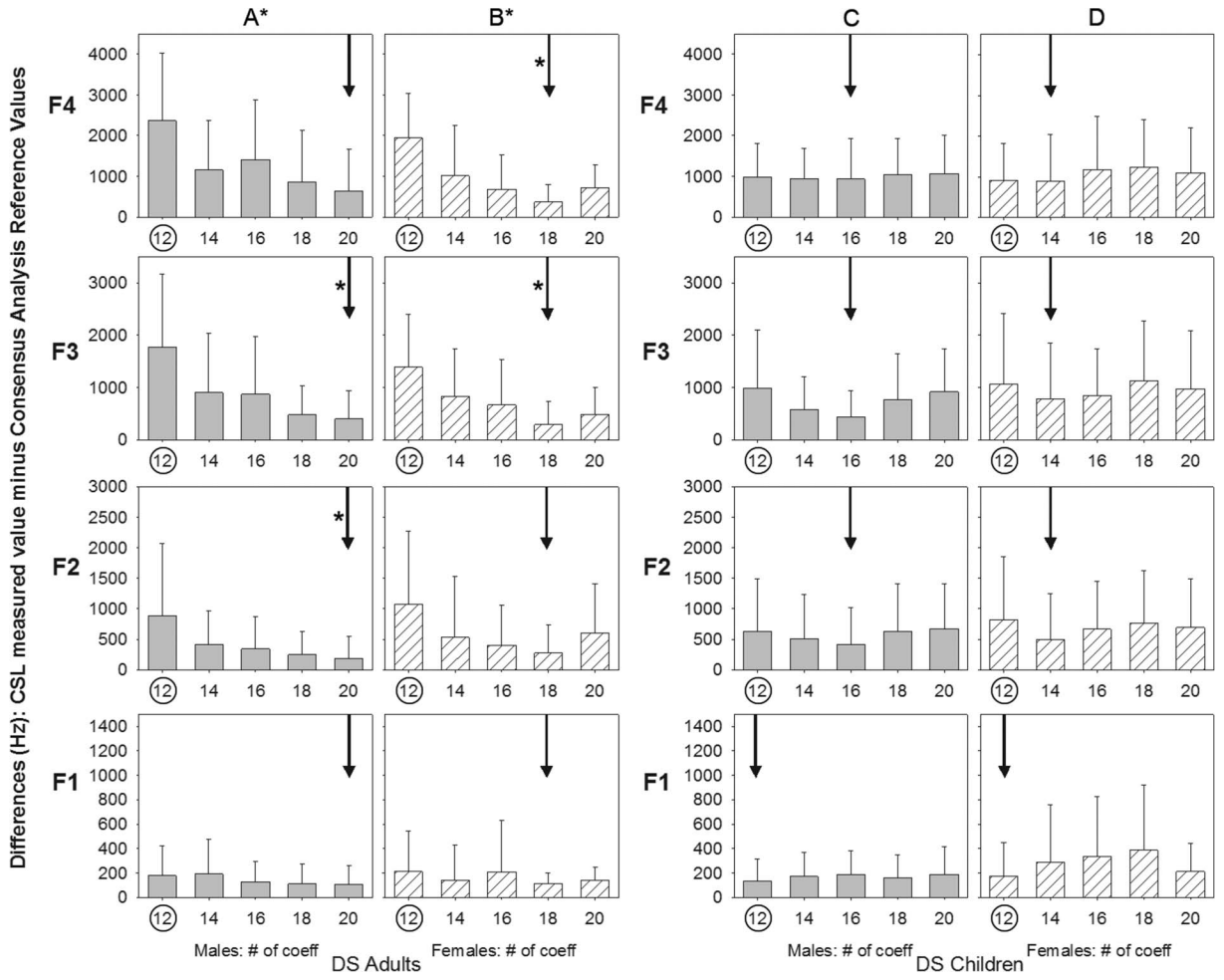
**Figure 2.** Difference values for formant frequencies (F1–F4) using Computerized Speech Laboratory for typically developing (TD) male adults (panel A), TD female adults (panel B), TD male children (panel C), and TD female children (panel D). The vertical bar charts display the mean absolute difference (calculated by subtracting the consensus analysis reference values (CARVs) from Speech Acoustic Software Package (SAASP)-generated formant measurement at default and with all permissible parameter manipulations) pooled across the four corner vowels with the simple error bar representing the standard deviation from the mean. The circled number on the *x*-axis indicates the default analysis parameter and the arrow the optimal analysis parameter (the smallest difference between the SAASP-measured value and the CARVs). Statistical significance between default and optimal analysis parameter is denoted by an asterisk (*), where an asterisk to the left of the arrow indicates statistically significant differences for the individual formant, and an asterisk next to the panel letter indicates statistically significant differences across all four formants.




for the differences appeared implausible. The results are summarized in Tables A2–A5, for all eight speaker groups with statistically significant differences in measurement accuracy denoted by an asterisk.

For CSL, Figure 2 displays differences values for TD male (panel A) and female (panel B) adults, and TD male (panel C) and female (panel D) children for F1–F4; and Figure 3 displays differences in F1–F4 measurements for male adults with DS (panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). The circle around 12 on the *x*-axis indicates that 12 is the default analysis parameter, and the arrows above each of the vertical bar charts indicate the optimal analysis parameter. For TD adults and adults with DS, increasing the number of coefficients and smoothing improves formant-frequency measurements across the four formants. Figures 2 and 3 show that increasing the number of LPC coefficients for adult speakers results in a systematic reduction in the mean absolute difference across the four formants (i.e., a decrease in the heights of the vertical bars). This improvement in measurement accuracy in CSL as tested by the Wilcoxon signed-ranks test was statistically significant for all adult speaker groups (summarized in Table A2): TD male adults ($Z = -9.433$; $p < .001$); TD female adults ($Z = -3.206$; $p = .001$); male adults with DS ($Z = -5.079$; $p < .001$); and female adults with DS ($Z = -4.948$; $p < .001$), with notable improvements for DS speakers, specifically F2 and F3 measurements for male adults with DS (F2: $Z = -2.651$; $p = .008$; F3: $Z = -3.360$; $p = .001$) and F3 and F4 measurements for female adults

**Figure 3.** Difference values for formant frequencies (F1–F4) using Computerized Speech Laboratory for male adults with Down syndrome (DS; panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.

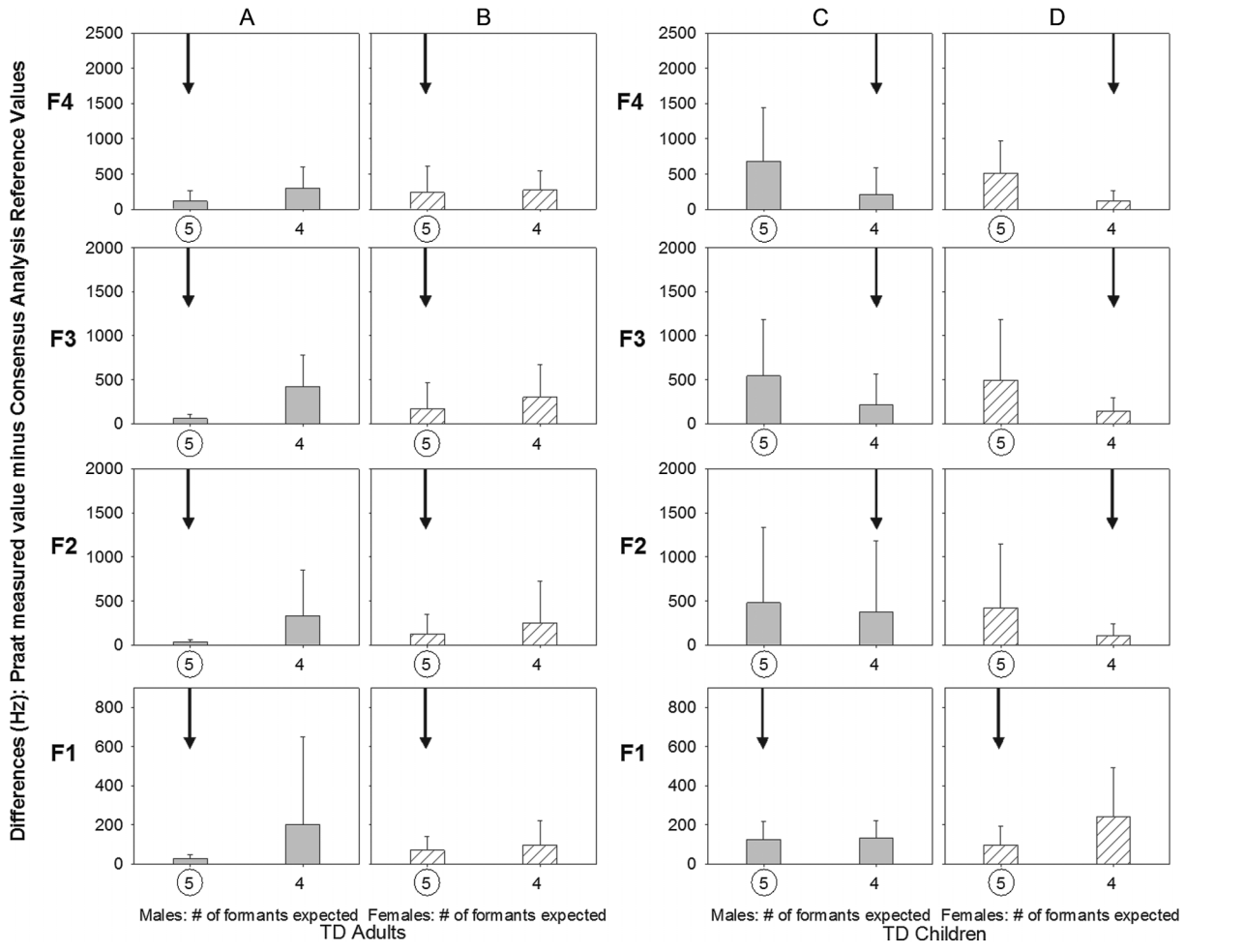



with DS (F3: $Z = -3.375$; $p = .001$; F4: $Z = -2.824$; $p = .005$). For children, although there was no systematic decrease in the mean absolute difference as analysis parameters were changed, there was a decrease at particular analysis settings that varied by speaker group and formant. For example, Figure 2, panel C, shows that TD child male speakers' F2–F4 measurements improve somewhat when the number of coefficient setting is 18. However, improvements in measurement accuracy were not statistically significant for any of the child speaker groups across the four formants or for the individual formants (see Table A2).

Results from Praat (see Figures 4 and 5, summarized in Table A3) show that default analysis parameters performed as well as any other parameter values for all adult groups. For children—both male and female—there is a systematic reduction in the mean absolute difference for F2–F4 measurements. Although statistical assessment of improvements in measurement accuracy was not significant for the individual formants, it was significant across the four formants for both male and female children with DS (boys: $Z = -2.045$; $p = .041$; girls: $Z = -2.867$; $p = .004$).

TF32 results (see Figures 6 and 7, summarized in Table A4) show that the direction of manipulation to the number of analysis coefficients does not yield a consistent effect on measurement accuracy. For TD male and female adults, a uniform effect is not observed, but for TD male and female children, the mean absolute differences are generally smaller with decreases in the number of coefficients. In a similar manner, a uniform pattern does not emerge for male and female adults with DS, but decreasing the number of coefficients usually results in smaller mean absolute differences for male and female children with DS. Manipulations to analysis parameters, irrespective of direction of change in number of coefficients, appear to be most effective across the four formants for the TD female adults (TD female adults: $Z = -3.430$; $p = .001$) and the child speaker groups, particularly the children with DS (male children: $Z = -2.721$; $p = .007$; and female children:

**Figure 4.** Difference values for formant frequencies (F1–F4) using Praat for typically developing (TD) male adults (panel A), TD female adults (panel B), TD male children (panel C), and TD female children (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




$Z = -3.034$; $p = .002$). Additional improvements on the individual formants were noted for F1 and F3 for TD female adults (F1: $Z = -3.335$; $p = .001$; F3: $Z = -2.764$; $p = .006$), and F3 for female children with DS ($Z = -2.999$; $p = .003$).

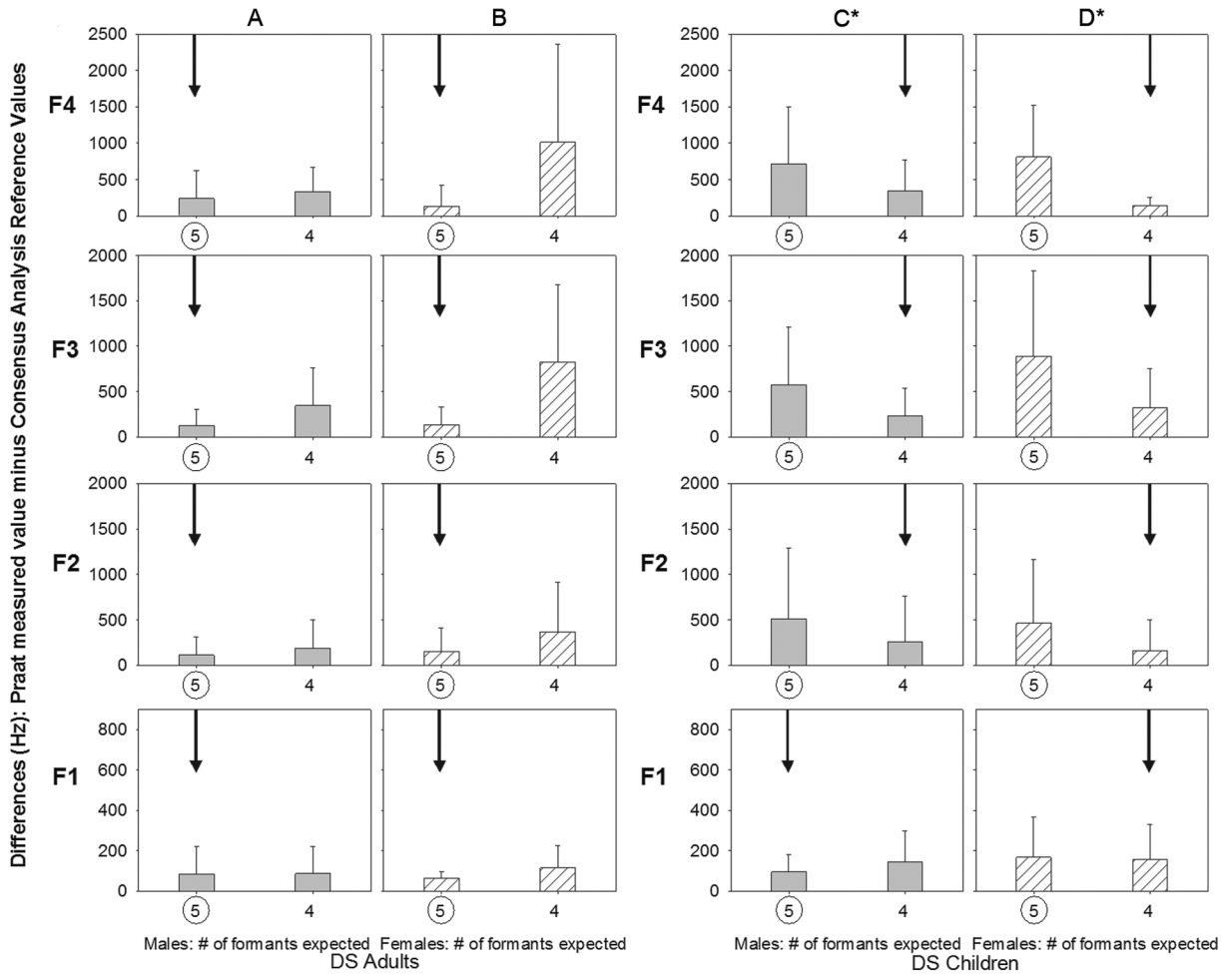
Results from WaveSurfer (see Figures 8 and 9, summarized in Table A5) show that in general, except for TD male adults, measurement accuracy is not enhanced with manipulations of default settings across speaker groups and the four formants. That is, the mean absolute differences are the smallest at the default settings and either remain stable or increase as the number of analysis coefficients is manipulated. An exception to this applies to all the formants in TD adult male speakers, in that manipulations yielded minor improvements in measurement accuracy. However, these improvements were not statistically significant.

The F0 measurements made manually and obtained from each SAASP are presented in Table A6, which shows that in general, across speaker groups the SAASPs-generated F0 measurements are comparable to the manual measurements. These data confirm the expected differences in F0 among speaker groups, especially between children and adults and between male and female adults. The most divergent results are seen for TD children and children with DS.

## Discussion

This study systematically assessed the effects of commonly reported analysis parameter manipulations on vowel formant measurements for diverse speaker groups using four SAASPs. The results confirmed the expectation that parameter manipulations can lead to enhanced formant-frequency measurements for some speaker/formant combinations. However, the benefit was not uniform across SAASPs, speakers, and formants. In general, manipulations to CSL were effective for all adult speaker groups, both TD and those with DS. Manipulations to TF32 yielded more

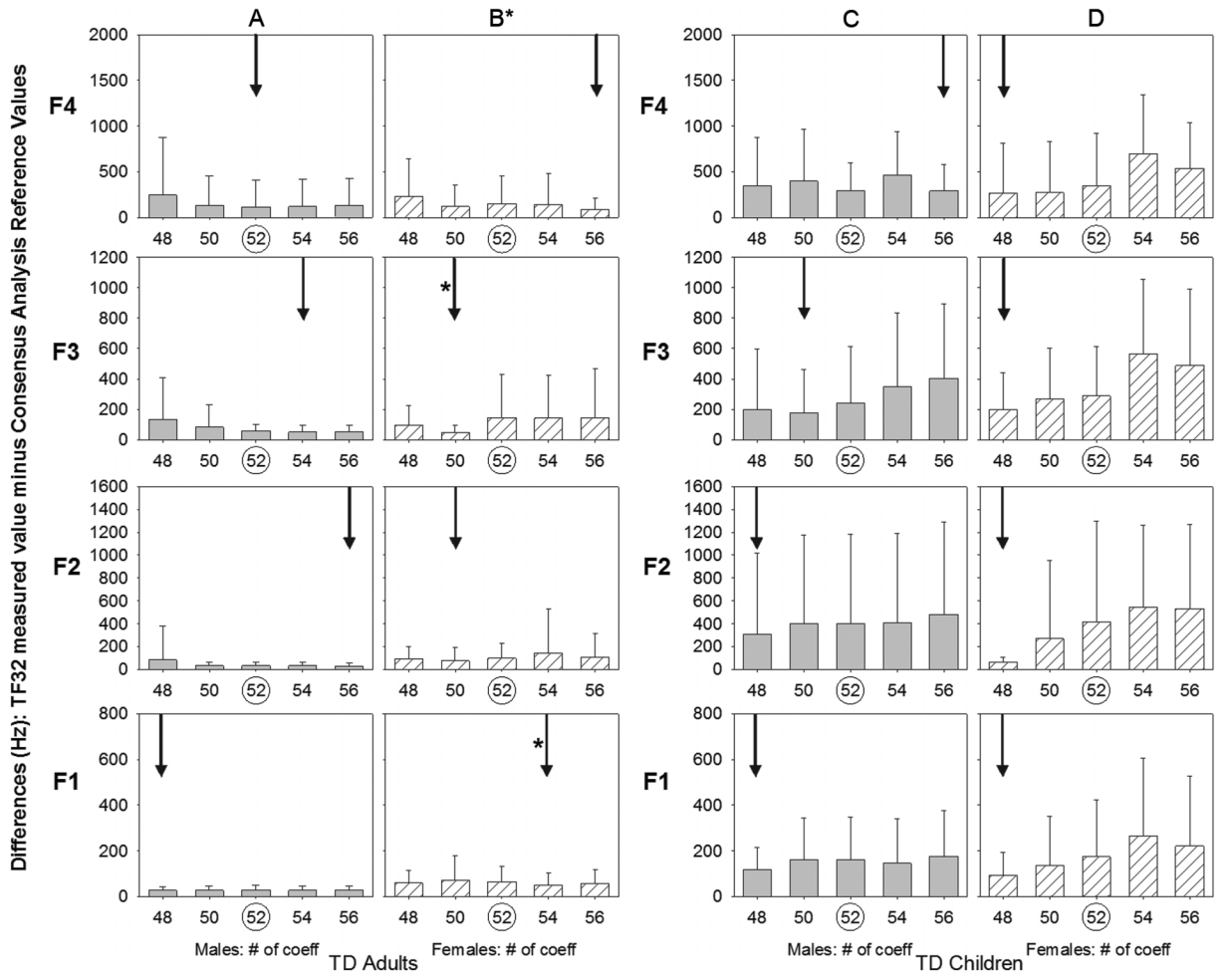
**Figure 5.** Difference values for formant frequencies (F1–F4) using Praat for male adults with Down syndrome (DS; panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




accurate formant-frequency measurements for TD female adults and children with DS—male and female. In Praat, manipulations improved analysis of speech for children with DS, both male and female. Although most speaker groups did not benefit from manipulations in WaveSurfer, TD male adults had a slight improvement with an increase in coefficients. Overall, parameter manipulations were effective in improving the accuracy of formant measurements, but only for select speaker groups, and mostly F2, F3, and F4. This outcome was unexpected given that the general advice provided to users of the SAASP is to alter the default analysis settings for diverse speaker groups such as those studied here.

The results from this study provide general guidelines for individuals who use the SAASP to make vowel formant measurements. To achieve the most accurate results, users should manipulate the analysis settings for the speaker and formant(s) of interest, while also taking into account the 10 acoustic analysis caveats listed in the following section. In a study directed to forensic applications, Harrison (2013) drew a similar conclusion: "The guidance [to the user] suggested that understanding the principles of LPC analysis, how it was implemented in specific software and the influence of analysis parameters were important when making formant measurements" (p. 299). The number of such manipulations or adjustments varies considerably across SAASPs but usually includes the analysis bandwidth used in spectrograms and the number of coefficients used in LPC analysis. It should be noted that SAASPs that require few if any changes in these settings are not necessarily more accurate than other systems, but only that manipulations make little improvement towards measurement accuracy. The use of the consensus analysis procedure implemented here (that takes into account the caveats listed in the following section), or an approach similar to it, is a more accurate method of making formant measurements than relying solely on the LPC values automatically generated by SAASPs. The consensus analysis method relies on the user's knowledge

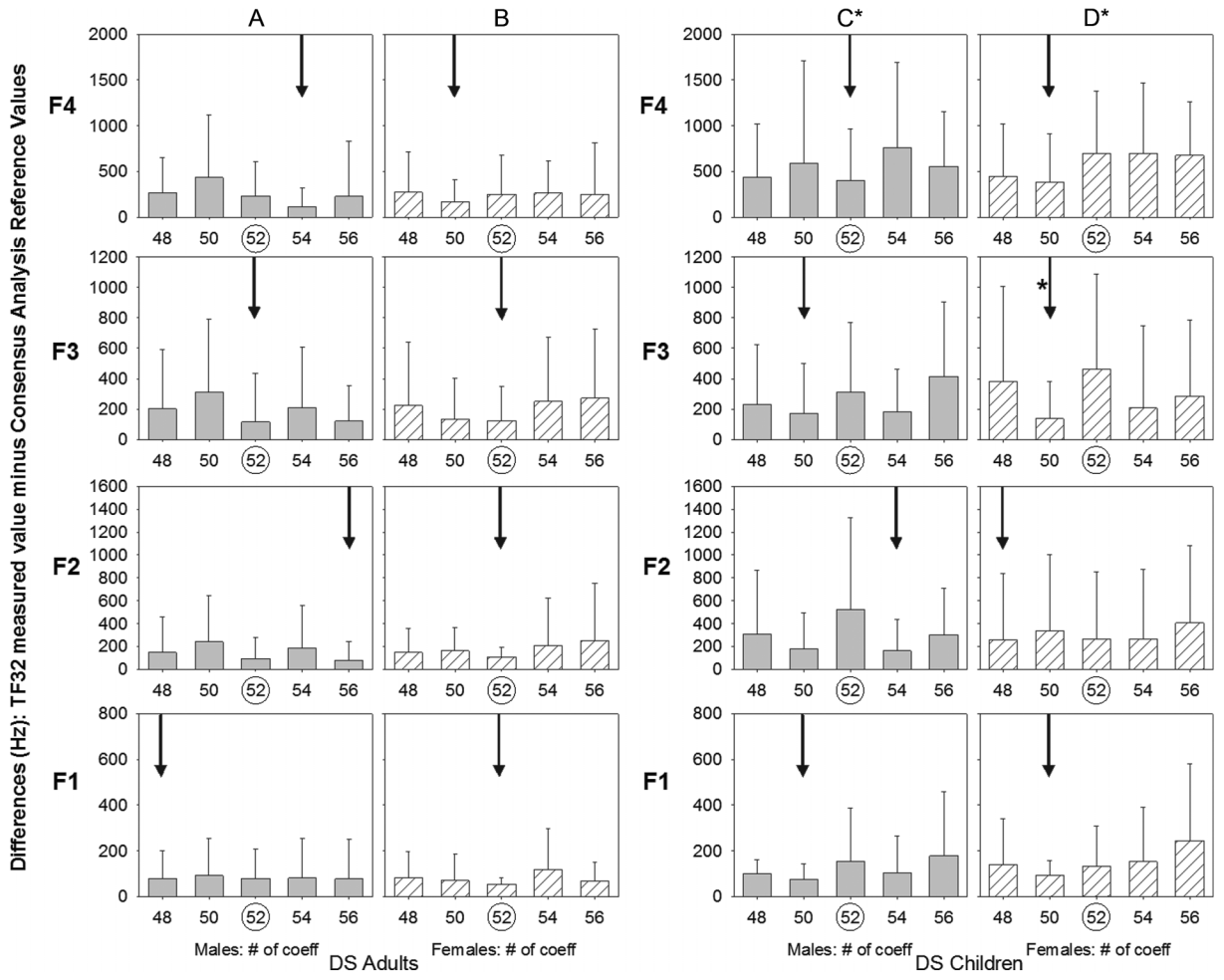
**Figure 6.** Difference values for formant frequencies using TF32 for TD male adults (panel A), typically developing (TD) female adults (panel B), TD male children (panel C), and TD female children (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




of speech acoustics as well as on a combination of SAASP displays—spectrogram, LPC formant tracks, LPC and FFT spectrum—to obtain accurate formant-frequency measurements. However, this method of formant measurement may not be suited to a clinical setting or other settings where users have limited time to examine multiple acoustic displays to determine an optimal analysis. The convenience of automatic formant measurements by the SAASP needs to be balanced against tolerance for error. In the case of formant bandwidth measurements, Burris et al. (2014) concluded that the errors are too large to be acceptable. Tolerance for errors in formant frequencies depends on the specific application. For purposes such as acoustic documentation of the outcome of clinical intervention (or worsening of a speech disorder as in the case of degenerative diseases), it is important that analysis tools are both easy to use and sensitive to changes in formant pattern particularly given the increased use of vowel formant measures in speech-language pathology. Findings here reveal that the current LPC algorithms do not appear to yield optimal F1–F4 measurements for all of the speaker groups examined in this study, as judged by the discrepancies between LPC values and the CARVs. Thus, this study and a related study by Burris et al. (2014) addressing aspects of the accuracy and efficiency of acoustic analysis for applications in clinical and research settings highlight a definite need for the future development of speech acoustic analysis approaches including the development of LPC algorithms that can be used with the speech patterns of children and individuals with disordered speech.

Until such advances are made in acoustic analysis approaches, it is important that all current users of these systems recognize that default analysis parameters do not necessarily yield optimal formant-frequency measurements in all SAASPs and for all speakers and that manipulations to analysis parameters may need to be made depending on the type of speaker being analyzed, the SAASP utilized, and even the particular formant(s) of interest. Aside from

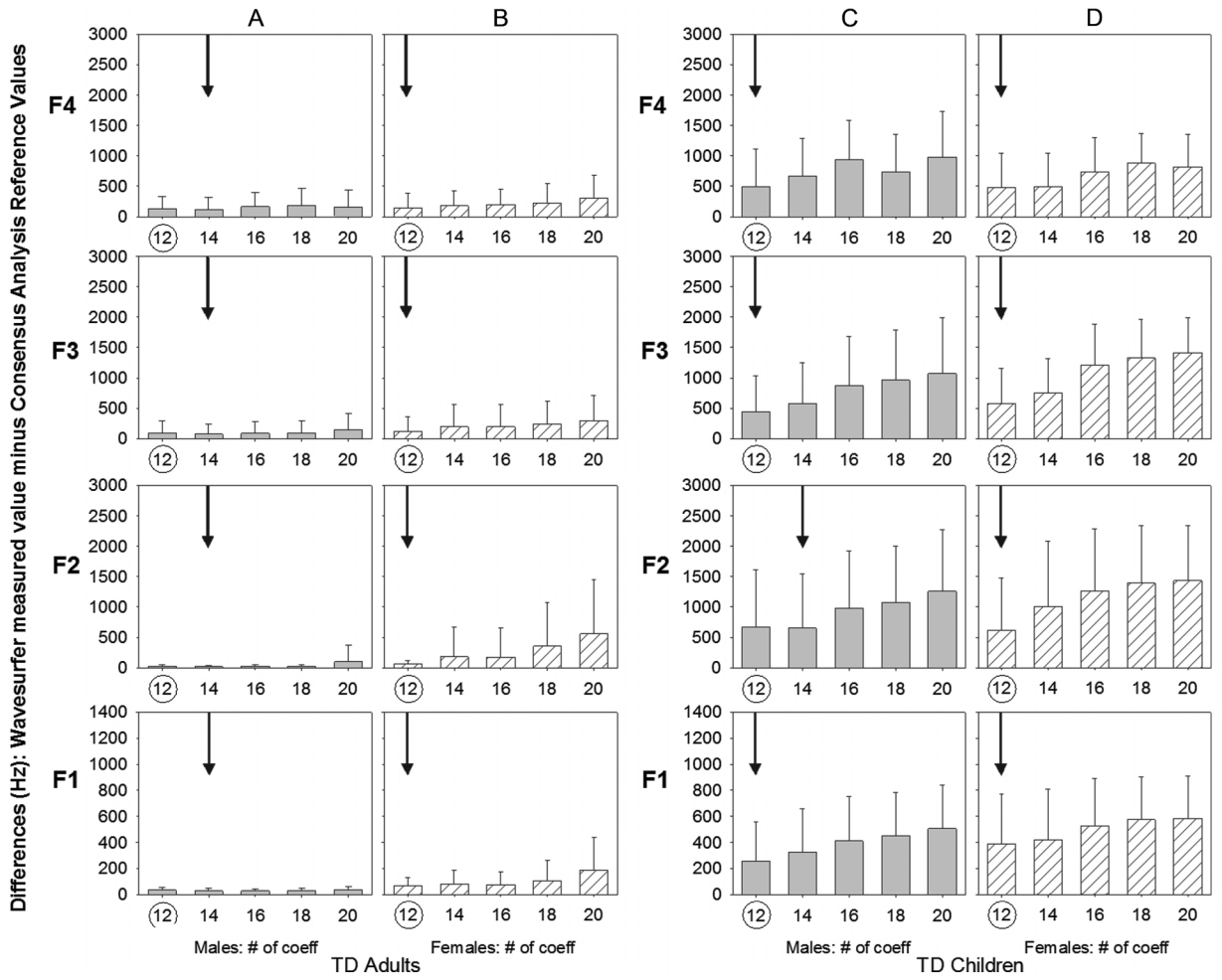
**Figure 7.** Difference values for formant frequencies (F1–F4) using TF32 for male adults with Down syndrome (DS; panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




the general guidelines that this study provides to making formant measurements, what follows are some additional pointers to making more accurate and representative vowel formant measurements.

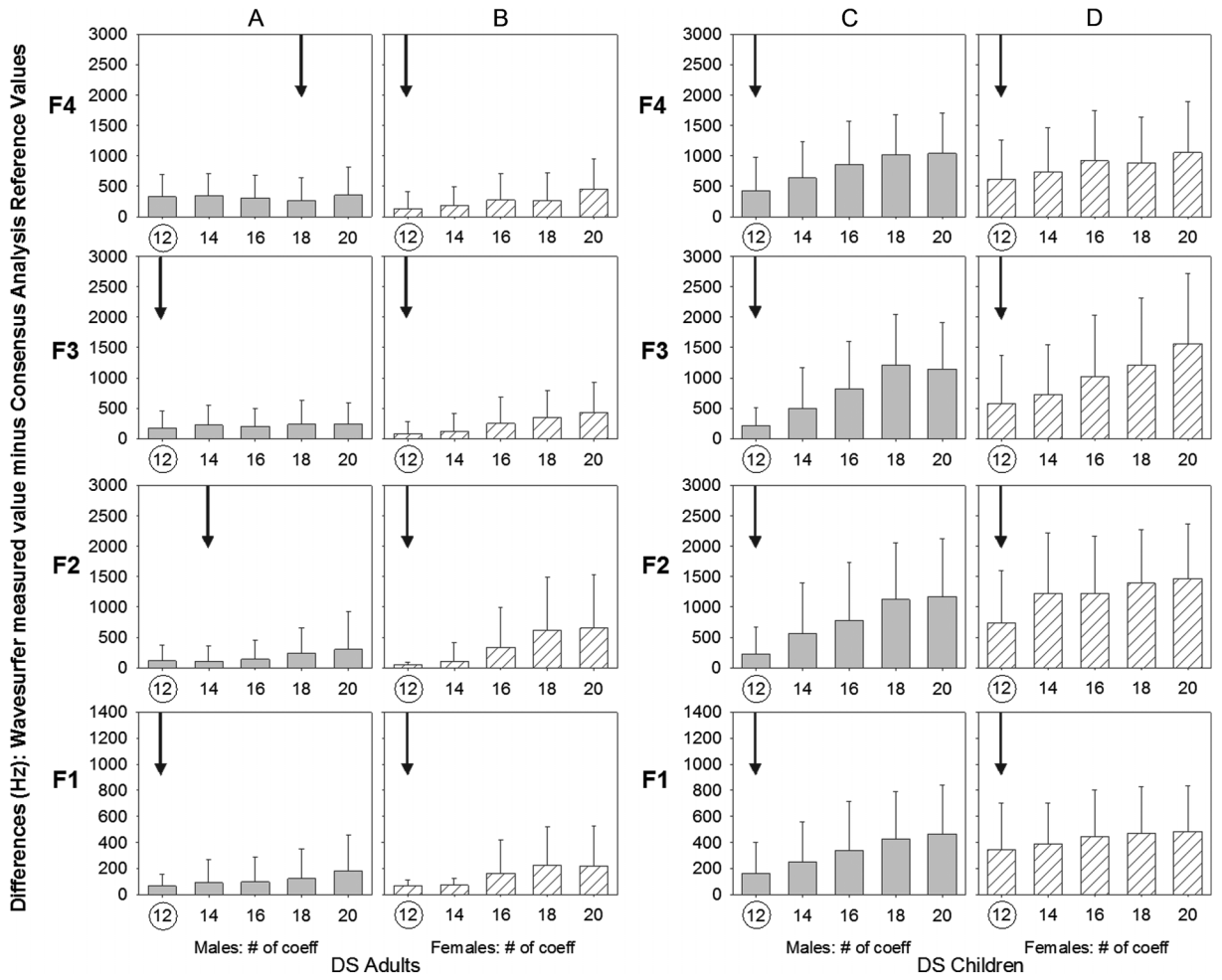
Although measurement of vowel formants may seem to be one of the most straightforward types of acoustic analysis of speech, there are 10 caveats to be noted: (1) The time point for analysis should be selected according to the purpose of the study. Although the temporal midpoint of a segmented vowel is relatively easy to select and is reliable for reanalysis, it may not be representative. For example, even though /u/ and /ae/ are traditionally classified as monophthongs, they often are produced as diphthongs, especially in the Midwestern dialect. If the goal is to capture the articulatory-acoustic extremes of the vowel quadrilateral, then the vowel midpoint is not necessarily suitable. For /u/, the F2 frequency regularly falls during the vowel, reaching its lowest value at the end of the segment. For /ae/, breaking is common so that an initial /æ/ gives way to a low-back vowel, such as /ɑ/. Therefore, it is important to establish criteria for selecting the point of formant-frequency measurement. Some possible criteria are the following: vowel /i/ —point of highest frequency of F2; vowel /u/—point of lowest frequency of F2; vowel /ɑ/—point of least separation of F1 and F2 frequencies; and vowel /æ/—point of most evenly spaced formants, taking care to avoid measurement at a point of decreasing F2–F1 difference (which reflects backing of the vowel). Criteria of this kind are particularly important in obtaining formant-frequency values for computation of vowel space area or similar metrics. (2) For all its power, LPC often fails to resolve formants for some speakers. Errors are especially likely for back vowels in which LPC may not distinguish closely spaced F1 and F2. A similar problem can occur for vowel /i/, in which F2 and F3 are close in frequency and therefore not always resolved by LPC. To guard against these and other errors, it is helpful to compare the LPC formant tracks with a wide-band FFT spectrogram. Most

**Figure 8.** Difference values for formant frequencies (F1–F4) using WaveSurfer for typically developing (TD) male adults (panel A), TD female adults (panel B), TD male children (panel C), and TD female children (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




SAASPs facilitate this process, as the LPC formant tracks are superimposed on a spectrogram. In cases of ambiguity, an additional step is to compare LPC and narrow-band FFT spectra for selected time points in the formant patterns. (3) If the LPC analysis appears to miss a formant, as in the case of F1 and F2 for back vowels just discussed, the analysis often can be improved by increasing the filter order (number of coefficients) in small increments depending on the choices permitted in a particular SAASP. If the LPC analysis appears to identify a spurious formant (such as may occur with interformant energy or a strong harmonic), then decreasing the filter order may be useful. (4) The analyzing filter of the FFT spectrogram should be matched to speaker characteristics. A bandwidth of 300 Hz generally works well for male adults, but a larger bandwidth of 400 or 500 Hz is often more effective in formant displays for female adults or children. With very high F0, the formant–harmonic interaction may be evident when individual harmonics are analyzed even when a wideband filter is used. (5) The higher formants F3 and F4 can be difficult to measure with either LPC or FFT, especially if the energy is weak or if there is substantial noise energy associated with voice qualities of breathiness or hoarseness. It can be helpful to inspect the overall pattern of the vowel, including onset and offset, for evidence of these formants. Because formants generally change gradually during a syllable nucleus, it often is possible to discern formant movements during the syllable. (6) If there is very little energy in the area of the higher formants, one can increase the dynamic range of the spectrogram so that a greater range of energies is displayed. However, such an adjustment may result in a less favorable signal-to-noise ratio. (7) It is important to note any perceptual or acoustic evidence of nasalization during the vowel. Most LPC algorithms are based on an all-pole model, which is not suited to oral–nasal resonance. (8) The spectrogram can be helpful in detecting changes

**Figure 9.** Difference values for formant frequencies (F1–F4) using WaveSurfer for male adults with Down syndrome (DS; panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to the Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter, and statistical significance.




in phonation that can affect the spectral analysis. Such changes include diplophonia, voice breaks, intervals of roughness or breathiness, or loss of periodicity. Inspection of a narrow-band spectrogram to display harmonic structure may be helpful in noting changes in phonation. (9) Formant bandwidths derived from LPC should be regarded with skepticism; it is wise to confirm these values with another method, such as the manual method used by Burris et al. (2014). (10) If measurements are made of F4, care should be taken to ensure that the frequency range of analysis is adequate to include the F4 frequency, which can approach 5 kHz in female adults and children.

Above all, the user of an SAASP should keep in mind the expected formant values for a given speaker, taking into account both sex and age. If a measured formant frequency is at odds with expectations on the basis of normative data, the analysis should be reconsidered. Duckworth et al. (2011) made a similar recommendation in formant analysis for forensic purposes. Our results showing that the most effective manipulations of acoustic analysis parameters are specific to both speaker and formant confirm earlier reports on typical speech (Harrison, 2013; Yao et al., 2010) and extend that conclusion to the disordered speech in individuals with DS. Additional studies with other speaker groups, especially groups with other speech disorders, are needed to assess the effect of analysis parameter settings. An additional need is establishing a developmental normative database, particularly one that includes the higher formants to serve both as a clinical reference as well as capture developmental trends. The higher formants F3 and F4, though often neglected in studies of developing and disordered speech, convey important information on vocal tract shape. For example, it is well known that a decrease in F3 frequency is an acoustic cue for /r/ and /r/-colored vowels (Hagiwara, 1995; Hamilton, Boyce, Scholl, & Douglas, 2014), and the center frequency of F4 has been associated with

the hypopharyngeal cavity (Takemoto, Adachi, Kitamura, Mokhtari, & Honda, 2006), a cavity that undergoes significant growth (Vorperian et al., 2009), and may be affected in various craniofacial morphologies. The present study included measurements of the first four formants, indicating that establishing a database of F1-F2-F3-F4 formant-frequency values for male and female speakers of different ages is an achievable goal, one that will facilitate the clinical application of acoustic methods.

Clinicians and others now have access to powerful software available for free or at modest cost that permits efficient data collection on various acoustic features. But accuracy and reliability cannot be assumed and must be demonstrated under conditions that are clinically relevant. The present study represents the first empirical assessment of the effect of analysis parameter manipulations on the accuracy of formant measurements when analyzing speech from children, female adults, and a population with speech disorders. We hope that this article encourages further tests with more diverse populations and the establishment of a normative developmental database that includes higher formants. In addition, the results of the present study, along with those of Burris et al. (2014), should be helpful in the development and refinement of systems or software for the acoustic analysis of speech. Particular needs are for improved robustness of analysis for diverse speaker groups, increased accuracy of formant bandwidths, greater convenience in generating multiple displays such as those used in the consensus analysis of this report, and ease of use particularly by clinicians.

## Acknowledgments

This work was supported by National Institute of Health Research Grant R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract) from the National Institute on Deafness and other Communicative Disorders (NIDCD) and by a core grant P-30 HD03352 to the Waisman Center from the National Institute of Child Health and Human Development (NICHHD). We declare no financial conflict of interest with any of the software systems considered in this study. We thank Drs. Jan R. Edwards and Gary G. Weismer for their suggestions on the design and conduct of this research. We also are indebted to Peggy Rosin, Erin Douglas, Carlyn Burris, Erin Nelson, Alyssa Wild, Michael P. Kelly, and Elaine Romenesko for their help at various stages of this project and the anonymous reviewers for comments on earlier versions of this article. This research was originally submitted by the first author as a master's thesis for the Department of Communication Sciences and Disorders at the University of Wisconsin–Madison. Portions of this research were presented in 2013 at the American Speech-Language-Hearing Association Convention in Chicago, IL.

## References

**Boersma, P., & Weenink, D.** (2010). *Praat* (Version 5.1.31). Available from http://www.fon.hum.uva.nl/praat

**Bunton, K., & Leddy, M.** (2011). An evaluation of articulatory working space in vowel production of adults with Down syndrome. *Clinical Linguistics & Phonetics, 25,* 321–334.

**Burris, C., Vorperian, H. K., Fourakis, M., Kent, R. D., & Bolt, D. M.** (2014). Quantitative and descriptive comparison of four acoustic analysis systems: Vowel measurements. *Journal of Speech, Language, and Hearing Research, 57,* 26–45.

**Chenausky, K., MacAuslan, J., & Goldhor, R.** (2011). Acoustic analysis of PD speech. *Parkinson's Disease, 2011.* doi:10.4061/2011/435232

**Deng, L., & Dang, J.** (2007). Speech analysis: The production-perception perspective. In C.-H. Lee, H. Li, L.-S. Lee, R.-H. Wang, & Q. Huo (Eds.), *Advances in Chinese spoken language processing* (pp. 3–32). Singapore: World Scientific Publishing Co.

**Diehl, J. J., & Paul, R.** (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics, 34,* 135–161.

**Duckworth, M., McDougall, K., de Jong, G., & Shockey, L.** (2011). Improving the consistency of formant measurement. *International Journal of Speech, Language, and the Law, 18,* 35–51.

**Dunn, O. J.** (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56,* 52–64.

**Fant, C. G. M.** (1960). *Acoustic theory of speech production*. The Hague, the Netherlands: Mouton Publishers.

**Fox, R. A., & Jacewicz, E.** (2009). Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America, 126,* 2603–2618.

**Gold, B., & Rabiner, L. R.** (1968). Analysis of digital and analog formant synthesizers. *IEEE Transactions on Audio and Electroacoustics, 16,* 81–94.

**Hagiwara, R.** (1995). *Acoustic realizations of American /r/ as produced by women and men* (Unpublished doctoral thesis). University of California at Los Angeles.

**Hamilton, S., Boyce, S., Scholl, L., & Douglas, K.** (2014). An acoustic threshold for third formant in American English. *The Journal of the Acoustical Society of America, 135,* 2389.

**Harrison, P. T.** (2013). *Making accurate formant measurements: An empirical investigation of the influence of the measurement tool, analysis settings and speaker on formant measurements* (Unpublished doctoral dissertation). University of York, York, England.

**Higgins, C. M., & Hodge, M. M.** (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech-Language Pathology, 10,* 271–277.

**Hillenbrand, J. M.** (1995). *James M. Hillenbrand*. Retrieved from http://homepages.wmich.edu/~hillenbr/voweldata.html

**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,* 3099–3111.

**Hodge, M., Gotzke, C., & Daniels, J.** (2009). *TOCS+ Project*. Retrieved from http://www.tocs.plus.ualberta.ca/

**Hustad, K. C., Gorton, K., & Lee, J.** (2010). Classification of speech and language profiles in 4-year-old children with cerebral palsy: A prospective preliminary study. *Journal of Speech, Language, and Hearing Research, 53,* 1496–1513.

**KayPentax.** (1996). *Computerized Speech Lab* (500). Lincoln Park, NJ: Author. Available from http://www.kayelemetrics.com/index.php?option=com_product &controller=product&Itemid=3&cid[]=11&task=pro_details

**Kent, R. D., & Read, C.** (2001). *Acoustic analysis of speech* (2nd ed.). San Diego, CA: Singular.

**Kent, R. D., & Vorperian, H. K.** (2013). Speech impairments in Down syndrome: A review. *Journal of Speech, Language, and Hearing Research, 56,* 1044–1092.

**Lansford, K. L., & Liss, J. M.** (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research, 57,* 57–67.

**Mahler, L. A., & Ramig, L. O.** (2012). Intensive treatment of dysarthria secondary to stroke. *Clinical Linguistics & Phonetics, 26,* 681–694.

**Mehta, D. D., & Wolfe, P. J.** (2015). Statistical properties of linear prediction analysis underlying the challenge of formant bandwidth estimation. *The Journal of the Acoustical Society of America, 137*(2), 944–950.

**Milenkovic, P.** (2010). *TF32* (Alpha). Available from http://userpages.chorus.net/cspeech/

**Moura, C. P., Cunha, L. M., Vilarinho, H., Cunha, M. J., Freitas, D., Palha, M., . . . Pais-Clemente, M.** (2008). Voice parameters in children with Down syndrome. *Journal of Voice, 22,* 34–42.

**Olive, J. P.** (1971). Automatic formant tracking by a Newton-Raphson technique. *The Journal of the Acoustical Society of America, 50,* 661–670.

**Roy, N., Nissen, S. L., Dromey, C., & Sapir, S.** (2009). Articulatory changes in muscle tension dysphonia: Evidence of vowel space expansion following manual circumlaryngeal therapy. *Journal of Communication Disorders, 42,* 124–135.

**Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., . . . Ruzicka, E.** (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease. *The Journal of the Acoustical Society of America, 134,* 2171–2181.

**Sapir, S., Polczynska, M., & Tobin, Y.** (2009). Why does the vowel space area as an acoustic metric fail to differentiate dysarthric from normal vowel articulation and what can be done about it? *Poznan Studies in Contemporary Linguistics, 45,* 201–311.

**Saz, O., Simon, J., Rodriguez, W.-R., Lleida, E., & Vaquero, C.** (2009). Analysis of acoustic features in speakers with cognitive disorders and speech impairments. *EURASIP Journal on Advances in Signal Processing, 2009,* 159234.

**Sjolander, K., & Beskow, J.** (2005). WaveSurfer (1.8.5) [Computer software]. Stockholm, Sweden: Centre for Speech Technology at KTH. Available from http://www.speech.kth.se/wavesurfer/

**Sjolander, K., & Beskow, J.** (2010). *WaveSurfer* (1.8.8). Available from http://sourceforge.net/projects/wavesurfer/

**Skodda, S., Gronheit, W., & Schlegel, U.** (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS ONE 7*(2): e32132. doi:10.1371/journal.pone.0032132

**Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., & Honda, K.** (2006). Acoustic roles of the laryngeal cavity in vocal tract resonance. *The Journal of the Acoustical Society of America. 120,* 2228–2238.

**Vallabha, G. K., & Tuller, B.** (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication, 38,* 141–160.

**Vallabha, G., & Tuller, B.** (2004). Choice of filter order in LPC analysis of vowels. In J. Slifka, S. Manuel, & M. Matthies (Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. B148–B163). Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology.

**Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., . . . Gentry, L. R.** (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America, 125,* 1666–1678.

**Wenke, R. J., Cornwell, P., & Theodoros, D. G.** (2010). Changes to articulation following LSVT and traditional dysarthria in non-progressive dysarthria. *International Journal of Speech-Language Pathology, 12,* 203–220.

**Yao, Y., Tilsen, S., Sprouse, R. L., & Johnson, K.** (2010). Automated measurement of vowel formants in the Buckeye Corpus. *UC Berkeley Phonology Lab Annual Report, 2010,* 80–94.

## Appendix A

**Table A1.** A summary of each Speech Acoustic Analysis Software Package's default settings in relation to permissible parameter manipulations, as well as the sequential order in which manipulations were made as denoted by column number: 1 = first parameter manipulated; 2 = second, and 3 = last. CSL = Computerized Speech Laboratory.

| Manipulation sequence | 1 | 2 | | 3 | |
|---|---|---|---|---|---|
| | Downsampling and other | # of coefficients | | Smoothing | |
| | | Default | Manipulation | Default | Manipulation |
| CSL | Must manually downsample to 16 kHz. Must add voice period marks. | 12 | Yes (12–20) | No | Yes |
| Praat | Automatic downsampling to 10 kHz. | 10* | No | N/A | No |
| | | * Indirectly changed # of coefficients to 8 by changing expected number of formants from default of 5 to 4 formants. | | | |
| TF32 | No downsampling. Number of coefficients is adjusted automatically to the sampling rate. | 52 | Yes (48–56) | No | Yes |
| WaveSurfer | Automatic downsampling to 10 kHz. | 12 | Yes (12–20) | Yes | No |

**Table A2.** Results from Computerized Speech Laboratory are presented per speaker group, where the underlined text indicates the default analysis parameter and a plus sign (+) indicates the optimal parameter per formant, and across the four formants where optimal per formant is defined as the smallest difference between the Speech Acoustic Analysis Software Package's measured value from the consensus analysis reference values; across the four formants, optimal parameter is defined as the particular analysis setting that was optimal for at least two individual formants. Asterisk denotes statistical significance between default and optimal analysis parameter assessed using the Wilcoxon signed rank sum test.

| Speaker group | # of coefficients | 12 | 14 | 16 | 18 | 20 | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| Typically developing | | | | | | | | |
| Male adults | F4 | | | | + | | –3.771 | .000* |
| | F3 | | | | | + | –4.967 | .000* |
| | F2 | | | | | + | –5.282 | .000* |
| | F1 | | | | | + | –4.852 | .000* |
| | F4–F1 | | | | | + | –9.433 | .000* |
| Female adults | F4 | | | | + | | –1.999 | .046 |
| | F3 | | | | + | | –2.470 | .014 |
| | F2 | | | | + | | –1.331 | .183 |
| | F1 | | | | + | | –0.887 | .375 |
| | F4–F1 | | | | + | | –3.206 | .001* |
| Male children | F4 | | | | + | | –0.621 | .535 |
| | F3 | | | | + | | –0.893 | .372 |
| | F2 | | | | + | | –0.299 | .765 |
| | F1 | | + | | | | –0.187 | .852 |
| | F4–F1 | | | | + | | –0.585 | .559 |
| Female children | F4 | | | | | + | –0.534 | .594 |
| | F3 | | | + | | | –0.187 | .852 |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Down syndrome | | | | | | | | |
| Male adults | F4 | | | | | + | –2.310 | .021 |
| | F3 | | | | | + | –3.360 | .001* |
| | F2 | | | | | + | –2.651 | .008* |
| | F1 | | | | | + | –1.344 | .179 |
| | F4–F1 | | | | | + | –5.079 | .000* |
| Female adults | F4 | | | | + | | –2.824 | .005* |
| | F3 | | | | + | | –3.375 | .001* |
| | F2 | | | | + | | –2.464 | .014 |
| | F1 | | | | + | | –0.597 | .550 |
| | F4–F1 | | | | + | | –4.948 | .000* |
| Male children | F4 | | | + | | | –0.893 | .372 |
| | F3 | | | + | | | –2.427 | .015 |
| | F2 | | | + | | | –1.083 | .279 |
| | F1 | + | | | | | | |
| | F4–F1 | | | + | | | –1.295 | .195 |
| Female children | F4 | | + | | | | –0.260 | .795 |
| | F3 | | + | | | | –0.597 | .550 |
| | F2 | | + | | | | –1.568 | .117 |
| | F1 | + | | | | | | |
| | F4–F1 | | + | | | | –0.708 | .479 |

*Note.* Statistical significance between default and optimal analysis parameter, assessed using the Wilcoxon signed-ranks test, is denoted by an asterisk (*).

**Table A3.** Results from Praat are presented per speaker group. Refer to the Table A2 caption for additional information regarding default analysis parameter and optimal analysis parameter.

| Speaker group | # of formants expected | 5 | 4 | *Z* | *p* |
|---|---|---|---|---|---|
| Typically developing | | | | | |
| Male adults | F4 | + | | | |
| | F3 | + | | | |
| | F2 | + | | | |
| | F1 | + | | | |
| | F4–F1 | + | | | |
| Female adults | F4 | | + | –2.042 | .041 |
| | F3 | + | | | |
| | F2 | + | | | |
| | F1 | + | | | |
| | F4–F1 | + | | | |
| Male children | F4 | | + | –0.973 | .331 |
| | F3 | | + | –1.023 | .306 |
| | F2 | | + | –0.037 | .970 |
| | F1 | + | | | |
| | F4–F1 | | + | –0.539 | .590 |
| Female children | F4 | | + | –1.689 | .091 |
| | F3 | | + | –1.083 | .279 |
| | F2 | | + | –0.523 | .601 |
| | F1 | + | | | |
| | F4–F1 | | + | –0.092 | .927 |
| Down syndrome | | | | | |
| Male adults | F4 | + | | | |
| | F3 | + | | | |
| | F2 | + | | | |
| | F1 | + | | | |
| | F4–F1 | + | | | |
| Female adults | F4 | + | | | |
| | F3 | + | | | |
| | F2 | + | | | |
| | F1 | + | | | |
| | F4–F1 | + | | | |
| Male children | F4 | | + | –1.293 | .196 |
| | F3 | | + | –2.240 | .025 |
| | F2 | | + | –1.381 | .167 |
| | F1 | + | | | |
| | F4–F1 | | + | –2.045 | .041* |
| Female children | F4 | | + | –1.481 | .139 |
| | F3 | | + | –2.277 | .023 |
| | F2 | | + | –1.717 | .086 |
| | F1 | | + | –0.933 | .351 |
| | F4–F1 | | + | –2.867 | .004* |

*Note*. Statistical significance between default and optimal analysis parameter, assessed using the Wilcoxon signed-ranks test, is denoted by an asterisk (*).

**Table A4.** Results from TF32 are presented per speaker group. Refer to the Table A2 caption for additional information regarding default analysis parameter and optimal analysis parameter.

| Speaker group | # of coefficients | 48 | 50 | 52 | 54 | 56 | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| Typically developing | | | | | | | | |
| Male adults | F4 | | | + | | | | |
| | F3 | | | | + | | –1.815 | .069 |
| | F2 | | | | | + | –1.038 | .299 |
| | F1 | + | | | | | –1.228 | .219 |
| | F4–F1 | | | | | | | |
| Female adults | F4 | | | | | + | –1.061 | .289 |
| | F3 | | + | | | | –2.764 | .006* |
| | F2 | | + | | | | –2.098 | .036 |
| | F1 | | | | + | | –3.335 | .001* |
| | F4–F1 | | + | | | | –3.430 | .001* |
| Male children | F4 | | | | | + | –0.511 | .609 |
| | F3 | | + | | | | –1.303 | .193 |
| | F2 | + | | | | | –0.348 | .727 |
| | F1 | + | | | | | –0.853 | .393 |
| | F4–F1 | + | | | | | –0.718 | .473 |
| Female children | F4 | + | | | | | –0.114 | .910 |
| | F3 | + | | | | | –0.840 | .401 |
| | F2 | + | | | | | –1.307 | .191 |
| | F1 | + | | | | | –0.829 | .407 |
| | F4–F1 | + | | | | | –1.259 | .208 |
| Down syndrome | | | | | | | | |
| Male adults | F4 | | | | + | | –0.795 | .426 |
| | F3 | | | + | | | | |
| | F2 | | | | | + | –0.141 | .888 |
| | F1 | + | | | | | –1.025 | .306 |
| | F4–F1 | | | | | | | |
| Female adults | F4 | | + | | | | –0.483 | .629 |
| | F3 | | | + | | | | |
| | F2 | | | + | | | | |
| | F1 | | | + | | | | |
| | F4–F1 | | | + | | | | |
| Male children | F4 | | | + | | | | |
| | F3 | | + | | | | –1.147 | .251 |
| | F2 | | | | + | | –1.428 | .153 |
| | F1 | | + | | | | –2.223 | .026 |
| | F4–F1 | | + | | | | –2.721 | .007* |
| Female children | F4 | | + | | | | –1.720 | .085 |
| | F3 | | + | | | | –2.999 | .003* |
| | F2 | + | | | | | –0.213 | .831 |
| | F1 | | + | | | | –0.710 | .478 |
| | F4–F1 | | + | | | | –3.034 | .002* |

*Note*. Statistical significance between default and optimal analysis parameter, assessed using the Wilcoxon signed-ranks test, is denoted by an asterisk (*).

**Table A5.** Results from WaveSurfer are presented per speaker group. Refer to the Table A2 caption for additional information regarding default analysis parameter and optimal analysis parameter.

| Speaker group | # of coefficients | 12 | 14 | 16 | 18 | 20 | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| Typically developing | | | | | | | | |
| Male adults | F4 | | + | | | | –0.833 | .405 |
| | F3 | | + | | | | –0.282 | .778 |
| | F2 | | + | | | | –1.761 | .078 |
| | F1 | | + | | | | –2.433 | .015 |
| | F4–F1 | | + | | | | –1.718 | .086 |
| Female adults | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Male children | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | | + | | | | –1.755 | .079 |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Female children | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Down syndrome | | | | | | | | |
| Male adults | F4 | | | | + | | –0.682 | .496 |
| | F3 | + | | | | | | |
| | F2 | | + | | | | –0.187 | .852 |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Female adults | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Male children | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |
| Female children | F4 | + | | | | | | |
| | F3 | + | | | | | | |
| | F2 | + | | | | | | |
| | F1 | + | | | | | | |
| | F4–F1 | + | | | | | | |

**Table A6.** The mean fundamental frequency (F0) in Hz and standard deviation pooled across vowels for a given speaker group are presented for both the manual F0 measurements and the Speech Acoustic Analysis Software Package default-generated F0 measurements. CSL = Computerized Speech Laboratory.

| | Manual measurements | | CSL | | Praat | | TF32 | | WaveSurfer | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Speaker group** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Typically developing | | | | | | | | | | |
| Male adults | 116.88 | 26.18 | 114.81 | 23.73 | 125.43 | 39.93 | 116.17 | 24.44 | 111.86 | 23.89 |
| Female adults | 218.20 | 33.31 | 213.24 | 31.25 | 215.39 | 34.53 | 218.72 | 32.37 | 214.28 | 31.33 |
| Male children | 239.49 | 33.61 | 234.07 | 33.70 | 233.46 | 47.58 | 240.69 | 35.54 | 234.58 | 46.04 |
| Female children | 238.04 | 18.29 | 240.58 | 21.42 | 237.45 | 12.02 | 238.65 | 17.89 | 238.47 | 17.60 |
| Down syndrome | | | | | | | | | | |
| Male adults | 146.42 | 17.78 | 141.63 | 23.89 | 142.16 | 16.42 | 144.80 | 17.68 | 144.57 | 17.47 |
| Female adults | 205.25 | 35.50 | 207.73 | 31.79 | 193.21 | 49.44 | 208.65 | 34.77 | 202.65 | 43.02 |
| Male children | 248.72 | 40.79 | 228.38 | 49.31 | 235.86 | 39.20 | 245.07 | 42.38 | 241.98 | 51.38 |
| Female children | 273.32 | 45.95 | 239.37 | 33.78 | 262.39 | 33.36 | 273.66 | 46.15 | 239.65 | 45.08 |

## Appendix B

TF32 Design Rationale

TF32 takes a different approach to sampling rate than the other three SAASPs considered in this report. The usual procedure for formant estimation by LPC analysis is to downsample the speech signal to a lower rate that is appropriate to the analysis objective. With either an original sampling at 10 kHz or downsampling to that value, the half-sampling rate or Nyquist bandwidth of 5 kHz is matched to the expected frequency range of the first five formants. This approach is common to most SAASPs. There are three main reasons for this procedure: (a) A significant computational burden arises with a large number of LPC coefficients required at a higher sampling rate. (2) The absence of voice source excitation energy at higher frequencies can result in a singular covariance matrix giving numerical indeterminacy. (3) The serial-pole acoustic model commonly used in LPC analysis is not physically valid at higher frequencies because of the presence of vocal tract cross-mode resonances. Each of these issues is addressed in the following discussion, which is based on personal communication with Paul Milenkovic (September 19, 2014), developer of TF32:

1. Moore's Law, which essentially states that processor speeds (or overall processing power) for computers doubles every 2 years, greatly reduces concerns about computational burden in low-cost desktop computers used in clinical speech analysis.

2. The absence of energy at high frequencies is addressed in the LPC algorithm used in the TF32 software. Applying the least-squares covariance algorithm to the pre-emphasized acoustic signal, a constant diagonal term is added to the covariance matrix according to a matrix regularization procedure that is well known in solving numerical least-squares problems. This procedure prevents the numerical indeterminacy resulting from a singular matrix. The diagonal term is scaled to represent applying a flat floor to the pre-emphasized LPC spectrum that is 30 dB below the energy of the acoustic signal. The effects of this floor at higher frequencies are readily confirmed in the time-slice spectrum plot in the TF32 program.

3. The acoustic processes generating any observed formants beyond F4 should not be a concern provided LPC analysis of speech at high sampling rates properly matches the lower formants. The higher formant correction to the low-frequency range of the serial-pole model (Olive, 1971), along with the particular way that LPC represents this in a discrete time or digital filter model (Gold & Rabiner, 1968), leads to a theoretical justification for LPC analysis at higher sampling rates, one which is borne out by the experimental results in the present study.

As explained by Olive (1971), a narrowband pole pair in the serial-pole model not only contributes a local spectrum peak, but also has shoulders that contribute to spectrum shaping far from the frequency of a particular formant. Furthermore, an analog serial-pole synthesizer as used in early work by Fant (1960) and others has a pronounced rolloff in the spectrum that needs to be corrected, as explained by Olive (1971) and Gold and Rabiner (1968). This correction represents the effect in the lower frequencies of the higher order formants of an idealized acoustic tube. Gold and Rabiner noted that a digital serial-pole model, of which LPC is an example, has a built-in higher pole correction to the spectrum. This correction is derived from considering the spectrum from minus the Nyquist rate to plus the Nyquist rate to be periodically continued, counteracting the rolloff seen in an analog synthesizer in the absence of the higher pole or formant correction.

This automatic higher pole correction can cause problems at low sampling rates. Considering a 10-kHz sampling rate (5-kHz Nyquist limit), an LPC model matching a formant at 4.75 kHz will have a mirror image pole at 5.25 kHz. If the true vocal tract does not have a pole at this location, the automatic higher pole spectrum correction in the 0–5 kHz range will deviate from what it should be, and the formant poles may be displaced by small amounts to compensate in achieving the least-square fit. At higher sampling rates, the mirror image poles in the LPC model will be many kHz away from the formant poles of interest.

### *Conclusion*

Simply because prior published studies downsample before performing LPC does not mean this procedure should be a continued or necessary practice. In light of the capabilities of modern computers, the steps taken to guard against a singular matrix in the LPC analysis in the TF32 software program, and the potential mismatch of the automatic higher pole correction in LPC when sampling at lower rates, higher rates merit consideration, as done in the present report. Conducting LPC at a higher sampling rate is borne out by the data in this report, showing that the LPC estimates become more, not less consistent, with the reference (CARV) data.