

TITLE

Optimizing vowel formant measurements in four acoustic analysis systems for diverse speaker groups

Ekaterini Derdemezis

430 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison
Wisconsin 53705

Houri K. Vorperian^a

427 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison
Wisconsin 53705

Ray D. Kent

491 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison
Wisconsin 53705

Marios Fourakis

Department of Communication Sciences and Disorders, University of Wisconsin-Madison, 466
Goodnight Hall, 1975 Willow Dr., Madison, Wisconsin 53706

Emily L. Reinicke

430 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison
Wisconsin 53705

Daniel M. Bolt

Department of Educational Psychology, University of Wisconsin-Madison, 1086 Educational
Sciences Building, 1025 W. Johnson Street, Madison, Wisconsin 53706

^a) Electronic mail: vorperian@waisman.wisc.edu

ABSTRACT:

Purpose: This study systematically assessed the effects of select LPC analysis parameter manipulations on vowel formant measurements for diverse speaker groups using four trademarked Speech Acoustic Analysis Software Packages (SAASP): CSL, Praat, TF32, and WaveSurfer.

Methods: Productions of four words containing the corner vowels were recorded from four speaker groups with typical development (male and female adults, male and female children) and four 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 four SAASP (using both the default analysis parameters and with adjustments/manipulations to select parameters). Smaller differences between values obtained from the SAASP 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 SAASP.

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 report 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 (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 (SAASP) in the study of disordered speech in

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

SAASP 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. Generally, SAASP 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 on-line user group exchanges) for the adjustment of analysis parameters to analyze the speech of women, children, or populations with speech disorders of either sex and/or different ages. The parameters that are most likely to be adjusted in SAASP 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 non-nasalized 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 based on 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 non-optimal. 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) two times the desired number of formants plus 2 (e.g., 10 coefficients to achieve an analysis of 4 formants) (Deng & Dang, 2007; Vallabha & Tuller, 2002). The basis for the latter is to include a sufficient number of model poles (2 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 non-formant effects associated with vocal tract excitation or lip radiation. Vallabha and Tuller (2002, 2004) proposed a method of adjusting the LPC filter order based on 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 SAASP 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 SAASP 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, & Bolt (2014) examined the accuracy and comparability of vowel formant measurements made by four SAASP 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 speech and natural speech. Substantial errors were noted for some measures. For example, Burris et al. observed that bandwidth measures 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 SAASP.

The natural speech samples used by Burris et al. were from typically developing children and healthy adults, and the speech samples used by Yao et al. were produced by healthy adults. To our knowledge, no study has examined how SAASP 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 SAASP with a diverse group of speakers of both sexes, including typically developing children and adults as well as children and adults with DS. The manipulations are not equivalent across the SAASP because of differences in software design including default settings. The aim of this evaluation was not to compare the performance of different SAASP 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.

Methods

Source Materials for Acoustic Analysis

This study was approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board (IRB). The speech samples, obtained from the Acoustics Database of the Vocal Tract Development Lab were monosyllabic words containing the vowels /i/, /u/, /ae/, /a/ as in eat,

hoot, hat and hot. Recordings were obtained from 42 individuals with DS (26 males and 16 females, ages 4 to 36 years) and 122 typically developing (TD) individuals (52 males and 70 females, ages 4 to 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 unidirectional Shure SM48 microphone, 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 that digitizes speech at 48 kHz with 16-bit resolution on a SanDisk Ultra II flashcard. 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 (5 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 an adult male) 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 to 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 (5 male; 5 female) ages 5-10 years; (b) 10 children with DS (5 male; 5 female) ages 5-10 years; (c) 10 adults with DS (5 male; 5 female) ages 21-35

years and; (d) 20 TD adults (10 male; 10 female) 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. Down syndrome 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 SAASP (the same software versions as those evaluated by Burris et al.): (a) CSL (Computerized Speech Laboratory model 4500, version 3.4.1 by Kay Elemetrics); (b) Praat (version 5.1.31 by Boersma and 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 by the Centre for Speech Technology at KTH in Stockholm, Sweden; 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. 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 SAASP 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 and spectrogram of the segmented vowel to: (1) select the steady state segment of the vowel (i.e., the most stable 25-50 ms of the total vowel duration), and (2) determine the F1-F4 frequencies, henceforth referred to as consensus analysis reference values (CARV). CARV were measured in TF32 because this SAASP allows the overlay of the LPC spectral slice on the FFT spectral slice; furthermore, TF32's spectrogram is synchronized with the LPC spectrum so that when the user moves the cursor across the LPC spectrum, a slaved cursor appears on the spectrogram (Figure 1A&B). Manipulations to the spectrogram were made to help establish the most accurate CARV and included: (a) changing the analysis bandwidth of the FFT based on 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: (i) two formants were close in frequency, (ii) strong harmonics precluded certainty of formant tracking, (iii) there were oral-nasal resonance imbalances or, (iv) vocal quality was atypical. Of the total number of CARV, 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 CARV was confirmed by remeasuring 10% of the tokens three months later. Paired t-test results were as follows: for TD speakers: $t = 1.643$; $p = .117$; and speakers with

DS: $t= 1.914$; $p=.061$. Approximately 80% of the remeasured values were within 50Hz 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 ten Hillenbrand et al., 1995, tokens representative of the four corner vowels) that were downloaded from Hillenbrand's homepage (<http://homepages.wmich.edu/~hillenbr/voweldata.html>) and used in Burris et al. study (2014). 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 25-50 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 25-50 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 the: (a) output measurements at the manufacturer's default settings (presumably optimal for adult male speech). F0 was 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 1 in Appendix A, 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 1 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 SAASP 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 women 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, first 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 based on 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 8, and was the only coefficient manipulation made given our interest in F1-F4 measurements. The time step was also changed to .00625, and for adult males, the maximum formant was changed to 5500 Hz, because the default 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 CARV. These values (Hz) are displayed in Figures 2-9 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 rank sum 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 seventeen 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 rank sum test (instead of a parametric paired t-test), since the normality assumption for the differences appeared implausible. The results are summarized in Tables 2 to 5 in Appendix A, 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 adult male (panel A) and female (panel B), and TD child male (panel C) and female (panel D) speakers for F1-F4 ; and Figure 3

displays differences in F1-F4 measurements for adult males with DS (panel A), adult females with DS (panel B), male children with DS (panel C), and female children with DS (panel D). The circle around 12 on the x-axes 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 rank sum test was statistically significant for all adult speaker groups (summarized in Table 2 in Appendix A): TD adult males ($Z = -9.433$; $p < .001$); TD adult females ($Z = -3.206$; $p = .001$); adult males with DS ($Z = -5.079$; $p < .001$); and adult females with DS ($Z = -4.948$; $p < .001$); with notable improvements for DS speakers, specifically F2 and F3 measurements for DS adult males (F2: $Z = -2.651$; $p = .008$; F3: $Z = -3.360$; $p = .001$) and F3 and F4 measurements for adult females 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 which 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 2, Appendix A).

Results from Praat (Figures 4 and 5, summarized in Table 3 in Appendix A) 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 (males: $Z = -2.045$; $p = .041$; females: $Z = -2.867$; $p = .004$).

TF32 results (Figures 6 and 7, summarized in Table 4 in Appendix A) 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. Similarly, 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 adult females (TD adult females: $Z = -3.430$; $p = .001$) and the child speaker groups, particularly the children with DS (males: $Z = -2.721$; $p = .007$; and females: $Z = -3.034$; $p = .002$). Additional improvements on the individual formants were noted for F1 and F3 for TD adult females (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 (Figures 8 and 9, summarized in Table 5 in Appendix A) show that in general, except for TD adult males, 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 6 in Appendix A, which shows that in general, across speaker groups the SAASP 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 SAASP. 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 SAASP, 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 accurate formant-frequency measurements for TD adult females 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 adult males 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 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 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 ten 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 SAASP but usually includes the analysis bandwidth used in spectrograms and the number of coefficients used in LPC analysis. It should be noted that SAASP 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 employed 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 SAASP. The consensus analysis method relies on the user’s knowledge 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 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 CARV. 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 SAASP 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 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 ten 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 re-analysis, it may not be representative. For example, even though /u/ and /æ/ 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 /æ/, breaking is common so that an initial /æ/ gives way to a low-back vowel, such as /a/. Therefore, it is important to establish criteria for selecting the point of formant-frequency measurement. Some possible criteria are: vowel /i/ -- point of highest frequency of F2; vowel /u/ -- point of lowest frequency of F2; vowel /a/ -- 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 SAASP 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 inter-formant 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 adult males, but a larger bandwidth of 400 or 500 Hz is often more effective in formant displays for women or children. With very high F0, 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 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 5kHz in women and children.

Above all, the user of a 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 based on normative data, the analysis should be reconsidered. Duckworth et al. (2011) make 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, & Honda, 2006), a cavity that undergoes significant growth (Vorperian, Wang, Chung, Schimek, Durtschi, Kent, Ziegert & Gentry, 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 males 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 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, women, and a population with speech disorders. We hope that this report encourages further tests with more diverse populations, also 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/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.

ACKNOWLEDGEMENTS

This work was supported by NIH 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 (NICHD). 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 paper. 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 (5.1.31). Amsterdam, The Netherlands: Publisher.
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.
- Deng, L., & Dang, J. (2007). Speech analysis: the production-perception perspective. In C. Lee, H. Li, L. Lee, R. 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.
- Fox, R.A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, 126, 2603-2618.
- Hagiwara, R. (1995) Acoustic realizations of American /r/ as produced by women and men. Unpublished PhD 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. *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 Ph.D. 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.
- Kay Elemetrics. (1996). Computer Speech Lab (500). Lincoln Park, NJ: KayPentax. Available from [http://www.kayelemetrics.com/index.php?option=com_product&controller=product&Itemid=3&cid\[\]=11&task=pro_details](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* (2 ed.). San Diego: Singular.
- Kent, R.D., & Vorperian, H.K. (2013). Speech impairments in Down syndrome: A review. *Journal of Speech, Language & 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.
- Milenkovic, P. (2010). TF32 (Alpha). Madison, WI: Publisher. Retrieved January 21, 2010.
Available from <http://userpages.chorus.net/cspeech/>
- Moura, C.P., Cunha, L.M., Vilarinho, H., Cunha, M.J., Freitas, D., Palha, M., Puschel, S.M., Pais-Clemente, M. (2008). Voice parameters in children with Down syndrome. *Journal of Voice*, 22, 34-42.
- 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., Picmausova, J., Roth, J., & Ruzicka, E. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease. *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*. Volume 2009, Available online: Article ID 159234.
- Sjolander, K. & Beskow, J. (2005). WaveSurfer (1.8.5). Stockholm, Sweden: Publisher.
Available from <http://www.speech.kth.se/wavesurfer/>
- Sjolander, K. & Beskow, J. (2010) WaveSurfer (1.8.8). Stockholm, Sweden: Publisher.
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, Makhtari P, Honda K. Acoustic roles of the laryngeal cavity in vocal tract resonance. *The Journal of the Acoustical Society of America*. 2006; 120:2228-38.
- 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). [Compact disk]. Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology.
- Vorperian HK, Wang S, Chung MK, Schimek EM, Durtschi RB, Kent RD, Ziegert AJ, Gentry LR. 2009. Anatomic development of the oral and pharyngeal portions of the vocal tract: an imaging study. *J Acoust Soc Am* 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. *Gengo Kenkyu*, 138, 99-113.

Figure Captions:

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 between cursors (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 LPC spectrum on FFT spectrum) of the marked analysis segment of the vowel, with vertical gridline markings in 1000 Hz intervals, and horizontal lines in 10 dB increments.

Figure 2: Difference values for formant frequencies (F1-F4) using CSL for 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 CARV from 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 CARV). 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.

Figure 3: Difference values for formant frequencies (F1-F4) using CSL 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). Refer to Figure 2 caption for additional information regarding vertical bar charts, default analysis parameter, optimal analysis parameter and statistical significance.

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

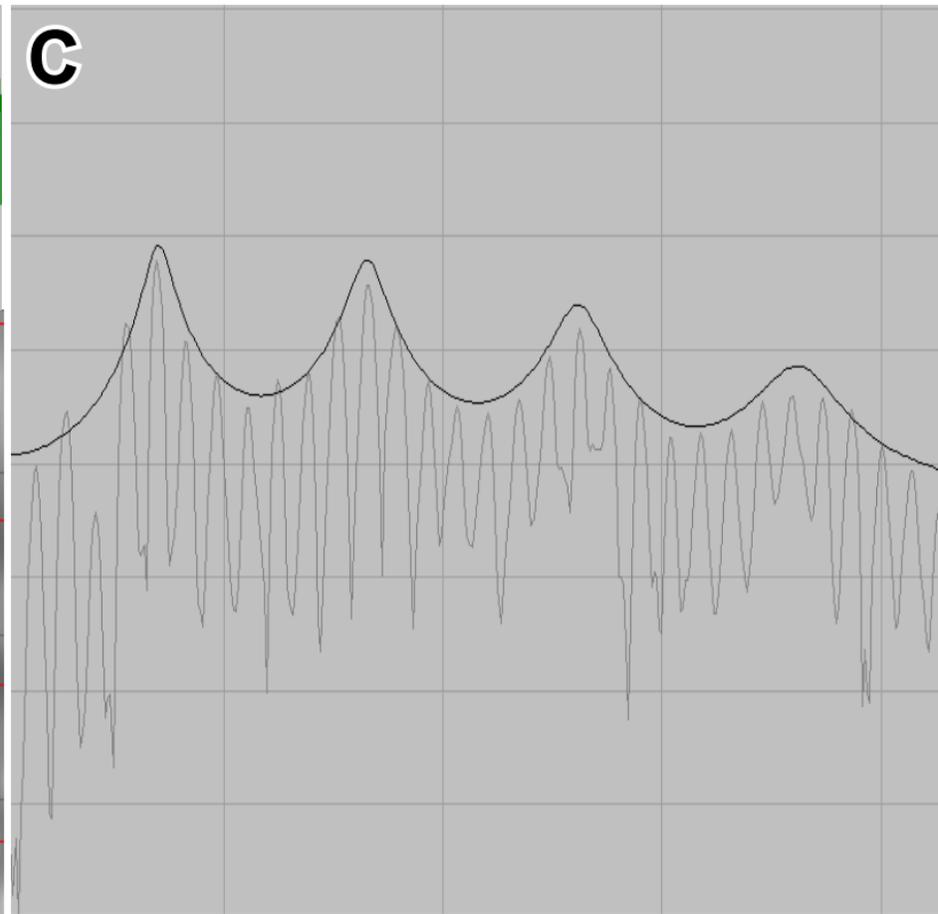
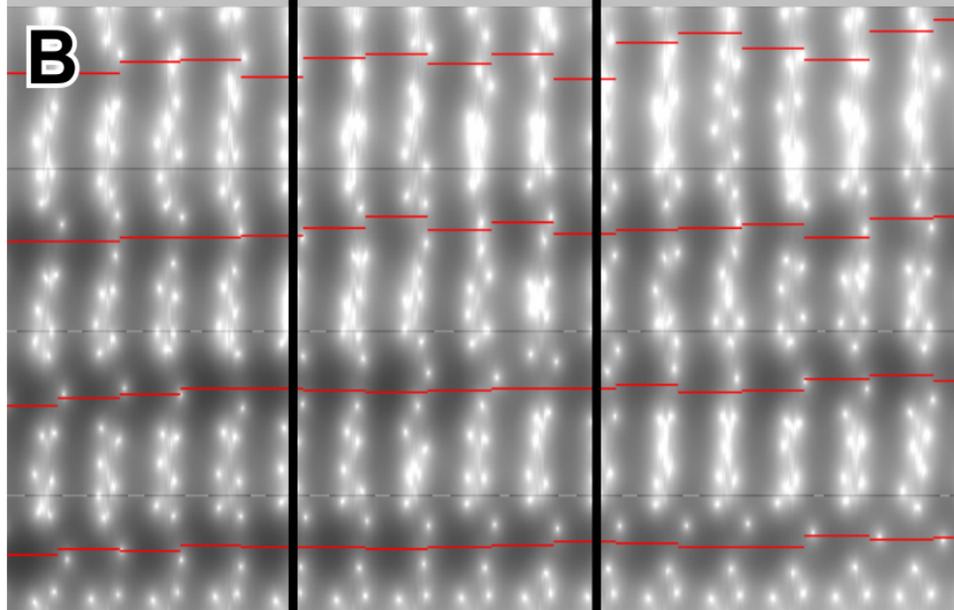
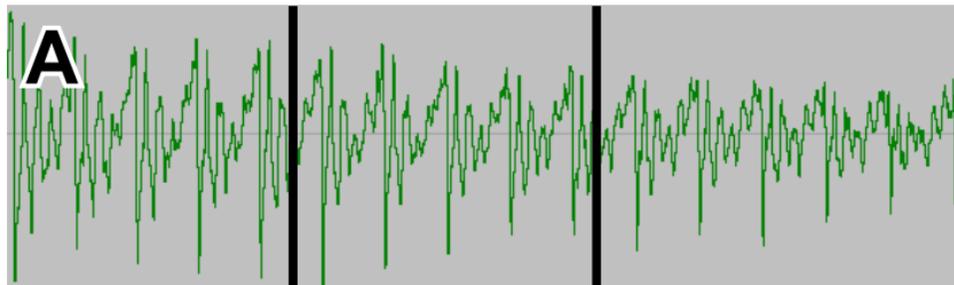
Figure 5: Difference values for formant frequencies (F1-F4) using Praat for male adults with DS (panel A), female adults with DS (panel B), male children with DS (panel C) and female children with DS (panel D). Refer to Figure 2 caption for additional information regarding vertical bar charts, default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

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

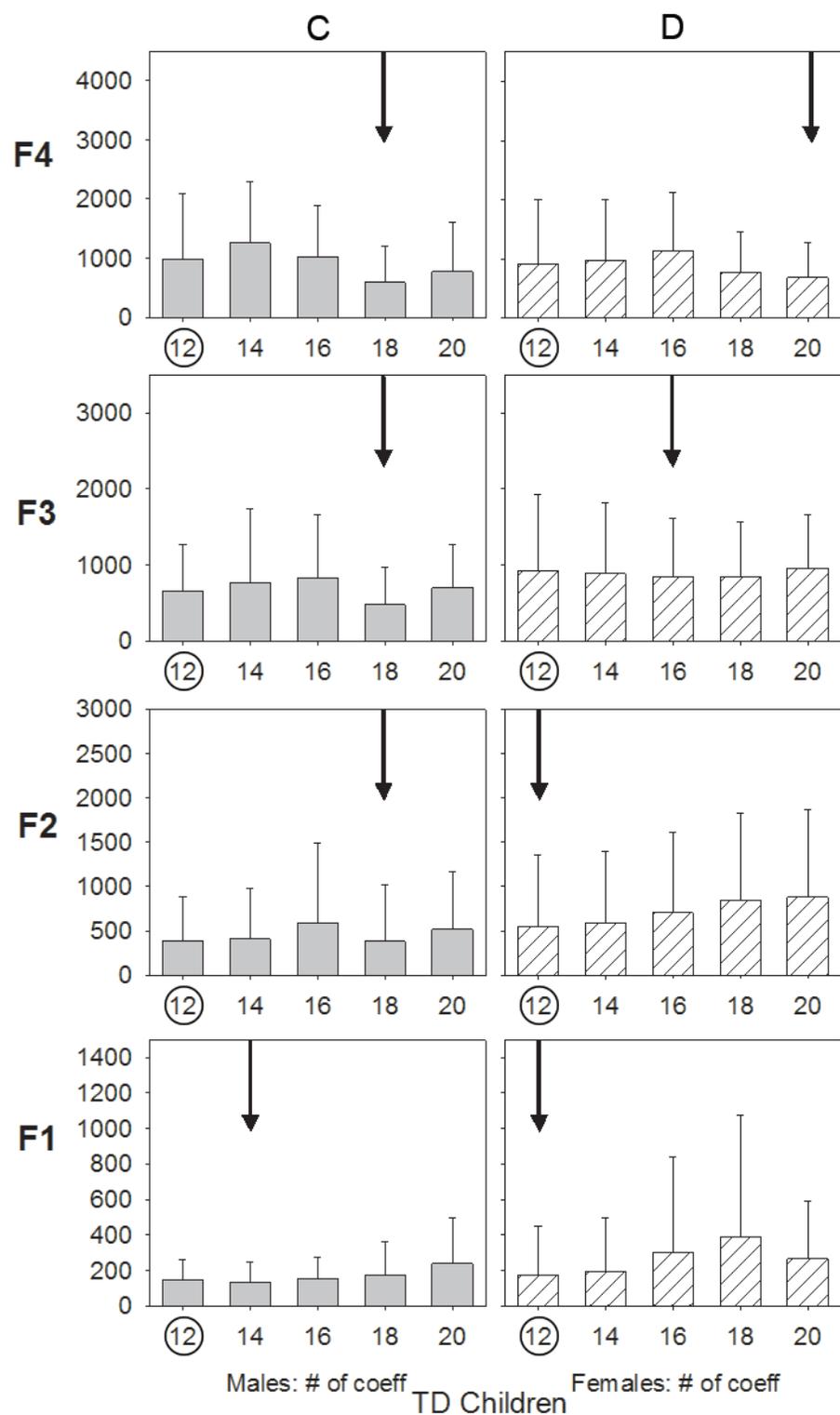
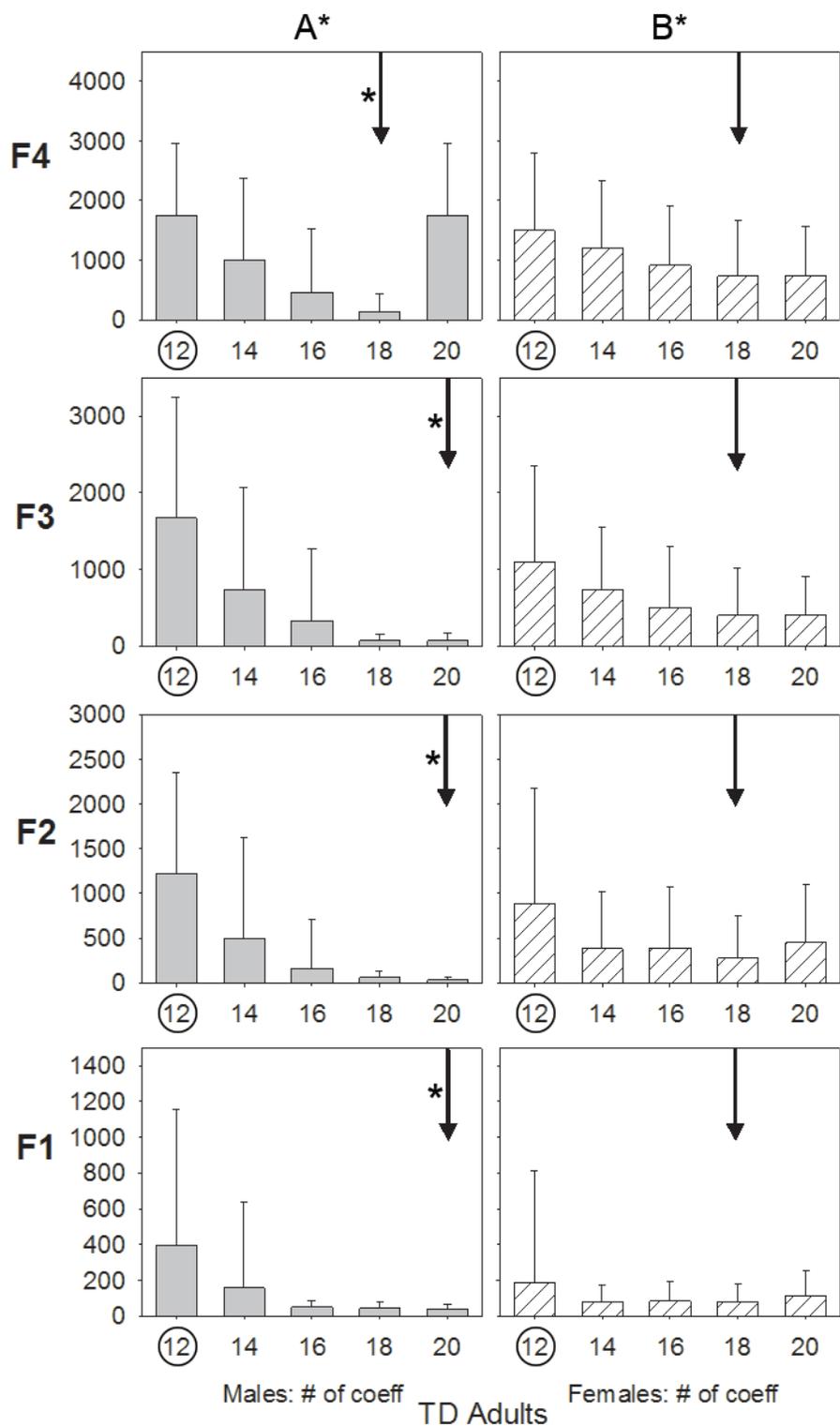
Figure 7: Difference values for formant frequencies (F1-F4) using TF32 for male adults with DS (panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to Figure 2 caption for additional information regarding vertical bar charts, default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

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

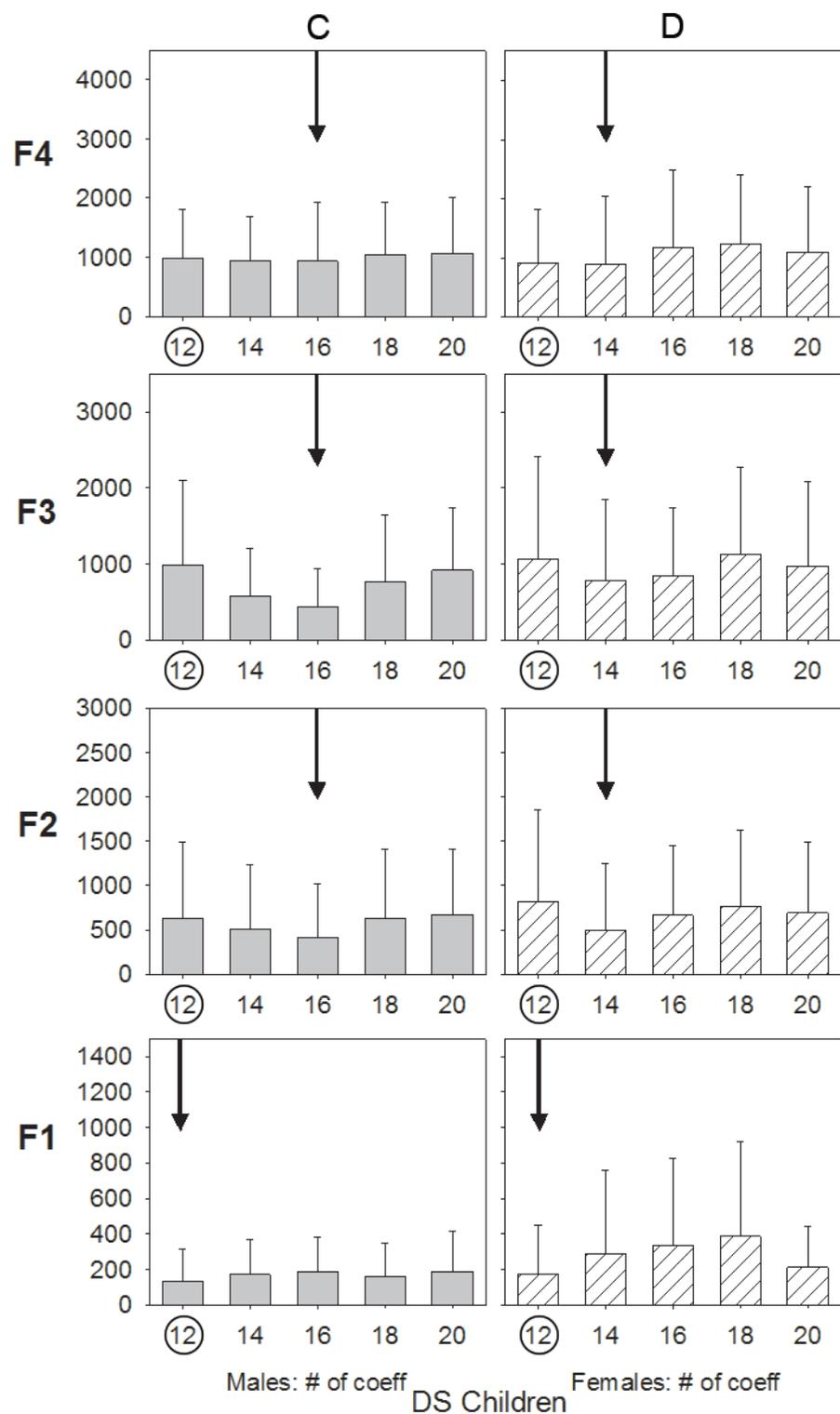
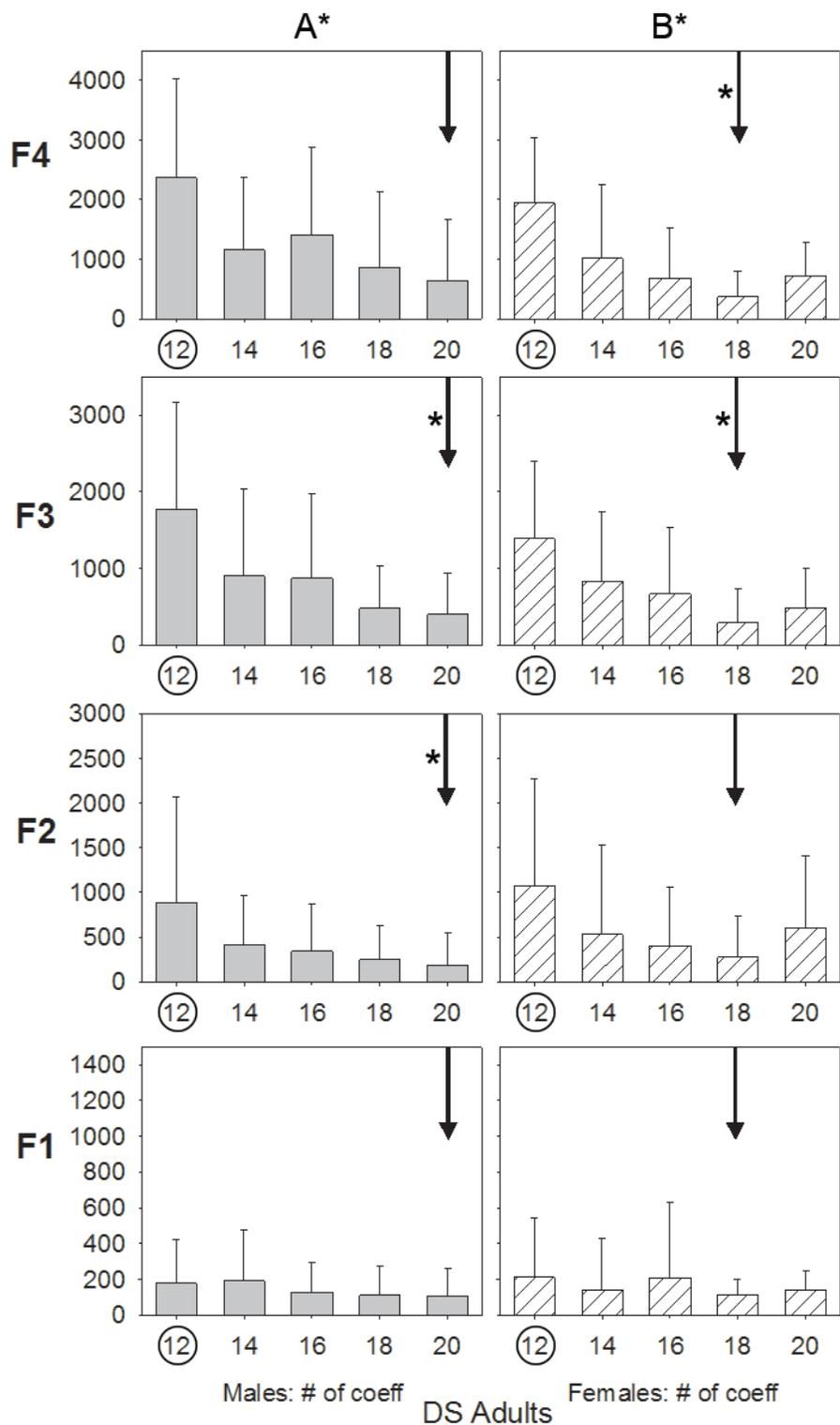
Figure 9: Difference values for formant frequencies (F1-F4) using WaveSurfer for male adults with DS (panel A), female adults with DS (panel B), male children with DS (panel C), and female children with DS (panel D). Refer to Figure 2 caption for additional information regarding vertical bar charts,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



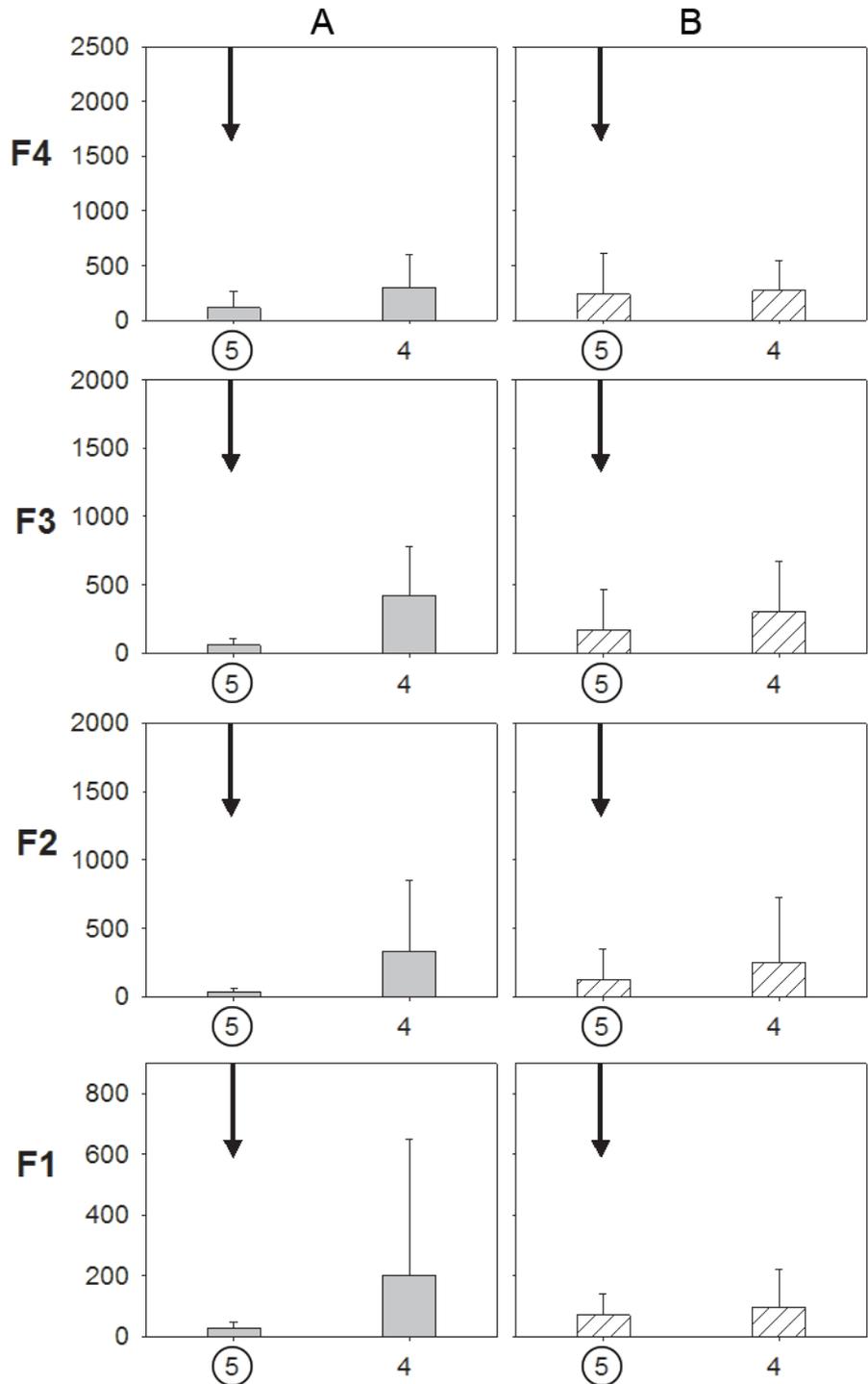
Differences (Hz): CSL measured value minus Consensus Analysis Reference Values



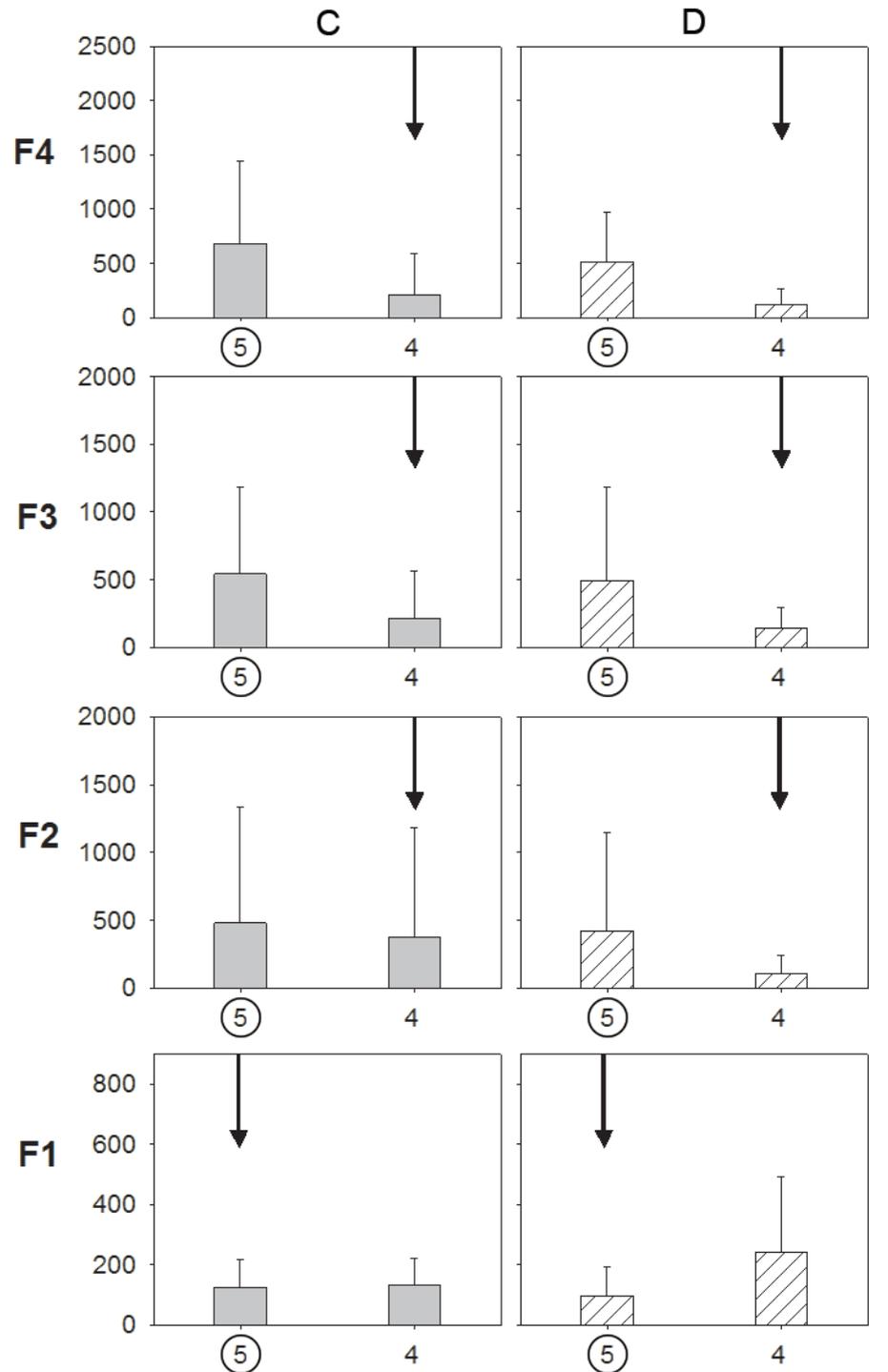
Differences (Hz): CSL measured value minus Consensus Analysis Reference Values



Differences (Hz): Praat measured value minus Consensus Analysis Reference Values

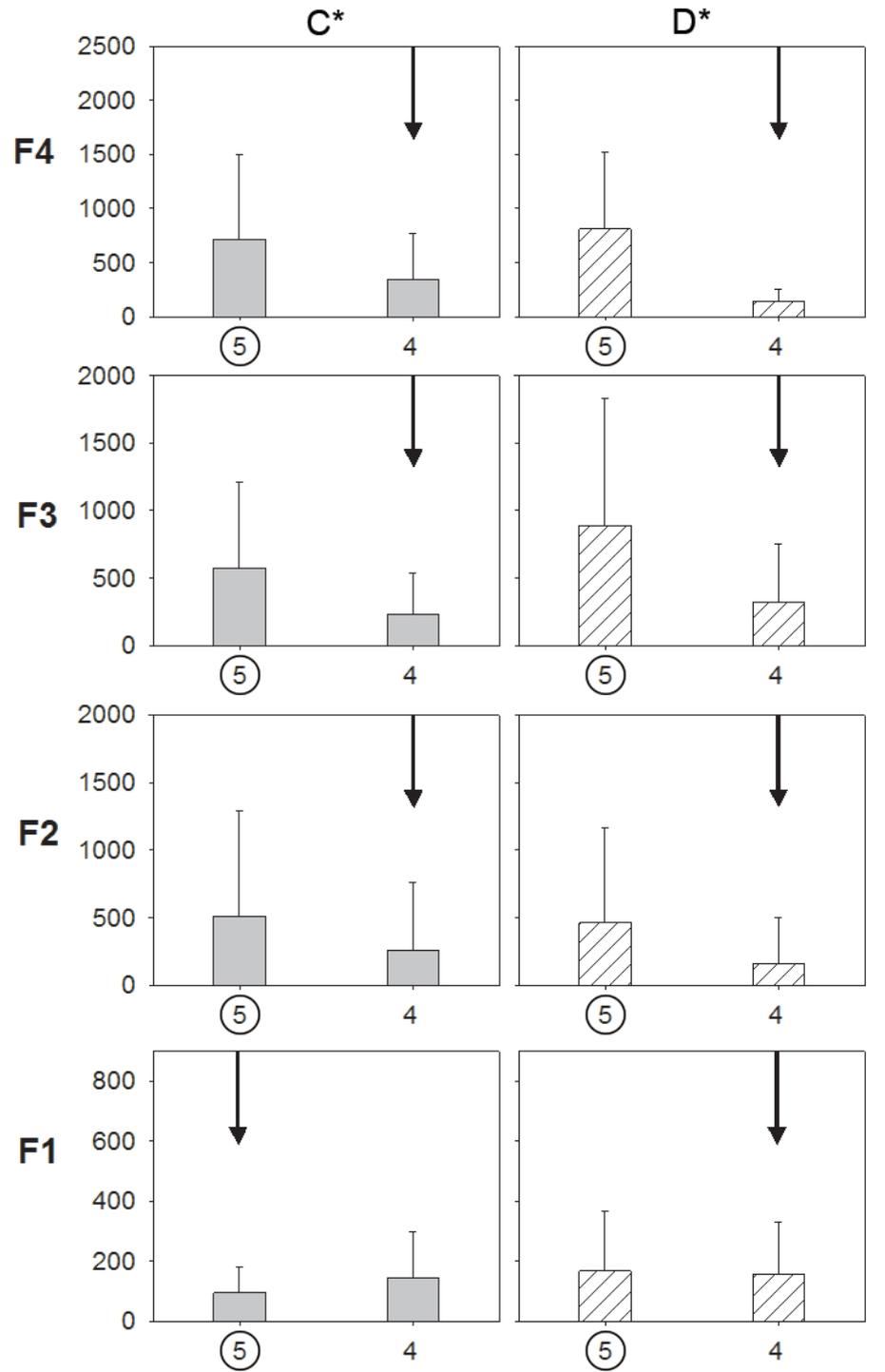
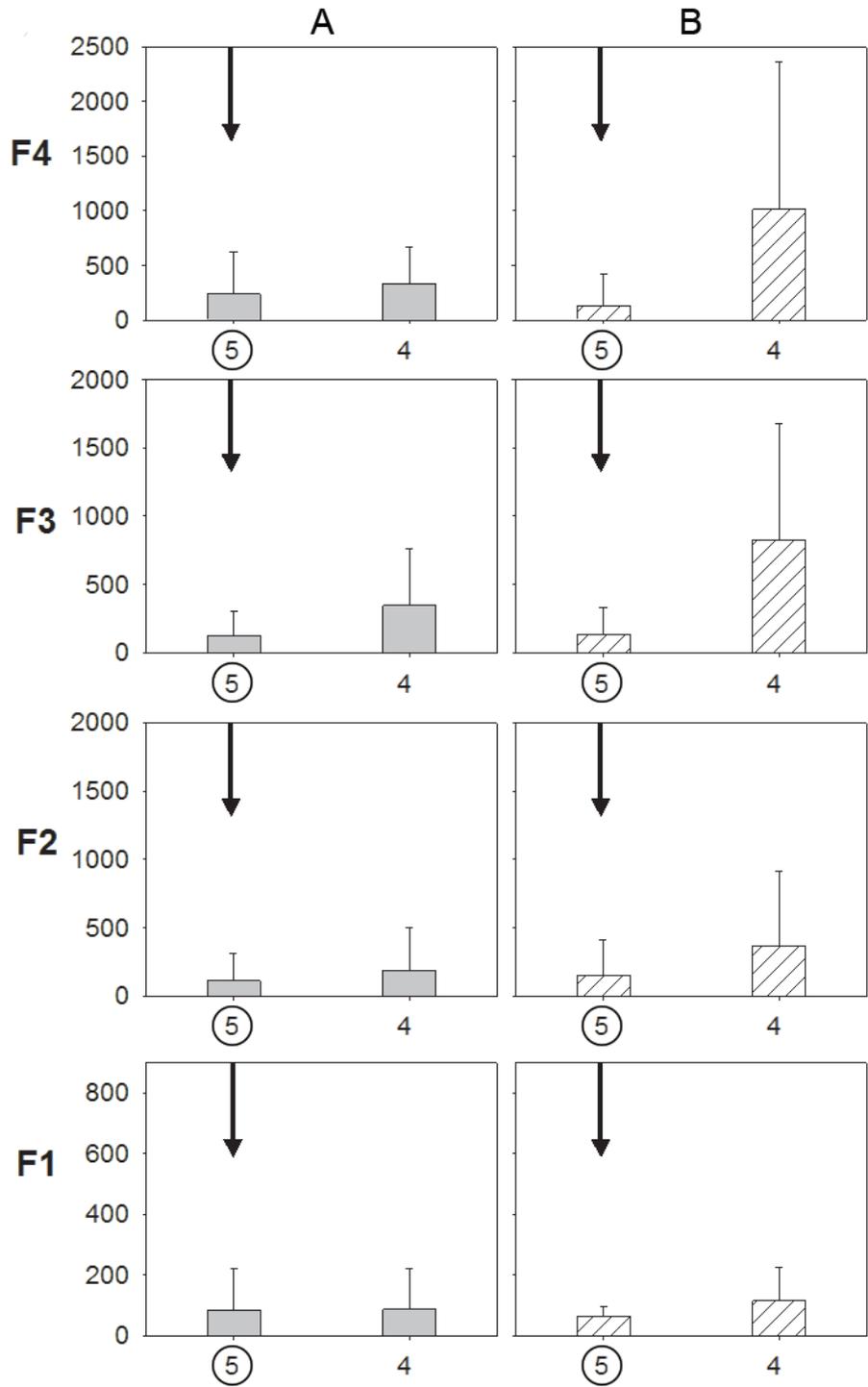


Males: # of formants expected Females: # of formants expected
TD Adults



Males: # of formants expected Females: # of formants expected
TD Children

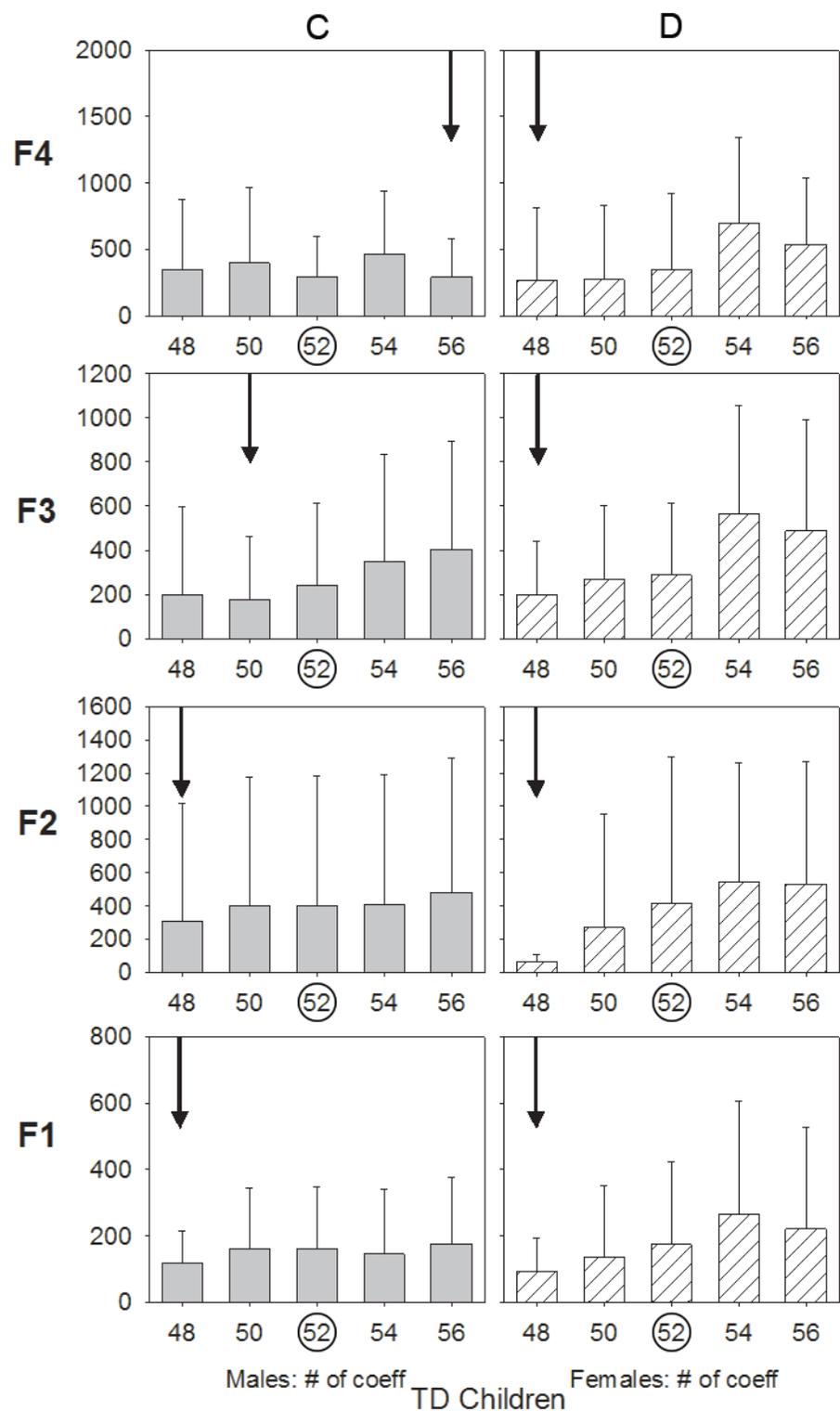
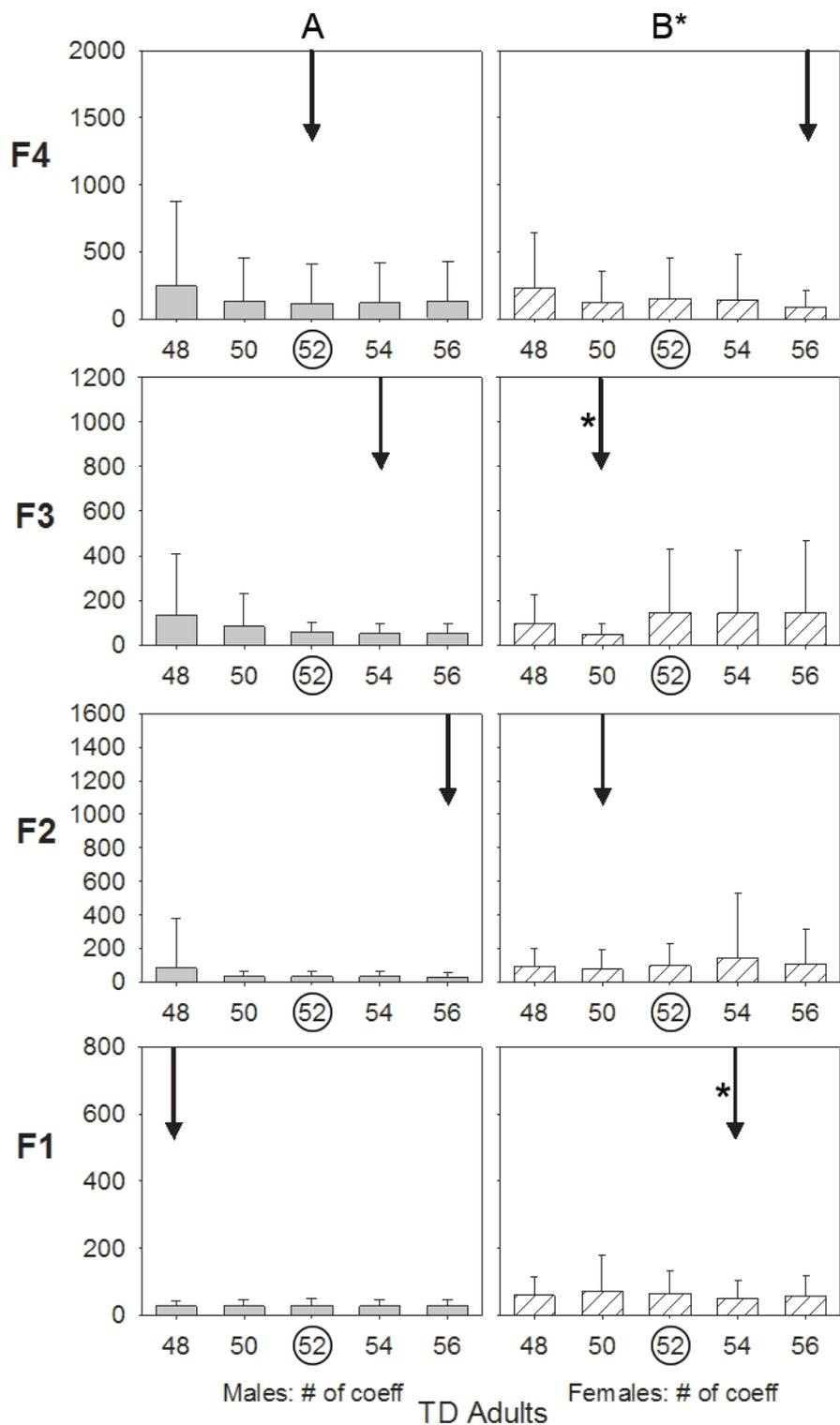
Differences (Hz): Praat measured value minus Consensus Analysis Reference Values



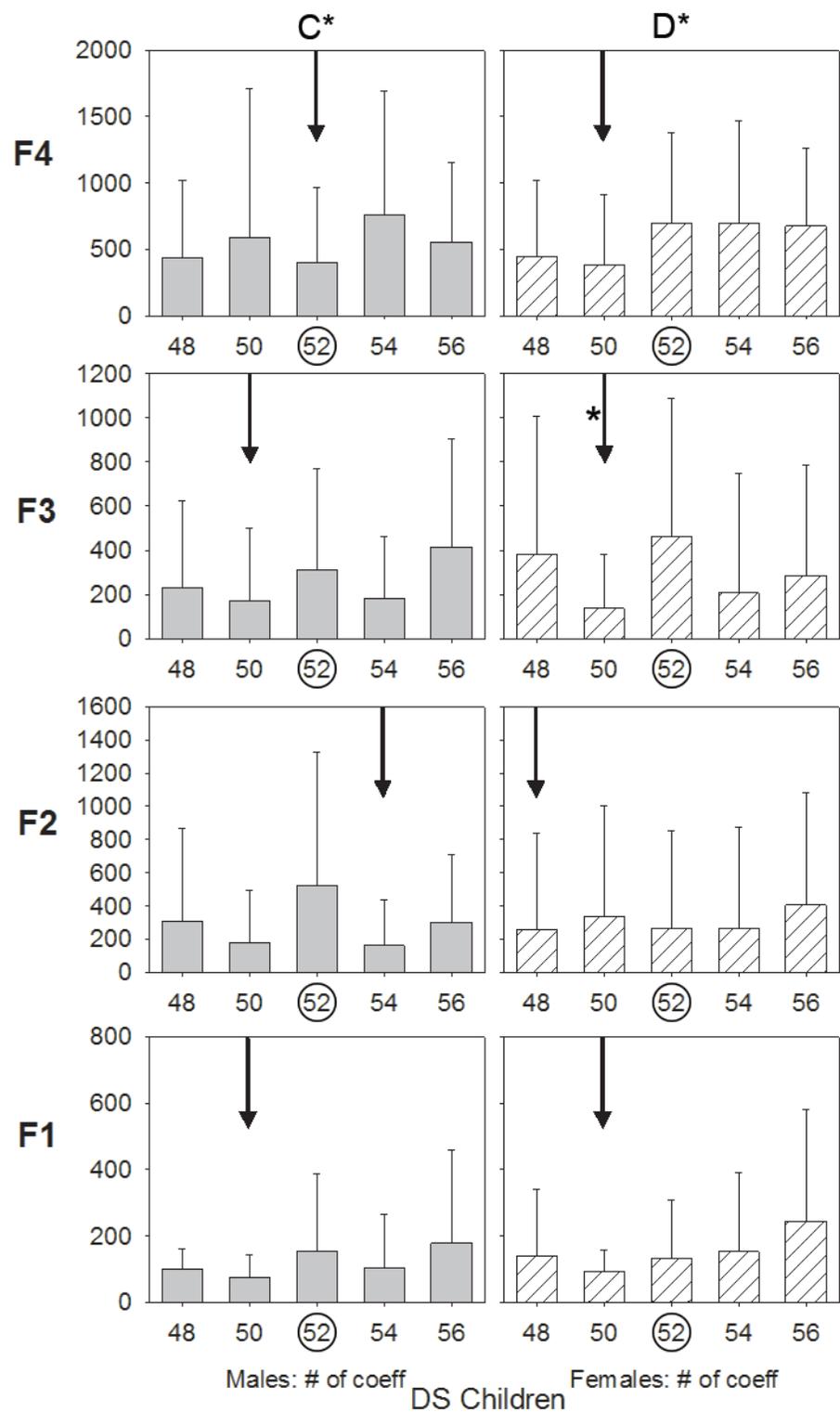
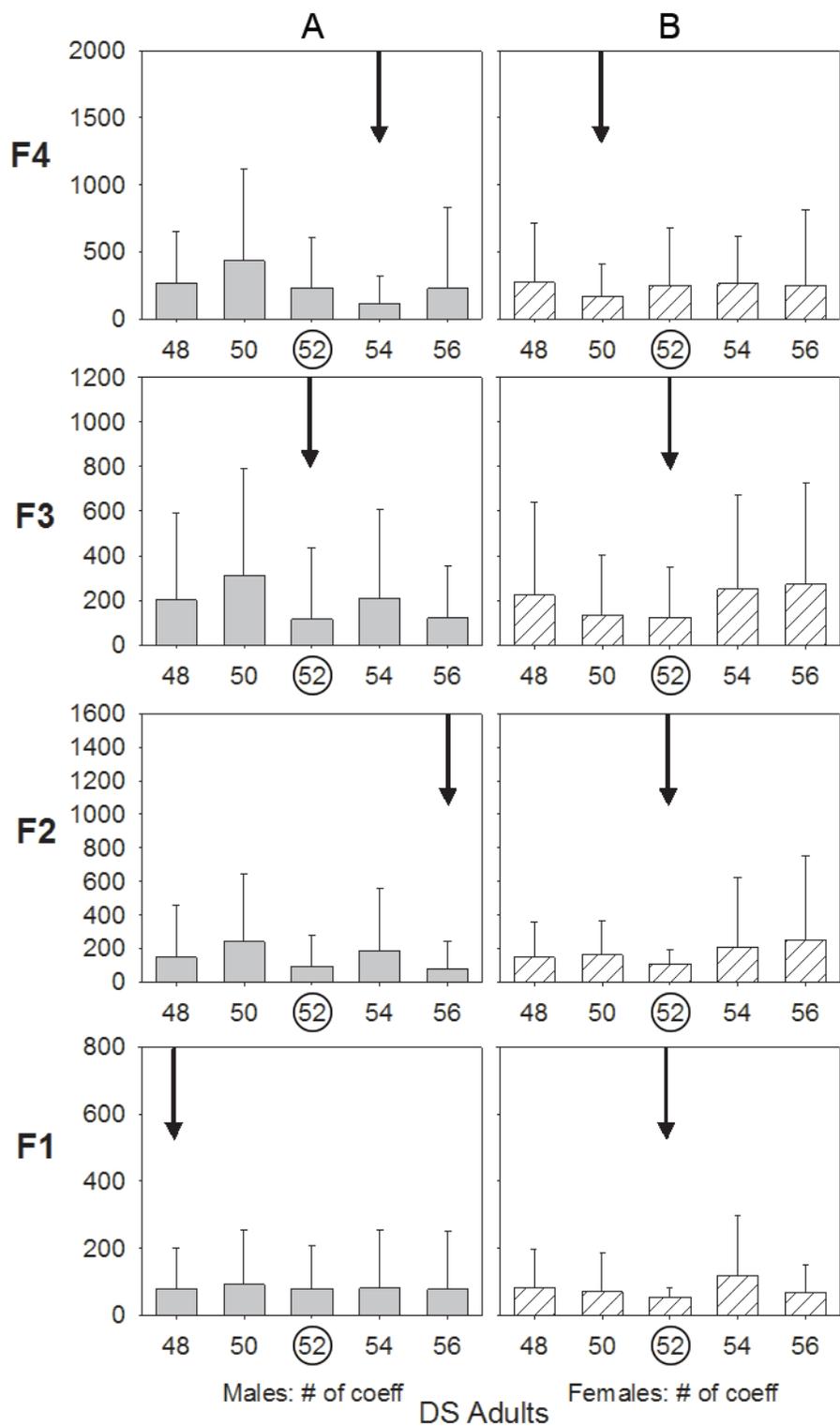
Males: # of formants expected Females: # of formants expected
DS Adults

Males: # of formants expected Females: # of formants expected
DS Children

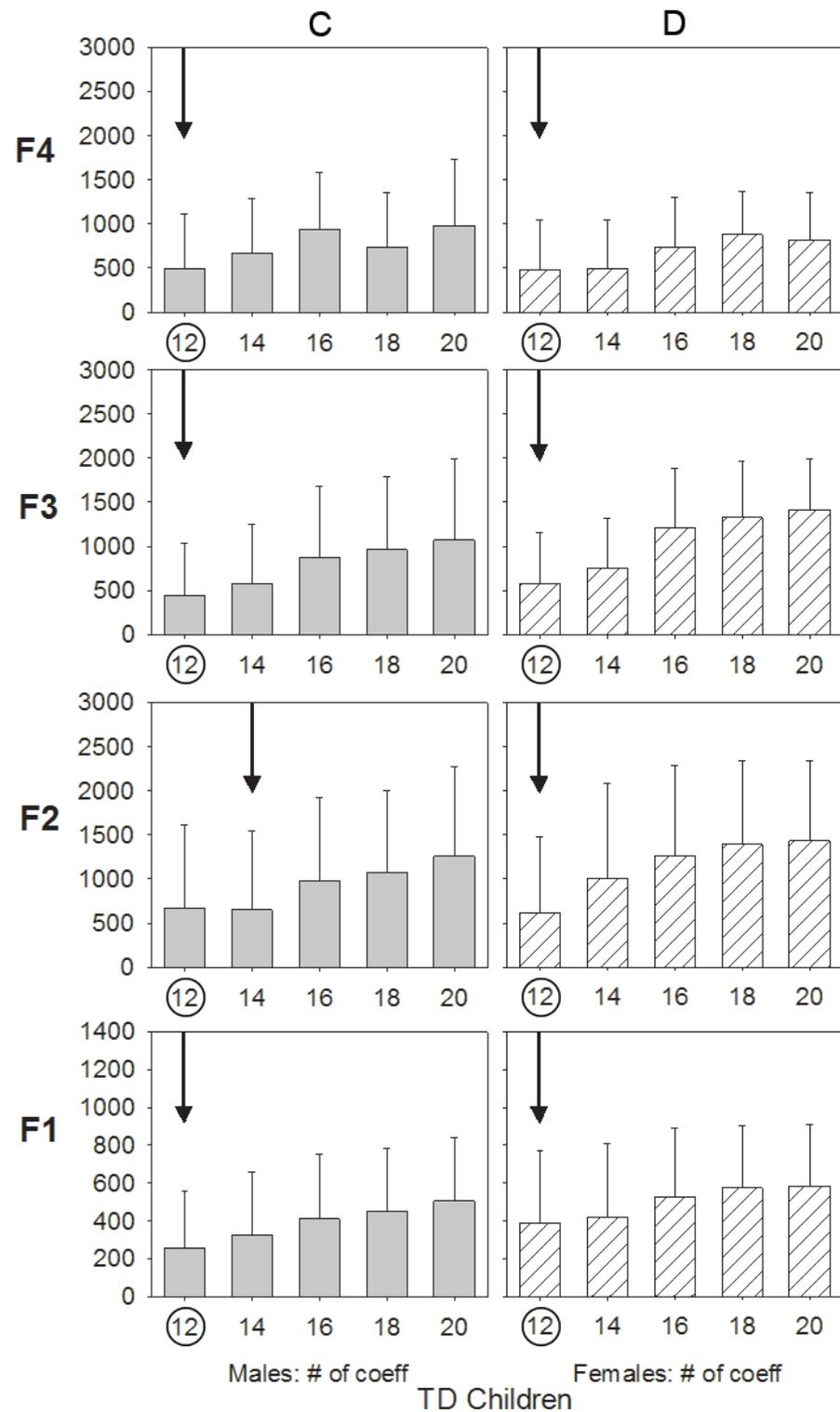
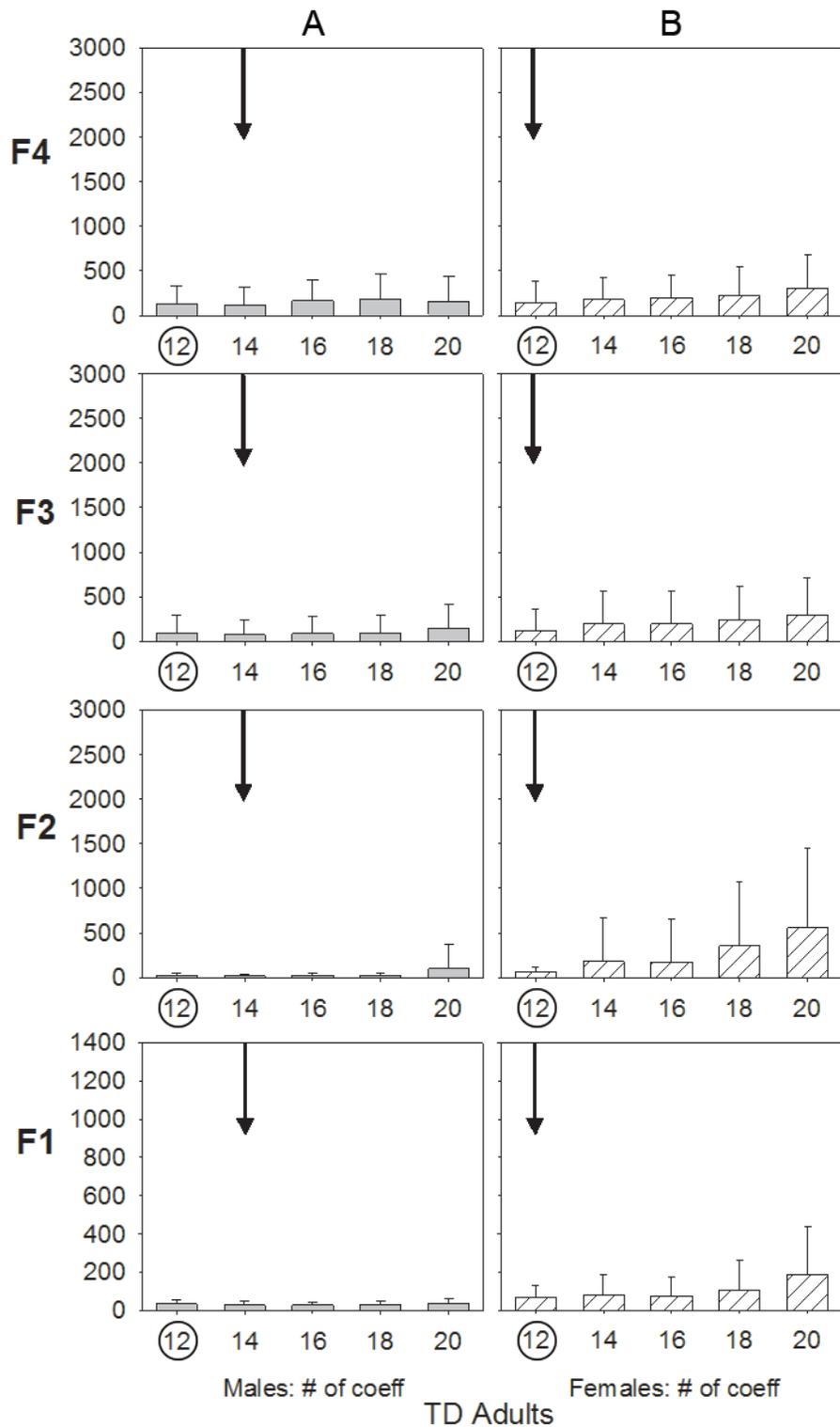
Differences (Hz): TF32 measured value minus Consensus Analysis Reference Values



Differences (Hz): TF32 measured value minus Consensus Analysis Reference Values



Differences (Hz): Wavesurfer measured value minus Consensus Analysis Reference Values



Differences (Hz): Wavesurfer measured value minus Consensus Analysis Reference Values

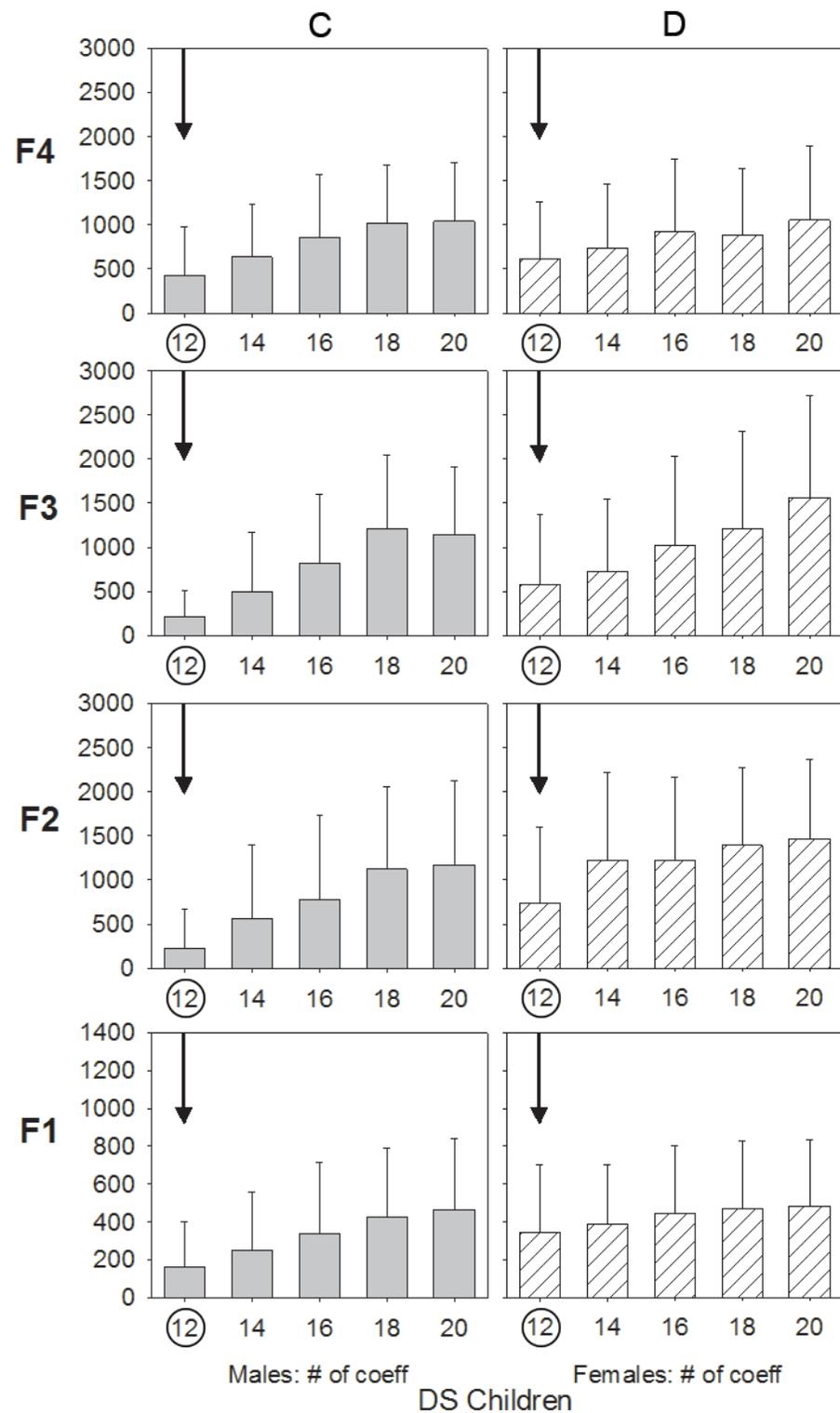
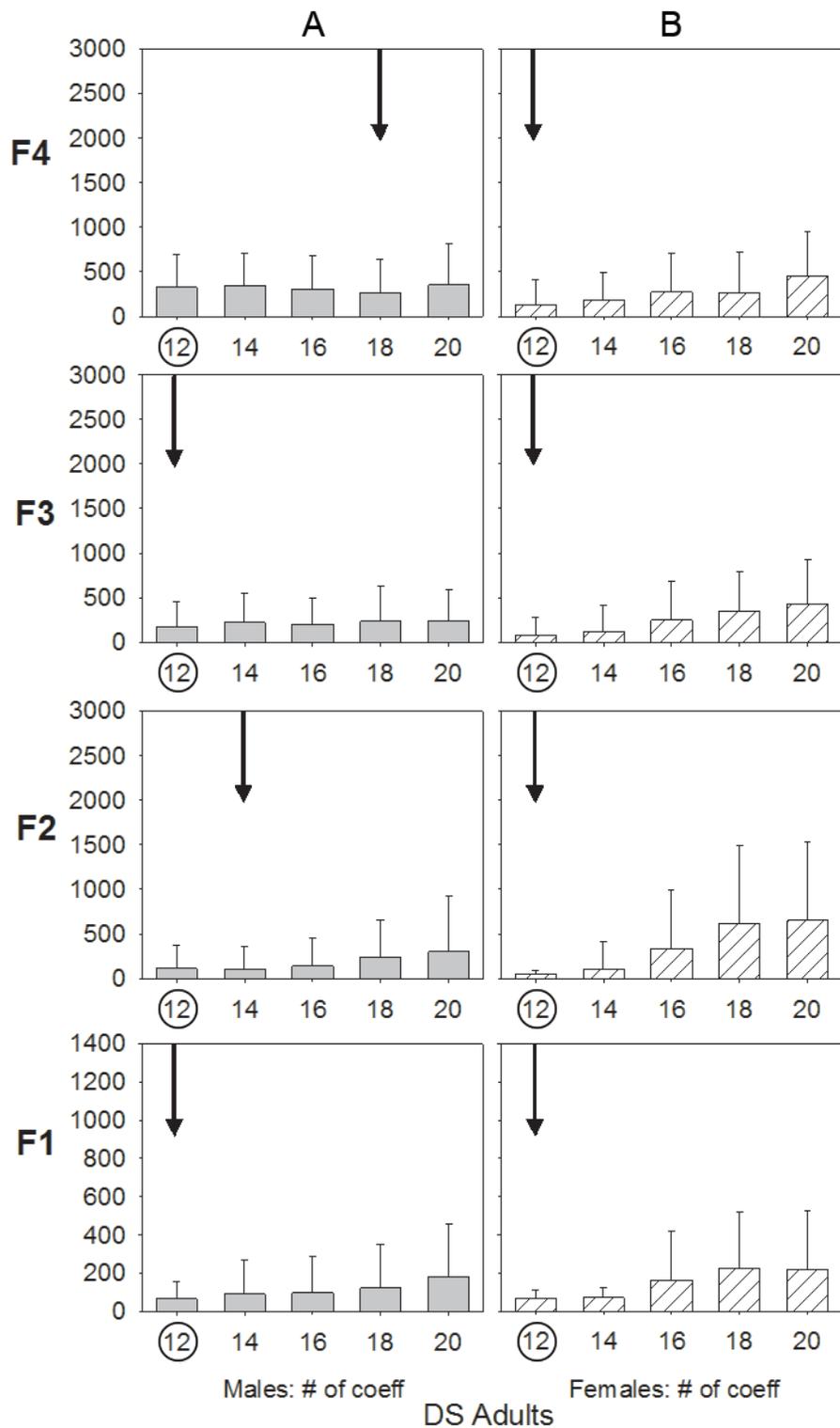


Table 1

Manipulation sequence	1	2	3
	Downsampling & other	# of coefficients <i>Default Manipulation</i>	Smoothing <i>Default Manipulation</i>
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 * Indirectly changed # of coefficients to 8 by changing expected number of formants from default of 5 to 4 formants.	N/A No
TF32	No downsampling. Number of coefficients is adjusted automatically to the sampling rate	52 Yes (48-56)	No Yes
Wavesurfer	Automatic downsampling to 10kHz	12 Yes (12-20)	Yes No

Table 2

CSL									TYPICALLY DEVELOPING								
ADULTS	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4				+		-3.771	.000*		F4				+		-1.999	.046
	F3					+	-4.967	.000*		F3				+		-2.47	.014
	F2					+	-5.282	.000*		F2				+		-1.331	.183
	F1					+	-4.852	.000*		F1				+		-.887	.375
	F1-F4					+	-9.433	.000*		F1-F4				+		-3.206	.001*
CHILDREN	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4				+		-.621	.535		F4					+	-.534	.594
	F3				+		-.893	.372		F3			+			-.187	.852
	F2				+		-.299	.765		F2	+						
	F1		+				-.187	.852		F1	+						
	F1-F4				+		-.585	.559		F1-F4	+						
MALE									FEMALE								
DOWNSYNDROME																	
ADULTS	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4					+	-2.310	.021		F4				+		-2.824	.005*
	F3					+	-3.360	.001*		F3				+		-3.375	.001*
	F2					+	-2.651	.008*		F2				+		-2.464	.014
	F1					+	-1.344	.179		F1				+		-.597	.550
	F1-F4					+	-5.079	.000*		F1-F4				+		-4.948	.000*
CHILDREN	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4			+			-.893	.372		F4		+				-.260	.795
	F3			+			-2.427	.015		F3		+				-.597	.550
	F2			+			-1.083	.279		F2		+				-1.568	.117
	F1		+							F1		+					
	F1-F4			+			-1.295	.195		F1-F4		+				-.708	.479
MALE									FEMALE								

Table 3

Praat										
TYPICALLY DEVELOPING										
ADULTS	# of formants expected	5	4	Z	p	# of formants expected	5	4	Z	p
	F4	+				F4		+	-2.042	.041
	F3	+				F3	+			
	F2	+				F2	+			
	F1	+				F1	+			
	F1-F4	+				F1-F4	+			
CHILDREN	# of formants expected	5	4	Z	p	# of formants expected	5	4	Z	p
	F4		+	-.973	.331	F4		+	-1.689	.091
	F3		+	-1.023	.306	F3		+	-1.083	.279
	F2		+	-.037	.970	F2		+	-.523	.601
	F1	+				F1	+			
	F1-F4		+	-.539	.590	F1-F4		+	-.092	.927
MALE					FEMALE					
DOWNSYNDROME										
ADULTS	# of formants expected	5	4	Z	p	# of formants expected	5	4	Z	p
	F4	+				F4	+			
	F3	+				F3	+			
	F2	+				F2	+			
	F1	+				F1	+			
	F1-F4	+				F1-F4	+			
CHILDREN	# of formants expected	5	4	Z	p	# of formants expected	5	4	Z	p
	F4		+	-1.293	.196	F4		+	-1.481	.139
	F3		+	-2.240	.025	F3		+	-2.277	.023
	F2		+	-1.381	.167	F2		+	-1.717	.086
	F1	+				F1		+	-.933	.351
	F1-F4		+	-2.045	.041*	F1-F4		+	-2.867	.004*
MALE					FEMALE					

Table 4

TF32		TYPICALLY DEVELOPING															
ADULTS	# of coeff.	48	50	52	54	56	Z	p	# of coeff.	48	50	52	54	56	Z	p	
	F4			+					F4					+	-1.061	.289	
	F3				+		-1.815	.069	F3		+				-2.764	.006*	
	F2					+	-1.038	.299	F2		+				-2.098	.036	
	F1	+					-1.228	.219	F1				+		-3.335	.001*	
	F1-F4								F1-F4		+				-3.430	.001*	
CHILDREN	# of coeff.	48	50	52	54	56	Z	p	# of coeff.	48	50	52	54	56	Z	p	
	F4					+	-.511	.609	F4	+					-.114	.910	
	F3		+				-1.303	.193	F3	+					-.840	.401	
	F2	+					-.348	.727	F2	+					-1.307	.191	
	F1	+					-.853	.393	F1	+					-.829	.407	
	F1-F4	+					-.718	.473	F1-F4	+					-1.259	.208	
MALE								FEMALE									
DOWNSYNDROME																	
ADULTS	# of coeff.	48	50	52	54	56	Z	p	# of coeff.	48	50	52	54	56	Z	p	
	F4				+		-.795	.426	F4		+				-.483	.629	
	F3			+					F3			+					
	F2					+	-.141	.888	F2			+					
	F1	+					-1.025	.306	F1			+					
	F1-F4								F1-F4			+					
CHILDREN	# of coeff.	48	50	52	54	56	Z	p	# of coeff.	48	50	52	54	56	Z	p	
	F4			+					F4		+				-1.720	.085	
	F3		+				-1.147	.251	F3		+				-2.999	.003*	
	F2				+		-1.428	.153	F2	+					-.213	.831	
	F1		+				-2.223	.026	F1		+				-.710	.478	
	F1-F4		+				-2.721	.007*	F1-F4		+				-3.034	.002*	
MALE								FEMALE									

Table 5

WS		TYPICALLY DEVELOPING															
ADULTS	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4		+				-.833	.405	F4	+							
	F3		+				-.282	.778	F3	+							
	F2		+				-1.761	.078	F2	+							
	F1		+				-2.433	.015	F1	+							
	F1-F4		+				-1.718	.086	F1-F4	+							
CHILDREN	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4	+							F4	+							
	F3	+							F3	+							
	F2		+				-1.755	.079	F2	+							
	F1	+							F1	+							
	F1-F4	+							F1-F4	+							
MALE								FEMALE									
DOWNSYNDROME																	
ADULTS	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4				+		-.682	.496	F4	+							
	F3	+							F3	+							
	F2		+				-.187	.852	F2	+							
	F1	+							F1	+							
	F1-F4	+							F1-F4	+							
CHILDREN	# of coeff.	12	14	16	18	20	Z	p	# of coeff.	12	14	16	18	20	Z	p	
	F4	+							F4	+							
	F3	+							F3	+							
	F2	+							F2	+							
	F1	+							F1	+							
	F1-F4	+							F1-F4	+							
MALE								FEMALE									

Table 6

	Manual measurements		CSL		Praat		TF32		Wavesurfer	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
TD Adult Males	116.88	26.18	114.81	23.73	125.43	39.93	116.17	24.44	111.86	23.89
TD Adult Females	218.20	33.31	213.24	31.25	215.39	34.53	218.72	32.37	214.28	31.33
TD Child Males	239.49	33.61	234.07	33.70	233.46	47.58	240.69	35.54	234.58	46.04
TD Child Females	238.04	18.29	240.58	21.42	237.45	12.02	238.65	17.89	238.47	17.60
DS Adult Males	146.42	17.78	141.63	23.89	142.16	16.42	144.80	17.68	144.57	17.47
DS Adult Females	205.25	35.50	207.73	31.79	193.21	49.44	208.65	34.77	202.65	43.02
DS Child Males	248.72	40.79	228.38	49.31	235.86	39.20	245.07	42.38	241.98	51.38
DS Child Females	273.32	45.95	239.37	33.78	262.39	33.36	273.66	46.15	239.65	45.08