Published in final edited form as:
J Speech Lang Hear Res. 2014 February 1; 57(1): 26–45. doi:10.1044/1092-4388(2013/12-0103).

Quantitative and descriptive comparison of four acoustic analysis systems: vowel measurements

Carlyn Burris,
429 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

Marios Fourakis,
Department of Communicative Disorders, University of Wisconsin-Madison, 466 Goodnight Hall, 1975 Willow Dr., Madison, Wisconsin 53706

Ray D. Kent, and
491 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

Abstract

Purpose—This study examines accuracy and comparability of four trademarked acoustic analysis software packages (AASP): Praat, Wavesurfer, TF32 and CSL using synthesized and natural vowels. Features of AASP are also described.

Methods—Synthesized and natural vowels were analyzed using each of AASP's default settings to secure nine acoustic measures: fundamental frequency (F0), formant frequencies (F1-F4), and formant bandwidths (B1-B4). The discrepancy between the software measured values and the input values (synthesized, previously reported, and manual measurements) was used to assess comparability and accuracy. Basic AASP features are described.

Results—Results indicate that Praat, Wavesurfer, and TF32 generate accurate and comparable F0 and F1-F4 data for synthesized vowels and adult male natural vowels. Results varied by vowel for adult females and children, with some serious errors. Bandwidth measurements by AASPs were highly inaccurate as compared to manual measurements and published data on formant bandwidths.

Conclusions—Values of F0 and F1-F4 are generally consistent and fairly accurate for adult vowels and for some child vowels using the default settings in Praat, Wavesurfer, and TF32. Manipulation of default settings yields improved output values in TF32 and CSL. Caution is recommended especially before accepting F1-F4 results for children and B1-B4 results for all speakers.

[a]vorperian@waisman.wisc.edu.

I. INTRODUCTION

Software for digital acoustic analysis of speech offers unprecedented opportunities for the analysis of speech samples for different purposes, including education, clinical practice, and research. Use of these software systems is almost certain to lead to a substantial increase in the application of acoustic measures and to the further development of acoustic databases for speech. However, neutral and objective evaluations of these systems' measurements have not been reported, so that potential users have little guidance in selecting a system for their use. Furthermore, there is no assurance that the accumulating data gathered from these different systems can be assumed to be accurate and comparable, for healthy or disordered speech, for males or females, or for children as well as adults.

Previous studies of acoustic analysis software packages (AASP) for speech have been of two major types. First, a small number of studies reported on comparisons of features across systems (Read, Buder & Kent, 1990, 1992) or described general approaches to signal acquisition and analysis without comparing systems (Ingram, Bunta & Ingram, 2004; Read, Buder, & Kent, 1990, 1992; Vogel & Maruff, 2008). Second, a few studies examined the accuracy and/or reliability of measures of voice, such as the perturbation measures of jitter and shimmer (Bielamowicz, Kreiman, Gerratt, Dauer & Berke, 1996; Deliyski, Evans, & Shaw, 2005; Karnell, Hall & Landahl, 1995; Smits, Ceuppens & De Bodt, 2005). The studies in the latter group raise a concern that values generated by different systems are not always comparable and that care should be taken in managing and interpreting data from these systems. In one previous study that compared analysis systems for the measurement of vowel formant frequencies, Woehrling and Mareuil (2007), reported that there were some "substantial differences" in the values of F1 obtained from Praat and Snap. To our knowledge, there has not been a systematic comparison across AASPs for the measurement of formant frequencies and bandwidths, despite the general interest in these entities for research on typical and atypical speech.

Speech AASPs generally afford the capability for FFT and LPC analysis of speech. LPC analysis has been particularly powerful and convenient because it generates numeric data for formant frequencies and bandwidths that can be displayed in patterns such as formant tracking. Vallabha and Tuller (2002) discuss four sources of error in LPC analysis that are relevant to this study, especially because LPC data are commonly used to generate formant tracks and to populate a data table. The first is quantization of the signal owing to the fundamental frequency, which results in an error estimated to be about 10% of F0. This error is particularly important for the speech of women or young children, who often have an F0 higher than 250 Hz. The second error is choice of an incorrect order for the LP filter. Users of analysis software should be aware of appropriate adjustments for filter order, taking into consideration characteristics of both the speaker and the speech sample to be analyzed. In general applications, many users adjust the filter order to the speaker only. There are two common guidelines to set the filter order: (a) set it to the number of formants expected plus two, or (b) set it to sampling frequency in kHz. Vallabha and Tuller (2004) describe a heuristic that can be used to determine the optimal filter order for either a corpus of vowels or a single vowel. The third error is an exclusive reliance on root solving in the LPC algorithm. The problem is most serious when the roots are close together. The fourth error relates to the 3-point parabolic interpolation that compensates for the coarse spectrum. A parabolic interpolation is particularly problematic when formants are in close proximity. Users probably cannot do much to overcome the third and fourth errors; rather, these errors should be taken as forewarnings of the limitations of analysis.

The purpose of this study was twofold: (1) to make quantitative assessment of the accuracy and comparability of data generated by four AASPs that are most commonly reported in the

speech literature; and (2) to describe the major features of those four AASPs. Assessment of accuracy and comparability of acoustic measurements included comparison of acoustic measures including fundamental frequency (F0), first through fourth formant frequencies (F1-F4), and first through fourth formant bandwidths (B1-B4). More specifically, accuracy was assessed by comparing software-generated measurements of synthesized vowels to the input values used to synthesize them. Comparability and accuracy were assessed by reanalyzing the speech samples of a previously published paper (Hillenbrand, Getty, Clark, & Wheeler, 1995) using the four aforementioned software packages and comparing data secured in each package to the reported findings. Descriptive comparisons were based mostly on an inventory of features.

II. METHODS

The four AASPs considered in this study are: Praat (*version 5.1.31 by Boersma and Weenink*), Wavesurfer (*version 1.8.5 by the Centre for Speech Technology at KTH in Stockholm, Sweden*), TF32 (*alpha test version 1.2 - see Appendix A; formerly CSpeech by Milenkovic*), and CSL (*Computerized Speech Laboratory model 4500, version 2.7.0 by Kay Elemetrics, currently known as KayPentax*; CSL is a combined hardware-software system, but for convenience it is classified as AASP for the purposes of this report). Quantitative comparison results reflect the analysis outcome of each acoustic variable in each software package, while descriptive comparisons reflect mostly an inventory of each software package's features highlighting particular features that may be beneficial to users.

Stimuli

Two types of male and female speech were analyzed to determine the accuracy and comparability of acoustic measurements across the four AASPs: 1) Synthesized adult vowels (corner vowels /i/, /ae/, /u/, and /a/); and 2) natural adult and child speech samples/ words containing one of the four corner vowels.

The synthesized speech sounds were the vowels used in Fourakis, Preisel, and Hawks (1998) and, as reported, were generated with the Klatt synthesizer using the cascade synthesis method with a quasi-periodic vibratory voicing source and a declining F0. Formant frequencies' input values were based on the formulas in Nearey (1989) and Hawks & Miller (1995). Also, because Fourakis and colleagues used the cascade method to synthesize vowels, the amplitude values were determined based on the formant pattern of each vowel, as per Fant's (1962) acoustic theory, rather than being chosen explicitly and orthogonally as they would have been using the parallel synthesis method (Klatt, 1980). To eliminate bias, vowel synthesis input values were made available to the present investigators only after all analyses were completed. The eight synthesized vowels used in this study (four synthesized with an adult male F0 and four synthesized with an adult female F0) were chosen from the larger pool of synthesized vowels used in Fourakis et al. (1998). Vowel selection was based on perceptual ratings by three listeners. The listeners heard the sound files with F1 and F2 values that fell within the range of values that are consistent with a particular corner vowel (see Table 1) and the listeners indicated the vowels they perceived to be the purest form of the four corner, or extreme, vowels in English (/i/, /a/, /ae/, or /u/). Thus, the vowel files chosen for acoustic analysis were the ones that the listeners determined to be the best representation of each vowel. The software-generated measurements were then compared to the known input values for each synthesized vowel, as listed in Table 2, to determine each software package's accuracy (see Table 2 and Figures 1AB and 2AB).

The natural speech samples were the four corner vowel productions ("heed," "had," "hawd," and "who'd") that were originally collected by Hillenbrand et al. (1995) and are available on Hillenbrand's homepage (http://homepages.wmich.edu/~hillenbr/voweldata.html).

Randomly selected samples from five adult males, five adult females, five male children and five female children were used and are referred to as the Hillenbrand stimuli. Hillenbrand's speech recordings were used because of the large number of participants and because present fundamental frequency and formant frequency findings can be compared against a reference of previously published data using digital signal processing methods, such as linear predictive coding (LPC) with manual corrections as needed. Hillenbrand and colleagues did not provide bandwidth measurements. Therefore, manual bandwidth measurements were made the Hillenbrand stimuli, as described in the following sections, for comparison to the software generated measurements.

Procedures

The average vowel length of the selected Hillenbrand stimuli was approximately 300ms as measured from vowel onset to offset while viewing the waveform and the spectrogram and confirmed by listening to the sound file. The middle 150ms of the vowel segment of each sound file was selected as the analysis segment. To create this analysis segment, the synthesized vowels and Hillenbrand stimuli were opened in Praat and the middle 150ms of the vowel, the longest steady state segment present in all stimuli, was segmented and saved separately. Segmentation was done using Praat because this software allows the user to annotate and label sections of the sound file (such as the vowel) and save specific sections as separate files (such as the 150 ms mid-vowel segment) that can be accessed later by all four software packages.

Analysis

All analyses were performed without knowledge of the input values of the synthesized vowels or the published measurements of the Hillenbrand stimuli. To ensure consistency of analyses across the four AASPs, a protocol was established and followed for each software package, which called for the use of a 150ms mid-vowel segment for analysis, as described previously. The default settings for each AASP (e.g., analysis window type, window length) were maintained with the assumption that the software developers selected default settings that are suited for the analysis of a large set of voices. Users may adjust the settings to improve the analysis for some types of speakers, for example women versus men, children versus adults. However, the software packages are highly variable in this regard; and their manuals do not provide explicit guidance on how settings should be adjusted to accommodate differences in speaker characteristics. Table 3 lists specific default values in each AASP examined.

Furthermore, to eliminate 'edge phenomena' effects or measurement errors that commonly occur at the very beginning and very end of an analysis segment, the first and last acoustic measurements from the list of measurements generated by each software package were excluded.

Manual measurements were also made for the synthesized vowels to compare against synthesis input values and AASP generated values. F0, and F1-F4 measurements were made manually in each AASP; however B1-B4 bandwidth measurements were done using TF32 because of the software's ease of use to display each formant peak, determine the 3dB point below each peak to secure bandwidth measurement, and scroll along the peak to take more accurate bandwidth measurements. Manual F0 measurements were used as the input value since the vowels were synthesized with a declining F0 (manual male F0 = 117Hz, manual female F0 = 225Hz). The manual F0 calculation entailed measuring the duration of the three most medial cycles in the analysis segment, dividing that duration by three, and then calculating the inverse. Manual F1-F4 and B1-B4 measurements were made from the LPC spectrum generated at the approximate midpoint of the analysis segment. The manual

measurements of B1-B4 made for each Hillenbrand stimulus were used as the input value to compare AASP generated values since Hillenbrand et al (1995) did not report bandwidth measurements. To clarify, manual F1-F4 measurements were made for the synthesized vowels by noting the value at each of the four formant peaks in each AASP. Manual B1-B4 measurements were made, for the synthesized vowels and the Hillenbrand vowels, from the LPC spectrum in TF32 by measuring the width or shoulder of each formant curve at 3dB below the peak (i.e. the half-power point on each side of the peak; Baken, & Orlikoff, 1999, and Kent & Read, 2001). In the event that a formant curve did not fall symmetrically on both sides, commonly due to close proximity to another formant curve, the bandwidth measure was determined by measuring the width from the peak to 3dB down on one side to the peak and doubling this value to represent the full formant bandwidth. It is important to note that the manual measurement cannot be assumed to be completely accurate, but was used as an alternative to the bandwidths calculated algorithmically by the different software packages. The manual method accords with long-standing practice in bandwidth estimation, which is based on the half-power points surrounding a resonant peak. To assess the accuracy and reliability of manual measurements a second researcher repeated 10% of these measurements.

Assessment of accuracy and comparability

Assessment of accuracy and comparability of the four AASPs was based on comparing the discrepancy scores secured. Discrepancy score is defined as the difference between values measured by each of the four AASPs (or measured manually) and the input values for synthesized vowels, or the reported values by Hillendbrand et al. (1995; henceforth the Hillenbrand input values). Comparisons are displayed graphically for the synthesized vowels (Figures 1 and 2) and the Hillenbrand vowels (Figures 3 to 7) where the zero reference line reflects maximum measurement accuracy (no difference between the measured value and the input value). The gray region above and below the zero reference line reflects a range of values (±5 to 10% of the input value, as indicated in the following section) that are considered to be reasonable deviations from the input value. Manual measurements discrepancy scores for the synthesized vowels are displayed as stars in Figures 1 and 2.

To verify our graphical interpretations in comparing software results, a series of nonparametric Friedman tests were conducted on the Hillenbrand stimuli discrepancy scores, but not the synthesized vowels because of the limited number of synthesized vowel tokens (one male and one female vowel token for each of the four vowels compared). Nonparametric Friedman tests were selected rather than parametric tests, such as ANOVA, for two reasons. First, it is apparent that the metric of the outcomes may not have interval level meaning, making comparisons of mean or absolute mean scores across software of questionable value. Second, most parametric tests, such as those based on ANOVA, require homogeneity and normality assumptions that would clearly not be appropriate given the distributions of the outcomes. The Friedman test compares AASP in terms of ranks (1= closet to the zero reference value i.e. most accurate, 4= farthest from the no difference reference value i.e. least accurate), and was applied using speakers and vowels as independent observations by which the four packages could be compared. Separate tests were performed for F0 and F1-F2 collapsed across vowels, speaker type and gender.

The accuracy and reliability of the manual measurements were assessed by comparing a subset of measurements made by two researchers. Paired t-test results, evaluating potential bias in measurements across researchers, was not significant ($t=1.034$; $df=19$; $p=0.314$). In addition, an index of the absolute average relative error (ARE) between measurements of the two researchers was calculated by averaging the absolute difference between their measurements divided by the mean measurement. The resulting ARE was 0.0004, suggesting very small differences in measurement, reflecting good inter-rater reliability.

III. RESULTS

Synthesized vowels

Discrepancy scores for the analysis of synthesized vowels are displayed in Figures 1 and 2. The box and whisker plots reflect discrepancy scores from the AASPs, and the stars for manual measurements. Interpretation of findings is based on these graphic displays without additional statistical analysis given the limited number of synthesized vowel tokens (one of each of the four corner vowels synthesized with a male F0 and a female F0). The F0 discrepancy scores, as shown in Figure 1A and 1B, indicate that three of the four AASPs (Praat, Wavesurfer, and TF32) are highly comparable and results are within the ±5% accuracy of the input value for all four vowels synthesized with male and female F0. For CSL, the discrepancy scores were more variable and contained outliers; and the criterion of ±5% accuracy was met for vowels /i/, /u/, and /ae/ synthesized with male F0, and the back vowels /u/ and /a/ synthesized with female F0.

Measurement accuracy for F1-F4 across the four AASPs did not appear to be comparable, as seen in Figures 1A and 1B. More specifically for F1, only one of the four AASPs (TF32) yielded values that are within ±5% accuracy of the input value for all four vowels synthesized with male and female F0. Similar accuracy was noted for Praat for all four vowels synthesized with a male F0; and for Wavesurfer but only for the low vowels /a/ and /ae/ synthesized with male and female F0. The remaining values varied in accuracy across software based on speaker and vowels. For F2, all four AASPs yielded F2 values that are within ±5% of the input value for vowels synthesized with the male F0; however, there was variability in synthesized vowels with female F0 in two of the software for the low vowels /a/ and /ae/ in Praat and vowels /i/, /u/ and /ae/ in CSL. As for F3, all four AASPs yield values that are within ±5% of the input value for vowels with male F0. The same applies for vowels synthesized with female F0 except for the vowels /i/, /a/ and /ae/ in CSL. Finally, for F4, all four AASPs yield values that are within ±5% of the input value for vowels synthesized with male F0. The same applies for vowels with female F0 except for the vowels /i/ in TF32 and /i/ and /a/ in CSL.

As for manual measurements, findings show highly accurate measurements with a few exceptions where most of the discrepancy scores, displayed as stars, are within the 5% accuracy region (shaded region) and even close to the zero reference line. Most errors occur for F1 and F2 of vowel /u/, which is not surprising given the proximity of the two formants.

To summarize, with a few exceptions, all AASP yield F0 and F1-F4 values that are within ±5% of the input value for vowels with male F0, however there is more software- and vowel-specific variability in vowels synthesized with the female F0.

The manual F1-F4 measurements for both males and females vowels made from the LPC spectrum in each of the four software packages, marked with a star, yield accurate results for most vowels as compared to the input values used for vowel synthesis.

Regarding the bandwidth results for synthesized vowels, findings displayed in Figures 2A and 2B, show that in general all AASP yield B1 to B4 values that vary and are not comparable across the four AASP. Also, they all yield B1-B4 values that exceed the ± 10% of the input value (a criterion that is less stringent than the ±5% criterion used for the F0 and F1-F4 formant values). The exceptions are Praat and Wavesurfer for B4 where three of the four vowels synthesized with male F0s fall within the ±10% of the input value criterion. Overall, the discrepancy scores for TF32 appear to yield B1-B4 measurements that are the least variable across vowels and speakers, and the closest to the synthesis input values (i.e. the zero reference line).

Manual bandwidth measurements marked with a star in Figures 2A and 2B indicate that manual B1-B4 measurements for the male synthesized vowels are generally accurate and close to the synthesis input values. Manual B1-B4 measurements for the female synthesized vowels, however, are more varied but are generally more consistent across vowels and synthesis input values than software generated values.

Hillenbrand vowels (natural speech)

Figures 3 to 7 display the discrepancy scores for the Hillenbrand stimuli. Each figure is specific to one AASP with the adult male and female speakers' discrepancy scores displayed in the left panel, and the child male and female speakers' discrepancy scores in the right panel. The scores were compared statistically by applying a series of nonparametric Friedman tests to assess differences in the output of the software packages. This assessment was done for each speaker group (adult males, adult females, child males, child females) and for each of the following acoustic measurements: F0, F1, F2, F3 and F4 across all four vowels (/i/, /u/, /ae/ and /a/). A Bonferroni correction was applied to account for repeated testing and to control for Type I error. Findings were as follows: F0 was significantly different across software for the adult males (χ^2= 14.70, p=.002) and child females (χ^2= 18.06, p=.000). Comparison of the mean ranks (1=best, 4=worst) for the adult males revealed that three of the four AASPs had similar mean ranks (mean rank for Praat =2.30; Wavesurfer=2.10; TF32=2.15 and CSL=3.45), and based on the mean ranks the AASPs could be ordered as follows: Wavesurfer, TF32, Praat and CSL. The mean rank values indicate that CSL ranks as the lowest among the four AASPs. Similarly, comparison of the mean ranks for the child females revealed that three of the four AASPs had similar mean ranks (Praat =2.50; Wavesurfer=1.95; TF32=2.05 and CSL=3.50) and could be rank ordered as follows: Wavesurfer, TF32, Praat and CLS. This again indicates that CSL ranks as lowest across the four AASPs. However, F0 findings in Figures 3 to 6 further reveal that despite the similarities in discrepancy values for three of the four AASPs, most of the F0 values exceed the ±10% input value criterion (i.e. falls outside the shaded region).

Measurement accuracy for F1-F4 across the four AASPs also suggested differences (see Figures 3 to 6). A Friedman test revealed F1 was significantly different across AASP for adult males (χ^2= 22.20, p=.000). Comparison of the mean ranks revealed that three of the four AASPs (Praat, Wavesurfer and TF32) are similar (mean rank for Praat =2.35; Wavesurfer=2.05; TF32=1.95 and CSL=3.65). Of those, the discrepancy values for the adult male and female speakers are within the ±10% input value criterion for all vowels with the exception of /ae/. See Figures 3 to 6. As for the child speakers, comparison of the discrepancy scores also reveal that in general the values are mostly within the ±10% input value criterion with some speaker and vowel specific exceptions. As for F2, a Friedman test showed no significant differences between the four AASPs. Comparison of the discrepancy values for the adult speakers in Figures 3 to 6, left panel, show that three of the four AASPs (Praat, Wavesurfer and TF32) are mostly within the ±10% input value criterion for all vowels with the exception of /ae/. As for the child speakers, comparison of the discrepancy scores reveals more variability than for the adult speakers across the four vowels. For F3, a Friedman test revealed c significant differences across AASP for both the adult males (χ^2= 26.34, p=000) and adult females (χ^2= 31.98, p=.000). Comparison of the mean ranks revealed that for three of the four AASPs (Praat, Wavesurfer and TF32) are again similar (adult males; Praat =2.00; Wavesurfer=2.35; TF32=1.90 and CSL=3.75; adult females; Praat =1.85; Wavesurfer=2.10; TF32=2.15 and CSL=3.90). Additional comparison of the discrepancy scores for these three AASPs, as displayed in figures 3 to 6 left panel, further reveal that almost all F3 values fall within the ±10% input value criterion. As for the child speakers, while there are no significant differences amongst the four AASPs, comparison of the discrepancy scores show speaker and vowel specific differences that exceed the ±10%

input value criterion. Finally, for F4, a Friedman test revealed significant differences across AASP for all groups except child females (adult male χ^2= 36.24, p=.000; adult female χ^2= 34.02, p=.000; child male χ^2= 17.34, p=.001). Consistent with the findings reported earlier in this section, comparison of the mean ranks revealed that for three of the four AASPs (Praat, Wavesurfer and TF32) results are highly comparable (adult males; Praat =2.00, Wavesurfer=2.10, TF32=1.90 and CSL=4.00; adult females; Praat =2.0, Wavesurfer=1.90, TF32=2.15 and CSL=3.95; child males; Praat =1.95, Wavesurfer=1.89, TF32=2.82 and CSL=3.34). Additional comparison of the discrepancy scores, in Figures 3 to 6, reveal most F4 values to fall within the ±10% input value criterion for adult males and females; and similarly for the child speakers, the values are within the ±10% input value criterion with some speaker and vowel and specific exceptions.

To summarize, comparison of the discrepancy scores for F0 and F1-F4, findings reveal differences between AASP's accuracy and comparability. Closer examinations of the mean ranks reveal three of the four software (Praat, Wavesurfer and TF32) to be comparable despite speaker group and vowel specific variations. Figures 3 to 6 show that none of the AASP yield highly accurate results for all speaker groups and acoustic measurements (F0 and F1-F4). Also, there are formant specific differences with some AASP performing better for higher or lower formants based on speaker group. In general, based on comparison of mean ranks and as displayed graphically, F0 discrepancy scores for all AASPs exceeded the ±10% of the input value criterion for accuracy (shaded in gray in Figures 3 to 6) for all speaker groups and across all vowels, except for child male speakers across most vowels. As for the formant values, in general, the three comparable software (Praat, Wavesurfer and TF32) had F1 to F4 discrepancy scores for the adult speaker groups that were mostly within the ±10% of the input value criterion for accuracy with some low vowel specific challenges evident particularly /ae/. The child speaker groups showed more variability and decreased accuracy. Overall, it appears that TF32 performance was optimal for adult males (F0, and F1-F4), and adult females (F0, and F1-F2). On the other hand, Praat performance was optimal for child male speakers, and Wavesurfer performance optimal for child female speakers.

Hillenbrand et al. (1995) did not report bandwidth values so it was necessary to identify another procedure to establish reference values that could be used to derive discrepancy values. Based on the finding that the manual measurements were accurate for the synthesized vowels with male F0, manual BW measurements for the adult male speakers were used as the input values to calculate discrepancy scores. The manual BW measurements were averaged across the five speakers and the four vowels, and ±10% of the averaged value was used as the input criterion (displayed as the shaded region in Figures 7A and 7B). The discrepancy values displayed in Figures 7A and 7B are the differences of the AASP measured value minus their respective manual measurements. Comparison of the discrepancy scores shows that the results across the four AASPs are highly variable and frequently not comparable. Also, they indicate that none of the bandwidths discrepancy scores fall within the ±10% input value criterion although B1 values come close for two of the four AASPs (Wavesurfer and TF32). The bandwidth discrepancy scores for the remaining speakers (adult females and child speakers) are not displayed given that the findings were even more variable and frequently in excess of the ±10% input value criterion.

Descriptive features

Descriptive features are listed in Tables 3 and 4, which outline similarities and differences between the four AASPs in terms of default settings (Table 3) and software features (Table 4). Aside from differences in default settings between the four software packages, such as analysis window size and type (Table 3), there are differences in the ability to manipulate manufacturer settings in terms of which settings can be manipulated, and the number of

settings available for manipulation (Table 4). Although Praat does not permit manipulation of window type, it does however, along with Wavesurfer and CSL, allow for a great number of setting manipulations, such as window length and pre-emphasis. TF32 on the other hand, allows the user to manipulate fewer settings such as window type and analysis bandwidth. Table 4 is similar to the tables presented in Read et al. (1990, 1992). Although most of the software packages have very similar features, they differ in how those features are displayed (e.g. time readout and the extent of zoom/scroll). CSL provides, with the purchase of the Multi-Dimensional Voice Profile (MDVP) module, a detailed analysis of the speakers' voice. A similar Voice Report analysis is available in Praat.

IV. DISCUSSION

This study made quantitative and descriptive comparisons of four acoustic analysis software packages (Praat, Wavesurfer, TF32, and CSL) to assess the accuracy of each software package as well as the comparability between software. Quantitative assessment entailed analyzing the same synthesized vowels and natural vowels – produced by healthy adults and typically developing children – to secure nine acoustic measurements in each of the four software packages without manipulating any of the manufacturers' default settings. Therefore, the following discussion is limited to results using default settings. Findings highlight the need for users to exercise caution and consider adjustments to default settings so as to achieve optimal analysis results. Descriptive features also are summarized.

Accuracy and Comparability

The acoustic measures for the synthesized vowels, as referenced to the input values, revealed that F0 values obtained by Praat, TF32, and Wavesurfer were accurate and comparable for all corner vowels synthesized with a male and female F0. CSL output values were more variable particularly for vowels synthesized with a female F0. As for formant measurements, results revealed that in general all four AASPs have discrepancy scores that are within the ±5% accuracy of the input value for most formants, and for all vowels synthesized with male and female F0 except for CSL for vowels synthesized with a female F0. In other words, the results are fairly accurate and comparable for vowels synthesized with a male F0.

Bandwidth measurements, on the other hand, were neither accurate nor comparable across the four AASPs. Manual bandwidth measurements B1-B4, though tedious, were a more accurate approach to bandwith analysis. Overall, the manual bandwidth values for the vowels synthesized with a male F0 were more accurate than the software-generated values. For vowels synthesized with a female F0, only B4 manual measurements were consistently accurate. Many B1 and B2 manual measurements were difficult to measure manually because of the narrow bandwidths. Although the manual measurements are varied, these findings, particularly for the adult male synthesized vowels, indicate that manual measurements were the most comparable to the synthesis input values and should be used to calculate bandwidth values rather than using software-generated values with the four software packages studied here.

For natural speech, statistical comparison of the discrepancy scores from the four AASPs for the Hillenbrand stimuli show that the four AASPs were comparable only for the second formant measurement across speaker groups and vowels. The adult male speaker group had the highest number of acoustic measurements that were significantly different across the four AASPs. This finding is somewhat surprising given that most default settings are reportedly set for the typical adult male speaker. Additional comparisons of the mean ranks, however, revealed that when there are significant differences across the four AASPs, in general three of the four AASPs (Praat, TF32, and Wavesurfer) are fairly comparable.

Interestingly, additional comparison of the mean ranks across the different speaker groups, did not reveal any of the software packages to rank consistently better for a particular speaker group or acoustic measure. There was a pattern, however, with TF32 ranking higher (mean rank= 1.9 to 2.15) for adult male speakers for all acoustic measures, and also adult female speakers (mean rank= 1.85 to 2.15) for the first two formants; and Praat ranking higher (mean rank= 1.7 to 2.3) for all acoustic measures for child male speakers. For the adult female speakers, TF32 scored higher for F0 and the first two formant frequencies (mean rank= 1.85 to 2.15), but Praat and Wavesurfer yieleded higher mean ranks for the higher formant frequencies F3 and F4 (mean rank= 1.85 to 2.10). Based on these findings, manual correction of formant measurements or applying a smoothing function (as is standard in Wavesurfer and available in TF32) iswarranted in all software packages and, furthermore, it appears that the default settings in CSL are not optimal for analyzing F0 or formant frequencies in adult females or male and female child speakers. Informal assessments revealed that adjustment of the default settings in CSL can lead to more consistent and comparable measurements of F0 and F1-F4 for the different speaker groups. Comparison of formant values derived from the software to formants inferred from FFT spectrograms is advisable, as such comparison helps detect erroneous formant frequency values particularly for vowels with closely spaced formants. The results for the adult male Hillenbrand vowel bandwidth analysis, displayed in Figure 7A and 7B, suggest that, although each AASP can produce some accurate bandwidth measurements, no single AASP produces consistently accurate bandwidth results. Therefore, clinicians and researchers should not assume the validity of any bandwidth values generated from any of the four AASPs in this study. As was noted for the synthesized vowels, manual bandwidth measurements are the most reliable.

Present findings suggest that using Praat, Wavesurfer, or TF32 (with the manufacturers' default settings) to analyze typical adult speech will generate fundamental frequency and formant frequency measurements that are reliable and comparable across the three packages and comparable to the manual measurements. The agreement for adult female vowels is perhaps surprising, given that settings were not necessarily optimized for female speakers. However, measurements made with CSL (using the manufacturer's default settings) are frequently varied and inaccurate and cannot be reliably compared to those measurements taken in Praat, Wavesurfer, or TF32. Formant bandwidth data indicate that there is not one clearly superior AASP. While each of the four AASPs demonstrates the ability to accurately generate accurate data for certain bandwidths for certain vowels, none of them produced values that compared closely to the manual measurements. This consistency and comparability decreases drastically for the other types of speakers considered in this study. Current results show that the most accurate method to secure formant bandwidth data is the manual method. However, caution must be taken when manually measuring B1 and occasionally B2 as the very small values are often difficult to measure on the LPC spectrum. Additional caution must be taken when manually measuring the bandwidths of closely positioned formants (e.g., F1 & F2 for vowels /a/ and /u/) because the interaction between these formants could lead to inaccurate manual measurements.

Very few studies have reported data on bandwidth for any group of speakers. Presumably, this lacuna in the speech acoustics database could be quickly remedied given that AASP routinely provide bandwidth data along with data on formant frequencies. However, as shown in this report, bandwidth values should not be accepted uncritically. Some of the errors in measurement are serious enough that it is safer to ignore bandwidth data completely than to report erroneous values. This is an unfortunate situation, given that accurate bandwidth data may be useful for several purposes including the study of speech development (Robb, Chen, & Gilbert, 1997; Whiteside & Hodgson, 1999), sleep apnea (Robb, Yates, & Morgan, 1997), and linear prediction modeling of the vocal tract (Mokhtari

& Clemont, 2000). Ideally, software developers should tackle this challenge because aside from the fact that the currently available erroneous bandwidth measurements can be misleading, having accurate bandwidth measurements will be of value for researchers and clinicians.

The decision to use each software package's default settings in this study was based on the assumption that manufacturers set the default settings to the optimal settings for their particular software package. Ease of use is also an important consideration, especially for clinicians to use in a busy clinical setting, and also for those without the in-depth knowledge of acoustics to manipulate these settings. In this study, additional preliminary manipulations of select settings in each of the four AASPs revealed improvements in accuracy over measurements made with the default settings, particularly for CSL. Interestingly, in CSL, not placing the voice period marks, contrary to the tutorial's advice, and adjusting the filter order yields more accurate F1-F3 formant tracking. Similarly, the application of the smoothing function in TF32, which is standard in Wavesurfer, improves software-generated measurements for F3-F4. However, the use of a smoothing function should be done selectively and with caution because it may hinder accurate assessment, particularly for disordered speech.

Descriptive Features

Table 4 contains comparative information on some features of the AASPs examined in this report. Users may find this summary to be a helpful first step in choosing among AASPs to meet their particular needs. The cost of the AASPs varies considerably, with at least two being free at the time of writing this report. Potential users are advised to visit the websites of each system to determine prices.

Considerations on Choosing a Software Package

Accuracy is the sine qua non of any analysis, particularly laboratory analysis. The quantitative assessments reported here should help users to make preliminary decisions about the choice of AASP for analysis. The decision of what AASP to use should be based not only on software availability and quantitative comparisons, but also on the user's needs. Qualitative findings revealed that AASP can benefit different users for different purposes. For users who are in a clinical setting, or who are new to acoustic analysis, TF32 is a good choice due to the userfriendly interface, fast generation of data, and small number of menus to navigate. Other AASPs offer different advantages: Praat has the advantages of annotating and labeling speech files and simplifying the analysis procedures, it can also provide a Voice report; Wavesurfer allows for a considerable range of setting manipulations; and CSL supports the use of additional modules, such as the MDVP, all of which would be useful in some research and clinical settings. The decision of which AASP to use should also be based on the user's knowledge of and familiarity with acoustic analysis and acoustic science as well as the individuals' goals of analysis. Table 4 outlines features of each AASP, which support the recommendation that each user must be familiar with the features of each AASP and determine which will best fit his or her individual needs. Thus, no blanket recommendations on what AASP to use can be made. However, it is necessary for users of these AASPs to be aware that values obtained from acoustic analysis using default settings are not necessarily accurate, however tempting it may be to accept the values unquestioned.

Conclusions and Recommendations

The results for measurements of F0 and F1-F4 frequencies show that Praat, Wavesurfer, and TF32 generate accurate and comparable data for both the synthesized vowels and vowels produced by adult males. For all four AASPs evaluated in this study, accuracy and comparability were less satisfactory for vowels produced by adult females and children.

Results also varied with vowel. Bandwidth measurements by AASPs were highly inaccurate as compared to manual measurements and published data on formant bandwidths. Some of the discrepancies observed in this study are considered to be serious threats to the accuracy of analysis. Users should exercise caution in accepting the values generated by LPC algorithms used in AASPs. To be sure, LPC is a powerful tool that has been extremely valuable in the study of speech acoustics. But users should be aware that LPC algorithms are vulnerable to errors. Cross-checks with other forms of analysis, such as FFT spectrograms, help to guard against analysis errors.

Although it is beyond the scope of this report to discuss in detail the steps that should be taken to obtain the most accurate and complete analyses using speech analysis software, users and developers can take certain steps toward improvements, as explained next.

Users should take care to understand the settings and requirements for any analysis system they use. It is particularly important to note (1) requirements and consequences related to sampling rate of input signals, (2) adjustment of LPC filter order for different speakers, and possibly different speech segments, (3) optional controls for smoothing of formant tracks, (4) adjustment of dynamic range to obtain the highest quality spectrograms, and (5) the ability to compare LPC-derived formant tracks with FFT spectrograms to detect errors in analysis, especially in the case of closely spaced formants. Above all, users should be aware of the possibility that certain data derived from AASPs can be inaccurate, sometimes seriously so. Clinical users of these systems should carefully follow instructions provided in manuals and take special care to examine data for flaws or inconsistencies. But it also should be noted that manuals do not necessarily provide the information needed to perform an optimal analysis of all speakers. It is recommended that researchers note and report analysis settings used in research projects, including downsampling, LPC filter order, and special adjustments made for speaker characteristics. We encourage the systematic collection of data to establish shared databases on measures of vowel formant frequencies and bandwidths. However, given the serious problems with bandwidth estimation in the four AASPs examined here, work on bandwidth must await improvements in the analysis or reliance on alternative means of bandwidth measurement. Until such improvements are made, users of these AASPs should recognize that certain measures, such as formant bandwidths, run a considerable risk of error.

Software developers should strive to improve the accuracy of analysis, or, at the minimum, caution users about the likelihood of errors in certain measures, such as formant bandwidth. Vallabha and Tuller (2004) pointed out that LPC algorithms confront two major sources of error: exclusive reliance on root solving, and use of the 3-point parabolic interpolation to compensate for the coarse spectrum. Refinement of the algorithms should be a major consideration for future development. It is also recommended that software developers guide users in making the most effective adjustments of analysis parameters for different speakers and different speech materials. Effective application of acoustic analysis systems requires a close collaboration of users and software developers.

The present report is just one step in a larger process of evaluating software for the acoustic analysis of speech. The long-term goals are to encourage continued refinement of AASPs, to ensure the validity of acoustic speech databases, and to promote the use of speech analysis for different speakers and different purposes. We agree with Woehrling and Mareuil (2007) conclusion that there can be substantial differences in the values generated for the same speech sample by different AASPs. Accuracy of measurement is essential to establishing a valid database for acoustic measurements and for the application of acoustic measures for various purposes.

Acknowledgments

This work was supported by NIH Research Grant R01-DC 006282 from the National Institute of Deafness and other Communicative Disorders, and Core Grant P-30 HD03352 from the National Institute of Child Health and Human Development. We declare no financial conflict of interest with any of the software systems considered in this review. We did contact all software developers to seek clarification or obtain additional information. In particular, we requested and obtained from the developer of TF32 software updates that provided formant bandwidth and fourth formant measurements that enabled us to meet the purpose of this study in terms of having comparable measurements to assess across all four software packages. We thank Drs. Jan R. Edwards and Gary G. Weismer for providing comments on previous versions of this paper. We also thank Erin Nelson for assistance with graphing, Ekaterini Derdemezis for verifying revision accuracy, Allison G. Petska for securing references, and Michael P. Kelly for assisting with statistical analysis. This research was originally submitted by the first author as a Master's thesis for the Department of Communicative Disorders at the University of Wisconsin-Madison. Portions of this research were presented in 2011 at the American Speech Language Hearing Association Convention in San Diego, CA.

REFERENCES

Baken, RJ.; Orlikoff, RF. Clinical Measurement of Speech & Voice (Speech Science). 2 ed.. San Diego: Singular; 1999.

Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. Journal of Speech and Hearing Research. 1996; 39:126–134. [PubMed: 8820704]

Boersma, P.; Weenink, D. Praat (5.1.32). Amsterdam, The Netherlands: Publisher; 2010. Available from http://www.fon.hum.uva.nl/praat

Fant CG. Descriptive analysis of the acoustic aspects of speech. Logos. 1962; 5:3–17. [PubMed: 13891546]

Fourakis M, Preisel C, Hawks JW. Perception of vowel stimuli synthesized with different fundamental frequencies. Journal of the Acoustical Society of America. 1998; 104:1778.

Deliyski DD, Evans MK, Shaw HS. Influence of data acquisition environment on accuracy of acoustic voice quality measurements. Journal of Voice. 2005; 19:176–186. [PubMed: 15907432]

Hawks JW, Miller JD. A formant bandwidth estimation procedure for vowel synthesis. Journal of the Acoustical Society of America. 1995; 97:1343–1344. [PubMed: 7876453]

Hillenbrand JM, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97(5 Pt 1):3099–3111. [PubMed: 7759650]

Hillenbrand, JM.; James, M.; Hillenbrand. Retrieved Sept 30, 2009, from http://homepages.wmich.edu/~hillenbr/voweldata.html

Ingram K, Bunta F, Ingram D. Digital data collection and analysis: application for clinical practice. Lang Speech Hear Serv Sch Language, Speech, and Hearing Services in Schools. 2004; 35:112–121.

Karnell MP, Hall KD, Landahl KL. Comparison of fundamental frequency and perturbation measurements among three analysis systems . Journal of Voice. 1995; 9:383–393. [PubMed: 8574304]

Kay Elemetrics. Computer Speech Lab (4300). Lincoln Park, NJ: KayPentax; 1996. Available from http://www.kayelemetrics.com/index.php?option=com_product&controller=product&Itemid=3&cid[]=11&task=pro_details

Kent, RD.; Read, C. Acoustic Analysis of Speech. 2 ed.. San Diego: Singular; 2001.

Klatt DH. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America. 1980; 67:971–995.

Milenkovic, P. TF32 (Alpha). Madison, WI: Publisher; 2010. Retrieved January 21, 2010. Available from http://userpages.chorus.net/cspeech/

Mokhtari, P.; Clermont, F. New perspectives on linear-prediction modeling of the vocal tract: Uniqueness, formant-dependence and shape parameterization. In: Barlow, M., editor. Proceedings of the eighth Australian international conference on speech science and technology. Canberra, Australia: Australian Speech Science and Technology Association; 2000. p. 478-483.

Nearey TM. Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America. 1989; 85:2088–2113. [PubMed: 2659638]

Read C, Buder EH, Kent RD. Speech analysis systems: a survey. Journal of Speech and Hearing Research. 1990; 33:363–374. [PubMed: 2193195]

Read C, Buder EH, Kent RD. Speech analysis systems: an evaluation. Journal of Speech and Hearing Research. 1992; 35:314–332. [PubMed: 1573872]

Robb MP, Chen Y, Gilbert HR. Developmental aspects of formant frequency and bandwidth in infants and toddlers. Folia Phoniatrica et Logopaedica. 1997; 49:88–95. [PubMed: 9197091]

Robb MP, Yates J, Morgan EJ. Vocal tract resonance characteristics of adults with obstructive sleep apnea. Archives of Otolaryngology. 1997; 117:760–763.

Sjolander, K.; Beskow, J. WaveSurfer (1.8.5). Stockholm. Sweden: Publisher; 2005. Available from http://www.speech.kth.se/wavesurfer/

Smits I, Ceuppens P, De Bodt MS. A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. Journal of Voice. 2005; 19:187–196. [PubMed: 15907433]

Vallabha GK, Tuller B. Systematic errors in the formant analysis of steady-state vowels. Speech Communication. 2002; 38:141–160.

Vallabha, G.; Tuller, B. Choice of filter order in LPC analysis of vowels. In: Slifka, J.; Manuel, S.; Matthies, M., editors. From sound to sense: 50+ years of discoveries in speech communication. Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology; 2004. p. C203-C208.[Compact disk]

Vogel AP, Maruff P. Comparison of voice acquisition methodologies in speech research. Behavior Research Methods. 2008; 40:982–987. [PubMed: 19001389]

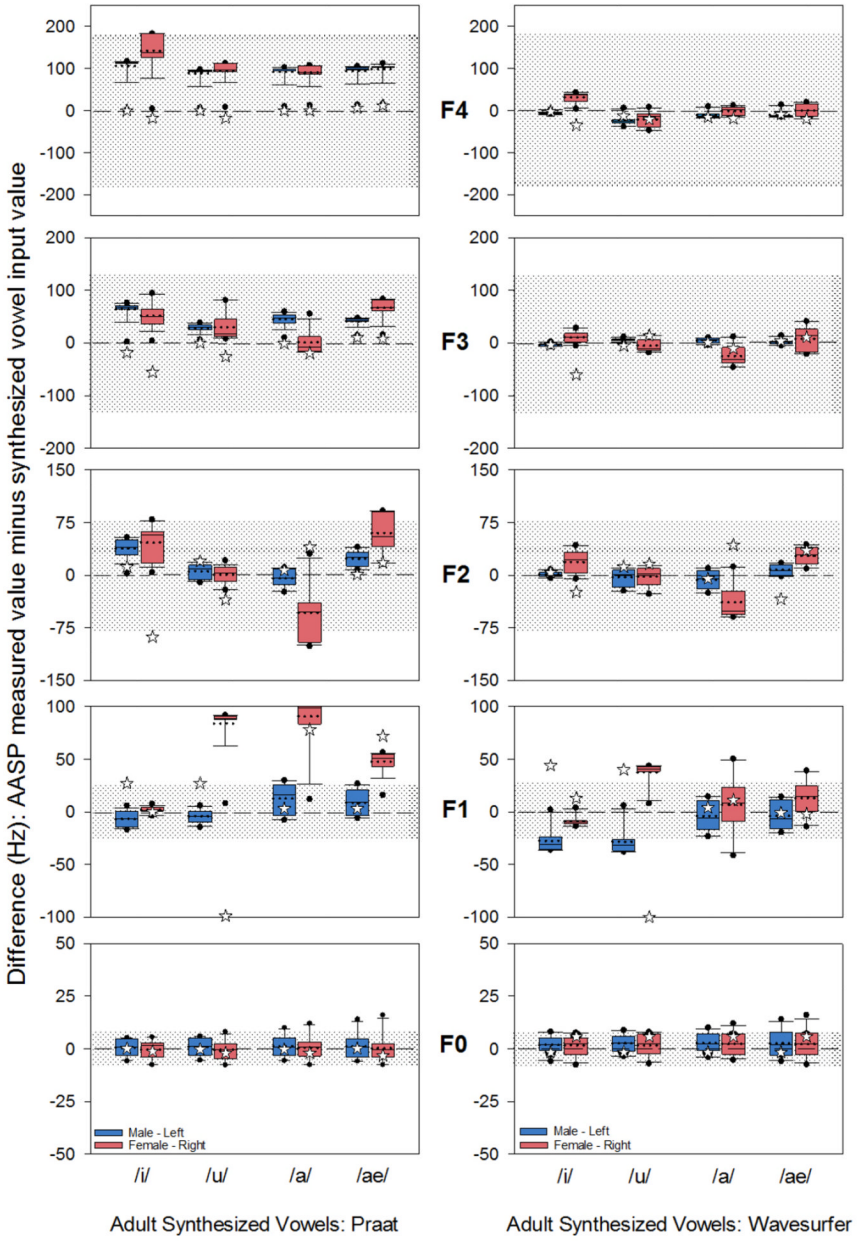
Whiteside SP, Hodgson C. Acoustic characteristics in 6–10-year-old children's voices: some preliminary findings. Logopedics, Phoniatrics, Vocology. 1999; 24:6–13.

Woehrling, C.; Mareuil, P.; Boula de. Antwerp, Belgium. Interspeech-2007, 8th Annual Conference of the International Speech Communication Association; Autust 27–31; 2007. ISSN 1990–9772l; ISCA Archive http://www.isca-speech.org/archive/interspeech_2007

APPENDIX A

The currently available version of TF32 for free download (revised July 26, 2005) does not include software-generated values for F4 or B1-B4. For the purposes of this study, the developer of the TF32 software program P. Milenkovic implemented modifications to that software to track the fourth formant and to provide numeric output for the fourth formant as well as all four formant bandwidths. We refer to the modified program as TF32 alpha version 1.2, which determine formants from LPC analysis according to a covariance-method algorithm giving equal least-squares weights to the samples within a pitch-synchronous analysis interval. The available version designated as revised July 26, 2005 uses varying least-squares weights to reduce the amount of ripple during the closed-glottis interval of the analysis interval, both for inverse filter analysis of the voice source along with formant calculation. As for formant bandwidth calculations, the modified TF32 program uses a method similar to the manual bandwidth measurement method used i.e. the formant bandwidths are computed based on the 3dB bandwidth of the spectrum shoulder. The software can also compute the frequency span from the formant peak to the nearest shoulder and then double this value when the peak does not fall equally on both sides. When the software is unsuccessful in its calculations, it reports the value zero.

A



B

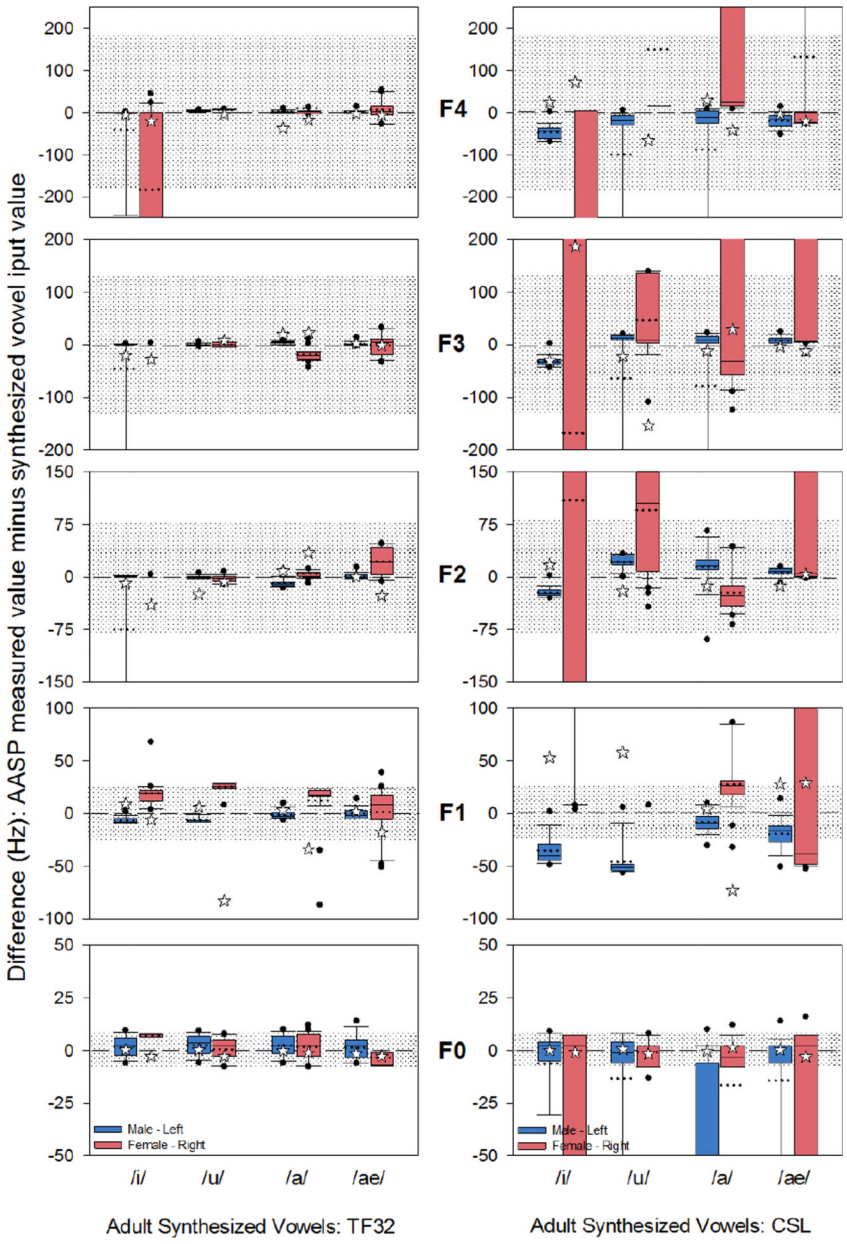


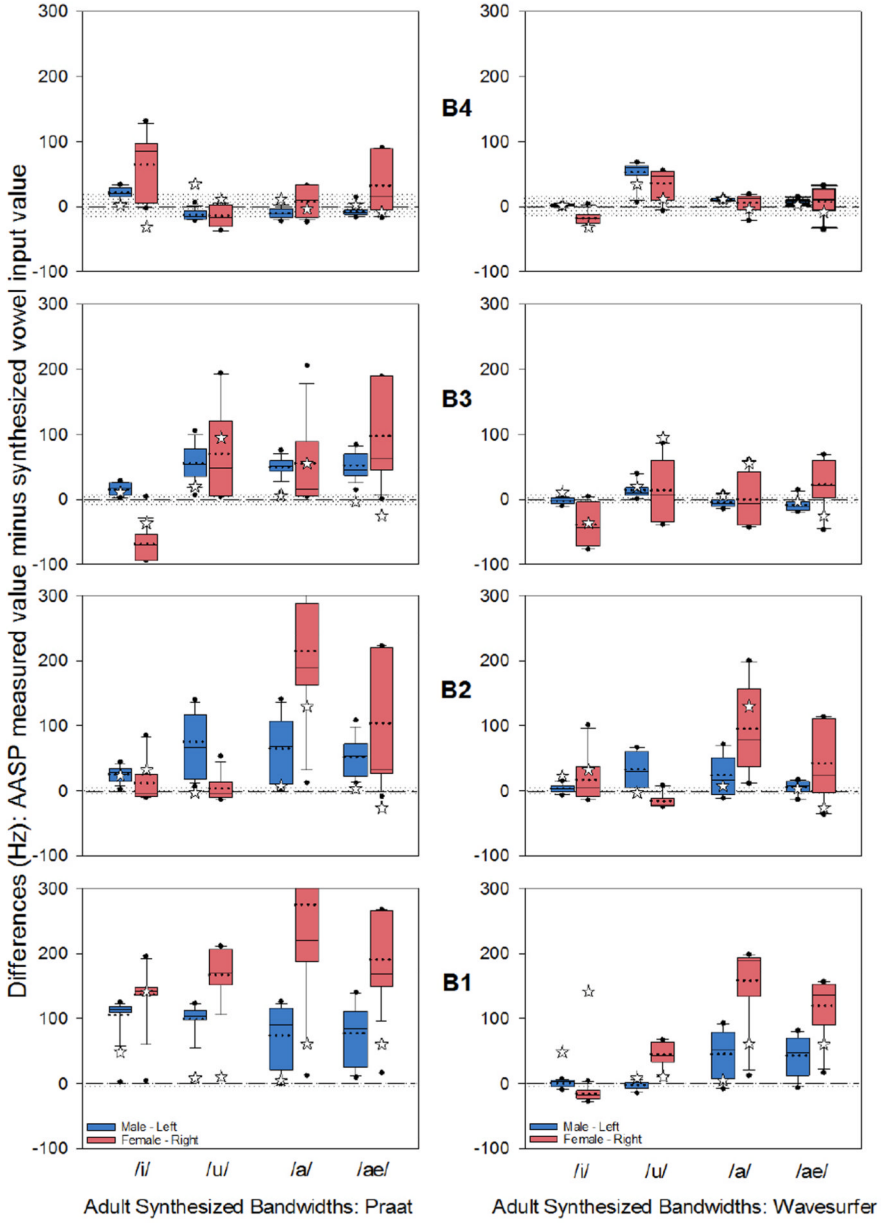
Figure 1.

A. Discrepancy scores for the fundamental frequency (F0) and formant frequencies (F1 to F4) for the four synthesized vowels. The box plots display the 25th and 75th percentile of the discrepancy scores, as well as the mode (solid line) and the median (dotted line). The whiskers display the 5th and 95th percentiles with the outlying data displayed as dots. The zero reference line is the measurement accuracy reference where zero discrepancy indicates no difference between the acoustic analysis software package (AASP) measured value and the input value for the synthesized vowel. The gray region above and below the zero reference line reflects ±5% range of synthesis input value. Manually measured F0 and F1-F4

are displayed with a star symbol. Left panel displays discrepancy scores using Praat, and the right panel Wavesurfer

B. Discrepancy scores for the fundamental frequency (F0) and formant frequencies (F1 to F4) for the four synthesized vowels, using TF32 (left panel) and CSL (right panel). The zero or accuracy reference line refers to no difference between the measured value and the input value for the synthesized vowel. Manually measured F0 and F1-F4 are displayed with a star symbol. For additional information regarding box plot or shaded region, refer to Figure 1A caption.

A



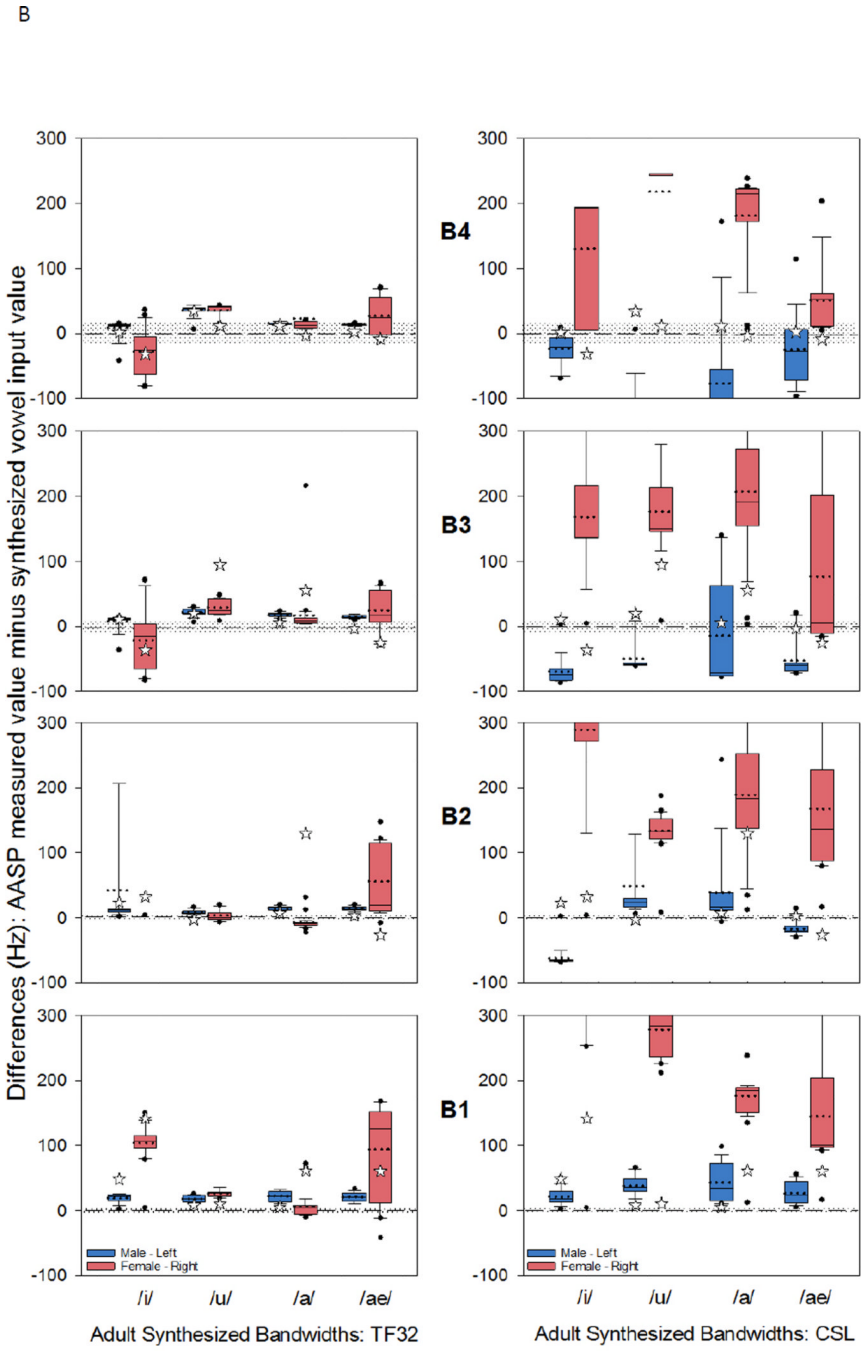


Figure 2.

A. Discrepancy scores for bandwidth (in Hz) for the four synthesized vowels. The box plots display the 25th and 75th percentile of the discrepancy scores, as well as the mode (solid line) and the median (dotted line). The whiskers display the 5th and 95th percentiles with the outlying data displayed as dots. The zero reference line is the measurement accuracy reference where zero discrepancy implies no difference between the acoustic analysis software package (AASP) measured value and the input value for the synthesized vowel. The gray region above and below the zero reference line reflects ±10% range of synthesis

input value. Manually measured B1-B4 are displayed with a star symbol. Left panel displays discrepancy scores using Praat, and the right panel Wavesurfer

B. Discrepancy scores for bandwidth (in Hz) for the four synthesized vowels. The box plots display the 25^{th} and 75^{th} percentile of the discrepancy scores, as well as the mode (solid line) and the median (dotted line). The whiskers display the 5^{th} and 95^{th} percentiles with the outlying data displayed as dots. The zero reference line is the measurement accuracy reference where zero discrepancy implies no difference between the acoustic analysis software package (AASP) measured value and the input value for the synthesized vowel. The gray region above and below the zero reference line reflects ±10% range of synthesis input value. Manually measured B1-B4 are displayed with a star symbol. Left panel displays discrepancy scores using Praat, and the right panel Wavesurfer

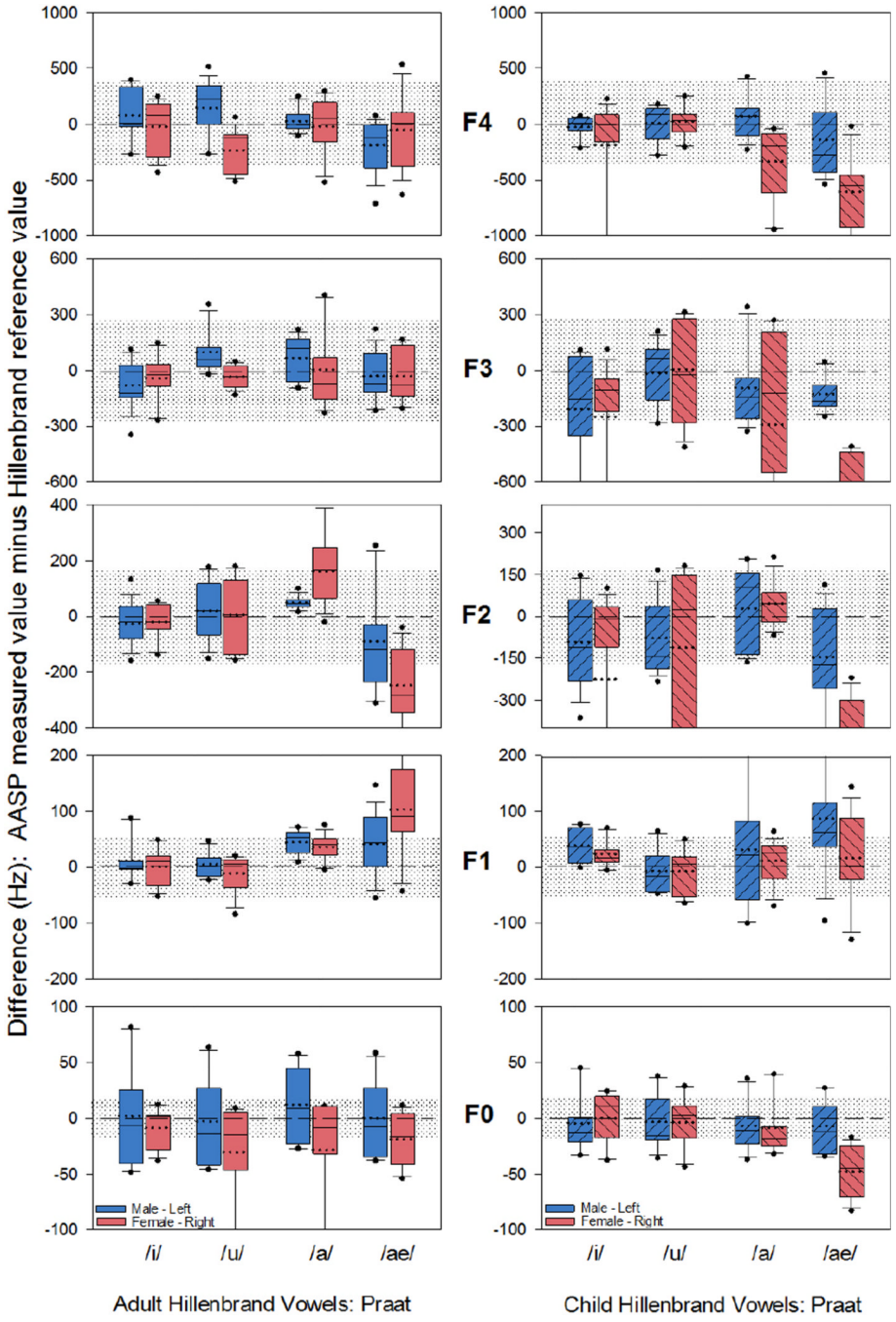


Figure 3.

Discrepancy scores using Praat for the adult male and female (left panel) and child male and female (right panel) speaker's fundamental (F0) and formant frequencies (F1-F4) for the four Hillenbrand vowels. The box plots display the mean, 25th percentile value, and 75th percentile value for F0 and F1-F4, in addition to the mode (solid line) and the median (dotted line). The zero or accuracy reference line represents the Hillenbrand et al (1995) reported values averaged across the five speakers analyzed, and the shaded region reflects ± 10% of the this averaged value.

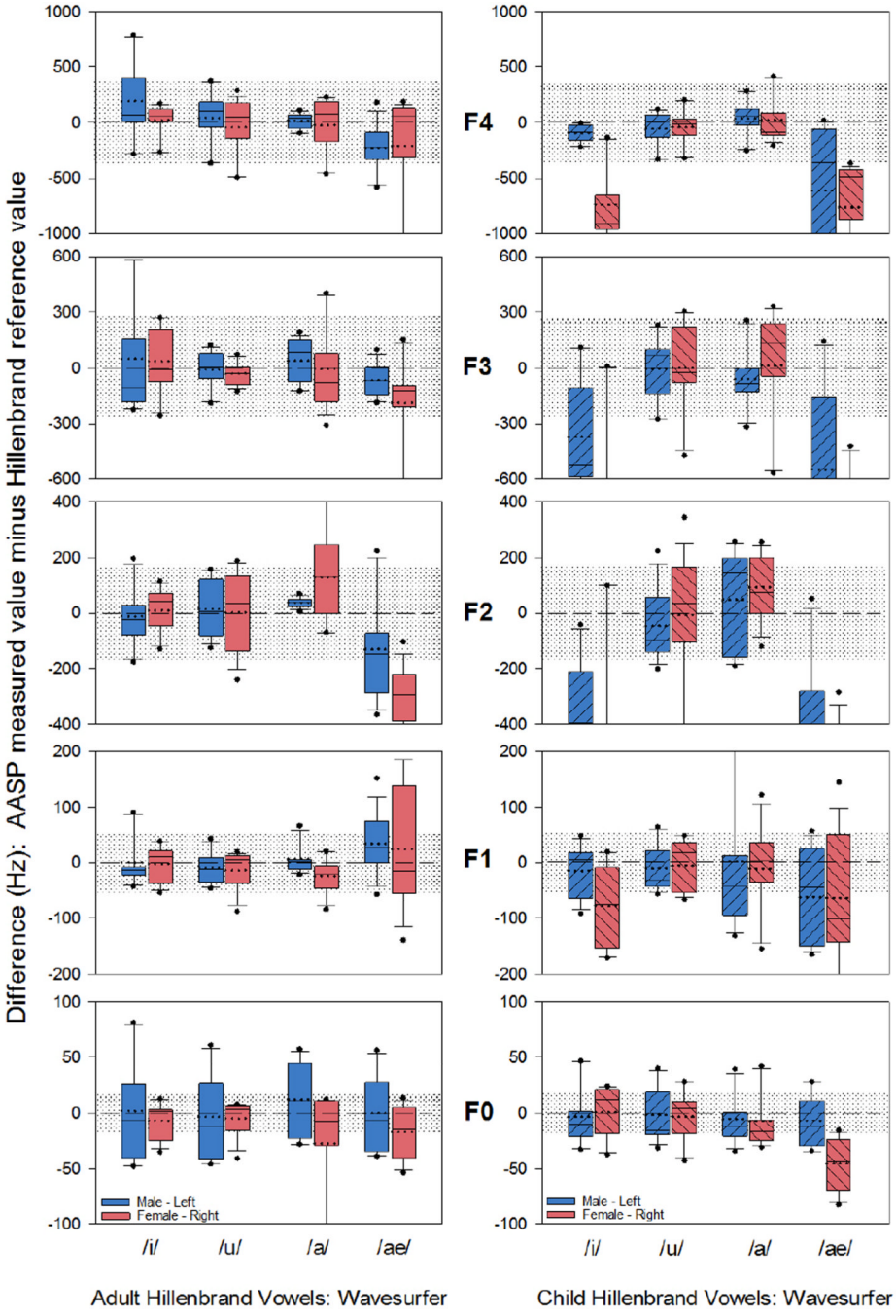


Figure 4.

Discrepancy scores using Wavesurfer for the adult male and female (left panel) and child male and female (right panel) speaker's fundamental (F0) and formant frequencies (F1-F4) for the four Hillenbrand vowels. For additional information regarding box plot or shaded region, refer to Figure 3 caption.

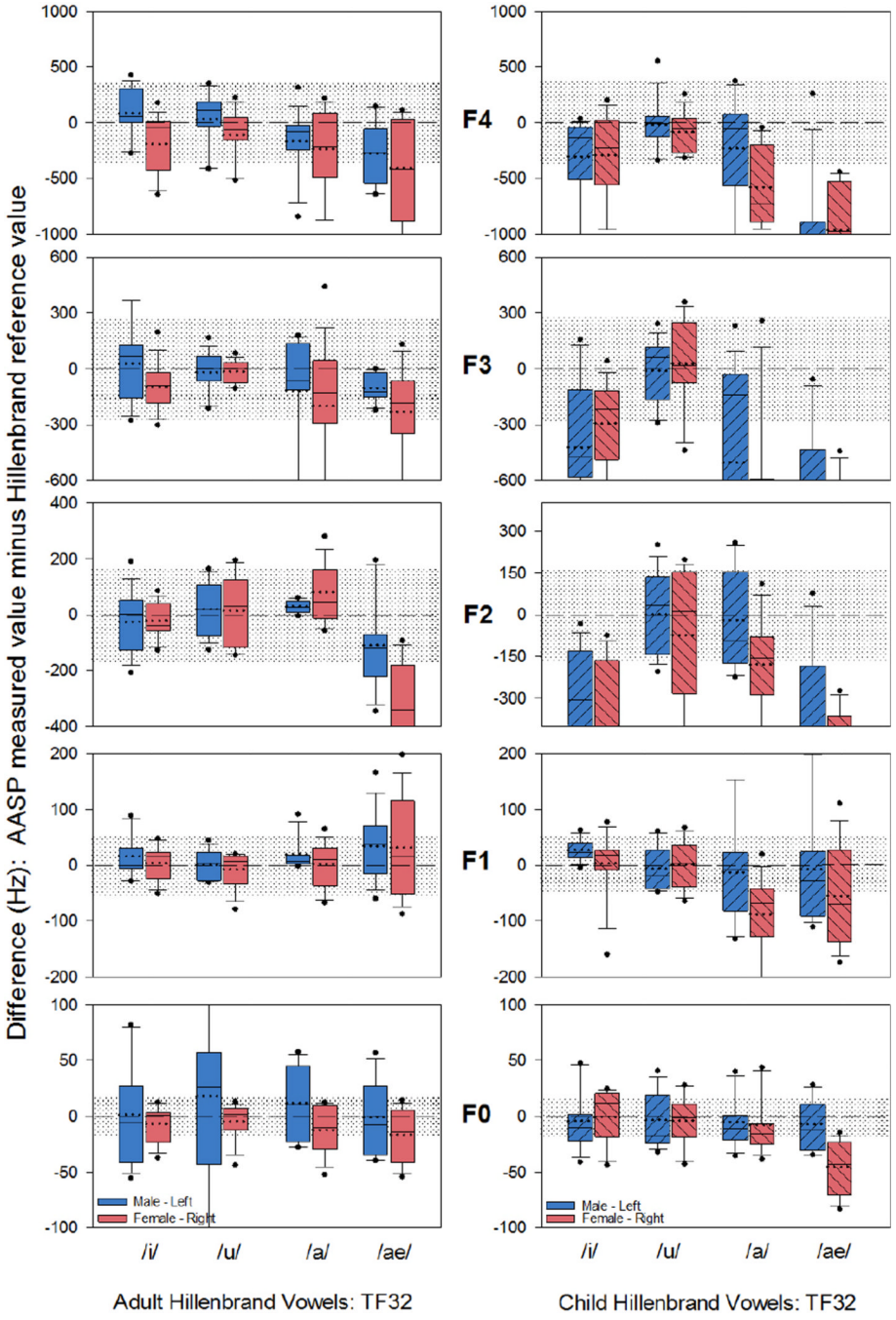


Figure 5.

Discrepancy scores using TF32 for the adult male and female (left panel) and child male and female (right panel) speaker's fundamental (F0) and formant frequencies (F1-F4) for the four Hillenbrand vowels. For additional information regarding box plot or shaded region, refer to Figure 3 caption.

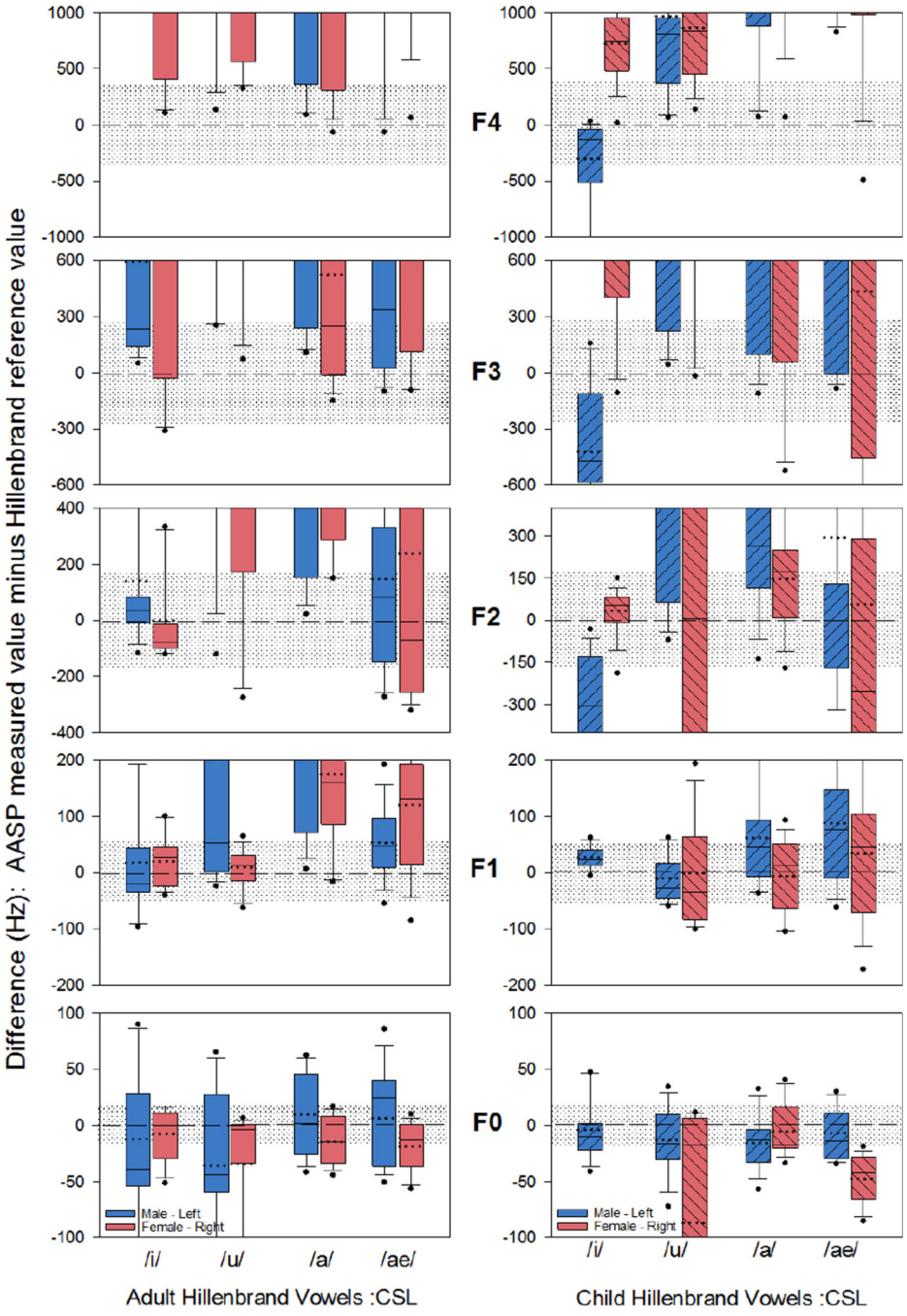
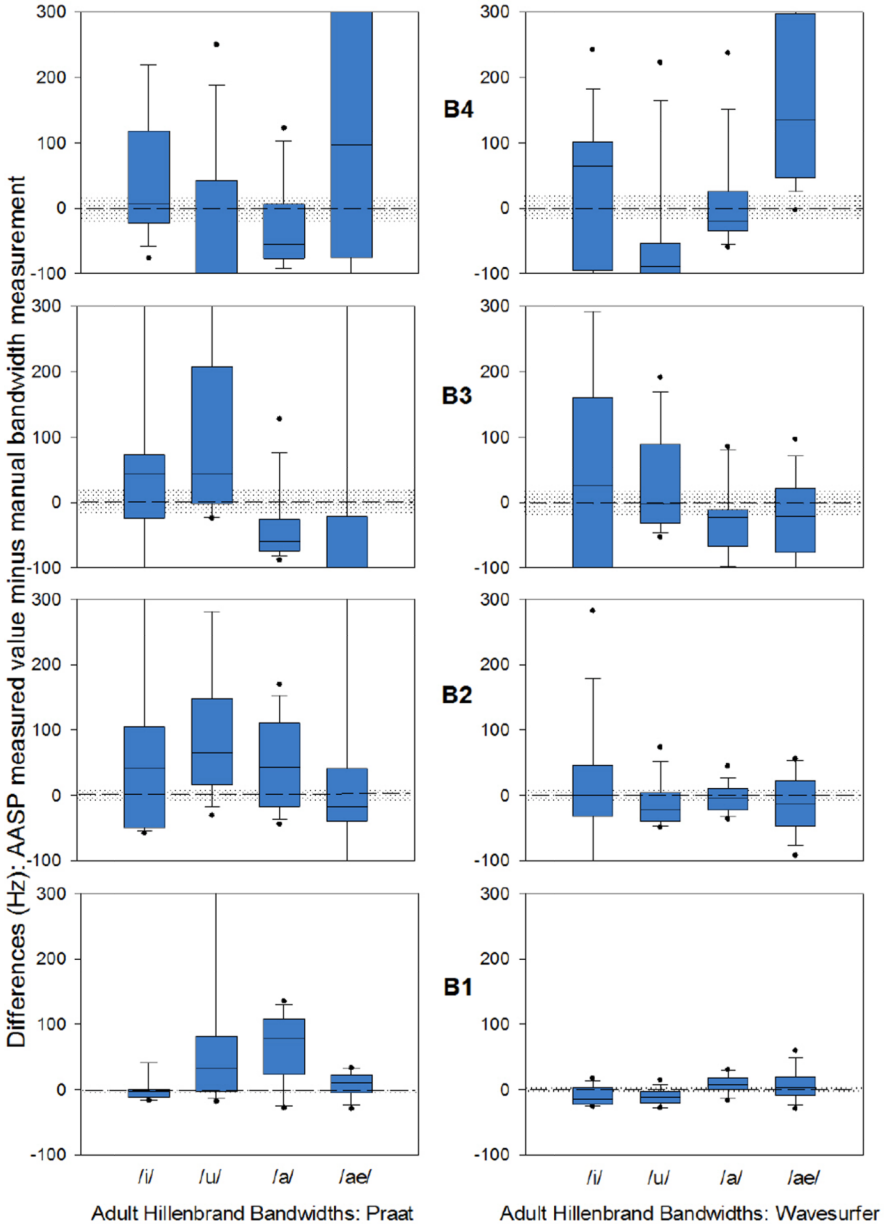


Figure 6.

Discrepancy scores using CSL for the adult male and female (left panel) and child male and female (right panel) speaker's fundamental (F0) and formant frequencies (F1-F4) for the four Hillenbrand vowels. For additional information regarding box plot or shaded region, refer to Figure 3 caption.

A



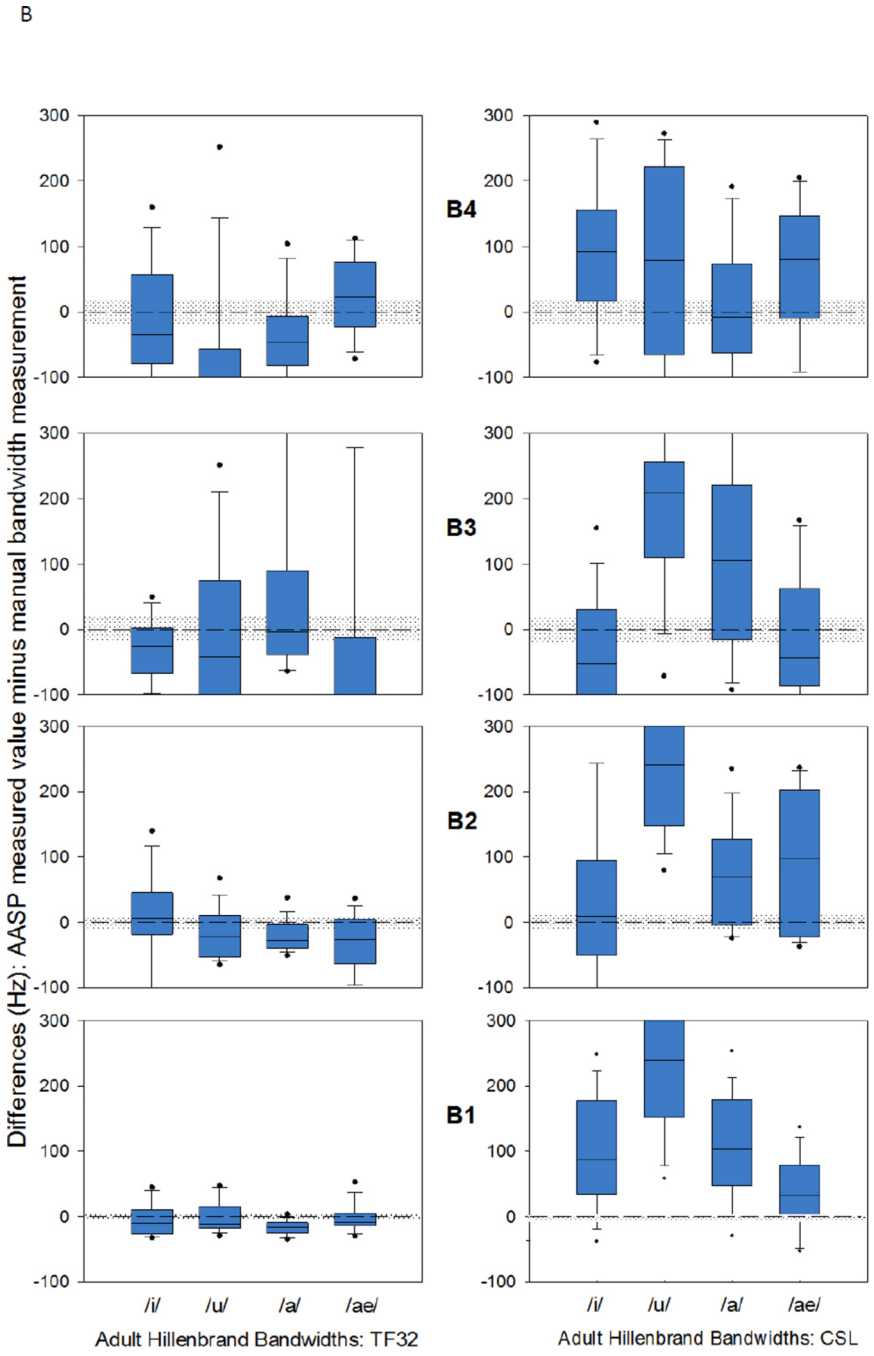


Figure 7.

A. Discrepancy scores for the bandwidth of the four synthesized vowels, using Praat (left panel) and Wavesurfer (right panel). The zero or accuracy reference line refers to no difference between the manually measured value and the AASP measured value. The gray region above and below the zero reference line reflects ±10% range of manually measured bandwidth value. For additional information regarding box plot refer to Figure 2A caption.

B. Discrepancy scores for the bandwidth of the four synthesized vowels, using TF32 (left panel) and CSL (right panel). For additional information refer to Figure 7A caption.

TABLE 1

Formant ranges used to determine the corner vowels from the array of synthesized vowels.

Vowel:	F1 range (Hz):	F2 range (Hz):
/i/	200–400	2200–2400
/ae/	550–750	1700–1900
/a/	550–750	1000–1200
/u/	200–400	750–950

TABLE 2

Input values for the chosen synthesized vowels. F0, F1, F2, and F4 values (in Hz) were collected from M. Fourakis' archives, F3 values (in Hz) were determined from Nearey (1989), and B1-B4 values (in Hz) were determined from Hawks & Miller (1995). Because the F0 was synthesized with a falling pitch, the F0 measure listed includes the highest value followed by the lowest value.

Vowel:	F0	F1	F2	F3	F4	B1	B2	B3	B4
Male /i/	145→105	300	2400	3086	3700	52	85	137	194
Male /u/	145→105	300	900	2248	3700	52	41	75	194
Male /a/	145→105	750	1110	2526	3700	41	42	93	194
Male /ae/	145→105	750	1750	2543	3700	41	53	94	194
Female /i/	245→205	250	2700	3419	3700	63	105	167	194
Female /u/	245→205	350	900	2288	3700	44	41	78	194
Female /a/	245→205	800	1200	2532	3700	41	43	93	194
Female /ae/	245→205	800	1950	2809	3700	41	61	114	194

TABLE 3

Settings for formant or spectrogram analyses in the four softwares.

	Praat	**Wavesurfer**	**TF32**	**CSL**
File types accepted	WAV, AIFF, AIFC, NeXT/Sun AU, NIST, FLAC, MP3	WAV, Sun AU, AIFF, MP3, CSL, SD	CSpeech, NCVS92 (new CSpeech), WAV, NSP, TIMIT database (CMU and SPHERE), UW Microbeam Database compressed (ACC)	NSP, WAV, RAW
Window type with default listed first other choices in parentheses	Gaussian	Hamming (others: Hanning, Bartlett, Blackman, Rectangular)	Hamming (other: Rectangular)	Blackman (others: Rectangular, Triangular, Hamming, Hanning)
Window length	25 ms for formants 5 ms for spectrograms	49 ms	6.7 ms for wide-band 22 ms for narrow-band	10 ms
Frame interval	3 ms	10 ms	Frame = 1 pitch period (no frame overlap)	10 ms
LPC analysis method	Burg	Modified Burg	Weighted least-squares	Autocorrelation
LPC order (number of coefficients) with default and range	10 (1 – 400+)	12 (1 – 40)	26 (5 – 240)	12 (2 – 36)
Smoothing	None	Automatic	Optional	Optional

TABLE 4

Comparison of select features of the four acoustic analysis software packages.

Feature **Version**	**Praat** **V. *5.1.31***	**Wavesurfer** **V. *1.8.5***	**TF32** **V. *alpha 1.2***	**CSL** **V. *3.4.1***
Compatible operating system(s) required	MacOS X, Windows® (2000, XP, Vista, 7) (32 or 64-bit) and Linux (32 or 64-bit)	Windows® XP or Vista 7	32-bit Windows® (98/NT/2000/XP)	Current CPU greater than or equal to 2 GHz, 1GB RAM 2. Windows® 7, Vista, or XP
Users' manual	• Online PDF of tutorial (27 pages) • Active user group to answer questions and offer advice	• Manual in help menu (6 pages with links to additional pages)	• No help menu Online PDF of user's manual (116 pages)	• In-software tutorial • Manual in help menu (464 pages)
Zoom/scroll	• Zoom to selection, zoom in/out, zoom out full • Scrollbar	• Zoom to selection, zoom in/out, zoom out full • Scrollbar	• Zoom to selected, zoom out full • Scrollbar • Keyboard arrows to move along waveform	• Select and out full. • Scrollbar • Keyboard arrows to move along waveform
Time readout	• 0.001ms	• 1ms	• 0.001ms	• 0.01ms
Amplitude readout	• Yes	• Yes	• No	• Yes
Frequency readout	• Yes	• Yes	• Yes	• Yes
View and play selected segments	• Yes	• Yes	• Yes	• Yes
Settings held when open multiple files	• Yes	• No	• Yes	• Yes
FFT	• Yes	• Yes	• Yes	• Yes
LPC	• Yes	• Yes	• Yes	• Yes
Waterfall	• No	• No	• No	• Yes
Pitch	• Yes	• Yes	• Yes	• Yes
Simultaneous displays	• Yes	• Yes	• Yes	• Yes
Labeling/annotation of segments	• Yes	• No	• No	• Yes
Settings maintained from one file to another during continued use	• Yes	• No	• Yes	• Yes

Feature **Version**	**Praat** **V. *5.1.31***	**Wavesurfer** **V. *1.8.5***	**TF32** **V. *alpha 1.2***	**CSL** **V. *3.4.1***
External proprietary hardware required	• No	• No	• No	• Yes
Database(s) Available with system	• No	• No	• No	• Yes
Additional features and information	• Labeling and segmenting, available at no cost • Voice report offers voice measurements	• Automatic smoothing function	• Optional smoothing function	• Multiple additional modules, including MDVP, must be purchased
Edit formant tracking on spectrogram	• Yes	• Yes	• Yes	• Not specified