Measures of Vocal F_0 from Continuous Speech Samples: An Interprogram Comparison

Mesures de la F_0 vocale à partir d'échantillons continus de la parole : une comparaison interlogiciels

Shaheen N. Awan & Shelley E. Scarpino

Abstract

Three commercially available computer programs for voice analysis were used to compute measures of fundamental frequency (F_0) from samples of the second sentence of the "Rainbow Passage" in groups of normal speakers (adult males, adult females, and children). Results indicated that interprogram correlations for mean F_0 estimation were strong (r's > .90) and all of the programs produced mean F_0 estimates within 5% of each other. However, based on the present data analyses, differences in the acoustic estimation of the voices of adult males suggest that care must be taken when analyzing and interpreting obtained fundamental frequency data. Significant differences in estimating F_0 standard deviations were also observed when analyzing adult female voices. Correlations of equivalence among the programs for measures of F_0 standard deviation were weaker in nature than those observed for estimates of mean speaking F_0 and were nonsignificant for the analysis of adult male voices. Results indicated that measures associated with speaking F_0 variability and range may be particularly influenced by gross F_0 extraction errors. Details of computer program algorithms, description of possible sources of gross F_0 extraction errors, and suggestions for recognizing and revising gross errors in F_0 extraction are provided.

Abrégé

On a employé trois logiciels commerciaux d'analyse de la voix pour obtenir des mesures de fréquence fondamentale (F_0) à partir d'échantillons de la deuxième phrase du « *Rainbow Passage* » provenant de groupes de locuteurs normaux (hommes adultes, femmes adultes et enfants). Les résultats obtenus ont indiqué de fortes corrélations interlogiciels d'estimation de la F_0 moyenne (r > 0,90) et tous les logiciels ont produit des estimations de la F_0 moyenne ayant des écarts inférieurs à 5 p. 100 l'un de l'autre. Toutefois, en se fondant sur les analyses de données actuelles, des différences d'estimation acoustique des voix des hommes adultes suggèrent qu'il faut user de prudence dans l'analyse et l'interprétation des données de fréquence fondamentale. On a également observé d'importantes différences dans l'estimation des écarts F_0 type dans l'analyse des voix de femmes adultes. Les corrélations d'équivalence d'un logiciel à l'autre pour ce qui est des mesures des écarts F_0 type étaient moindres en nature que celles observées pour les estimations de la F_0 moyenne de la voix et qu'elles étaient non significatives dans l'analyse des voix des hommes adultes. Les résultats ont indiqué que les mesures associées à la variabilité et à la plage de la F_0 de la voix peuvent être grandement influencées par des erreurs grossières d'extraction de la F_0. On présente des détails des algorithmes des logiciels, une description des sources possibles des erreurs grossières d'extraction de la F_0 et quelques suggestions visant à reconnaître et à réviser erreurs grossières d'extraction de la F_0.

Key words: Fundamental frequency; fundamental frequency variability; speech analysis systems; fundamental frequency extraction algorithms.

Introduction

Descriptions of pathological vs. normal voice characteristics often have focused on analyses of sustained vowel samples, since many of the traditional methods for quantifying perturbations in the acoustic signal (e.g., jitter, shimmer, and harmonic-to-noise ratio) are only valid when applied to signals of expected frequency and intensity stability (Baken, 1987; de Krom,

Shaheen N. Awan
Shelley E. Scarpino
Dept. of Audiology & Speech Pathology
Bloomsburg University
Bloomsburg, PA USA

1994; Horii, 1979). However, in everyday communicative settings, judgements regarding the normality of a speaker's voice are generally made based upon the perception of continuous speech (Parsa & Jamieson, 2001). A number of studies have reported that acoustic measures extracted from continuous speech samples may be useful in the discrimination between normal vs. disordered voice and contribute information more relevant to listener perception than measures obtained from sustained vowels alone (Askenfelt & Hammarberg, 1986; de Krom, 1995; Parsa & Jamieson, 2001).

Two commonly used measures derived from continuous speech samples are mean speaking fundamental frequency (F_0) and F_0 standard deviation, and both have been observed to be important discriminators for various aspects of normal and disordered voices. Measures of mean speaking F_0 have been used to discriminate among various aspects of normal voice such as gender (Aronson, 1990; Boone & McFarlane, 1988), race (Awan & Mueller, 1996; Hudson & Holbrook, 1982; Wheat & Hudson, 1988), and aging (Awan & Mueller, 1992; Honjo & Isshiki, 1980; Liss, Weismer, & Rosenbek, 1990; Pedersen, Moller, Krabbe, & Bennett, 1986). In addition, changes in mean speaking F_0 have also been associated with disordered states such as hyperfunctional voice and laryngeal fatigue (Morrison & Rammage, 1993; Stemple, Stanley, & Lee, 1995), neurological conditions (Canter, 1963; Dromey, Ramig, & Johnson, 1995), and the presence of mass lesions or distributed tissue change (Gilbert & Weismer, 1974; Hirano, Tanaka, Fujita, & Terasawa, 1991; Murry, 1982). F_0 standard deviation (i.e., the average variation of F_0) is referred to as a measure of long-term variability in which variations in frequency occur more slowly than the glottal vibration itself (Hartelius, Buder, & Strand, 1997). Changes in the F_0 standard deviation in speech may reflect characteristics such as monopitch, excessive pitch variability, or reductions in control of speaking F_0. Measures of F_0 variability in speech have been observed to be an important discriminator of normal aging populations (Awan, 2001; Morris & Brown, 1994) and have also been used to characterize vocal function in cases of psychological disturbances (Leff & Abberton, 1981), dysarthric states (Countryman, Hicks, Ramig, & Smith, 1997; Zwirner, Murry, & Woodson, 1991), mass lesions and distributed tissue change (Hirano et al., 1991; Murry, 1982), and cases of deafness/hearing impairment (Horii, 1982; Monsen, 1979). Increased F_0 variability has been described as one of the key characteristics of disordered voice (Callan, Kent, Roy, & Tasko, 1999; Wolfe & Steinfatt, 1987).

Measures of mean speaking F_0 and F_0 standard deviation obtained via computer software analysis have become common additions to data collected in voice diagnostic and therapeutic situations. While several studies have provided information on the capabilities and usability of several software packages (Mann, 1987; Read, Buder, & Kent, 1990; Ryalls & Baum, 1990; Thomas-Stonell, 1989), or provided objective evaluation of the validity of F_0 extraction methods used in sustained vowel analyses (Bielamowicz, Kreiman, Gerratt, Dauer, & Berke, 1996; Karnell, Hall, & Landahl, 1995; Parsa & Jamieson, 1999; Titze & Liang, 1993), relatively few have attempted to assess the ability to extract F_0 from continuous speech samples. It is essential that F_0 extraction algorithms be robust to factors such as amplitude and frequency modulations and the presence of noise if they are to have valid and reliable clinical usage. This robustness is a requirement for many measures of perturbation that may be extracted from a continuous speech sample (Parsa & Jamieson, 2001), but also for the measurements of mean F_0 and F_0 standard deviation. Studies of note which have evaluated the effectiveness of speech/voice analysis programs to extract acoustic measures such as speaking F_0 are by Read, Buder, and Kent (1992), Morris and Brown (1996), and Parsa and Jamieson (2001). Read et al. (1992) evaluated the ability of eight speech/voice analysis programs to derive various acoustic characteristics of the speech signal. For F_0 analysis, recordings of a single male speaker producing the word "yes" at various intonation patterns were analyzed. Parameters for the various programs were adjusted according the instructions provided in their respective manuals – the authors also experimented with "best" settings wherever necessary. Read et al.'s results indicated that, even under the best of circumstances, "the most troublesome application for the systems reviewed ... was that of fundamental frequency analysis" (p. 328). Read et al. felt that failure of F_0 extraction algorithms was "likely" in the cases of pathological voices, or the voices of women and children, though particular reasons for this "likely" failure were not provided. In their reviews of the various programs, Read et al. indicated that F_0 extraction errors could often be attributed to factors such as trouble analyzing unvoiced sections of a signal and changes in amplitude of the signal throughout the sample.

Morris and Brown (1996) evaluated aspects of reliability and agreement among six speech analysis systems in determining vocal F_0. Sustained vowels and reading samples from five male and five female adults were analyzed using the documented directions for F_0 extraction as detailed in each of the program's respective manuals. Computer analyzed results from sustained vowels were also compared to manual estimations of F_0 from sonograms – no manual estimations were made for the spontaneous speech samples. Results indicated excellent within system (intraprogram) reliability, but considerable variation in agreement across analysis systems (interprogram). These authors did not provide speculations regarding the reasons for the observed variability among systems. Agreement among systems was high for male sustained vowels and female reading samples, but poor agreement was observed for oral reading samples of males and for sustained vowels produced by females. According to these authors, F_0 extraction errors ("spurious output") may occur, but tend to be rare. Due to variability in speech analysis system agreement, Morris and Brown concluded that caution must be exercised when comparing F_0 data reported from different software systems.

Parsa and Jamieson (2001) emphasized the importance of the accuracy of F_0 extraction algorithms in a study which compared the ability of acoustic measures derived from sustained vowels versus continuous speech to discriminate between normal and pathological speakers. Though the focus of the Parsa and Jamieson study was not solely on the ability to extract speaking F_0 measures, these authors did observe that acoustic measures which were dependent upon the accuracy of the F_0 extraction algorithm were not as effective in classifying normal versus disordered speech samples as those which did not require precise measurements of the F_0 contour. Since F_0 extraction algorithms must have the capability to accurately identify voiced sections of a speech sample, any errors in this task will result in erroneous estimates of F_0. F_0 errors may be attributed to poor voice/unvoiced/silence detection and, particularly for disordered voice samples, the presence of noise components in the voice signal itself.

As observed in the Parsa and Jamieson (2001) study, analysis of the time-based acoustic waveform and identification of the cycle boundaries necessary for F_0 extraction would be expected to be a difficult proposition in disordered voice samples due to the influence of characteristics such as the presence of additive noise (ex. breathiness), periods of irregular vocal fold vibration (ex. roughness), some combination of the aforementioned characteristics (hoarseness), and rapid fluctuations in vocal frequency and intensity. However, the aforementioned studies by Read et al. (1992) and Morris and Brown (1996) appear to indicate that the analysis of continuous speech samples may present difficulties for computer methods of F_0 extraction *even in normal speech/voice cases.* Variation in the ability of F_0 extraction algorithms to effectively derive acoustic characteristics such as mean speaking F_0 and F_0 standard deviation may affect the degree of equivalence between available software packages, resulting in a lack of consistency in normative F_0 expectations. It is important that users of various computer programs for the measurement of speaking F_0 and speaking F_0 variability know whether normative expectations derived from one computer program are applicable to data collected via alternative programs. In addition, equivalence among programs is necessary if speaking F_0 data are to be shared across laboratories and/or clinics (Morris & Brown, 1996). With these issues in mind, the present study provides data regarding the equivalence of speaking F_0 data from three computer programs: CSpeechSP, CSL 4300, and Dr. Speech. In addition, this study provides information relevant to the operation of the various F_0 extraction algorithms used in the aforementioned programs. Finally, this study provides information regarding the possible sources of F_0 extraction differences that may occur among available computerized methods.

Method

Participants

Subjects were 10 adult males and 10 adult females ranging in age from 18-30 years (mean age in years = 24.1 vs. 23.7, respectively). In addition, 10 children (5 males and 5 females) ranging in age from 5-9 years (mean age in years = 7.8) were also included (Total N = 30). This subject group was felt to provide a relatively wide and representative range of normal voice types in terms of both age and gender. All subjects were native of northeastern Pennsylvania, USA, and were native speakers of English. All subjects were nonsmokers and were judged by a trained speech pathologist experienced in the assessment of speech and voice disorders to have voice characteristics within normal expectations for pitch, loudness, and quality. In addition, all subjects were required to pass a hearing screening (25 dB HL at .5, 1, 2, and 4 kHz). The use of human subjects for this research study was approved by the Institutional Review Board of the host institution.

Speech Sample

Each subject was instructed to read the "Rainbow Passage" (Fairbanks, 1960) at a comfortable pitch and loudness level. The use of the second sentence[1] of the "Rainbow Passage" has been commonly used for analysis of continuous reading samples for the following reasons:

1. An unpublished study reported in Shipp (1967) and research by Horii (1975) indicate that mean F_0 calculated from the second sentence correlates well (r's= 0.99 and 0.985, respectively) with measured F_0 from the entire paragraph.
2. By using an embedded sentence, we hopefully retain the naturalness of the patient's speaking style but avoid possible initial or final sentence effects (Horii, 1975).
3. The second sentence is linguistically "simpler" than other sentences in the passage and is easily read or repeated by most subjects.

Samples were recorded using a high-quality microphone (Shure SM10A) and preamplifier (Audio Technologies Model M200) and digitized directly to hard disk (.wav format) at 44.1 kHz with 16-bit resolution using a soundcard (SoundBlaster Live 5.1, Creative Labs, Milipitas, CA) and WaveLab v.1.6 computer software (Steinberg Media Technologies AG, Hamburg, Germany). The second sentence was edited and saved to disk for later analysis.

Programs

Three commercially available software programs were used to analyze the speech/voice samples for measures of mean F_0, F_0 standard deviation and range (maximum and minimum speaking F_0's). These analysis programs included *Computerized Speech Lab (CSL) - Base Module Model 4300* (Kay Elemetrics, Pine Brook, NJ); *CSpeechSP* (P. Milenkovic, Madison, WI); and *Dr. Speech v.3.0 - Speech Analysis Module*

[1] "The rainbow is a division of white light into many beautiful colors" (Fairbanks, 1960)

(Tiger DRS Inc., Seattle, WA). All of the programs provide for automatic F_0 extraction (i.e., they do not require an estimate of the expected F_0 to be manually input by the user). Two of the three programs used in this study (CSpeechSP and Dr. Speech) are designed to analyze .wav format signals. Prior to analysis using the CSL program, a utility provided within the CSL program was used to import .wav format signals and convert them to .nsp format. The utility program requires the sampling rate (44.1 kHz) and header size (44 bytes for .wav format files) for the digitized signal to be imported. In addition, the user manual for the CSL program stresses that the polarity of the signal be adjusted such that the significant impulse in each cycle be in a positive direction (i.e., a "positive" peak). Since the CSL program would appear to be searching for "positive" peaks, "negative" going impulses may result in poor F_0 extraction. Therefore, all acoustic waveforms were adjusted (i.e., polarity reversed if necessary) for the presence of positive peaks.

The three speech/voice analysis software systems were configured according to the instructions for continuous speech F_0 extraction provided in their respective manuals. In the event that instructions for parameter settings for various voice types were unavailable, modification of key parameters was performed to reduce the presence of obvious errors in F_0 extraction. It was not our purpose to make *a priori* decisions for the various parameters of these programs which would result in the "best" analysis. In fact, specific parameter settings for "best" analysis would have been almost impossible to ascertain due to (a) the multiple possible combinations of parameter settings and (b) the fact that "best" settings for one voice sample would most probably be quite different for another. Instead, it was our goal to use each of the specified programs in a manner which would provide a reasonable degree of accuracy and efficiency in analysis for each voice type analyzed (adult male, adult female, and child voices) and still be useable for clinicians who may have limited experience in the use of the programs and their various parameter settings as described in this study.

Algorithm Types

F_0 extraction algorithms generally fall into two categories: event-detectors or short-time average methods (Titze & Liang, 1993). Event detectors commonly utilize either peak-picking or zero-crossing algorithms. In contrast, short-time average methods typically window the signal and estimate an average F_0 for the cycle(s) contained within the analysis window via techniques such as autocorrelation or cepstral analysis. Review of the documentation for the programs used in this study, combined with discussions with the developers or representatives for each of the programs, provided further insight into the algorithms utilized, a summary of which is provided below:

CSpeechSP: The PITCH subprogram (cpitch.exe) was used for the analysis of the speech samples. The algorithm for the PITCH subprogram first downsamples the digitized signal, applies an analysis window, and center-clips the signal. Downsampling is performed to reduce the number of data points in the analysis, and thereby increase the efficiency of the F_0 extraction algorithm. A form of center clipping is applied in which the positive and negative peaks within each cycle are retained and the digitized speech signal between predetermined positive and negative amplitude levels is zeroed. The purpose of center clipping is to reduce formant information in the windowed speech segment and retain information which primarily relates to periodicity (Papamichalis, 1987). Following these procedures, autocorrelation of the windowed segment is performed.

As described in Papamichalis (1987, p. 163), if *s(n)* is a windowed segment of the speech signal starting at $n = 0$ and having a length of N samples, then the autocorrelation function is defined by the following equation:

$$R(k) = \sum_{n=0}^{N-1} s(n)\, s(n+k) \qquad k = 0,1,2,...$$

Variable k is the time index, also referred to as the *lag*. In voiced speech segments, a relatively sharp peak will occur in the autocorrelation function at the lag which corresponds to the average fundamental period of the windowed speech segment. According to Milenkovic[2], a modified autocorrelation function is computed by correlating each positive peak with each successive positive peak within the analysis window. A similar process is carried out for the negative peaks. Finally, all of the component correlations are summed. This process is repeated for each windowed segment of the speech sample under analysis.

The CSpeechSP PITCH subroutine provides a number of parameters that may be adjusted by the user to optimize F_0 extraction. Unfortunately, the CSpeechSP manual does not provide details of optimal settings for various voice types. Experimentation with the program parameters indicated that particular adjustments resulted in reasonable F_0 contours with a reduced number of obvious F_0 extraction errors for most speech samples. For the purposes of this study, the following parameters were adjusted:

Frequency Range: 70-250 Hz (Males); 150-350 Hz (Females); 150-450 Hz (Children)

Window Length: 30 ms (Males); 15 ms (Females); 15 ms (Children)

Update Interval: 10 ms (Default value used for all subjects)

Downsampling Factor: 3 (Used for all subjects)

According to the program developer, no interframe smoothing is applied prior to graphical display of the F_0

[2]Personal communication, July 9, 2002

contour and derivation of statistical results[3]. In addition, the reported F_0 values are restricted to the range designated by the user – all other results are zeroed. The polarity of the signal should not affect the results of the autocorrelation technique used in the CSpeechSP PITCH algorithm.

CSL Model 4300: The PITCH EXTRACTION algorithm used in the CSL program is actually a hybrid of short-time averaging and event detection methods. As previously described, autocorrelation is applied to windowed segments of the speech signal. However, the results of autocorrelation are next used to locate the sample closest to zero which immediately precedes the positive peak ("most significant impulse") of each cycle. The differences between these approximate "zero-crossing" points are converted to F_0 values for each identified cycle. Because there may be more than one cycle identified for each windowed segment, the CSL pitch tracking algorithm selects "the value that best correlates with the result computed in both the preceding and the following voiced frame" (Kay Elemetrics Corp., 1991, p. 333). The CSL manual (Kay Elemetrics Corp.) indicates that "No correction is attempted through averaging or interpolating adjacent frame values" (p. 333).

The CSL Model 4300 PITCH subroutine provides a number of parameters that may be adjusted by the user to optimize F_0 extraction based on expected fundamental frequencies. The following parameters were adjusted for the various voice types analyzed in this study according to suggested settings described in the CSL manual:

Frame Length: 20 ms (Males); 15 ms (Females); 10 ms (Children)

Frame Advance: 15 ms (Males); 10 ms (Females); 5 ms (Children)

Analysis Range: 60-250 Hz (Males); 70-350 Hz (Females); 150-450 Hz (Children)

In addition to the aforementioned parameter settings, the CSL Model 4300 manual stresses that the polarity of the waveforms analyzed during F_0 extraction may have a substantial effect. Therefore, the polarity of the waveform was set such that the most significant impulse of each cycle was positive going (i.e., a "positive peak"). The CSL program provides a utility that can be applied to the waveform under analysis to reverse polarity if necessary.

Dr. Speech 3.0 – Speech Analysis: The PITCH EXTRACTION algorithm used in the Dr. Speech 3.0 program is also a hybrid of short-time averaging and event detection methods. Though specific information regarding the F_0 extraction algorithm is absent from the Dr. Speech 3.0 manual, discussion with the program developer[4] revealed the following information. The speech signal is first downsampled to 11 kHz, low pass filtered, and then transformed using three-level clipping (a special form of center clipping). The sample is next segmented using a fixed 24 ms window. Autocorrelation of the windowed and clipped signal is then performed. The results of autocorrelation provide estimates by which approximate zero-crossing points for each cycle are identified. Finally, a bilinear interpolation technique (Huang, Minifie, Kasuya, & Lin, 1995) is used to identify the "true" zero-crossing point identifying the cycle boundary for each cycle. Prior to providing the graphical display of the F_0 contour and statistical results, the F_0 estimates are median smoothed as a means of accounting for gross voiced/unvoiced errors or other F_0 estimation errors (Papamichalis, 1987).

Unlike the CSL and CSpeechSP programs, most of the necessary parameters for F_0 extraction have already been "set" within the coded F_0 extraction algorithm of the Dr. Speech 3.0 program. Therefore, a smaller number of parameters were adjusted for the various voices to be analyzed:

Low Limit (Hz): 70 Hz (Males); 150 Hz (Females); 150 Hz (Children)

High Limit (Hz): 250 Hz (Males); 350 Hz (Females); 450 Hz (Children)

Reliability within each computer program was not assessed since computer algorithms analyzing the same digitized samples using identical parameter settings will always result in similar analysis results.

Results

A series of 1-Within (3 Levels of Program Type) ANOVAs were computed for each of the male, female, and child groups to analyze interprogram differences in computing mean F_0, F_0 standard deviation, maximum speaking F_0, and minimum speaking F_0. *A priori* statistical probability levels were set at $p < .01$ to avoid significant findings with only marginal F_0 differences. In the event of significant ANOVA findings, post-hoc means comparisons were conducted using the Bonferroni *t* procedure for multiple comparisons. Table 1 provides various mean F_0 results for the adult male, adult female, and child groups.

Mean Speaking F_0

Comparisons between the various computer programs on measures of mean speaking F_0 were found to be significantly different for the adult male group, $F\,[2, 18] = 7.08, p < .01$. Post-hoc analysis using Bonferroni *t* indicated that CSpeechSP produced a significantly lower estimate of mean speaking F_0 than the CSL or Dr. Speech programs (see Table 1). Comparisons of mean speaking F_0 within the adult female group and the child group were nonsignificant. It should be noted that all of the programs produced estimates of mean F_0 within 5% of each other.

A series of Pearson's *r* correlations were computed to investigate the presence and strength of relationships among the various programs for the estimation of mean speaking F_0 within each of the subject groups. When correlating two (or more) different forms of a measure of the same attribute,

[3]P. Milenkovic, personal communication, June 26, 2002

[4]D. Huang, personal communication, June 30, 2002

Table 1

Mean speaking F_0, F_0 standard deviation, and maximum and minimum F_0's computed for each of the adult male (M), adult female (F) and child (C) groups using various computer software programs. Group standard deviations are provided in brackets.

		CSL	CSpeechSP	Dr. Speech
Mean Speaking F_0	M	115.64 (11.80)	111.14 (14.24)	115.28 (13.59)
	F	218.05 (17.23)	219.62 (17.38)	219.81 (18.23)
	C	260.89 (16.70)	257.99 (16.68)	260.73 (17.01)
F_0 Standard Deviation	M	19.26 (7.27)	14.38 (5.69)	15.71 (6.05)
	F	34.03 (3.78)	33.28 (6.10)	29.22 (4.14)
	C	33.94 (8.57)	34.26 (9.49)	32.68 (9.07)
Maximum Speaking F_0	M	202.10 (36.07)	162.38 (33.98)	201.88 (51.92)
	F	309.00 (27.83)	332.82 (28.83)	307.76 (37.17)
	C	422.70 (29.40)	372.39 (52.69)	383.23 (56.97)
Minimum Speaking F_0	M	77.00 (14.63)	73.19 (5.92)	90.96 (9.69)
	F	102.10 (18.27)	159.17 (9.03)	148.05 (34.82)
	C	157.40 (8.67)	160.75 (16.01)	158.91 (45.66)

the resulting correlation coefficient has been referred to as a "coefficient of equivalence" (Schiavetti & Metz, 2002, p. 117). Results are presented in Table 2. All correlations of equivalence were significant ($p < .001$, ranging from 0.93 to 0.99), indicating strong associations among these programs.

F_0 Standard Deviation

None of the computer programs provided significantly different estimates of F_0 standard deviation for the adult male voices or child voices. Within the adult female group, a significant difference was observed, $F[2, 18] = 7.32, p < .01$. Post-hoc analysis indicated that the CSL and CSpeechSP programs produced significantly higher estimates of F_0 standard deviation than did the Dr. Speech program (see Table 1).

A series of Pearson's *r* correlations was computed among the various programs for the estimation of F_0 standard deviation for each of the subject groups. Results are presented in Table 3. Significant correlations ($p < .01$) ranging from 0.78 to 0.88 were observed among the various programs within the child group. Within the adult female group, a significant correlation ($p < .01$) was observed between the Dr. Speech and CSL programs

Table 2

Pearson *r* matrix showing bivariate correlations between the various methods of computing mean speaking F_0.

		CSL	CSpeechSP	Dr. Speech
Males	CSL	1.00	0.97	0.93
	CSpeechSP		1.00	0.97
	Dr. Speech			1.00
Females	CSL	1.00	0.97	0.99
	CSpeechSP		1.00	0.99
	Dr. Speech			1.00
Children	CSL	1.00	0.99	0.98
	CSpeechSP		1.00	0.98
	Dr. Speech			1.00

All correlations are significant at $p < .001$

Table 3

Pearson *r* matrix showing bivariate correlations between the various methods of computing speaking F_0 standard deviation.

		CSL	CSpeechSP	Dr. Speech
Males	CSL	1.00	-0.18	-0.42
	CSpeechSP		1.00	0.60
	Dr. Speech			1.00
Females	CSL	1.00	0.41	0.80**
	CSpeechSP		1.00	0.76
	Dr. Speech			1.00
Children	CSL	1.00	0.83**	0.88**
	CSpeechSP		1.00	0.78**
	Dr. Speech			1.00

** significant at $p < .01$

(r = 0.80). None of the interprogram correlations were significant for the analysis of F_0 standard deviation within the adult male group.

Maximum and Minimum Speaking F_0's

It has been our experience that obvious, gross F_0 extraction errors are often associated with the range of F_0 estimates produced by a particular program. Therefore, evaluations of significant differences in estimating maximum and minimum speaking F_0's were carried out. None of the computer programs provided significantly different estimates of maximum speaking F_0 when analyzing adult male, adult female, or child voices.

In terms of minimum speaking F_0's in the male subjects, post-hoc analysis of a significant ANOVA ($F[2, 18] = 9.51$, $p < .01$) indicated that Dr. Speech produced a significantly higher estimate of minimum speaking F_0 than CSpeechSP or CSL (see Table 1). Within the adult female group, post-hoc analysis of a significant ANOVA ($F\,[2, 18] = 18.90, p < .001$) indicated that CSL produced a significantly lower estimate of minimum speaking F_0 than CSpeechSP or Dr. Speech (see Table 1).

Discussion

A primary goal of the present study was to assess the equivalence of speaking F_0 data (mean speaking F_0, F_0 standard deviation, and measures of maximum and minimum speaking F_0's) from three computer programs (CSpeechSP, CSL 4300, and Dr. Speech). The results of this study indicate a high degree of correspondence in estimates of mean speaking F_0 among the three programs tested, with all of the programs observed to produce mean F_0 estimates within 5% of each other, regardless of gender or age of the subject group producing the samples. These results are comparable to those reported by Bielamowicz et al. (1996) for F_0 estimations obtained from sustained vowels. In addition, interprogram correlations for mean F_0 estimation were strong (r's > .90), indicating that measures of speaking F_0 produced by one program may be used to estimate values from another with a high degree of predictive accuracy. Program equivalence for measures of F_0 variability (standard deviation, maximum speaking F_0 and minimum speaking F_0) appear to be weaker, with significant differences in estimating F_0 standard deviations observed when analyzing adult female voices, and significant differences observed in estimating minimum speaking F_0 in both male and female adult voices. Correlations of equivalence among the programs for measures of F_0 standard deviation were weaker in nature than those observed for estimates of mean speaking F_0 and were nonsignificant for the analysis of adult male voices.

It may be unreasonable to expect that different programs and their respective algorithms will produce identical results in extracting mean speaking F_0. Variations in F_0 extraction algorithms such as the use of zero crossing versus peaks to guide the search for cycle boundaries, the use of F_0 estimates from analysis frames/windows vs. F_0 estimates from individual cycles, and the incorporation of F_0 smoothing algorithms (such as median smoothing, incorporated in the Dr. Speech program) will tend to produce minor variations in estimates of mean speaking F_0's between programs. However, the results of this study indicate that interprogram variations across gender and age may be expected to be well within 5% of each other. The only statistically significant difference observed in estimating mean speaking F_0 was found in the analysis of adult male speakers, with values derived from CSpeechSP significantly lower than those from CSL or Dr. Speech. This result may indicate that some algorithms may have difficulty with voices of a particular gender. The finding of the present study is consistent with that of Morris and Brown (1996), who also observed poor agreement among programs for adult male oral reading samples. F_0 extraction differences in the analysis of the male voice may be related to the increased complexity of the male voice signal in comparison to the higher pitched voices of normal adult females and children. Because the lower F_0 voice signal of the male subject contains a greater number of harmonic frequencies than higher pitched voices within the theoretical frequency range to be considered (up to approx. 22 kHz when using a 44.1 kHz sampling rate), the complexity of the signal (i.e., the number of "peaks" and zero-crossings) will also be expected to increase. This increased complexity may create difficulties in accurately identifying cycle boundaries, with false "peak" values resulting in variations in final mean F_0 estimates.

In contrast to interprogram measures of mean speaking F_0, the results of this study indicate poorer prediction between programs in terms of measures of speaking F_0 standard deviation. Again, some of the variability between programs may be attributable to several of the basic algorithm differences previously mentioned. However, it is our view that gross F_0 extraction errors may be a primary source of interprogram differences in measures of speaking F_0 variability such as F_0 standard deviation and F_0 range. Gross F_0 extraction errors may be due to factors such as waveform characteristics and/or program parameter settings. As previously stated, the complexity of the lower F_0 male voice may be problematic for many F_0 extraction algorithms, resulting in variability in F_0 values and the lack of significant intercorrelations for analysis of F_0 standard deviations. Program parameter settings may have resulted in the poor correlations between CSL and the CSpeechSP and Dr. Speech programs for the analysis of F_0 standard deviation in adult female voices. It may be that the extremely low F_0 parameter setting for the range of F_0 analysis (70 Hz) suggested in the CSL manual provides for the possibility of uncorrected low F_0 extraction errors. This view is supported by the observation that CSL produced a significantly lower estimate of minimum speaking F_0 in adult female speech samples than either CSpeechSP or Dr. Speech (see Table 1). In general, intercorrelations for measures of F_0 standard deviation were stronger when there was more similarity between programs in parameter settings for F_0 analysis range.

The possibility of weak interprogram correlations for measures of F_0 standard deviation may have important clinical implications, particularly when F_0 analyses are used in diagnostic decision making by those clinicians who may be less experienced with the interpretation of graphical and statistical results from F_0 analysis and/or the inherent complexities of F_0 extraction algorithms. When estimates of F_0 variability are artificially inflated by gross F_0 extraction errors, results may be misinterpreted as reflecting characteristics such as vocal F_0 control difficulties (ex. pitch breaks) or instances of aperiodic vocal fold vibration. The following suggestions are made by which the occurrence and influence of these gross F_0 extraction errors may be minimized.

Appropriate Adjustment of Program Parameters: F_0 extraction algorithms are quite sensitive to the adjustment of their various parameters. It is essential that program developers provide detailed information on parameters and optimal settings for various possible voice types, as well as information regarding the advantages and disadvantages of the core algorithm(s) used (Read et al., 1992). The results of this study indicate that parameter settings that include extreme F_0 analysis ranges may have an increased tendency for gross F_0 extraction errors which can inflate estimates of F_0 standard deviation. Therefore, programs requiring input regarding the frequency range to be analyzed should be restricted as much as possible to the expected F_0 range of the sample being analyzed. In addition, the use of a large frequency range in parameter settings may also result in algorithms inadvertently selecting the first formant of vowel productions (particularly in the case of high-front vowel productions) as the vocal F_0. Again, the result may be an error in estimating F_0. The possible effect of vowel type on F_0 extraction was not a specific focus of this study, and is a parameter which, generally, would not be controlled when analyzing samples of continuous speech. However, future studies that may attempt to extract measures of perturbation from continuous speech samples will need to carefully examine the effect of vowel type on F_0 extraction and subsequent perturbation measures (such as jitter) which are derived from F_0 extraction algorithms. It has been well documented that vowel type may have a significant effect on measures such as jitter and harmonic-to-noise ratio obtained from sustained vowel productions (Deem, Manning, Knack, & Matesich, 1989; Gelfer, 1995; Sussman & Sapienza, 1994).

When available as parameter settings, both frame length and frame advance should be restricted. Frame length should be approximately two times the period of the lowest frequency of interest, while small values for frame advance provides analysis overlap which may be advantageous in analyzing the highly variable F_0's of continuous speech samples (Kay Elemetrics Corp., 1991, p. 288). Frame lengths smaller than two times the lowest frequency of interest may produce F_0 estimates higher than expected values. In contrast, period doubling (i.e., an estimated period approximately two times as long as it actually is, resulting in F_0 estimates approximately half of the "true" F_0 value) can be a strong likelihood when using relatively long analysis windows, particularly when analyzing higher pitched voices. As an example, the possibility that the peak in the autocorrelation function will coincide with a cycle other than the immediately following cycle is increased if more than two cycles of vibration are contained within an analysis window. It should also be noted that period doubling may occur in voices which have some degree of amplitude modulation (as in a subharmonic) which causes the peak amplitude of successive cycles to vary. This problem will also result in low F_0 estimates in event detectors if the amplitude modulation causes a peak value to fall below the clip level for an analysis frame. Sudden shifts to lower frequencies in the graphical F_0 contour may be indicative of period doubling.

Unfortunately, restricting program parameter settings for a particular voice sample depends upon a prior knowledge of the F_0 characteristics of the sample. In addition, while restricting parameter settings may appear to be advantageous, continuous speech samples require parameter settings that can account for the relatively large degrees of F_0 variation that occur in normal intonation patterns (i.e., large F_0 analysis ranges and large analysis windows). A possible alternative is to divide the speech sample into smaller portions that present less variation to the F_0 extraction algorithm and thus allow for more restricted parameter settings. The F_0 results for each portion could then be summed and averaged using a weighted averaging method in which longer duration samples are weighted more heavily in the computation (Awan, 2001; Awan & Mueller, 1992).

Signal Review and Manual Correction: Regardless of the program or method used for F_0 extraction, it is essential that the graphical and numerical results be interpreted by the user. The presence of sudden maxima or minima in the graphical F_0 contour or unusually consistent range values (i.e., extremely similar or identical maximum and/or minimum F_0 estimates for different subjects) may be signs of gross F_0 extraction errors. These possible errors should be analyzed by focusing ("zooming") in on the questionable section of the speech signal and observing for evidence of excessive variability. All of the programs used in this study (CSpeechSP, CSL 4300, and Dr. Speech) allow the user to "zoom" in on a particular section and play back for review. In the event that probable F_0 extraction errors are observed, most programs provide some method by which these errors may be manually corrected or removed from analysis (the Dr. Speech program is the only program used in the current study that does not provide this option). CSpeechSP includes a subprogram (EDIT- F_0) in which the user can manually use cursors to measure cycle periods and insert these measurements into the F_0 contour. CSL requires the user to apply voice impulse markers to the segment under analysis which may then be manually adjusted.

While some may feel that manual correction introduces experimenter bias into the analysis, a representative of Kay Elemetrics[5] and the developer of CSpeechSP[6] both stressed that most automatic F_0 extraction algorithms should be considered a first step in F_0 analysis, followed by hand-marking/manual correction as the final decision. Hand-marking has been considered a valid method of assessing the accuracy of F_0 extraction algorithms in studies by Rabinov, Kreiman, Gerratt, and Bielamowicz (1995), Bielamowicz et al. (1996) and Parsa and Jamieson (1999). It should be noted that, in some cases, zeroing or discarding erroneous data (Morris & Brown, 1996) may be more advisable than correction. Since autocorrelation methods are based on the average period of an analysis frame, replacement with the F_0 from a single cycle may not be a valid replacement.

In the present study, we have attempted to assess the equivalence of three speech analysis programs and their respective F_0 extraction algorithms. However, several limitations of the current study should be noted. First, only certain parameters that were felt to have a major effect on F_0 extraction of the digitized speech samples used in this study were adjusted. Various other parameters (clip level, silence threshold, LP filtering, etc.) also could have been manipulated and may have resulted in improved analyses for some of the programs examined. Future studies that systematically vary the numerous parameters of F_0 extraction algorithms are needed to provide optimal suggestions to the user applying these programs to various normal and disordered speech samples. A second limitation is that several of these programs have newer versions available which may incorporate improvements in F_0 extraction. Future studies should continue to evaluate the validity of F_0 extraction algorithms incorporated in commercially available programs. Finally, the current study focused on the ability to extract F_0 from a variety of normal speech samples. It is clear that the ability of these various programs and algorithms to accurately estimate characteristics such as mean F_0 and F_0 standard deviation may be severely limited in the presence of the frequency, intensity, and voicing variations present in dysphonic speech. Future validity studies need to be extended to dysphonic speech samples to investigate and describe F_0 extraction difficulties that may arise, particularly in the clinical situation.

Clinical Implications and Conclusion

The results of this study have a number of important clinical implications, particularly for those clinicians with lesser experience and familiarity with the intricacies of F_0 extraction programs:

1. The results of this study indicate that clinicians may be fairly confident that measures of mean F_0 computed with a particular speech/voice analysis program will be readily comparable to mean F_0 values obtained with a different program. Any differences between programs in mean F_0 computation may be expected to be within 5% and clinically insignificant in most cases. In addition, strong correlations of equivalence indicate that the mean speaking F_0 estimates derived from one of these programs may be reliably estimated from any of the other programs tested.
2. Because speech/voice analysis programs and their respective algorithms may not correlate well for computations of F_0 standard deviation (particularly for the analysis of male voices), clinicians should be aware that clinical comparison to normative data collected using a program/algorithm different from that used by the clinician in his/her initial assessment may result in erroneous clinical decision-making. As an example, a highly variable speaking F_0 (and, therefore, increased speaking F_0 standard deviation) may be considered to be abnormal and possibly indicative of disorder. However, the result may be due to F_0 extraction differences or errors which may have inflated the computed F_0 variability as compared to another program.
3. Clinicians should be aware that errors in F_0 extraction may give the faulty impression of a possibly abnormal voice characteristic. As an example, a visual display of a patient's speaking F_0 contour which contains F_0 extraction errors may be erroneously interpreted as including pitch breaks (errors upwards in F_0), fry phonation (errors downwards in F_0) or other F_0 control problems not actually present in the perceived voice sample.
4. The clinician must be aware that certain programs may be prone to gross F_0 extraction errors that influence measures of F_0 variability (F_0 standard deviation and range) when their respective parameter settings are not adequately adjusted to suit the F_0 characteristics of the voice or voices to be analyzed.
5. Though not a focus of this paper, F_0 extraction errors may influence vocal quality measures derived from continuous speech samples. As noted by Parsa and Jamieson (2001), the noise inherent to many pathological voices may make the voiced/unvoiced segments of speech difficult to identify, thereby resulting in F_0 extraction difficulties and errors. Since measurable correlates of vocal quality such as jitter and harmonic-to-noise ratio are based upon the accuracy of F_0 extraction results, errors in F_0 extraction may render these measures invalid and unreliable. This problem may be avoided by using spectral-based measures such as long-term average spectrum (LTAS) which work on the entire speech sample and do not require the demarcation of cycle boundaries.

Overall, the results of this study are in agreement with Morris and Brown (1996), who stated that caution should be exercised in comparing F_0 data (particularly measures of F_0 variability) reported from different software systems. It is also suggested that intrasubject changes in F_0 variability should be assessed using the same computer program and

[5] B. Kiely, personal communication, June 26, 2002
[6] P. Milenkovic, personal communication, June 26, 2002

same parameter settings as was used in previous analyses. While computer-based measures of the voice signal will continue to be common and useful clinical tools, the results of this study indicate that careful interpretation of results and knowledge of the methods used to obtain those results will play an integral role in the appropriate application of these measures.

Acknowledgments

The authors of this paper gratefully acknowledge Dr. P. Milenkovic (Dept. of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI), Dr. D. Huang (Tiger DRS Inc., Seattle, WA), and Mr. B. Kiely (Kay Elemetrics, Pinebrook, NJ) for their informative comments and suggestions on various aspects of this paper.

References

Aronson, A. E. (1990). *Clinical voice disorders: An interdisciplinary approach* (3rd ed.). New York, NY: Thieme Medical Publishers.

Askenfelt, A., & Hammarberg, B. (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research, 29,* 50–64.

Awan, S. N. (2001). *The Voice Diagnostic Protocol: A practical guide to the diagnosis of voice disorders.* Austin, TX: Pro-Ed.

Awan, S. N., & Mueller, P. B. (1992). Speaking fundamental frequency characteristics of centenarian females. *Clinical Linguistics and Phonetics,* 6(3), 249–254.

Awan, S. N., & Mueller, P. B. (1996). Speaking fundamental frequency characteristics of White, African American and Hispanic kindergartners. *Journal of Speech and Hearing Research, 39,* 573–577.

Baken, R. J. (1987). *Clinical measurement of speech and voice.* Boston, MA: Little, Brown and Company.

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., & Berke, G. S. (1996). Comparison of voice analysis systems for perturbation measurement. *Journal of Speech and Hearing Research, 39,* 126–134.

Boone, D., & McFarlane, S. (1988). *The voice and voice therapy* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Callan, D. E., Kent, R. D., Roy, N., & Tasko, S. M. (1999). Self-organizing maps for the classification of normal and disordered female voices. *Journal of Speech and Hearing Research, 42,* 355–366.

Canter, G. (1963). Speech characteristics of patients with Parkinson's disease. I. Intensity, pitch and duration. *Journal of Speech and Hearing Disorders, 28,* 221–229.

Countryman, S., Hicks, J., Ramig, L. O., & Smith, M. E. (1997). Supraglottal hyperadduction in an individual with Parkinson disease: A clinical treatment note. *American Journal of Speech-Language Pathology,* 6(6), 74–84.

de Krom, G. (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *Journal of Speech and Hearing Research,* 37, 985–1000.

de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *Journal of Speech and Hearing Research,* 38, 794–811.

Deem, J. F., Manning, W. H., Knack, J. V, & Matesich, J. S. (1989). The automatic extraction of pitch perturbation using microcomputers: Some methodological considerations. *Journal of Speech and Hearing Research,* 32, 689–697.

Dromey, C., Ramig, L. O., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. *Journal of Speech and Hearing Research, 38,* 751–764.

Fairbanks, G. (1960). *Voice and articulation drill book* (2nd ed.). New York, NY: Harper Collins.

Gelfer, M. P. (1995). Fundamental frequency, intensity, and vowel selection: Effects on measures of phonatory stability. *Journal of Speech and Hearing Research, 38,* 1189–1198.

Gilbert, H. R., & Weismer, G. G. (1974). The effects of smoking on the fundamental frequency of women. *Journal of Linguistic Research,* 3(3), 225–231.

Hartelius, L., Buder, E. H., & Strand, E. A. (1997). Long-term phonatory instability in individuals with multiple sclerosis. *Journal of Speech, Language, and Hearing Research, 40,* 1056–1072.

Hirano, M., Tanaka, S., Fujita, M., & Terasawa, R. (1991). Fundamental frequency and sound pressure level of phonation in pathological states. *Journal of Voice,* 5(2), 120–127.

Honjo, I., & Isshiki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Archives in Otolaryngology, 106,* 149–150.

Horii, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research, 18,* 192–201.

Horii, Y. (1979). Fundamental frequency perturbation observed in sustained phonation. *Journal of Speech and Hearing Research, 22,* 5–19.

Horii, Y. (1982). Some voice fundamental frequency characteristics of oral reading and spontaneous speech by hard of hearing young women. *Journal of Speech and Hearing Research, 25,* 608–610.

Huang, D. Z., Minifie, F. D., Kasuya, H., & Lin, S. X. (1995). Measures of vocal function during changes in vocal effort level. *Journal of Voice, 9(4),* 429–438.

Hudson, A., & Holbrook, A. (1982). Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading. *Journal of Speech and Hearing Research, 25,* 25–28.

Karnell, M. P., Hall, K. D., & Landahl, K. L. (1995). Comparison of fundamental frequency and perturbation measurements among three analysis systems. *Journal of Voice,* 9(4), 383–393.

Kay Elemetrics Corp. (1991) Computerized Speech Lab (CSL) – Base Module Model 4300 [Computer Software]. Seattle, WA: Author

Kay Elemetrics Corp. (1991). *Operations manual: CSL Computerized Speech Lab Model 4300 Software Version 3.x.* Pinebrook, NJ: Kay Elemetrics.

Leff, J., & Abberton, E. (1981). Voice pitch measurements in schizophrenia and depression. *Psychological Medicine, 11,* 849–852.

Liss, J., Weismer, G., & Rosenbek, J.C. (1990). Selected acoustic characteristics of speech production in very old males. *Journal of Gerontology: Psychological Sciences, 45,* 35–45.

Mann, V. (1987). Review of DSPS Realtime Signal Lab. *ASHA, 29,* 64-65.

Milenkovic, P. (1994) CSpeechSP [Computer Software]. Madison, WI: Author.

Monsen, R. B. (1979). Acoustic qualities of phonation in young hearing-impaired children. *Journal of Speech and Hearing Research, 22,* 270–288.

Morris, R. J., & Brown, W. S. (1994). Age-related differences in speech variability among women. *Journal of Communication Disorders, 27,* 49–64.

Morris, R. J., & Brown, W. S. (1996). Comparison of various automatic means for measuring fundamental frequency. *Journal of Voice, 10(2),* 159–165.

Morrison, M. D., & Rammage, L. A. (1993). Muscle misuse voice disorders: Description and classification. *Acta Otolaryngologica, 113,* 428–434.

Murry, T. (1982). Phonation: Assessment. In N. Lass, L. McReynolds, J. Northern, & D. Yoder (Eds.), *Speech, language, and hearing, Vol. II* (pp. 477–488). Philadelphia, PA: W. B. Saunders Company.

Papamichalis, P. E. (1987). *Practical approaches to speech coding.* Englewood Cliffs, NJ: Prentice-Hall.

Parsa, V., & Jamieson, D. G. (1999). A comparison of high precision F_0 extraction algorithms for sustained vowels. *Journal of Speech and Hearing Research, 42,* 112-126.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathologic voice: Sustained vowels versus continuous speech. *Journal of Speech and Hearing Research, 44,* 327-339.

Pedersen, M. F., Moller, S., Krabbe, S., & Bennett, P. (1986). Fundamental voice frequency measured by electroglottography during continuous speech: A new exact secondary sex characteristic in boys in puberty. *International Journal of Pediatric Otorhinolaryngology, 11,* 21–27.

Rabinov, C. R., Kreiman, J., Gerratt, B. R., & Bielamowicz, S. (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research, 38,* 26–32.

Read, C., Buder, E., & Kent, R. (1990). Speech analysis systems: A survey. *Journal of Speech and Hearing Research,* 33, 363–374.

Read, C., Buder, E., & Kent, R. (1992). Speech analysis systems: An evaluation. *Journal of Speech and Hearing Research,* 35, 314–332.

Ryalls, J., & Baum, S. (1990). Review of three software systems for speech analysis: CSPEECHSP, BLISS, and CSRE. *Journal of Speech-Language Pathology and Audiology, 14,* 49-52.

Schiavetti, N., & Metz, D. E. (2002). *Evaluating Research in Communication Disorders, 4th Edition.* Boston, MA: Allyn & Bacon.

Shipp, T. (1967). Frequency, duration, and perceptual measures in relation to judgments of alaryngeal speech acceptability. *Journal of Speech and Hearing Research, 10,* 417-427.

Stemple, J. C., Stanley, J., & Lee, L. (1995). Objective measures of voice production in normal subjects following prolonged voice use. *Journal of Voice,* 9(2), 127–133.

Sussman, J. E., & Sapienza, C. (1994). Articulatory, developmental, and gender effects on measures of fundamental frequency and jitter. *Journal of Voice,* 8(2), 145–156.

Thomas-Stonell, N. (1989). Review of Speech Viewer. *Journal of Speech-Language Pathology and Audiology, 13,* 59-60.

Tiger DRS Inc. (1995) Dr. Speech v.3.0 for Windows – Speech Analysis Module [Computer Software]. Seattle, WA: Author.

Titze, I., & Liang, H. (1993). Comparison of F_0 extraction methods for high-precision voice perturbation measurements. *Journal of Speech and Hearing Research, 36,* 1120–1133.

Wheat, M. C., & Hudson, A. I. (1988). Spontaneous speaking fundamental frequency of 6-year-old black children. *Journal of Speech and Hearing Research, 31,* 723–725.

Wolfe, V., & Steinfatt, T. M. (1987). Prediction of vocal severity within and across voice types. *Journal of Speech and Hearing Research, 30,* 230–240.

Zwirner, P., Murry, T., & Woodson, G. (1991). Phonatory function of neurologically impaired patients. *Journal of Communication Disorders, 24,* 287–300.

Manuscript received: *Novermber 10, 2003*
Accepted: *April 10, 2004*