Comparison of voice acquisition methodologies in speech research

ADAM P. VOGEL AND PAUL MARUFF
University of Melbourne, Victoria, Australia

The use of voice acoustic techniques has the potential to extend beyond work devoted purely to speech or vocal pathology. For this to occur, however, researchers and clinicians will require acquisition technologies that provide fast, accurate, and cost-effective methods for recording data. Therefore, the present study aimed to compare industry-standard techniques for acquiring high-quality acoustic signals (e.g., hard drive and solid-state recorder) with widely available and easy-to-use, computer-based (standard laptop) data-acquisition methods. Speech samples were simultaneously acquired from 15 healthy controls using all three methods and were analyzed using identical analysis techniques. Data from all three acquisition methods were directly compared using a variety of acoustic correlates. The results suggested that selected acoustic measures (e.g., f_0 , noise-to-harmonic ratio, number of pauses) were accurately obtained using all three methods; however, minimum recording standards were required for widely used measures of perturbation.

Acoustic analysis of vocal functioning is a necessary part of the clinical and empirical voice investigation (Dejonckere et al., 2001); therefore, a high quality of measurements is integral for capturing reliable and useable acoustic signals. Until recently, the majority of the voice acoustic analysis literature has focused on pathological voice conditions and motor–speech disorders; however, recent advancements have led to the exploration of variant populations with disorders involving central nervous system functioning, including depression (Mundt, Snyder, Cannizzaro, Chappie, & Geralts, 2007) and schizophrenia (Alpert, Rosenberg, Pouget, & Shaw, 2000). As the possibilities for the use of voice analysis extend beyond areas concerning purely speech or vocal pathology, the need for accessible, cost-effective acquisition methodologies has increased. Current practices in voice acoustic technology are hindered by a lack of consensus concerning minimal software and hardware requirements.

The accuracy of acoustic measures is limited by several factors, including the acquisition environment (Deliyski, Evans, & Shaw, 2005). A number of researchers have attempted to quantify the variables that are inherent in data acquisition and analysis. From the type of software systems used to analyze data (Bielamowicz, Kreiman, Gerratt, Dauer, & Berke, 1996; Bough, Heuer, Sataloff, Hills, & Cater, 1996; Carson, Ingrisano, & Eggleston, 2003; Deliyski, Shaw, & Evans, 2005b; Karnell, Hall, & Landahl, 1995; Karnell, Scherer, & Fischer, 1991), to analysis algorithms relating to the size of the analysis window (Fette, Gibson, & Greenwood, 1980; Gerhard, 2003; Hirose, Fujisaki, & Seto, 1992; Karnell, 1991; Rabiner, 1977; Roa, Bennewitz, & Behnke, 2007), the effect of additional

noise (Carson et al., 2003; Deliyski, Shaw, & Evans, 2005a; Deliyski et al., 2005b; Hillenbrand, 1987), data acquisition environments (Deliyski, Evans, & Shaw, 2005; Scherer, Vail, & Guo, 1995; Smits, Ceuppens, & De Bodt, 2005), sampling rates (Deliyski et al., 2005b; Parsa, Jamieson, & Pretty, 2001), the microphone selection (Parsa et al., 2001; Titze & Winholtz, 1993; Winholtz & Titze, 1997), the recording hardware used (Mundt et al., 2007), population characteristics (Lei, Yang, Shen, & Gong, 2000; Mueller, 1997; Scherer et al., 1995; Xue & Fucci, 2000), and the appropriateness of stimuli employed (Parsa & Jamieson, 2001), almost every aspect of voice acoustics has been investigated in an effort to determine the minimum requirements for accurate and repeatable recording and analytical configurations. This body of work has yielded a broad set of criteria for acoustic analysis. The degree to which aspects of the acquisition and analysis process need to be adjusted or scrutinized, however, has yet to be determined.

To begin, it may be possible for the researcher or clinician to record and analyze specific acoustic data using less stringent recording and analysis paradigms. For example, in studies focusing on broad timing-based measures, such as those looking at central nervous system disturbances (Mundt et al., 2007), subtle perturbation measures are not necessarily required. Similarly, investigations that incorporate fundamental and formant frequencies—such as investigations of vowel profiles (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000)—and that do not look at rates of variation within those measures may not require high-grade recording configurations, because previous studies have demonstrated the rigor of some frequency-based measures.

A. P. Vogel, avogel@cogstate.com

The literature has also identified a number of acoustic correlates that are more resistant to recording and analysis inconsistencies than are others. Specifically, fundamental frequency (f_0) has proven a robust acoustic measure in a variety of recording and analytical environments (Deliyski et al., 2005a). Similarly, some researchers have highlighted the opportunities that timing-based measures can offer acoustic studies (Green, Beukelman, & Ball, 2004). Green et al. evaluated the validity and performance of automatic analytical processes that were designed to detect speech timing information in connected speech samples and compared those results with manual investigation. The authors found that both the automatic analysis and the manual analysis yielded significantly comparable results, suggesting that timing measures offer another resilient acoustic measure in speech studies.

Not only do some acoustic measures offer wider application in speech studies, evidence provided by Winholtz and Titze (1998) has illustrated the potential applications of alternative acquisition methodologies in voice acoustics. Direct comparison of data acquisition by minidisc and by digital audiotape recordings on a variety of perturbation measures and using synthetic test signals revealed significant similarities in dynamic range, frequency response, and distortion (Winholtz & Titze, 1998). More recent work by Kent, Vorperian, and Duffy (1999) evaluated the reliability of the Multi-Dimensional Voice Program (MDVP) (KayPENTAX, Lincoln Park, NJ) in a variety of vocal measures collected from individuals with dysarthria. They concluded that the analysis program was generally consistent on a number of acoustic features. Similar comparative studies have highlighted a degree of reliable crossover between distinct analysis programs. Smits et al. (2005) compared Doctor Speech (Tiger Electronics, Neu-Anspach, Germany) and the Computerized Speech Laboratory (KayPENTAX) on 120 normal voices and found a significant correlation between some measures, including f_0 and harmonics-to-noise ratio.

Despite an abundance of literature acknowledging the robust nature of some stimuli, methodologies, and analysis configurations, a substantial portion of the research maintains that a high standard of recording and analysis is mandatory for accurate and reliable acoustic investigation. Deliyski and colleagues have conducted a series of comprehensive studies investigating the influence of data acquisition environments (Deliyski, Shaw, Evans, & Veselinov, 2006), sampling rate (Deliyski et al., 2005a), and environmental noise (Deliyski, Evans, & Shaw, 2005). The authors have concluded that gender, intrasubject variability, microphone selection, environmental noise (type and level), data acquisition system, and analysis software all play a role in voice quality measurements. Furthermore, Deliyski, Evans, and Shaw (2005) reported that sound cards do not perform as well as specialized professional grade recording systems do; however, sound cards have the potential for accurate voice assessment if they are modified to secure balanced input and effective dynamic range. Nevertheless, contrary evidence provided by Carson et al. (2003) suggested that sound cards performed as

well as professional grade acquisition systems did when the effects of noise in computer-derived samples were compared across three hardware and software configurations. One was a stand-alone A/D converter (CSL Module 4300B) with a PC and the MDVP software. The other two consisted of a Creative Labs A/D converter (Creative Technology Ltd, Singapore) with the same PC and one of two different software systems.

Corresponding research has outlined the sensitive nature of both frequency and intensity perturbation measures, citing a variety of causal factors, including analysis software (Bough et al., 1996; Karnell et al., 1995; Perry, Ingrisano, & Blair, 1996; Perry, Ingrisano, & Scott, 1996) and environmental noise across different hardware and software combinations (Carson et al., 2003; Deliyski et al., 2005a; Ingrisano, Perry, & Jepson, 1998; Perry, Ingrisano, Palmer, & McDonald, 2000).

The use of alternative or lower grade recording technologies is not typically welcomed in acoustic investigation. Configurations that include generic computer sound cards and universal serial bus (USB)-connected microphones pose a number of technical problems for acoustic analysis, such as additive noise. Some features of the recording process can be controlled, including recording software, analytical software, sampling rate, quantization level, and sound-isolation environments; however, a myriad of extraneous issues can arise during computer-based recording: Personal computers often lack adequate sound-level meters, which can lead to difficulty when one is flagging overloaded signal input; they have issues relating to impedance and the loss of signal strength through loading; they are limited by restrictions associated with the connector and wiring systems of the microphone; noise and interference produced inside the computer may result in a loss of dynamic range; computer systems often possess preamplifiers with a limited range; they can contain poorly constructed automatic gain control circuitry; and they often have poor separation of channels for input and output, which can prevent the isolation of input signals from noise. Moreover, analog input signals can be corrupted as they enter the computer system, and inaccurate polarity can affect the pitch-extraction process.

The advantages provided by USB-based microphones offset some of the shortcomings of computer-based sound recordings. With USB-based microphones, the sound-recording process completely bypasses the sound card device, converting analog signals to digital signals before they reach the computer. This is especially valuable when users do not know the quality of their sound card, or when the quality of the sound card is poor. For experimental designs that incorporate large clinical assessments over several sites and with large participant numbers, USB microphones offer a unique means of experimental standardization that eliminates the need for traditional acquisition systems or extensive user expertise. Current accepted acquisition methodology places high demands on recording environments, hardware, and software setups, but alternative computer-based methods require less rigid recording regimes, are more widely

available, and are cost effective. Recent advances in voice technology and computer design have increased accessibility and created more opportunities for research in the field, since even basic systems have the potential to provide adequate acoustic investigation.

In light of the clear need for further investigation into the limits and benefits of more accessible recording configurations, the present study compared existing industry-standard techniques for acquiring high-quality acoustic signals with widely available and easy-to-use computer-based data acquisition methods. The study accomplished this by using simultaneous speech sampling with healthy controls and uniform analysis techniques.

METHOD

Participants

Healthy individuals (6 females and 9 males) between 24 and 52 years of age (mean = 32.87 ± 9.4 years) participated in this study. Participants gave informed consent prior to the commencement of the study, and they were recruited by members of the staff at an associated institution. Participants were screened to exclude smokers, heavy drinkers, heavy coffee drinkers, recreational drug or alcohol abusers, people with neurological trauma, and those with psychological illness.

Materials

Each participant performed the speech battery individually and in a quiet room. The speech samples were recorded simultaneously using three acquisition methods. The first method utilized a Marantz PMD671 Portable Solid State Recorder (D&M Holdings, Mahwah, NJ) and an AKG C 420 head-mounted cardioid condenser microphone (AKG Acoustics Harman ProGmbH, Munich). A distance of 8 cm was maintained between mouth and microphone to avoid distortion and to preserve intensity levels. The data were then transferred to a computer for preservation and analysis. The second method utilized an Alesis ML-9600 Masterlink disc recorder (Alesis, Cumberland, RI) (hard disk drive [HDD]) coupled with a Behringer Eurorack UB802 mixer (Willich, Germany) with an AKD Model C420 head-mounted cardioid condenser microphone positioned 8 cm to the side of the mouth. The AKG C 420 microphone has a minimum sensitivity of -43 dB and a frequency range of 20 Hz to 20 kHz. The third method involved the use of a laptop PC with basic factory settings (Sony) and a Logitech A-0374A head-mounted microphone with USB connection (Logitech, Hong Kong; minimum sensitivity of -38 dB and a frequency range of 100 Hz–16 kHz), which was positioned at a 45° angle, 8 cm from the mouth. All data were sampled at 44.1 kHz, with quantization at 16 bits for all methods.

All three acquisition methods are commonly employed in clinical and empirical settings; however, the degree of perceived accuracy and reliability of signal capture varies between hardware configurations. Of the three acquisition configurations used in this experiment, the HDD setup is considered the most precise, because the HDD has no movable parts, the high-quality condenser microphone uses phantom power (supplied by an AKG B 18V battery), a separate mixer controls sound levels, and all wiring is thickly insulated. All of these measures minimize the amount of extraneous noise that can enter the signal during the recording process. The SSR is also a popular recording device, and it provides good signal acquisition, but not to the standards of the HDD setup used in this study. The PC configuration is considered the most fallible acquisition technique of the three.

Procedure

All participants performed the same speech tasks and used the three acquisition methodologies. This battery took approximately

10–15 min to complete and consisted of three speech tasks presented in succession by the researcher. Participants were asked to speak using their normal or natural pitch and intensity when providing samples.

Stimuli

Experimental stimuli were chosen in accordance with criteria relating to ease of use and suitability for postrecording analysis. Participants were instructed to speak in a natural manner, using their typical speaking rate, and with a volume appropriate for speaking to one or two people in quiet surroundings. Three speech tasks were included in the battery, including the Grandfather passage (a phonetically balanced paragraph), which required participants to read a passage aloud. Participants were allowed to familiarize themselves with the passage before reading. Participants were also asked to sustain an open “ah” vowel in an even tone, at a comfortable pitch, and at constant amplitude, for approximately 6 sec. The final task required participants to produce a monologue on any topic of their choosing (e.g., work, family, or friends). The sample was recorded for up to 1 min. All participants completed the same speech tasks, using the same three acquisition devices simultaneously.

Acoustic Analysis

All speech samples were analyzed using the same analytical method. The speech samples were digitized at a sampling rate of 44.1 kHz, with a 16-bit conversion rate for each acquisition system. Each speech sample was then segmented and analyzed using Praat (Version 4.6.09; <http://www.praat.org/>). Silences were removed from the start and end of the sustained vowel task, and the sample was truncated, leaving 1.5 sec on each side of the temporal midpoint to allow automated analysis of each vowel segment. Silences were also removed from the start and end of the conversation samples and of the read passages.

Timing measures were taken from the conversation sample and the recited passage using an automated script run within Praat. The following measures were derived: total signal time, number of pauses, total pause time, mean pause time, standard deviation of pause, percent pause, total speech time, and speech:pause ratio. Mean pause time was calculated as a portion of the total utterance falling below 46 dB for longer than 0.1 sec.

Frequency-based measures were also derived from the same speech samples, using an automated script designed for use in Praat. The following measures were taken: f_0 , the standard deviation of f_0 , and the coefficient of variation of f_0 . Two separate voice-analysis programs were employed to investigate acoustic measures of the sustained vowel. Using Praat, the following measures were taken: f_0 , the standard deviation of f_0 , the coefficient of variation of f_0 , the first formant (F_1) mean, the standard deviation of F_1 , the coefficient of variation of F_1 , the second formant (F_2) mean, the standard deviation of F_2 , and the coefficient of variation of F_2 . The high-end frequency for formant analysis was set at 5000 Hz and targeted the first five formants. The Computerized Speech Laboratory's Multi-dimensional Voice Program (MDVP) Model 5105 (KayPENTAX) was used to derive a number of frequency and amplitude measures, including f_0 , its derivatives, and its measures of perturbation (e.g., jitter or shimmer).

RESULTS

All data collected from the speech samples were analyzed statistically, and the acquisition method was treated as the independent variable. To identify the variance inherent in acquisition methodologies, the same behavior (speech) was recorded in all participants using different acquisition configurations, and it was analyzed using a consistent analysis protocol. Variance between recording systems was determined using a one-way ANOVA.

Table 1
Means and Standard Deviations of Acoustic Analysis Resulting
From Different Acquisition Methodologies

Measure	HDD		SSR		PC	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of pauses	42.33	19.76	43.97	17.18	35.60	15.04
Percent pause	0.32	0.20	0.21	0.08	0.22	0.10
Mean pause length (sec)	0.39	0.17	0.27	0.07	0.31	0.08
<i>f</i> 0 (Hz)	141.31	39.59	139.43	38.32	141.85	40.38
<i>F</i> 1 (Hz)	721.38	146.08	695.82	173.38	752.78	175.21
<i>F</i> 2 (Hz)	1,321.75	112.03	1,315.03	117.01	1,361.34	113.88

Timing

The ANOVA yielded no significant differences for number of pauses [$F(2,86) = 1.9, p = .155$] or percent pause time [$F(2,77) = 2.27, p = .11$] on free-speech and recited-passage tasks on all three acquisition systems. On measures of mean length of pauses, however, comparison of acquisition methods yielded significantly varied results [$F(2,86) = 6.94, p = .0016$] (Table 1).

Fundamental Frequency

Similarly, the ANOVA for *f*0, the standard deviation of *f*0, and the coefficient variation of *f*0 yielded no significant differences on sustained-vowel, free-speech, and recited-passage tasks on all three acquisition systems [$F(2,133) = 0.046, p = .95$] (Table 1).

Formant Frequencies

The ANOVA for *F*1 [$F(2,42) = 0.448, p = .642$] and *F*2 [$F(2,42) = 0.719, p = .493$] yielded no statistically significant differences on the sustained-vowel task on all three acquisition methods. Similarly, the corresponding *F*1 standard deviations [$F(2,41) = 1.651, p = .848$] and *F*1 coefficient of variance [$F(2,40) = 1.66, p = .203$], as well as the *F*2 standard deviations [$F(2,41) = 1.778, p = .182$] and *F*2 coefficient of variance [$F(2,41) = 1.901, p = .162$], using the same tests, yielded no statistically significant differences between acquisition methods.

Perturbation

Perturbation measures were calculated using the MDVP standard protocol. The ANOVA revealed no significant differences between all acquisition methods for absolute jitter [$F(2,36) = 1.201, p = .313$], jitter percent [$F(2,38) = 1.433, p = .251$], noise-to-harmonic ratio [$F(2,38) = 0.511, p = .604$], pitch perturbation quotient [$F(2,38) = 1.704, p = .196$], relative average perturbation [$F(2,38) = 1.283, p = .289$], and shimmer in dB [$F(2,38) = 1.718, p = .193$].

The ANOVA for amplitude perturbation quotient [$F(2,37) = 3.571, p = .038$] and shimmer percent [$F(2,3) = 4.37, p = .0198$] yielded significant differences between acquisition methods. Similar results were found for the ANOVA between acquisition methods on measures of peak-to-peak amplitude variation [$F(2,38) = 6.94, p = .0027$], soft phonation index [$F(2,35) = 12.2, p < .0001$], voice turbulence index [$F(2,38) = 7.996, p = .0013$], and phonatory *f*0 range in semitones [$F(2,38) = 3.264, p = .049$].

DISCUSSION

The findings of this study suggest that a variety of data acquisition methods can yield comparable results on selected acoustic measures. Measures of *f*0, *F*1, and *F*2 and the corresponding standard deviations and coefficients of variance were all invariant across acquisition methods. Similarly, some timing measures—including the number of pauses and percent pause time—were not significantly different across acquisition methods. Surprisingly, some measures of frequency and intensity variation also proved stable across recording equipment. Measures derived from the MDVP analysis showed that absolute jitter, jitter percent, noise-to-harmonic ratio, pitch perturbation quotient, relative average perturbation, and shimmer in dB were invariant across methods.

Disagreement between acquisition methods was observed in a number of acoustic parameters (Table 2). Some aspects of timing, including the mean length of pauses, appeared to be affected by the choice of recording equipment. Moreover, features of pitch and intensity perturbation varied between methods. The acoustic parameters that

Table 2
Summary of Acoustic Stimuli That Are Accurately or
Inaccurately Obtained via PC Acquisition, Compared With
Higher Quality Configurations (i.e., HDD and SSR)

Measures	Accurately	Inaccurately
<i>f</i> 0	✓	
<i>f</i> 0 <i>SD</i>	✓	
<i>f</i> 0 coefficient of variance (CV)	✓	
<i>F</i> 1	✓	
<i>F</i> 1 <i>SD</i>	✓	
<i>F</i> 1 CV	✓	
<i>F</i> 2	✓	
<i>F</i> 2 <i>SD</i>	✓	
<i>F</i> 2 CV	✓	
No. of pauses	✓	
% Pause	✓	
Mean length of pause		×
Absolute jitter	✓	
% Jitter	✓	
Noise-to-harmonic ratio	✓	
Pitch perturbation quotient	✓	
Relative average perturbation	✓	
Shimmer in dB	✓	
Shimmer percent		×
Amplitude perturbation quotient		×
Peak-to-peak amplitude variation		×
Soft phonation index		×
Voice turbulence index		×
Phonatory <i>f</i> 0 range in semitones		×

varied significantly between methods included amplitude perturbation quotient, shimmer percent, peak-to-peak amplitude variation, soft phonation index, voice turbulence index, and phonatory f_0 range in semitones.

These results replicated or supported a number of studies that have demonstrated the robust and invariant nature of key frequency and timing-based measures (Mundt et al., 2007), and the sensitive character of supplementary acoustic parameters (Carding et al., 2004; Deliyski et al., 2006).

As the role of voice acoustics expands beyond purely diagnostic or clinical use, the need for the adoption of specialized recording equipment lessens. Admittedly, each clinical population presents with acoustic profiles that require a specific focus of investigation. Nevertheless, a portion of acoustic studies rely on a fairly narrow line of vocal parameters. For example, in a study by Mundt et al. (2007), the researchers looked at voice acoustic changes that are inherent in individuals undergoing treatment for clinical depression. The features of speech that were targeted by the investigators centered on basic timing and frequency measures, and they were collected via the telephone. Traditionally, voice acoustic studies have required specific, high-grade recording equipment to accurately and reliably collate data. After investigating the precision of interactive voice-response technology (telephone) (Cannizzaro, Reilly, Mundt, & Snyder, 2005), however, Mundt et al. collected data on 35 physician-referred patients with depression and found that some pitch- and pause-related features changed significantly over the course of treatment, thus demonstrating the applicability of alternative acquisition methods.

Alternative acquisition methodologies have also been used frequently in the somnolence literature. Whitmore and Fisher (1996) looked at fundamental frequency and word length in repeated phrases over the course of a 36-h sustained-wakefulness experiment. The researchers employed a Racal (Versasoft model) signal recorder to obtain participants' speech signals on videocassettes. Although data were recorded using low-quality recording equipment, which has been shown to lack the clarity and quality of digital recording devices (Jiang, Lin, & Hanson, 1998), the two acoustic measures that were selected proved adequate for gaining significant information on the status of participants.

These two studies illustrate the potential value of alternative acquisition methodologies when the type of acoustic parameter is appropriately chosen. Not all acquisition environments are suitable for widespread acoustic investigation, especially when certain analysis systems are employed (Deliyski, Evans, & Shaw, 2005; Deliyski et al., 2005b; Smits et al., 2005) or the signal-to-noise ratio is unsuitable (Deliyski et al., 2005a; Ma & Yiu, 2005; Qi, Hillman, & Milstein, 1999). The use of less specialized equipment such as PCs can limit a study design in other ways. In addition to introducing noise that is caused by a lack of isolation of wiring systems, interference that is produced inside the computer, and variation within and between analyses systems, the hardware configurations of PCs themselves can vary between research groups. Furthermore, voice acoustic research labs around the world

often vary in their choice of software and hardware, making it difficult to directly compare results. Fortunately, USB-connected microphones mitigate some of the variation among computer-based acquisition systems.

The results of the present study demonstrate the clear need for specialized acquisition hardware when one is investigating certain acoustic parameters, including features of frequency and intensity perturbation. The results also suggest that some timing measures require high-level hardware in order to obtain accurate and reliable results. Despite these limitations, the findings of the present study undoubtedly confirm the role of widely accessible recording equipment in empirical work. Key features of voice frequency and timing can be collected reliably using a variety of acquisition methodologies when the sampling and quantization levels are established at a minimum of 44.1 KHz and 16 bits, respectively, and when the data are analyzed using specific analytical programs—in this case, Praat and MDVP.

Future research of this nature could focus on a greater variety of acquisition devices, including personal digital assistants (PDAs) and mobile phones. The provision of methods that lessen the burden of equipment cost, user expertise, or complexity of components could revolutionize the speech science industry. Furthermore, the introduction and acceptance of new technologies in acoustic studies would greatly expand the audience and reach of speech science in general.

AUTHOR NOTE

The authors are grateful to Paul Kukiel for his assistance with the analysis scripts and to all those who participated in the research. Correspondence concerning this article should be addressed to A. P. Vogel, Level 7/21 Victoria Street, Melbourne, VIC 3000, Australia (e-mail: avogel@cogstate.com).

REFERENCES

- ALPERT, M., ROSENBERG, S. D., POUGET, E. R., & SHAW, R. J. (2000). Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry Research*, *97*, 107-118.
- 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 & Hearing Research*, *39*, 126-134.
- BOUGH JR., I. D., HEUER, R. J., SATALOFF, R. T., HILLS, J. R., & CATER, J. R. (1996). Intrasubject variability of objective voice measures. *Journal of Voice*, *10*, 166-174.
- CANNIZZARO, M. S., REILLY, N., MUNDT, J. C., & SNYDER, P. J. (2005). Remote capture of human voice acoustical data by telephone: A methods study. *Clinical Linguistics & Phonetics*, *19*, 649.
- CARDING, P. N., STEEN, I. N., WEBB, A., MACKENZIE, K., DEARY, I. J., & WILSON, J. A. (2004). The reliability and sensitivity to change of acoustic measures of voice quality. *Clinical Otolaryngology & Allied Sciences*, *29*, 538-544.
- CARSON, C. P., INGRISANO, D. R., & EGGLESTON, K. D. (2003). The effect of noise on computer-aided measures of voice: A comparison of CSpeechSP and the Multi-Dimensional Voice Program software using the CSL 4300B Module and Multi-Speech for Windows. *Journal of Voice*, *17*, 12-20.
- DEJONCKERE, P. H., BRADLEY, P., CLEMENTE, P., CORNUT, G., CREVIER-BUCHMAN, L., FRIEDRICH, G., ET AL. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Oto-Rhino-Laryngology*, *258*, 77-82.

- DELIYSKI, D. D., EVANS, M. K., & SHAW, H. S. (2005). Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice*, **19**, 176-186.
- DELIYSKI, D. D., SHAW, H. S., & EVANS, M. K. (2005a). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, **19**, 15-28.
- DELIYSKI, D. D., SHAW, H. S., & EVANS, M. K. (2005b). Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopedics, Phoniatrics, Vocology*, **30**, 55-62.
- DELIYSKI, D. D., SHAW, H. S., EVANS, M. K., & VESSELINOV, R. (2006). Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica*, **58**, 274-288.
- ESPY-WILSON, C. Y., BOYCE, S. E., JACKSON, M., NARAYANAN, S., & ALWAN, A. (2000). Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America*, **108**, 343-356.
- FETTE, B., GIBSON, R., & GREENWOOD, E. (1980). Windowing functions for the average magnitude difference function pitch extractor. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, **5**, 49-52.
- GERHARD, D. (2003). *Pitch extraction and fundamental frequency: History and current techniques* (Tech. Rep. No. TR-CS 2003-06). Regina, SK: University of Regina, Department of Computer Science.
- GREEN, J. R., BEUKELMAN, D. R., & BALL, L. J. (2004). Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. *Journal of Medical Speech-Language Pathology*, **12**, 149-154.
- HILLENBRAND, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals. *Journal of Speech, Language, & Hearing Research*, **30**, 448-461.
- HIROSE, K., FUJISAKI, H., & SETO, S. (1992). A scheme for pitch extraction of speech using autocorrelation function with frame length proportional to the time lag. *Acoustics, Speech, & Signal Processing*, **1**, 149-152.
- INGRISANO, D. R. S., PERRY, C. K., & JEPSON, K. R. (1998). Environmental noise: A threat to automatic voice analysis. *American Journal of Speech-Language Pathology*, **7**, 91-96.
- JIANG, J., LIN, E., & HANSON, D. G. (1998). Effect of tape recording on perturbation measures. *Journal of Speech, Language, & Hearing Research*, **41**, 1031-1041.
- KARNELL, M. P. (1991). Laryngeal perturbation analysis: Minimum length of analysis window. *Journal of Speech & Hearing Research*, **34**, 544-548.
- 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**, 383-393.
- KARNELL, M. P., SCHERER, R. S., & FISCHER, L. B. (1991). Comparison of acoustic voice perturbation measures among three independent voice laboratories. *Journal of Speech & Hearing Research*, **34**, 781-790.
- KENT, R. D., VORPERIAN, H. K., & DUFFY, J. R. (1999). Reliability of the Multi-Dimensional Voice Program for the analysis of voice samples of subjects with dysarthria. *American Journal of Speech-Language Pathology*, **8**, 129-136.
- LEI, K., YANG, X., SHEN, J. Z., & GONG, J. R. (2000). The study on the standardization of adults voice acoustic parameters by objective analysis. *Journal of Clinical Otorhinolaryngology*, **14**, 255-257.
- MA, E. P., & YIU, E. M. (2005). Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals. *Folia Phoniatrica et Logopaedica*, **57**, 38-47.
- MUELLER, P. B. (1997). The aging voice. *Seminars in Speech & Language*, **18**, 159-168.
- MUNDT, J. C., SNYDER, P. J., CANNIZZARO, M. S., CHAPPIE, K., & GERALTS, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics*, **20**, 50-64.
- PARSA, V., & JAMIESON, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *Journal of Speech, Language, & Hearing Research*, **44**, 327-339.
- PARSA, V., JAMIESON, D. G., & PRETTY, B. R. (2001). Effects of microphone type on acoustic measures of voice. *Journal of Voice*, **15**, 331-343.
- PERRY, C. K., INGRISANO, D. R., & BLAIR, W. (1996). The influence of recording systems on jitter and shimmer estimates. *American Journal of Speech-Language Pathology*, **5**, 86-90.
- PERRY, C. K., INGRISANO, D. R., PALMER, M. A., & McDONALD, E. J. (2000). Effects of environmental noise on computer-derived voice estimates from female speakers. *Journal of Voice*, **14**, 146-153.
- PERRY, C. K., INGRISANO, D. R., & SCOTT, S. R. (1996). Accuracy of jitter estimates using different filter settings on Visi-Pitch: A preliminary report. *Journal of Voice*, **10**, 337-341.
- QI, Y., HILLMAN, R. E., & MILSTEIN, C. (1999). The estimation of signal-to-noise ratio in continuous speech for disordered voices. *Journal of the Acoustical Society of America*, **105**, 2532-2535.
- RABINER, L. R. (1977). On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, **25**, 24-33.
- ROA, S., BENNEWITZ, M., & BEHNKE, S. (2007). Fundamental frequency estimation based on pitch-scaled harmonic filtering. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007 (pp. IV-397 to IV-400).
- SCHERER, R. C., VAIL, V. J., & GUO, C. G. (1995). Required number of tokens to determine representative voice perturbation values. *Journal of Speech & Hearing Research*, **38**, 1260-1269.
- SMITS, I., CEUPPENS, P., & DE BODT, M. S. (2005). A comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. *Journal of Voice*, **19**, 187-196.
- TITZE, I. R., & WINHOLTZ, W. S. (1993). Effect of microphone type and placement on voice perturbation measurements. *Journal of Speech & Hearing Research*, **36**, 1177-1190.
- WHITMORE, J., & FISHER, S. (1996). Speech during sustained operations. *Speech Communication*, **20**, 55-70.
- WINHOLTZ, W. S., & TITZE, I. R. (1997). Miniature head-mounted microphone for voice perturbation analysis. *Journal of Speech, Language, & Hearing Research*, **40**, 894-899.
- WINHOLTZ, W. S., & TITZE, I. R. (1998). Suitability of minidisc (MD) recordings for voice perturbation analysis. *Journal of Voice*, **12**, 138-142.
- XUE, S. A., & FUCCI, D. (2000). Effects of race and sex on acoustic features of voice analysis. *Perceptual & Motor Skills*, **91**, 951-958.

(Manuscript received December 23, 2007;
revision accepted for publication March 12, 2008.)