



Published in final edited form as:

Behav Res Methods. 2009 May ; 41(2): 318–324. doi:10.3758/BRM.41.2.318.

Standardization of pitch range settings in voice acoustic analysis

Adam P. Vogel^{1,*}, Paul Maruff¹, Peter J. Snyder², and James C. Mundt³

¹Centre for Neuroscience, University of Melbourne, VIC, Australia

²Rhode Island Hospital and the Warren Alpert Medical School of Brown University, Providence, RI, U.S.A

³Healthcare Technology Systems, 7617 Mineral Point Road, Madison, WI, USA

Abstract

Voice acoustic analysis is typically a labor intensive, time consuming process that requires the application of idiosyncratic parameters tailored to individual aspects of the speech signal. These processes limit the efficiency and utility of voice analysis in clinical practice as well as applied research and development. In the current study, we analyzed 1120 voice files using standard techniques (case by case hand analysis); taking roughly 8 weeks of personnel time complete. The obtained results were then compared to the analytic output of several automated analysis scripts that made use of pre-set pitch range parameters. The automated analysis scripts reduced processing time of the 1680 speech samples to less than 2.5 hours and produced results comparable to the hand analysis when pitch window were appropriately selected to account for known population differences (i.e., sex differences). Caution should be exercised when applying suggested settings to pathological voice populations.

Introduction

Acoustic analysis of the voice requires the selection of parameter settings specific to sample characteristics such as intensity, duration, frequency and filtering. Historically, acoustic experiments have relied on small sample sizes, where the laborious use of idiosyncratic parameters for individual samples on a case-by-case basis was feasible (Kent & Read, 2002). As a result, most voice acoustic studies lack standardized or automated analytical procedures (Titze, 1994). In addition, perceptual evaluation of the voice during clinical investigations remains the primary method for assessing vocal change (Kent & Read, 2002). However, Carding, Carlson, Epstein, Mathieson and Shewell (2001) have identified numerous limitations of perceptual voice analysis, citing poor reliability of analyses within and between raters (Bassich & Ludlow, 1986), disparity in the design of the available perceptual rating scales, variability of the human voice and difficulties interpreting complex acoustic signals as contributory factors. Despite these limitations, it has to be acknowledged that some clinical populations may not be suitable for standardized analysis procedures. Voice disorders such as glottal fry (where the voiced speech signal may contain non periodic signals) or individuals with atypical frequency profiles (e.g., puberphonia) may also be inappropriate for generic analysis procedures, and thus require a combination of perceptual and idiosyncratic evaluation.

The need for objective and repeatable evaluation of vocal change in large, generalizable experimental settings calls for greater accountability and quantitative judgment in voice acoustic research (Eskenazi, Childers, & Hicks, 1990; John, Sell, Sweeney, Harding-Bell, & Williams, 2006; Kent, 1996). Standardized acoustic analysis has the comparative advantage

of providing objective and repeatable measures of vocal change relative to procedures based on perceptual judgments.

Current procedures for acoustic voice analysis typically require idiosyncratic specifications of pitch range settings that determine window length, selection of intensity cut-offs, hand splicing of sound files, and expert knowledge of particular software/hardware configurations. Such case-by-case analytic procedures are time and resource intensive, and are an impediment to the growing need for fast and accurate voice analysis. Development of standardized, automated procedures could streamline this process, thus reducing the time, cost, and need for hands on intervention.

Standardizing the selection of the pitch range settings for measuring particular acoustic properties using known characteristics of the speakers could provide a means for enhancing the efficiency of the analytic process. Window frame lengths determine how pitch contours are displayed and computed and the application of inappropriate frame properties can result in inaccurate and unreliable data. However, to date, very few studies have evaluated the effects of systematically altering window frame length or pitch range settings on large batches of sound files and compared them to idiosyncratically selected gold standards. Attempts to employ automated pitch extraction and window length algorithms are being refined (Atkinson, Kondoz, & Evans, 1995; Fette, Gibson, & Greenwood, 1980; Karnell, Scherer, & Fischer, 1991; Rabiner, 1977), but this remains a complex process. Until fast and accurate techniques are developed to facilitate large, batch processing of voice files, preset frame sizes designed for distinct population groups will be required. Fortunately, tailored techniques that adapt to subtle pitch variations within an utterance are generally unnecessary, as the variability within an utterance is usually within one octave above or below the average pitch in a typical sample (Rabiner, 1977).

A number of studies have addressed analytic processes related to window size or type (Fette et al., 1980; Takagi, Seiyama, & Miyasaka, 2000), signal length requirements (Scherer, Vail, & Guo, 1995), timing (Green, Beukelman, & Ball, 2004), prosodic measures (Hirst, 2002) and appropriateness of stimuli used to elicit acoustic measures (Zraick, Wendel, & Smith-Olinde, 2005). Calls for better standardization and automation of voice analysis are prevalent in the literature, yet significant technical limitations remain. Consequently, objective studies of the voice continue to be dominated by relatively small, clinical experiments that do not require rapid processing of large batches of acoustic data. Understandably, there is little incentive to develop or apply generic analysis settings under such circumstances. Additionally, the use of generic analytical settings may not be appropriate for all acoustic measures. For example, measures of perturbation like jitter and shimmer may vary within and between analysis tools, as voices that violate assumptions of periodicity often require case-by-case evaluation (Karnell, Hall, & Landahl, 1995; Perry, Ingrisano, & Scott, 1996; Titze & Liang, 1993). Similarly, extraction of error free pitch measures from a voice that varies within a short period of time, and differences between speakers due to sex or emotional state can be problematic (Mendoza, Munoz, & Valencia Naranjo, 1996; Mueller, 1997). Consequently, it is important that window frame lengths be kept small, as fast excursions of pitch should not be smoothed over (Gerhard, 2003). However, at the same time, the window frame length should be long enough to a complete enough cycles of the periodic waveform (Fette et al., 1980; Karnell et al., 1991).

It is advantageous to the researcher that some acoustic measures appear to retain levels of consistency despite alterations in analysis methodology, especially those relating to fundamental frequency (f_0). F_0 has demonstrated robustness as an acoustic parameter in trials comparing software configurations (Deliyski, Shaw, & Evans, 2005b; Kent, Vorperian, & Duffy, 1999; Takagi et al., 2000), environments with poor signal to noise ratios (Deliyski, Shaw, & Evans, 2005a) acquisition environments (Deliyski, Evans, & Shaw, 2005), and

sampling rates (Deliyski, Shaw et al., 2005b). Aspects of motoric timing during speech production, such as measures of pause frequency and speaking rate have also been found to be important, replicable measures (Cannizzaro, Harel, Reilly, Chappell, & Snyder, 2004). Based on existing literature and past experiences, a limited range of upper and lower bounds for defining pitch ranges for voice acoustic analysis appear to be adequate for 'normal' vocal populations. Development of automated or semi-automated processes that would allow widespread application of generic pitch ranges, in the context of voice acoustic analysis, has the potential to lessen the use of idiosyncratic analytical settings. Evaluation of such a process, however, requires comparison to results obtained through laborious, resource-intensive gold-standard practices of individually analyzed files. Analysis of large batches of speech data using automated processes could dramatically decrease the time and resources currently required by acoustic studies, and could improve the reliability and repeatability of study results. Considering a single, fixed window frame length has been shown to be unsuitable for all speech samples (Rabiner, 1977). Therefore, this study investigated the potential use of generic pitch range and window frame settings, tailored to the sex of the speaker. The resulting acoustic measures are compared to those obtained through optimized gold-standard methods resulting from individual, case-by-case analysis.

Methods

Participants

Twenty participants (10 female and 10 male, mean age = 40.5 years) who had participated in a methodology study investigating speech changes associated with treatment for depression were randomly selected from a cohort of thirty-five. Subject demographics, clinical characteristics, participation criteria, and study procedures are described in Mundt, Snyder, Cannizzaro, Chappie and Geraltz (2007).

Data Collection

Speech samples produced in response to automated elicitation procedures were obtained over a standard office telephone using an interactive voice response (Mundt et al., 2007) system. Performance of the speech elicitation tasks produced samples of free speech (participants extemporaneously discussed their recent physical, emotional and general functioning lasting around one minute), prolonged vowels (/a:/, /ae:/, /u:/, /i:/), and reading of a standard passage (Grandfather Passage - 175 syllables). Validation of the telephony-based procedures for eliciting voice acoustic measures comparable to those obtained in a laboratory setting have been published (Cannizzaro, Reilly, Mundt, & Snyder, 2005).

Each study participant produced 56 speech samples, generating a total of 1120 recordings for vocal acoustic analysis. Each speech sample was individually scored by hand, requiring approximately 350 to 420 hours of total personnel time. The case-by-case, hand-scored acoustic measures provided the gold standards for evaluating the accuracy of the automated scoring procedures described below.

Measures

The primary measures of interest for the current investigation were f0, f0 standard deviation (SD) and f0 coefficient of variation (CV). These measures were selected because of their robust analytical potential, as frequency measurements have been shown to be resistant to acoustic artefaction and is largely independent of signal quality. Frequency profiles were calculated from the samples previously mentioned. F0, and its corresponding SD and CV were derived from all samples.

Procedure and acoustic analysis

All speech samples were segmented and analyzed using PRAAT version 5.0.32 (Boersma & Weenink, 2008) which employs a user-supplied estimate of analysis window length. Silences were removed from the start and end of the sustained vowel samples and each sample was truncated 1.5 seconds each side of the temporal midpoint. The other speech samples were not truncated. In order to determine window length two primary parameters were considered. *Time step* is a measurement interval (frame duration) in seconds and is calculated (when time step is set to 0) by dividing 0.75 by (*pitch floor*), (e.g. if the pitch floor is 75 Hz, the time step equals 0.1 seconds; prompting PRAAT to compute 100 pitch values per second). *Pitch floor* determines the length of the analysis window and also represents the lowest fundamental frequency targeted within each sample. For calculation of pitch, the window should be long enough to contain three periods (for pitch detection); for example, for a pitch floor of 75 Hz, the window will be effectively $3/75 = 0.04$ seconds long. Increasing the time step will speed up the editor window, however, it may lead to under sampling of the pitch and formant curves, which in turn, will influence the accuracy of selected measures. The *pitch ceiling* is a post-processing step that ignores candidates above the prescribed setting, and promotes the most efficient use of available data. A summary of the complete multi-parameter algorithm for calculating the fundamental frequency, as it is implemented into the speech analysis and synthesis program, PRAAT, is described in Boersma (1993).

Pitch range settings in PRAAT are the most important parameters in pitch analysis. As described, the pitch floor determines the window length and the pitch ceiling restricts the values being recruited during the analysis. Therefore, the question of automatically defining optimal values for floor and ceiling is not a trivial one and the output can be fairly different depending on the values used. The pitch range / window range settings could be displayed in time, however, for ease of translation, the current study have maintained units of measurement in hertz. Based on professional experience, recommended software settings, and clinical impressions of pitch variation between male and female speakers, *pitch floor* values (50, 70, and 100 Hz) were considered in conjunction with five *pitch ceiling* values (250, 300, 500, 600 and 625 Hz) as potential sex-specific, generic frequency window settings for automating acoustic analysis. Pitch floor settings dictate that candidates below this frequency will not be recruited; similarly, the pitch ceiling ensures that candidates above this frequency will be ignored. All f_0 plots were produced by an autocorrelation algorithm. The nine resulting sets of acoustic parameters defining potential floor/ceiling pitch range settings were evaluated; 50-250 Hz; 50-300 Hz; 50-500 Hz; 50-600 Hz; 70-250 Hz; 70-300 Hz; 70-500 Hz; 70-600 Hz; 70-625 Hz; 100-250 Hz; 100-300 Hz, 100-500 Hz and 100-600 Hz. The results from each pitch range setting were compared to data derived from the gold standard settings, based on unique sample characteristics identified during hand scoring of the data reported in Mundt et al. (2007).

The complete set of 1120 speech samples (20 subjects providing 56 samples each) were analyzed using each of the 13 pitch range settings identified above. Thus, a total of 15680 analytic cycles were performed using automated batch processing of the speech samples. The derived f_0 , f_0 SD, and f_0 CV measures obtained from each cycle were stored and compared to the gold standard measures obtained from the individually hand scored samples. Gold standard measures were determined by visually inspecting the waveform of each sound file and selecting parts of the signal that were periodic. It is easy for speech scientists to visually identify parts of the signal that conform to a regular pattern, however, this task is difficult for computer programs. The signal characteristics (e.g., frequency range of sample) were then identified manually and used for analysis. This process is very time consuming, however, it is currently considered the most accurate method of acoustic analysis.

Statistical analysis—Statistical differences of data sets were analyzed using one-way ANOVA and differences between groups were assessed by Dunnett's method, which compares independent variables (generic pitch settings) with control (gold standard) data using commercially available statistics software (JMP Statistical Discovery, SAS Institute Inc. Cary, NC, USA). Effect sizes were calculated using Cohens *d*. P-values <0.05 were considered statistically significant.

Results

Data obtained via gold standard analysis were divided into two groups; male and female (table 1). Independent group t-tests were employed to compare the means of each measure. Results revealed significant differences between males and females ($p < 0.05$) for f_0 and f_0 SD values.

Measures that varied significantly between sexes, when analyzed via gold standard practices, were then examined using the generic pitch range settings. Statistical differences between gold standard techniques and trial pitch range setting groups were analyzed by one-way ANOVA and Dunnett's multiple comparisons test using JMP 7.0 Statistical Discovery (SAS Institute: NC, USA). Resulting differences between gold standard methods and specific trial pitch frames were then further analyzed by Student's *t*-test.

Fundamental frequency

One-way ANOVA of the f_0 obtained by each of the pitch range settings revealed significant variance between groups for both female ($F(4635)=10.94, p < 0.0001$) and male groups ($F(3910)=92.1, p < 0.0001$). Similar findings were observed on measures of f_0 SD, female ($F(4635)=67.6, p < 0.0001$), male ($F(3910)=166.74, p < 0.0001$) and f_0 CV female ($F(4635)=66.77, p < 0.0001$) male groups ($F(3910)=150.92, p < 0.0001$). These results suggest that f_0 , f_0 SD and f_0 CV values require specific pitch range settings.

Dunnett's multiple comparison method revealed significant differences between methods on the mean scores of f_0 , f_0 SD and f_0 CV for both female (table 2) and male (table 3) groups.

Data in tables two and three illustrate the degree of overall variability inherent in the selection of frame range settings, despite controlling for a number of variables. All 1120 speech samples were analyzed repeatedly using each pitch range setting, using the same software program and analysis scripts. As only one aspect of the study was modified during each analysis cycle, all observed variance in f_0 was due to manipulation of pitch range settings.

Using f_0 as a guide, a number of generic settings appeared to provide accurate estimates of pitch, however, when the data was broken down by task, very few settings yielded equivalent results across all stimuli and measures compared to data derived from hand selected settings. In addition, speech samples that fell outside the prescribed generic pitch range settings were not automatically analyzed, and consequently were not included in the results. All female data was analyzed without exclusion, however 17 files were excluded within the male batch when the pitch floor was 100Hz. In light of these issues, of specific interest are the experimental conditions that, when broken down by sex and task, appear to be suitable for large batch analysis of all acoustic data. Statistical differences and significant effect sizes were not observed between gold standard techniques and specific pitch range lengths for pitch ranges between 100-300 Hz and 100-250 Hz for females, and 70 to 250 Hz for males.

Discussion

The use of sex-specific, generic pitch range lengths on more than 1100 speech files in an automated analysis script produced statistically equivalent results to those obtained from

idiosyncratic, individualized, gold standard practices. A strong relationship was observed among window ranges determined by pitch range settings that adapted to the higher and lower frequencies introduced by the sex of the speaker. Specifically, a clear set of pitch range settings suitable for male participants were identified that encompassed low pitch floors (e.g., 70 Hz) and mid level pitch ceiling (e.g., 250 Hz). Similarly, a pitch floor setting of 100 Hz and ceiling settings of 250 and 300 Hz for female speakers yielded comparable results to current gold standard practices. Identification of speaker sex, prior to analyzing the voice samples to obtain acoustic measures like f_0 , permits part-individualization of analysis procedures and promises to increase the efficiency and speed of analysis compared to current hand-scoring procedures.

These findings underscore the value of considering generic window lengths as a means of expediting large batch voice acoustic analysis for both male and female populations. Traditional analysis methods typically rely on labor intensive individually selected settings or slow adaptive analysis algorithms based on variation in frequency within the speech signal (Karnell et al., 1991; Morris & Brown, 1996; Rabiner, 1977). In the current study, hand scoring of the 1680 voice files required roughly 10 weeks of personnel time. Adapting the analysis settings based on known characteristics of the speaker offers an alternative to laborious, idiosyncratic analyses. Such adaptive techniques must be validated in acoustic environments that require fast and accurate analysis of several thousand speech files. Automated analysis of the same 1680 voice files using sex specific pitch range settings only required an average of five seconds per file.

The relationship between frame lengths and the f_0 of a speaker is complicated due to the inherent variation of frequency profiles from one speaker to the next (Hirose, Fujisaki, & Seto, 1992), but managing speaker specific analysis settings individuality requires extensive expertise and time and is impractical for large volumes of data. If the pitch floor is set too low, very fast f_0 changes will be missed, and if it is set too high, low f_0 values will be neglected. However, by separating populations into distinct vocal groups, for example, male and female, a set of cohort specific settings can be applied. Such procedures could greatly reduce the labor intensive process in current practices.

Inherent variability of the human voice precludes the use of generic frame lengths in all populations, but standardized sex-specific pitch ranges may be suitable for a wide range of 'typical' frequency profiles. Individuals with frequency profiles outside expected levels may not be accurately analyzed using automated acoustic analysis. However, establishing reasonable restrictions on the recording environment, software selection, and effectively screening participants could permit effective application of automated and standardized acoustic analysis.

Clinically and within the literature, consideration of the prescriptive aspects related to both procedure and analysis of voice acoustics, is rarely undertaken, despite calls for standardization and consensus. Titze (1994) put forth some arguments for the implementation of standardized recording techniques and analysis settings, stating that simplification of techniques will allow for a reduction in the number of processes involved. Titze also suggests that standardization can lead to conservation of resources. In this case, time, money and expertise.

In the current study 15680 analysis cycles used generic pitch range settings to obtain widely used acoustic measures like fundamental frequency in individuals with periodic speech signals. The measures obtained demonstrated that the use of appropriately selected, sex specific pitch settings yield comparable results to measures obtained from individualized analysis procedures. The acoustic data was analyzed in a fraction of the time, promising new opportunities for researchers to improve the accuracy and reliability of large studies investigating vocal acoustics while decreasing the labor and time required by current practices.

Acknowledgements

Data collection for this study was supported by a Small Business Innovation Research Grant from the National Institute of Mental Health (R43MH68950). Amy Fredrickson is acknowledged for her invaluable statistical advice and Paul Kukiel for his work on the analysis scripts.

References

- Atkinson IA, Kondo AM, Evans BG. Pitch detection of speech signals using segmented autocorrelation. *Electronics Letters* 1995;31(7):533–535.
- Bassich CJ, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. *Journal of Speech & Hearing Disorders* 1986;51(2):125–133. [PubMed: 3702360]
- Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sample sound. Institute of Phonetic Sciences, University of Amsterdam, *Proceedings* 1993;17:97–110.
- Boersma, P.; Weenink, D. Praat: doing phonetics by computer. Vol. Version 4.6.09. 2008.
- Cannizzaro MS, Harel B, Reilly N, Chappell P, Snyder PJ. Voice acoustical measurement of the severity of major depression. *Brain & Cognition* 2004;56(1):30–35. [PubMed: 15380873]
- Cannizzaro MS, Reilly N, Mundt JC, Snyder PJ. Remote capture of human voice acoustical data by telephone: A methods study. *Clinical Linguistics & Phonetics* 2005;19(8):649. [PubMed: 16147408]
- Carding PN, Carlson E, Epstein R, Mathieson L, Shewell C. Re: Evaluation of voice quality. *International Journal of Language & Communication Disorders* 2001;36(1):127. [PubMed: 11221429]
- Deliyski DD, Evans MK, Shaw HS. Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice* 2005;19(2):176–186. [PubMed: 15907432]
- Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice* 2005a;19(1):15–28. [PubMed: 15766847]
- Deliyski DD, Shaw HS, Evans MK. Influence of sampling rate on accuracy and reliability of acoustic voice analysis. *Logopedics, Phoniatrics, Vocology* 2005b;30(2):55–62.
- Eskenazi L, Childers DG, Hicks DM. Acoustic correlates of vocal quality. *Journal of Speech & Hearing Research* 1990;33(2):298–306. [PubMed: 2359270]
- Fette, B.; Gibson, R.; Greenwood, E. Windowing functions for the average magnitude difference function pitch extractor. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing; 1980.
- Gerhard, D. Pitch Extraction and Fundamental Frequency: History and Current Techniques. Regina, US: 2003. Technical Report No. 0 7731 0455 0
- Green JR, Beukelman DR, Ball LJ. Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. *Journal of Medical Speech-Language Pathology* 2004;12(4):149–154.
- Hirose K, Fujisaki H, Seto S. A scheme for pitch extraction of speech using autocorrelation function with frame length proportional to time lag. *Acoustics, Speech & Signal Processing* 1992;1:149–152.
- Hirst, D. Automatic Analysis of Prosody for Multi-lingual Speech Corpora. In: Keller, E.; Bailly, G.; Monaghan, A.; Terken, J.; Huckvale, M., editors. *Improvements in Speech Synthesis*. 2002. p. 320-327.
- John A, Sell D, Sweeney T, Harding-Bell A, Williams A. The Cleft Audit Protocol for Speech-Augmented: A Validated and Reliable Measure for Auditing Cleft Speech. *The Cleft Palate - Craniofacial Journal* 2006;43(3):272. [PubMed: 16681400]
- Karnell MP, Hall KD, Landahl KL. Comparison of fundamental frequency and perturbation measurements among three analysis systems. *Journal of Voice* 1995;9(4):383–393. [PubMed: 8574304]
- Karnell MP, Scherer RS, Fischer LB. Comparison of acoustic voice perturbation measures among three independent voice laboratories. *Journal of Speech & Hearing Research* 1991;34(4):781–790. [PubMed: 1956185]
- Kent RD. Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology* 1996;5(3):7–23.

- Kent, RD.; Read, C. *Acoustic Analysis of Speech*. Vol. 2nd ed.. Singular Thomson Learning; Albany, NY: 2002. p. 64
- Kent RD, Vorperian HK, Duffy JR. Reliability of the multi-dimensional voice program for the analysis of voice samples of subjects with dysarthria. *American Journal of Speech-Language Pathology* 1999;8(2):129–136.
- Mendoza E, Munoz J, Valencia Naranjo N. The long-term average spectrum as a measure of voice stability. *Folia Phoniatica et Logopaedica* 1996;48(2):57–64. [PubMed: 8765550]
- Morris RJ, Brown WS. Comparison of various automatic means for measuring mean fundamental frequency. *Journal of Voice* 1996;10(2):159–165. [PubMed: 8734391]
- Mueller PB. The aging voice. *Seminars in Speech & Language* 1997;18(2):159–168. [PubMed: 9195688]
- Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Geralts DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of Neurolinguistics* 2007;20(1):50–64.
- Perry CK, Ingrisano DR, Scott SR. Accuracy of jitter estimates using different filter settings on Visi-Pitch: A preliminary report. *Journal of Voice* 1996;10(4):337–341. [PubMed: 8943136]
- Rabiner LR. Use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics & Signal Processing* 1977;25(1):24–33.
- Scherer RC, Vail VJ, Guo CG. Required number of tokens to determine representative voice perturbation values. *Journal of Speech & Hearing Research* 1995;38(6):1260–1269. [PubMed: 8747819]
- Takagi T, Seiyama N, Miyasaka E. A method for pitch extraction of speech signals using autocorrelation functions through multiple window lengths. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 2000;83(2):67–79.
- Titze IR. The G. Paul Moore Lecture. Toward standards in acoustic analysis of voice. *Journal of Voice* 1994;8(1):1–7. [PubMed: 8167782]
- Titze IR, Liang H. Comparison of Fo extraction methods for high-precision voice perturbation measurements. *Journal of Speech & Hearing Research* 1993;36(6):1120–1133. [PubMed: 8114479]
- Zraick RI, Wendel K, Smith-Olinde L. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *Journal of Voice* 2005;19(4):574–581. [PubMed: 16301103]

Table 1
Acoustic measures obtained using gold standard case-by-case pitch settings.

	Males (N=10)		Females (N=10)		P value
	M	SD	M	SD	
f0 Mean	114.74	25.79	170.19	20.51	<0.0001
f0 SD	7.26	6.46	11.37	10.54	<0.0001
f0 CV	0.06	0.05	0.07	0.07	0.2854

Table 2

Actual estimated female means (+/- SD) obtained using preset pitch ranges and actual estimated gold standard settings.

Task	Measure	Gold Standard	50-250	50-300	50-500	50-600	70-250	70-300	70-500	70-600	70-625	100-250	100-300	100-500	100-600	
GRAN	F0		165.08	167.14	169.72 [†]	171.51 [†]	165.64	167.70	170.15 [†]	171.97 [†]	172.45 [†]	170.97 [†]	173.13 [†]	175.29 [†]	177.19 [†]	
			**	(18.78)	(19.21)	(18.85)	(15.97)	(17.68)	(18.11)	(17.61)	(17.63)	(12.3)	(14.25)	(14.64)	(13.93)	
			(17.07)	d=0.48	d=0.06	d<0.00	d=0.77	d=0.37	d=0.02	d<0.00	d<0.00	d=0.00	d<0.00	d=0.00	d=0.00	d=0.15
			d=0.89	1	1	1	1	1	1	1	1	2	1	1	2	
GRAN	F0 sd		28.86 ^{**}	30.14 [*]	37.44 [*]	45.03 [*]	27.61	28.96 ^{**}	36.41 [*]	44.49 [*]	46.57 [*]	20.85 [†]	22.31 [†]	30.59 [*]	39.96 [*]	
			(8.59)	(8.79)	(10.22)	(13.93)	(7.93)	(8.27)	(9.15)	(12.35)	(13.81)	(5.31)	(6.07)	(7.75)	(12.32)	
			d=1.58	d=2.14	d=2.14	d=2.14	d=1.08	d=1.63	d=2.14	d=2.14	d=2.14	d=2.14	d=0.02	d<0.00	d=2.14	d=2.14
														1		
GRAN	F0 cv		0.18 [*]	0.18 [*]	0.23 [*]	0.27 [*]	0.17 ^{**}	0.18 [*]	0.22 [*]	0.26 [*]	0.27 [*]	0.12 [†]	0.13 [†]	0.18 [*]	0.23 [*]	
			(0.07)	(0.07)	(0.08)	(0.1)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.03)	(0.03)	(0.04)	(0.07)	
			d=2.0	d=2.14	d=2.14	d=2.14	d=1.4	d=1.77	d=2.14	d=2.14	d=2.14	d<0.00	d<0.00	d=1.75	d=2.14	
													1	1		
FREE	F0		154.64 [*]	156.08 [*]	158.1 ^{**}	159.35 [*]	156.92 [*]	158.44 ^{**}	160.39 [†]	161.61	161.89	163.97 [†]	165.52 [†]	167.48	168.69	
			(19.84)	(21.18)	(22.08)	(16.81)	(18.24)	(18.77)	(19.3)	(18.73)	(18.73)	(14.31)	(15.63)	(16.53)	(16.09)	
			d=2.1	d=2.1	d=1.32	(21.86)	d=1.79	d=1.2	d=0.56	d=0.13	d=0.2	d<0.00	d<0.00	d=0.3	d=0.67	
					d=0.89								1	1		

Task	Measure	Gold Standard	50-250	50-300	50-500	50-600	70-250	70-300	70-500	70-600	70-625	100-250	100-300	100-500	100-600	
FREE	F0 sd	19.60 (7.46)	28.15*	29.40*	35.52*	41.09*	25.92*	27.23*	33.79*	39.47*	40.96*	17.69 [†]	19.15 [†]	26.88*	33.53*	
			(11.74)	(11.53)	(13.13)	(16.22)	(10.25)	(10.4)	(11.8)	(15.27)	(15.78)	(5.89)	(6.65)	(10.52)	(14.58)	
			d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=0.00	d=2.14	d=2.14
														1		
FREE	F0 cv	0.12 (0.06)	0.19*	0.2*	0.23*	0.27*	0.17*	0.18*	0.22*	0.25*	0.26*	0.11 [†]	0.11 [†]	0.16*	0.2*	
			(0.1)	(0.1)	(0.11)	(0.12)	(0.08)	(0.08)	(0.11)	(0.11)	(0.11)	(0.11)	(0.03)	(0.03)	(0.06)	(0.09)
			d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=0.00	d=0.00	d=2.14	d=2.14
														2		
Vowel	F0	174.10 (22.48)	166.93	165.89	171.68 [†]	168.06	167.41	168.66	168.72	168.72	168.72	172.32 [†]	173.37 [†]	173.46 [†]	173.50 [†]	
			(29.12)	(31.15)	(27.95)	(30.15)	(26.87)	(28.08)	(28.07)	(28.08)	(28.07)	(28.07)	(21.73)	(22.84)	(22.88)	(22.84)
			d=1.09	d=1.53	d=0.01	d=0.8	d=0.8	d=0.67	d=0.63	d=0.63	d=0.63	d=0.63	d=0.00	d<0.001	d<0.001	d<0.001
					7								3		1	
Vowel	F0 sd	2.05 (1.48)	7.39*	7.55*	7.24*	7.45*	7.44*	7.36*	7.42*	7.42*	7.42*	2.69 [†]	2.68 [†]	2.89 [†]	3.13 [†]	
			(12.76)	(12.88)	(13.06)	(13.11)	(12.52)	(12.43)	(12.65)	(12.65)	(12.65)	(5.17)	(5.08)	(5.9)	(7.04)	
			d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d<0.001	d<0.001	d=0.01	d=0.06
													1	1		

Task	Measure	Gold Standard	50-250	50-300	50-500	50-600	70-250	70-300	70-500	70-600	70-625	100-250	100-300	100-500	100-600
Vowel	F0 cv	0.1 (0.01)	0.05*	0.05*	0.05*	0.05*	0.05*	0.05*	0.05*	0.05*	0.05*	0.02 [†]	0.02 [†]	0.02 [†]	0.02 [†]
			d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d<0.00	d<0.00	d=0.00	d=0.02

Preset pitch ranges that differed significantly from gold standard measures

GRAN = Grandfather passage; FREE = extemporaneous speech; Vowel = /a:/, /i:/, /u:/, /ae:/, /e:/; sd = standard deviation; cv = coefficient of variation; d = Cohen's d (effect size).

- * p<0.001
- ** p<0.01
- *** p<0.05
- [†] p>0.5

Table 3

Actual estimated male means (\pm SD) obtained using preset pitch ranges and actual estimated gold standard settings.

Task	Measure	Gold Standard	50-250	50-300	50-500	50-600	70-250	70-300	70-500	70-600	70-625	100-250 [^]	100-300 [^]	100-500 [^]	100-600 [^]
GRAN	F0	114.93	114.94 [†]	115.40 [†]	121.74 [†]	127.06	115.70 [†]	116.10 [†]	121.47 [†]	126.41	127.98	128.07	128.95	154.07 [*]	167.29 [*]
		(23.71)	(24.15)	(24.29)	(24.4)	(25.09)	(23.54)	(23.68)	(23.23)	(23.36)	(23.23)	(23.23)	(15.41)	(15.25)	(46.36)
			d<0.001	d<0.001	d=0.1	d=0.64	d<0.001	d<0.001	d=0.08	d=0.57	d=0.74	d=0.75	d=0.85	d=2.14	d=2.14
GRAN	F0 sd	13.77	15.38 [†]	17.38 [†]	42.16 [*]	59.89 [*]	14.81 [†]	16.82 [†]	40.67 [*]	58.66 [*]	64.23 [*]	19.31 [†]	22.71	61.09 [*]	82.85 [*]
		(4.72)	(4.48)	(5.25)	(22.17)	(31.62)	(4.37)	(4.78)	(19.3)	(28.05)	(29.9)	(7.02)	(8.22)	(39.82)	(48.91)
			d<0.001	d=0.006	d=2.14	d=2.14	d<0.001	d=0.001	d=2.14	d=2.14	d=2.14	d=0.1	d=0.51	d=2.14	d=2.14
GRAN	F0 cv	0.12	0.14 [†]	0.15 [†]	0.35 [*]	0.47 [*]	0.13 [†]	0.15 [†]	0.34 [*]	0.47 [*]	0.51 [*]	0.15 [†]	0.18	0.38 [*]	0.48 [*]
		(0.03)	(0.04)	(0.04)	(0.18)	(0.23)	(0.03)	(0.04)	(0.17)	(0.22)	(0.23)	(0.06)	(0.07)	(0.18)	(0.2)
			d<0.001	d=0.08	d=2.14	d=2.14	d<0.001	d=0.02	d=2.14	d=2.14	d=2.14	d=0.07	d=0.58	d=2.14	d=2.14
FREE	F0	109.66	109.42 [†]	110.12 [†]	116.67	122.0 ^{**}	110.34 [†]	111.1 [†]	116.86	121.46 ^{**}	123.21 [*]	129.43 [*]	131.12 [*]	161.21 [*]	178.61 [*]
		(25.05)	(25.36)	(25.77)	(27.19)	(30.04)	(24.98)	(25.44)	(26.37)	(27.88)	(28.45)	(17.39)	(17.64)	(55.06)	(69.97)
			d<0.001	d<0.001	d=0.48	d=1.45	d<0.001	d<0.001	d=0.51	d=1.34	d=1.72	d=2.14	d=2.14	d=2.14	d=2.14
FREE	F0 sd	11.98	14.67 [†]	16.92	40.85 [*]	57.38 [*]	13.40 [†]	16.09 [†]	38.92 [*]	54.9 [*]	61.19 [*]	22.29 ^{**}	26.57 [*]	64.39 [*]	87.92 [*]
		(4.55)	(5.35)	(7.72)	(25.57)	(35.39)	(4.77)	(7.41)	(23.62)	(32.61)	(34.57)	(9.73)	(11.63)	(40.59)	(51.75)
			d=0.02	d=0.37	d=2.14	d=2.14	d<0.001	d=0.21	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14	d=2.14

Task	Measure	Gold Standard	50-250	50-300	50-500	50-600	70-250	70-300	70-500	70-600	70-625	100-250 [^]	100-300 [^]	100-500 [^]	100-600 [^]
FREE	F0 cv	0.11 (0.03)	0.14 [†] (0.05)	0.15 ^{***} (0.07)	0.35 [*] (0.19)	0.46 [*] (0.24)	0.12 [†] (0.12)	0.14 (0.06)	0.33 [*] (0.18)	0.45 [*] (0.23)	0.49 [*] (0.24)	0.17 [*] (0.07)	0.2 [*] (0.08)	0.38 [*] (0.18)	0.48 [*] (0.2)
			d=0.3	d=0.97	d=2.14	d=2.14	d=0.008	d=0.62	d=2.14	d=2.14	d=2.14	d=1.93	d=2.14	d=2.14	d=2.14
Vowel	F0	118.69 (26.27)	117.19 [†] (27.05)	117.24 [†] (27.04)	117.51 [†] (26.89)	117.85 [†] (26.94)	118.78 [†] (26.02)	118.82 [†] (26.0)	118.94 [†] (25.94)	119.24 [†] (25.89)	119.17 [†] (25.84)	128.01 ^{**} (22.93)	130.11 [*] (28.33)	132.06 [*] (32.32)	137.06 [*] (52.42)
			d<0.001	d<0.001	d<0.001	d<0.001	d<0.001	d<0.001	d<0.001	d<0.001	d<0.001	d=1.31	d=1.87	d=2.14	d=2.14
Vowel	F0 sd	1.9(2.49)	4.95 ^{***} (7.76)	5.18 ^{**} (8.14)	5.86 [*] (10.51)	6.73 [*] (15.28)	3.17 [†] (5.66)	3.42 [†] (6.16)	3.74 (7.65)	4.6 ^{***} (12.94)	4.52 (12.02)	2.47 [†] (4.39)	3.16 [†] (7.37)	4.26 (12.29)	5.67 ^{**} (20.68)
			d=1.17	d=1.34	d=1.87	d=2.14	d=0.1	d=0.23	d=0.41	d=0.94	d=0.89	d<0.001	d=0.09	d=0.68	d=1.62
Vowel	F0 cv	0.02 (0.02)	0.05 [*] (0.07)	0.05 [*] (0.08)	0.05 [*] (0.1)	0.06 [*] (0.12)	0.03 [†] (0.04)	0.03 [†] (0.05)	0.03 (0.06)	0.04 ^{***} (0.1)	0.04 ^{***} (0.1)	0.02 [†] (0.03)	0.02 [†] (0.04)	0.03 (0.08)	0.03 (0.1)
			d=1.87	d=2.14	d=2.14	d=2.14	d=0.13	d=0.29	d=0.51	d=1.04	d=1.04	d<0.001	d=0.008	d=0.39	d=0.78

Preset pitch ranges that differed from gold standard measures

GRAN = Grandfather passage; FREE = extemporaneous speech; Vowel = /a:/, /i:/, /u:/, /ae:/; sd = standard deviation; cv = coefficient of variation; d = Cohen's d (effect size)

* p<0.001

** p<0.01

*** p<0.05

† p>0.5

^ missing files as acoustic data was outside prescribed pitch range.