Research Article

A New Era in Acoustic Analysis: Use of Smartphones and Readily Accessible Software/ Applications for Voice Assessment

Cecyle K. Carson* and Jack Ryalls

Department of Communication Sciences and Disorders, University of Central Florida, USA

Abstract

Rapid advancements in technology have made the acoustic assessment of voice more convenient and less costly; thus, there are few reasons for speech language pathologists (SLPs) not to use acoustic measures to supplement perceptual ratings. Smartphones have been found to be comparable to external microphones in recording quality, and free software programs are available to download into computers to obtain acoustic analyses results. Suggestions on a protocol for capturing and analyzing voice signals using smartphones and computer freeware or smartphone applications will be provided.

ABBREVIATIONS

SLP: Speech Language Pathologist; OperaVOX: On Person Rapid Voice examiner; CPP(S): Cepstral Peak Prominence (smoothed); ADSV: Analysis of Dysphonia in Speech and Voice; MDVP: Multidimensional Voice Program

INTRODUCTION

A comprehensive voice assessment typically includes a case history, an oral-peripheral examination, an assessment of respiration, and the use of perceptual, acoustic, electro glottography, and aerodynamic data, in addition to endoscopic (and possibly stroboscopic) imaging results [1,2]. This evaluation involves at least a physician (preferably an otolaryngologist) and an SLP [3]. With an estimated 7.5 million Americans experiencing voice problems [4], the need for accurate and reliable estimates of voice production becomes evident.

Although perceptual assessment of voice is considered the gold standard in voice analysis [5,6], both intra-rater and inter-rater reliability issues remain problematic, with concerns such as, but not limited to, 1). The variability of judgments based on internal standards [7], 2.) Random errors due to factors such as inattentiveness or fatigue [8] 3) resolution of the rating scale (continuous versus equal appearing scale) [9], 4) speaking task [10]. The perceptual evaluation of voice will continue to remain a mainstay in assessment because voice is a perceptual phenomenon, and treatment outcomes are commonly determined by the sound of the patient's voice [5].

With the limitations previously discussed concerning perceptual ratings, it is not surprising that instrumental measures are commonly employed to augment and somewhat objectify perceptions. In particular, acoustic measures are attractive because they 1. Are noninvasive (especially important for young children); 2. Are readily available as open access, downloadable computer software (e.g., Praat, Paul Boersma and David Weenick; Institute of Phonetic Sciences, University of Amsterdam, The Netherlands, http://praat.org) or as low cost phone/tablet applications (e.g., OperaVOX (On Person Rapid Voice examiner), OwainRhyes Hughes and Anil Alexander; Oxford Research Wave Ltd, UK); and 3. Provide measures that are traditionally used and/or are newer to the field and have research to support their clinical application in diagnosis and treatment [11]. Acoustic analysis has been recommended as part of the European standard protocol for voice assessment [12]; however, a standardized protocol remains undefined in the United States [3]. This paper will describe research findings involving acoustic analysis and provide suggestions for capture and analysis of voice samples for SLPs (Speech Language Pathologists).

RESULTS AND DISCUSSION

The latest survey of voice assessment practices in the United States was published in 2005, and included 53 SLPs who had three or more years of recent experience in acoustic instrumentation and stroboscopy [13]. The survey was posted on a web-based discussion site sponsored by ASHAs Special Interest Division 3: Voice and Voice Disorders (SIG3). According to Behrman [13], at the time the survey was posted to solicit responses, there were over 1000 subscribed members of SIG3. Results found that all

53 respondents used clinical estimations of voice quality and patient self-perceptions for patients who were referred due to muscle tension dysphonia. Instrumental assessment "likely" to be used by participants included stroboscopy (81%) and acoustic measurements (75%). Of particular concern was the finding that 60% of respondents indicated that they would not modify the type of acoustic analysis performed when confronted with a voice that was highly irregular or highly dysphonic, even though the use of time-based acoustic perturbation measures (e.g., jitter and shimmer) has been considered unreliable in voices that are considered moderately to severely dysphonic [14,15]. Perturbation measures are estimates of the periodicity of the voice signal [16]. In 1995, Titze [17] identified three types of voice signals. He recommended the use perturbation measures for type 1 voice signals which were defined as nearly periodic and displayed clearly identifiable harmonics via a narrowband spectrogram [17,18]. According to Titze [17], when the voice signal contained modulations (subharmonics) or rapid qualitative changes (type 2), or if the signal appeared to have no consistent pattern of vibration (aperiodic) (type 3), then spectrographic and perceptual analysis were suggested. In contrast, the use of perturbation measures and signal-to-noise ratio has been recommended for type 1 and some type 2 signals [18]. Unfortunately, most dysphonic voices are type 2 and 3 signals [19]; therefore, these types of signals may not lend themselves to the accurate delineation of cycle-to-cycle boundaries necessary for time-based algorithms [20], which can limit their clinical usefulness.

A consideration in the accuracy of jitter and shimmer estimates relates to the way in which fundamental frequency and cycle amplitudes are determined. Of the two basic types of methods for establishing fundamental frequency, short-term averaging or wave-form matching has been reported to be more accurate than event-detection (peak-picking or zero-crossing) methods [16]. Thus, software/apps, such as Praat, that use algorithms based on wave-form matching may yield more accurate perturbation measures.

More recently, the use of cepstral/spectral algorithms has received a lot of attention in that the measures yielded from these algorithms are not reliant on accurate computation of fundamental frequency [21], and thus aren't prone to problems associated with time-based perturbation measures. One cepstral/spectral estimate that has received a lot of support in the literature is Cepstral Peak Prominence (CPP). This measure is based on the idea that periodic voices should display prominent harmonic peaks in the spectrum [22]. CPP is derived first by applying a Fourier transformation of the voice waveform, changing the representation to frequency based from time based; in other words, the amplitude of the signal is graphed across frequencies (termed spectrum), as opposed to amplitude across time. Then, a Fourier transformation of the spectrum is performed, creating a cepstrum. Prominent harmonics emerge from the noise on the cepstrum as elevated peaks (i.e., increased magnitude in dB). A regression line or line of best fit is created across all frequencies and the dB value located below the highest peak or Cepstral peak is subtracted from the magnitude (in dB) of the Cepstral peak to yield CPP [23]. See Heman-Ackah et al. [23], for more information and graphic displays. CPP been considered to be a very robust objective measure of the perceived severity of dysphonia and perhaps the most promising single measure [22]. Further, CPP has been found to have robust sensitivity and specificity in discriminating between those with normal voices from those with dysphonia [23], and has been recommended for use with voices in the moderate to severe range [22]. One further advantage of using CPP, or any other measure from a cepstral/spectral analysis, is that both vowels and connected speech can be used, in contrast to the analysis of only prolonged vowels imposed when measuring jitter, shimmer, and signal-to-noise ratio [24]. It should be noted that different formulas are used for computing CPP and produce different magnitudes of values. Importantly, values yielded from Praat using a smoothing formula for CPP (termed CPPS) were found to be highly correlated with CPPS results derived from the Analysis of Dysphonia in Speech and Voice (ADSV, PENTAX MEDICAL, Montvale, NJ) [25]. ADSV is a popular software program that yields CPPS (termed as CPP in ADSV) as well as other cepstral/spectral measures, and has numerous research support for its use clinically [24-27]. Note that the CPPS values cited in Watts et al. [25], required adjustments to the automatic procedures available in Praat and ADSV [25]. In another comparison of different methods of calculation, CPPS measures yielded from Praat and Speech Tool were reported to be equivalent [26]; thus, either program could be used to obtain a CCPS value. Due to the use of various algorithms in yielding CCP and CCPS values in software programs, it is recommended that the same software be utilized for repeated intra-client measurements [25].

Rapid advancements in technology have made the acoustic assessment of voice more convenient and less costly; thus, there are few reasons for not applying these measures to supplement perceptual ratings [6]. Free or low cost downloadable software programs for voice analysis via computer are available (e.g., Praat, SpeechTool) to obtain acoustic data that include spectrograms, time-based estimates (e.g., fundamental frequency, jitter, shimmer, harmonic-to-noise ratio) and may include cepstral/spectral measures. Applications (apps) for capturing voice signals on smartphones have proven to yield similar perturbation results when compared to digitized signals input directly into a computer (iPhone model A1303) [28] and Samsung Galaxy Note 3 [29]. A recent study compared perturbation measures yielded from Praat when recorded on an expensive smartphone (HTC One), an inexpensive smartphone (Wiko model CINK SLIM2) and direct recording via a microphone (Sennheiser model MD421U) [29]. Results revealed high correlations among the perturbation measures among the three recording devices. Based on research evidence, waveforms recorded via some smartphones (e.g., iPhones series 3 upward [28] have been deemed suitable to use for eventual acoustic voice analysis.

Apps for acoustic analysis on smartphones are currently receiving a lot of attention. Some provide specific real-time information about pitch and volume (e.g., Voice Analyst), pitch, loudness, and phonation time (Loudness: Sonneta Voice Monitor), or display the sound level in the environment (NIOSH Sound Level Meter). Recently, an app for iPhones, iPads, and iPod Touch has been developed and marketed under the name of OperaVOX (On Person RApidVOiceeXaminer, OperaVOX Ltd.). The app captures voice recordings at a rate of 44.1 KHz (number

Carson et al. (2018)
Email: cecyle.carson@ucf.edu

⬤SciMedCentral

of "snapshots" taken on a waveform per second) with a 16-bit quantization rate [30]. With this sample rate, frequencies up to 20 kHz are delineated (remember that the limit of human hearing is 20 kHz) [30]. OperaVOX provides a microphone icon for clients to view to help determine the appropriateness of the recording environment and the minimum vocal amplitude necessary for the recording. If too much ambient noise is detected, no recording can occur. The user is provided with visual instructions to provide 1). A sample of a prolonged /ah/ sound 2). Three attempts at maximum phonation time while prolonging a vowel with the maximum time saved, 3). A recording of a supplied standard reading passage (The Rainbow Passage). 4). A singing sample sliding from lowest to highest fundamental frequencies to estimate pitch range. Perturbation analyses are automatically generated on OperaVOX for the client to view, and recordings are also stored for future analysis by an SLP. Perturbation measures derived via OperaVOX have been found to be equivalent to those same measures when obtained from Praat [31] and consistent with values from a popular software program, Multidimensional Voice Program (MDVP, PENTAX Medical, Montvale, NJ), for the measures of fundamental frequency, jitter, and shimmer (but not comparable for noise-to-harmonic ratio) [32]. OperaVOX appears to be a viable choice for SLPs; it is very user friendly for clients and SLPs. The only limiting factors are the lack of cepstral/spectral measures for those voices that are moderately to severely dysphonic, limited spectrographic display with no ability to control bandwidth, etc., and acoustic measures (perturbations) that are only available for the prolonged vowel sample.

Portable devices for capture and analysis of voice signals allow ongoing longitudinal tracking of treatment protocols and provide a more accurate representation of real-life voice production during periods where extra load may be placed on the vocal mechanism (e.g., lecturing, singing, at the end of the day), or when monitoring changes occurring in progressive disorders. These advantages are in contrast with the typical one-time snapshot taken during an initial comprehensive evaluation for baseline data, with evaluation of the effects of medical or behavioral treatment limited to clinic visits [28]. Further, voice samples recordings have been proposed for voice screening purposes [29].

Even though presently there is not a set of standard methods recommended for voice assessment and ongoing tracking of treatment protocols in the U.S., the appropriateness of subjective and objective measures should continue to be an integral part of a comprehensive voice evaluation [3]. SLPs who serve clients with voice problems, but are not part of a medical practice or research facility, may not have access to costly recording and analysis equipment. However, there are some new, low cost/no cost opinions that can provide relatively accurate objective acoustic data. A cellphone could be used to record a voice sample at various times during the day or week, allowing sequential monitoring with the client following a standardized capture procedure (e.g., quiet room, specified mouth-to-microphone distance, etc.) [29]. The client would save the sample and email it to the SLP for analysis and possible feedback. Because there are many different makers and models of cellphones marketed, the SLP must make sure that the sample rate is at least 22.5K,

based on the sample rate limit for signals input into ADSV, and taking into account the Nyquist theorem on sampling rates. The information about sample rates is readily available by searching the web for the recording sample rate of a particular cellphone make and model. For example, iPhone models 3-6 sample at a rate of 48K, which is appropriate for capturing speech. Once receiving the wave file, the SLP could download it into a computer software program that analyzes speech signals. Praat appears to be a good computer software choice because it not only provides traditional time-related perturbation measures (e.g., jitter, shimmer, signal-to-noise ratio) using a preferred algorithm [16] and narrow band spectrograms for use with prolonged vowels, but also yields spectral/cepstral measures (e.g., CPP, CPPS) that can be used for analyzing running speech. Other attractive attributes of Praat are, it: 1). Is freeware and has automatic analyses, 2). Is downloadable to different computer operating platforms (e.g., Windows, Linux) [26]. Has instruction manuals available from different sources, 4). Has many apps that are downloadable, 5). Is very popular with researchers, and thus has published support for its applications, and 6). Does not have a restriction placed on capture rate. One limiting factor with Praat is that it may require time to learn its many functions, as it is not intuitive to novice users.

CASE EXAMPLE

Father Joe was referred for voice therapy due to hyper functional voice usage, resulting in what was perceived as a rough voice with noticeable strain. The referring otolaryngologist reported that the Father's vocal folds appeared normal, although they were difficult to visualize due to medial movement of the false folds with the true folds, which partially obscured viewing. Father reported that his voice became increasingly "hoarse" through the week, beginning with the evening mass on Saturday and the subsequent two masses on Sunday. He stated that his voice was normal when he had a break from his duties (e.g., during vacations).

Baseline data were collected during the diagnostic session, and included acoustic measures taken from samples using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) [33]. The CAPE-V requires the individual to sustain the vowels /ah/ and /ee/ for 3-5 seconds, read or repeat six sentences, and response to the question, "Tell me how your voice is functioning." The voice samples were analyzed using Praat as follows: 1). Narrowband spectrograms were used to on the two vowel samples to determine the type of voice signal [17-18]; and 2). CPP, shimmer, jitter, and fundamental frequency measures were extracted from the vowel samples, and 3). CPP values were generated from each of the six sentences for baseline purposes. During the following week, Father Joe was instructed to make short (one minute) recordings of his voice before and after mass, and during some of his usual weekly activities. All samples were to be recorded in a quiet environment (a room with the door closed and with minimal background noise) onto the Father's iPhone using the Voice Memos app, saved, and then emailed to the SLP for analysis. A mouth-to-microphone distance of 10 cm was used, and measured via a small piece of string provided during the assessment session. The phone was held at a 90degree angle from the microphone so that the client could use the start

SciMedCentral

and stop buttons on the screen and also view the length of the recording. This procedure allowed a relatively standardized capture routine for acquiring baseline data over several time periods, rather than be forced to assume that the samples taken during the initial voice assessment represented the client's typical voice. It is important, especially for professional voice users, to obtain data after periods of intense vocal use [34]. Treatment involved direct (physiologic methods) and indirect (counseling, education, facilitating techniques) approaches, with continued samples collected weekly with special focus on periods following strenuous vocal activities to document progress, along with perceptual.

CONCLUSION

In conclusion, SLPs are no longer limited in obtaining objective acoustic measures of voice to supplement perceptions due to technological advancements in recording devices and analysis systems. Even though there remains a lack of minimum standards for software and hardware requirements for acoustic analysis [35], results from research cited in this paper provide some suggestions. Smartphones have proven to be comparable to external microphones in recording voice signals from those with normal voices [36] and those with dysphonia of different levels of severity [29] and various etiologies [28,29]; and thus, they are suitable for use with voice analysis apps or software that yield acoustic measures, if the sample capture rates are at or above 25 kHz [37]. Minimum recording standards (e.g., mouth-to-microphone distance, maximal sound pressure level for recording) should be established. The mouth-to-smartphone microphone distance has varied from a minimum of 4 cm [36] to a maximum of 30 cm [32], and in between (10 cm [29]; 13 cm [28]. Research is needed to create a recommended range. Further, instructions must be provided to clients so that the quality of the voice capture is maintained across differing environments or settings in terms of environmental noise which affects perturbation measures by increasing the values (i.e., making the voice appear "more" dysphonic [38-40]. Free or low cost apps and software programs are readily available to download waveform files emailed by clients. Acoustic analyses typically include perturbation measures and narrowband spectrograms for visual interpretation. Some programs also provide cepstral/spectral measures such as CPP and CPPS, plus other multivariate measures (e.g., Acoustic Voice Quality Index [26]). When using an acoustic analysis software program or app, the rule is that voices that are moderately or severely dysphonic should not be analyzed utilizing time-based measures; instead, narrowband spectrograms and cepstral/spectral measures are recommended. Data collected should be compared on an intra-client basis [25], rather than inter-client, because we want to determine if change has occurred during/after surgical or behavioral treatment. Future trends will likely witness further development and refinement of apps for smartphones that analyze captured signals on-device.

REFERENCES

1. Voice Disorders: Assessment. American Speech-Language-Hearing Association.

2. Mehta DD, Hillman RE. Voice assessment: updates on perceptual, acoustic, aerodynamic, and endoscopic imaging methods. Current Opinion in Otolaryngology and Head and Neck Surgery. 2008; 16: 211–215.

3. Roy N, Barkmeier-Kraemer J, Eadie T, Sivesandar, Mehta D, Paul D, et al. Evidence-based clinical voice assessment: A systematic review. American Journal of Speech-Language Pathology. 2013; 22: 212-226.

4. What is Voice? What is Speech? What is Language?. National Institute on Deafness and Other Communication Disorders. 2003.

5. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research. J Speech Hear Res. J Speech Hear Res. 1993; 36: 21-40.

6. Oates J. Auditory-perceptual evaluation of disordered voice quality. Folia Phoniatrica et Logopaedica. 2009; 61: 49-56.

7. Gerratt BR, Kreiman J, Antonanzas-Barroso N. Berke GS. Comparing internal and external standards in voice quality judgments. J Speech Hear Res. 1993; 36: 14-20.

8. Shrivastav R, Sapienza C, Nandur V. Application of psychometric theory to the measurement of voice quality using rating scales. J Speech Hear Res. 2005; 48: 323–335.

9. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. J Acoust Soc Am. 2007; 122: 2354-2364.

10. Lu FL, Matteson S. Speech tasks and interrater reliability in perceptual voice evaluation. Journal of Voice. 2014; 28: 725-732.

11. Awan SM. The voice diagnostic protocol: A practical guide to the diagnosis of voice disorders. Austin, TX.: Pro-Ed, Inc; 2011.

12. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). European Archives of Otorhinolaryngology. 2001; 258: 77-82.

13. Behrman A. Common practices of voice therapists in the evaluation of patients. J Voice. 2005; 19: 454-469.

14. Rabinov CF, Kreiman J, Gerratt BR, Bielamowicz S. Comparing reliability of perceptual ratings of roughness and acoustic measure of jitter. J Speech Hear Res. 1995; 38: 26-32.

15. Bielamowicz S, Kreiman J, Gerratt B, Dauer M, Berke G. Comparison of voice analysis systems for perturbation measurement. J Speech Hear Res. 1996; 9: 126-134.

16. Titze IR, Liang H. Comparison of Fo extraction methods for high-precision voice perturbation measurements. J Speech Hear Res. 1993; 36: 1120-1133.

17. Titze IR. Workshop on acoustic voice analysis: Summary statement. National Center for Voice and Speech. The University of Iowa, Ames, IA. 1995.

18. Sprecher A, Olszewski A, Jiang J. Updating signal typing in voice: Addition of type 4 signals. The J Acoust Soc Am. 2010; 127: 3710-3716.

19. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice. J Speech Lang Hear Res. 2000; 44: 327-339.

20. Awan SM, Roy N. Toward the development of an objective index of dysphonia severity: A four-factor acoustic model. Clin Linguist and Phon. 2006; 20: 35-49.

21. Sapienza C, Hoffman Ruddy B. Evaluation. In: Voice disorders. 3rd ed. Plural Publishing; 2018.

22. Maryn Y, Roy N, DeBolt M, Van Cauwenberge P, Corthals P. Acoustic

measurement of overall voice quality: A meta-analysis. J Acoust Soc Am. 2009; 126: 2619-2634.

23. Heman-Ackah YD, Michael DD, Baroody MM, Ostrowski R, Hillenbrand J, Heuer RJ, et al. = Cepstral peak prominence: A more reliable measure of dysphonia. Ann Otol Rhinol Laryngol. 2003; 112: 324-333.

24. Awan SN. Analysis of dysphonia in speech and voice (ADSV): An application guide. Montvale, NJ: PENTAX; 2011.

25. Watts CR, Awan SN, Maryn Y. A comparison of cepstral peak prominence measures for two acoustic analysis programs. J Voice. 2017; 31: 387.e1-387. e10.

26. Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: Smoothed cepstral peak prominence and acoustic voice quality index. J Voice. 2015; 29: 35-43.

27. Watts CR. The effect of CAPE-V sentences on cepstral/spectral acoustic measures in dysphonic speakers. Folia Phoniatr Logop. 2015; 67: 15-20.

28. Lin E, Hornibrook J, Ormond T. Evaluating iPhone recordings for acoustic voice assessment. Folia Phoniatr Logop. 2012; 64: 122-130.

29. Uloza V, Padervinskis E, Vegiene A, Pribuisene R, Saferis V, Vaiciukynas E, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameter and voice pathology screening. Eur Arch Otorhinolaryngol. 2015; 272: 3391-3399.

30. Sample rates.

31. Rhys Hughes O, Alexander A. Developing OperaVOX – the on person rapid voice examiner app. ENT News (November). 2013.

32. Mati-Baki M, Wood G, Alston M, Ratcliffe P, Sandhu G, Rubin JS, et al.

Reliability of OperaVOX against Multidimensional Voice Program. Clin Otolaryngol. 2014; 40: 22-28.

33. Kempster GB, Gerratt BR, Verdolini-Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. American Journal of Speech-Language Pathology. 2009; 18: 124-132.

34. Grillo EU, Fugowski JM. Voice characteristics of female physical education student teachers. J Voice. 2011; 25: 149-157.

35. Vogel AP. Maruff P. Comparison of voice acquisition methodologies in speech research. Behav Res Methods. 2008; 40: 982-987.

36. Grillo EU, Brosious JN, Sorrell SL, Supraja, A. Influence of smartphones and software on acoustic voice measures. Int J Telerehabil. 2016; 8: 9-14.

37. Deliyski DD, Shaw HS, Evans MK. Influence of sampling rate on accuracy and reliability of acoustic voice analysis. Logoped Phoniatr Vocol. 2005; 30: 55-62.

38. Ingrisano DR-S, Perry CK, Jepson KR. Environmental noise: a threata to automatic voice analysis. American Journal of Speech-Language Pathology. 1998; 7: 91-96.

39. Perry CK, Ingrisano DR, Palmer MA, McDonald EJ. Effects of environmental noise on computer-derived voice estimates from female speakers. J Voice. 2000; 14: 146–153.

40. Carson CP, Ingrisano DR, Eggleston KD. 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. J Voice. 2003; 17: 12–20.