Mobile Communication Devices, Ambient Noise, and Acoustic Voice Measures

*,†,‡Youri Maryn, §Femke Ysenbaert, *Andrzej Zarowski, and *Robby Vanspauwen, *‡*Antwerp and* †*Ghent, Belgium, and* §*Utrecht, The Netherlands*

Summary: Objectives. The ability to move with mobile communication devices (MCDs; ie, smartphones and tablet computers) may induce differences in microphone-to-mouth positioning and use in noise-packed environments, and thus influence reliability of acoustic voice measurements. This study investigated differences in various acoustic voice measures between six recording equipments in backgrounds with low and increasing noise levels.

Methods. One chain of continuous speech and sustained vowel from 50 subjects with voice disorders (all separated by silence intervals) was radiated and re-recorded in an anechoic chamber with five MCDs and one high-quality recording system. These recordings were acquired in one condition without ambient noise and in four conditions with increased ambient noise. A total of 10 acoustic voice markers were obtained in the program *Praat*. Differences between MCDs and noise condition were assessed with Friedman repeated-measures test and posthoc Wilcoxon signed-rank tests, both for related samples, after Bonferroni correction.

Results. (1) Except median fundamental frequency and seven nonsignificant differences, MCD samples have significantly higher acoustic markers than clinical reference samples in minimal environmental noise. (2) Except median fundamental frequency, jitter local, and jitter rap, all acoustic measures on samples recorded with the reference system experienced significant influence from room noise levels.

Conclusions. Fundamental frequency is resistant to recording system, environmental noise, and their combination. All other measures, however, were impacted by both recording system and noise condition, and especially by their combination, often already in the reference/baseline condition without added ambient noise. Caution is therefore warranted regarding implementation of MCDs as clinical recording tools, particularly when applied for treatment outcomes assessments.

Key Words: Voice–Acoustic analysis–Ambient noise–Mobile recording systems.

INTRODUCTION

Acoustic analysis and documentation of recorded speech signals is among the most frequently used clinical voice assessment methods[1] and covers the most often studied voice measuring tool category.[2] Typical statistics on the acoustic voice signal are fundamental frequency (f0; ie, as a measure related to vocal pitch) and sound intensity (ie, as a measure related to vocal loudness), as well as for example jitter, shimmer, various ratios between the spectral amplitudes of harmonic and noise energy, and first harmonic emergence (ie, as measures related to vocal sound quality). Such measures have been considered especially interesting in the voice clinic because of their relation to vocal fold physiology and pathology, noninvasiveness, ease of application, relatively low cost, and quantitative output as basis for theories on physiological/vocal phenomena (eg, vocal vibration periods, amplitudes, and regularities).[3] Over 100 algorithms for acoustic analysis of the voice signal have been developed and described in scientific literature, as tabulated by Buder.[4]

However, the level of noise present in the recording environment has proven to be one of many factors that affect accuracy, reliability, and validity of at least some of these acoustic estimates. Such environmental noise includes all surrounding signals that contaminate the direct speech signal before being captured by the microphone. It concerns signals that are not meant to be recorded and analyzed and not relevant in the clinical assessment of voice; for example, computer fan noise and air-conditioning noise. With computer analysis systems being commonly available and having penetrated nonresearch clinical settings, especially with mobile computer and communication devices (ie, smartphones and tablet computers) equipped to record sound, voice samples may not be recorded in most recommendable circumstances.

Environmental noise

The influence of surrounding noise on the outcome of acoustic voice measures is generally investigated by (1) determining the average sound intensity level of the voice/speech signal (ie, the "signal" or "S"), (2) determining the average sound intensity level of the environmental noise in the absence of speech (ie, the "noise" or "N"), and (3) subtracting the noise level from the speech signal level to obtain the so-called "signal-to-noise ratio" (SNR). Ingrisano et al[5] found that jitter and shimmer from *KayPENTAX*'s *Multi-Dimensional Voice Program* (*MDVP*; KayPENTAX Corp., Lincoln Park, NJ) increased in conditions with elevated noise, both in one synthesized and in one normal male vocal signals. f0 on the other hand was hardly affected by changes in SNR. Perry et al[6] elaborated on that study and included voice recordings of five women with normophonia, one

Accepted for publication July 26, 2016.

Conflict of interest: The authors declare that they do not have conflicts of interest.

From the *Department of Otorhinolaryngology and Head and Neck Surgery, European Institute of ORL, Sint-Augustinus General Hospital, Antwerp, Belgium; †Department of Speech-Language Pathology and Audiology, University College Ghent, Ghent, Belgium; ‡Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; and the §Center of Paramedical Studies, Utrecht University College, Utrecht, The Netherlands.

Address correspondence and reprint requests to Youri Maryn, Department of Otorhinolaryngology and Head and Neck Surgery, European Institute for ORL, Sint-Augustinus General Hospital, Oosterveldlaan 24, Wilrijk, Antwerp 2610, Belgium. E-mail: youri.maryn@gza.be

Journal of Voice, Vol. 31, No. 2, pp. 248.e11–248.e23

0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.07.023

man with normophonia, and one woman with mild dysphonia. They also found that two of *MDVP*s perturbation estimates (ie, relative average perturbation [RAP] and amplitude perturbation quotient [APQ]) almost systematically increased per addition of computer fan noise, whereas f0 remained relatively stable across SNR conditions. Furthermore, Carson et al[7] expanded these studies with 10 women with normophonia and found that f0 changed significantly at SNR = 15 dB, whereas RAP changed significantly at SNR = 15 dB or SNR = 20 dB, depending on the computer system. All comparisons of APQ data across SNR levels showed significant changes, irrespective of computer system. Finally, Deliyski et al[8] examined the impact of environmental noise on f0 and various jitter and shimmer markers of 20 participants with normophonia (10 men and 10 women) while controlling for gender, age, intersubject variability, intrasubject variability, microphone, computer hardware, software, and noise type. Their findings indicated SNR < 30 dB to be unacceptable, SNR ≥ 30 dB to be acceptable, and SNR ≥ 42 dB to assure reliable and valid measures with relative error within 1%.

From these studies, involving altogether 38 subjects with normophonia and only 1 subject with dysphonia, it can be concluded that voice perturbation measures fail to retain accuracy and reliability when SNR drops below certain levels, and consequently that the clinician should strive to optimal recording conditions when determining such estimates.

Mobile communication devices

As of August to September 2015, adoption of tablet computers in the Flanders market reached 58.3%, whereas smartphones penetrated 68.5%. Google Android and Apple iOS are the most frequently used platforms, with 52.8% and 31.9% for smartphones, and with 38.4% and 50.3% for tablets, respectively.[9] These "mobile devices" carry a multitude of built-in sensors (eg, microphone, camera, global positioning system receiver, accelerometer, and light sensor) and have numerous measurement software applications (ie, apps) installed. With built-in microphone and sound-related apps, these mobile devices can be used to record, store, and even analyze sound signals. Clinical use of such mobile communication devices (MCDs) in acoustic voice and speech assessment protocols is therefore hypothesized to grow among voice and speech clinicians.

However, how different are the outcomes for acoustic measures when different mobile data acquisition systems and acoustic analysis programs are used? In other words, are MCDs as operationally potent as the more traditional nonmobile recording systems? Before addressing this question, however, it is interesting to probe what can be learned from the literature on variability between nonmobile clinical desktop/laptop recording systems and analysis software. Karnell et al[10] compared jitter and shimmer results from three computer systems (ie, *Voice Analysis Program* on Kay Elemetrics 5500 DSP SonaGraph, *CSpeech* on Zenith Z-200 computer, and *AUDED/SEG program* on a Digital LSI 11–23 computer). They found correlation coefficients ranging from 0.29 to 0.64 for jitter and from 0.26 to 0.75 for shimmer. There were significant differences between various systems, and these authors concluded that the perturbation programs clearly do not result in comparable outcomes. Bielamowicz et al[11] also compared perturbation measures of four computer programs (ie, *CSpeech 4.0*, *Computerized Speech Laboratory* of Kay Elemetrics, *SoundScope 1.09*, and an interactive hand-marking program). Correlation coefficients are as follows: 0.33–0.80 for jitter, 0.81–0.89 for shimmer, and 0.23–0.81 for HNR. Statistically significant differences were found for several measures. For shimmer, they concluded that there is reasonable interprogram reliability across different severity levels. For jitter and HNR, however, much less reliable results were obtained across computer programs. Comparing three commercial software packages for clinical voice analysis (ie, *MDVP 4305* vs *Computerized Speech Lab 4300B*, *MDVP 4305* vs *Multi Speech 3700*, and *CSpeech*), Carson et al[7] found no significant difference in RAP nor APQ under conditions of optimal SNR ratio. However, with increasing ambient noise, differences became significant, especially for the shimmer data. Smits et al[12] obtained voice samples with a Sony TCD-D100 DAT (Sony Corp., Tokyo, Japan) recorder and compared jitter, shimmer, and HNR derived from two computer programs (ie, *Dr. Speech* of Tiger Electronics, and *MDVP 4305* of Kay Elemetrics; Tiger DRS, Inc., Seattle, WA). They found correlations of 0.26, 0.69, and 0.74, respectively. Furthermore, Maryn et al[13] investigated both intersystem variability and interprogram variability of several acoustic voice measures. They found that jitter measures (ie, absolute jitter, percent jitter, RAP, and pitch perturbation quotient) of *MDVP 5105* version 2.6.2 and *Praat* version 4.4.01 (Paul Boersma and David Weenink, Institute of Phonetic Sciences, University of Amsterdam, The Netherlands) correlated between 0.36 and 0.48 for voice samples recorded by the two separate systems in which these two programs function, and between r = 0.37 and r = 0.47 for voice samples made by the same recording system. Furthermore, shimmer measures (ie, shimmer in dB, percent shimmer, and APQ) correlated from 0.33 to 0.46 between recording systems and from 0.54 to 0.87 between analysis programs. All measures differed significantly between systems and programs. They concluded that perturbation measures across voice analysis programs and recording systems could hardly be compared. Amir et al[14] also compared the measures of mean f0, jitter, shimmer, noise-to-harmonics ratio (NHR), and percentage of unvoiced segments of *MDVP* and *Praat* in [a:] and [i:] recordings of 58 women with dysphonia. Except for mean f0 on both vowels and shimmer on [a:], all comparisons revealed significant differences. Although reasonably high correlations (mostly because of outlying/extreme points in the dataset) between the two programs emerged, they discouraged combined use of the two programs for clinical purposes. In another similar study contrasting jitter measures, shimmer measures, and NHR of *MDVP model 5105* version 2.7.0 and *Praat* version 4.2.17, Oğuz et al[15] obtained rather opposite results: strong correlations between 0.89 and 0.92 but differing data for the jitter measures, weaker correlations between 0.69 and 0.77 without significantly different data for the shimmer measures, and correlation of 0.80 and different data for NHR. Mat Baki et al[16] recorded sustained [a:] of 50 subjects with normophonia and 50 subjects with dysphonia with an iPod Touch 4 (Apple Inc., Cupertino, CA), and juxtaposed fundamental frequency, jitter percent, shimmer percent and NHR of the *MDVP*, and *OperaVOX* (OperaVOX Ltd., Oxford, UK) programs. They found strong intraclass correlation coefficients (ICCs) for f0, jitter, and shimmer, but low ICC for NHR, and concluded that both programs are statistically comparable[a]. Finally, Hanschmann et al[17] simultaneously recorded multiple vowels of 30 participants with normophonia with two systems working on separate computers: head-mounted microphone at 3 cm in system with *rpSzene* version 7.0e (Rehder/Partner GmbH, Hamburg, Germany), and sound level meter at 30 cm in system with *lingWAVES* version 2.50.0043 (WEVOSYS, Forchheim, Germany). After applying various statistical reliability methods, they concluded that relative jitter, relative shimmer, and glottal-to-noise excitation ratio (GNE) cannot be compared properly between these two systems, and that valid comparison can only be made within recordings systems. Synopsized, acoustic voice quality measures derived from different recording systems vary significantly, and even when alleged to derive similar results, the outcomes of distinct voice recording systems and programs should not be compared.

One constant observation through most of these studies,[7,10–12,14,16] on the other hand, is the nearly perfect correlation and similarity in measures of f0 across recording systems or voice analysis programs. This demonstrates acoustic measures of vocal f0 to be exceptionally stable and robust for aspects related to recording hardware or analysis software.

Such variability due to data acquisition hardware or acoustic analysis software is additionally important to consider when utilizing MCDs for clinical voice assessment purposes. Their mobility may induce inconstancies in microphone-to-sound source distance and use in noise-loaded environments. In the past lustrum, multiple research projects focused on comparability and clinical feasibility of MCD-derived acoustic measures. Lin et al[18] investigated the difference in various acoustic voice markers (ie, f0, jitter percent, shimmer percent, SNR, amplitude difference between the first two harmonics, singing power ratio, central frequency of first formant, and central frequency of second formant) between an iPhone and a customized sound recording system. Their results showed significant differences only for percent shimmer, SNR, amplitude difference between the first two harmonics and singing power ratio. There was no sound recorder effect for the other measures. Kardous and Shaw[19] investigated the accuracy of a representative set of 10 smartphone or tablet computer applications in measuring sound pressure level for occupational noise measurement purposes. Their results indicate that only certain, and surely not all, apps may be appropriate for determining sound pressure levels. Vogel et al[20] calculated the differences (ie, root mean square errors) in various acoustic voice quality measures of different voice analysis programs across four recording systems: hard disc recorder with mixer and table-mounted microphone, landline telephone coupled to a computer, iPhone, and laptop computer with head-mounted microphone. With exception of f0 analyses in *MDVP* across the four recording system and f0 analyses in *Praat* across most of the systems, root mean square error exceeded the 10% threshold for almost all other acoustic measures across the four systems. They concluded that acoustic voice quality analysis could not be assumed comparable if speech is taped with different recording systems. Uloza et al[21] recently found significant differences in jitter, normalized noise energy, SNR, and HNR between 118 (ie, 34 subject with dysphonia and 84 subjects with normophonia) [a:] samples simultaneously recorded with a computer-with-microphone system and a Samsung Galaxy Note 3. However, f0 and shimmer did not differ significantly, and correlation as well as various classifier statistics revealed reasonable outcomes in terms of proportional relationship between and categorization accuracy of both recording systems. Uloza et al[21] therefore concluded that smartphones are reliable for the purpose of clinical voice assessment. Finally, Manfredi et al[22] synthesized voice signals with three different voice deviance levels, and recorded them with two smartphones and a reference acquisition system in three different noise conditions (ie, $3 \times 3 \times 3 = 27$ tokens). They subsequently evaluated the correlation in jitter, shimmer, and NHR between recording systems. Although bivariate and ICCs as well as Bland-Altman plots demonstrated strong agreement across recording systems, inspection of the scatterplots at least suggests relevant differences between the two smartphones for jitter on 120-Hz samples and for NHR on 200-Hz samples, as well as between the reference system and a smartphone for jitter and NHR on 200-Hz samples. Manfredi et al[22] concluded that using smartphones is "a convenient way of increasing the sample size in single-case designs and multiple-baseline designs" and "of particular relevance when investigating efficacy of treatments." However, given significant influences of factors like microphone-to-mouth placement[23,24] and environmental noise (see under previous subtitle), vehemently standardized methods should be considered to minimize their impact on these time-domain acoustic markers.

Other acoustic markers

To our knowledge, however, the degree with which findings regarding ambient noise and mobile recording devices also pertains to a cepstrum-based marker such as the smoothed cepstral peak prominence (CPPS),[25,26] the GNE,[27,28] or a multivariate model such as the Acoustic Voice Quality Index (AVQI)[29,30] has not been investigated before. Yet, these acoustic methods have been described as promising estimates of dysphonia severity with acceptable feasibility,[31] reliability,[32] or validity.[33] The CPPS represents the distance between the first harmonic peak and the point with equal quefrency (between 3.03 milliseconds and 16.7 milliseconds) on the regression line through the smoothed cepstrum. The rationale behind this acoustic marker is that the more periodic a complex voice signal, the more the Fourier spectrum displays a well-defined harmonic configuration, and consequently, the more the first harmonic emerges in the smoothed cepstrum. Since its introduction in the field of dysphonia severity measurement by Hillenbrand et al,[25] and Hillenbrand and Houde,[26] CPPS has proven to be a reliable and valid estimate of overall voice quality, and especially breathiness, across a multitude of studies.[29,31,34–37] In the program *Praat* (Paul Boersma and David Weenink, Institute for Phonetic Sciences, University of Amsterdam, The Netherlands), an estimate of the CPPS can be obtained using the commands: "`To PowerCepstrogram. . .`" and

[a]It must be noted, however, that although recordings were made with an MCD, the present study only compared two programs (and not two recording systems).

"Get CPPS. . .". The parameters of these command can be found in Maryn and Weenink.[30] The GNE quantifies the amount of excitation due to vocal fold oscillations versus excitation due to turbulent noise, and has been shown to correlate with auditory perceptions of breathiness. In healthy voices with sufficient glottic closures, all pre-defined frequency channels of the inverse-filtered voice signal are simultaneously excited, and therefore their Hilbert transforms are hypothesized to have equal shape and to correlate well with each other. In breathy voices with turbulent noise due to insufficient glottic closures, however, these frequency channels are excited by narrowband noise, leading to variant and uncorrelated Hilbert envelopes. The more noise in the voice signal, the less Hilbert envelopes correlate and the lower GNE becomes.[27] In the program *Praat*, an estimation of the GNE is yielded via the "To Harmonicity (gne). . ." command with default settings and the "Get maximum" query.[38] Finally, the AVQI is a multivariate, accessible, feasible, and reasonably valid method to clinically estimate overall dysphonia severity in chained recordings of both a sustained vowel and continuous speech. It actually represents a statistical model with the following six constituents: CPPS, shimmer local (SL), shimmer local dB (SLdB), HNR, general slope of the long-term average spectrum, and tilt of the regression line through the long-term average spectrum. Since its creation, the AVQI has shown to correlate well with auditory-perceptual ratings of "G"[39] or "overall severity" from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol[40] and to carry acceptable diagnostic accuracy across multiple external validity studies with different recording methods, laryngological populations, and listener panel: in Dutch,[41] in an English pediatric population,[42] in German,[43] in Dutch, German, English and French,[33] in Korean,[44] and in Finnish.[45] In the program *Praat*, the AVQI can be obtained using the script appended in Maryn and Weenink.[30]

Study goals

Analogous to earlier studies mentioned in this Introduction, the present investigation aimed at assessing the impact of (1) environmental noise, (2) recording device, and (3) their combination (as MCDs are expected to be used outside clinical assessment rooms with less controlled environmental noise) on the following of *Praat*'s acoustic voice measures in a clinically representative set of voice recordings: median f0, jitter local (JL), jitter rap, jitter ppq5, SL, SL dB, shimmer apq3 (APQ3), shimmer apq5, shimmer apq11, HNR, CPPS, GNE, and AVQI.

METHODS

Subjects

Voice samples from 50 Flemish Dutch-speaking subjects were employed in this investigation. Their voices were recorded at the beginning of the standard voice assessment as part of routine clinical practice. Primary laryngological diagnoses included in the sample, using an Olympus ENF-V flexible transnasal chip-on-tip laryngostroboscope (Olympus Corporation, Tokyo, Japan), were the following: 12 normal vocal folds, 13 vocal fold nodules, 9 unilateral vocal fold paralysis, 5 posthead and neck cancer treatment, 4 muscle tension dysphonia, 3 laryngitis, 2 polypoid mucosa, 1 presbylarynx, and 1 leukoplakia. This group consisted of 29 women and 21 men, with ages ranging from 10 to 77 years (mean = 44.9 years, standard deviation = 19.2 years). This sample was considered to be adequately representative of a voice clinic population, reflecting different ages, genders, different types and degrees of voice quality, and voice-related disability.[33]

Original voice recordings

The methods of initial audio recording are identical to those described previously in Maryn et al.[29,41] At the beginning of a standard voice assessment, every participant was asked to sustain the vowel [a:] for at least 5 seconds and to read aloud a phonetically balanced text using a comfortable pitch and loudness. Both voice samples were recorded in an anechoic audiometric booth using an AKG C420 head-mounted condenser microphone and digitized at a sampling rate of 44.1 kHz and 16 bits of resolution using the *Computerized Speech Lab* (CSL model 4500; KayPENTAX Corp., Lincoln Park, NJ). All samples were saved as WAV files. The vowel samples used in this study were trimmed to include only the middle 3 seconds. The read text/connected speech samples were edited to include only the first two sentences.

Further formatting of these original voice recordings consisted of the following two steps. First, for every participant, the two types of voice samples were concatenated in the following order using *Praat*: a pause of 1 second, the continuous speech segment, a pause of 2 seconds, the 3-second sustained vowel segment, and a pause of 1 second. Except for the continuous speech segments, all these parts had constant duration. Boundary markers were inserted as well as pauses and vowels with constant and known length to minimize synchronization error (ie, to maximize the boundary marker duration of 0.002 seconds) in manual segmentation procedures at a stage of this study. Second, all these 50 concatenated sound signals were chained to one long sound file (ie, to enable a single audio presentation of all 50 concatenations after one another); however, they were separated by a short acoustic marker (ie, to indicate the boundary between two concatenated sound files for later segmentation of the long sound files). Every in-between acoustic token to demarcate the margins between two concatenated sound files consisted of two sinusoidal cycles of 0.001 seconds and between –1 and +1 Pascal. Both editings of the original sound signals are illustrated in Figure 1.

Recording systems

First, the reference (ie, clinically representative, referred to as RS1) recording system consisted of an AKG C544L head-mounted condenser microphone (AKG Acoustics, Vienna, Austria), an MPA V L mini-male-to-standard-female XLR connection with phantom power adapter (AKG Acoustics), a Focusrite Forte USB 2.0 professional-grade audio interface (Focusrite Audio Engineering, Bucks, UK), and a MacBook Air with OS Mountain Lion 10.8.5 and *Praat* for MacIntosh version 5.3.53 (Apple Inc., Cupertino, CA). Following its user manual (AKG Acoustics[56], pp. 20–21), the AKG C544L has a cardioid directivity and a frequency range from 20 Hz to 20 kHz. Its frequency response is flat (ie,

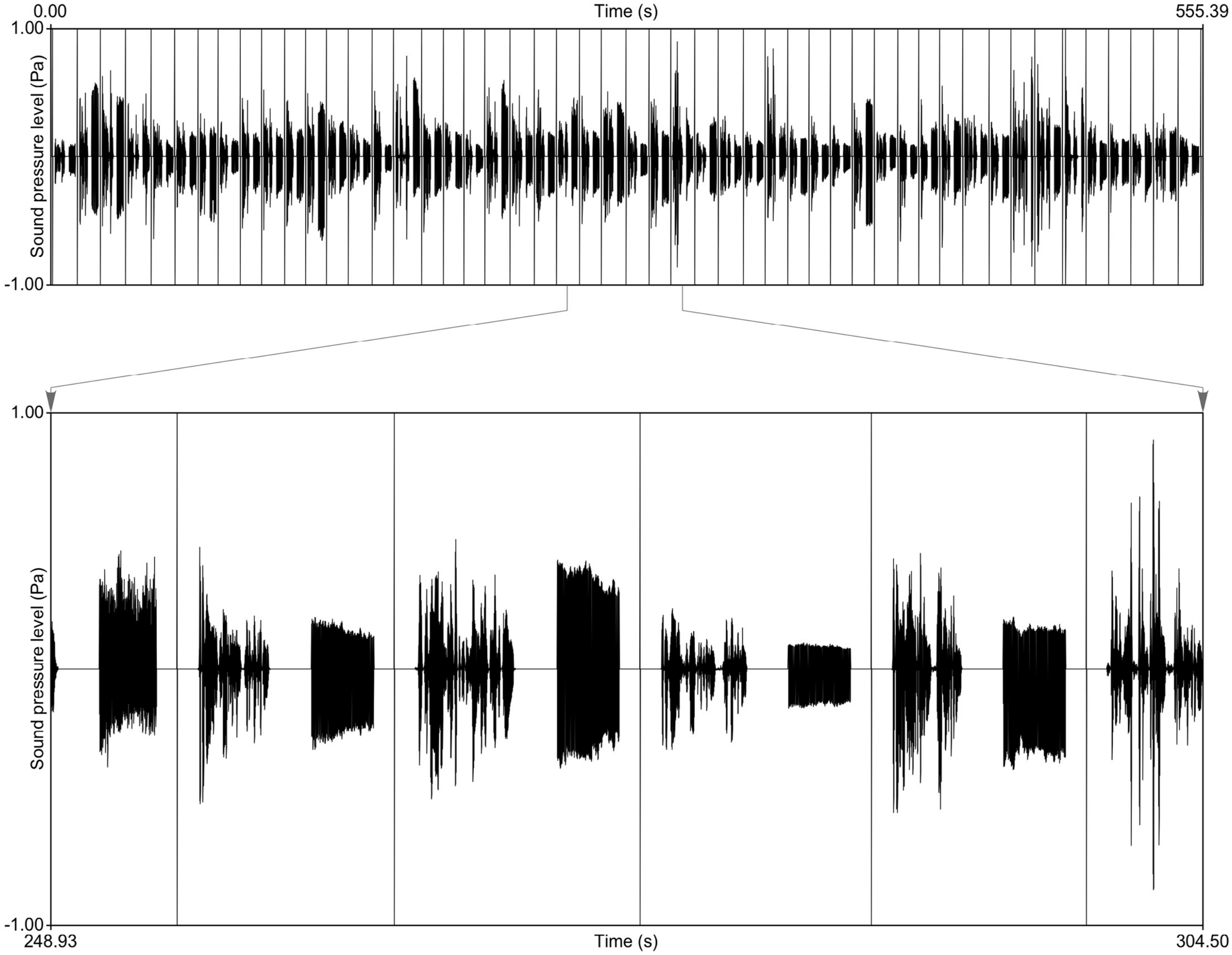

FIGURE 1. Illustration of the experimental sound signal chaining in this study. Upper oscillogram: sequence of all sound signals to one long sound chain of 555.39 seconds. Lower oscillogram: medial four concatenated sound files (pause, text segment, pause, vowel segment, and pause) with their bounds designated by an imprinted acoustic mark.

varies with less than 2 dB) between circa 35 Hz and circa 6.5 kHz. Its equivalent noise level of 22 dB_A is more than 15 dB_A below softest voice level, which is expected to be approximately 45 dB_A.[46] Furthermore, this microphone reaches a total harmonic distortion of 1% at 126 dB_{SPL}, which exceeds maximal vocal sound pressure levels of approximately 115 dB_{SPL}.[47] The Focusrite Forte external analog-to-digital converter has a 24-bits resolution and a sampling rate of 192 kHz (Focusrite Audio Engineering[57], p. 5). Conform the recommendations of Švec and Granqvist,[24] both AKG C544L condenser microphone and Focusrite Forte analog-to-digital converter (connected to each other with MPA V L and to the computer with USB 2.0) are considered especially suitable for clinical voice assessment.

Second, five MCDs were employed as all-inclusive portable recording devices (ie, encompassing microphone, analog-to-digital converter and data storage). There were two tablet computers and three smartphones. Table 1 summarizes the five models that were used in the present study together with their operating system and software application for audio recording (ie, RS2, RS3, RS4, RS5, and RS6). At the time of the present study, this sample of devices was considered representative as these five MCDs were among the most used MCDs in Belgium.[9]

Signal-to-noise ratio

Radiation and recording of the experimental voice samples took place in an anechoic audiometry booth (IAC Acoustics, Niederkrüchten, Germany) conform ISO 8253-1:2010 (Acoustics—Audiometric test methods—Part 1: pure-tone air and bone conduction audiometry). However, to assess the influence of ambient noise level on selected acoustic voice measures, the experimental voice signals (see next paragraph) were radiated or captured equally in five different ambient noise conditions. These five conditions were created by playing audiometric speech noise (ie, acoustic noise specifically created from speech material to mask the spoken stimuli during speech audiometry)

TABLE 1.
List of Mobile Audio Recording Devices With Their Software Applications and the Format of the Recorded Files

Abbreviation	Type	Devices	Operating System	Software Application	File Format
RS2	Tablet	iPad 2	iOS 8.2	*Voice Recorder HD 1.0.5*	WAV
RS3	Tablet	Google Nexus 9	Android Lollipop 5.0	*Voice Recorder HD 1.0.5*	WAV
RS4	Smartphone	iPhone 5S	iOS 8.2	*Voice Recorder HD 1.0.5*	WAV
RS5	Smartphone	Samsung Galaxy S5	Android Jelly Bean 4.2	*Voice Recorder HD 1.0.5*	WAV
RS6	Smartphone	Nokia Lumia 520	Windows Phone 8.1	*Voice Recorder Pro+ 1.4.0.0*	WAV

derived from all monosyllabic words from the audiometry lists of the Dutch Association for Audiology[48,49] presented in free field with the audiometer set at five different levels: 0 dB_{SPL}, 30 dB_{SPL}, 40 dB_{SPL}, 50 dB_{SPL}, and 60 dB_{SPL}. These five levels corresponded with 20.5 dB_A, 27.6 dB_A, 36.9 dB_A, 47.0 dB_A, and 56.7 dB_A, respectively, as determined with a CR:832C integrating averaging class 2 sound-level meter (Cirrus Research plc, Hunmanby, North Yorkshire, UK). The five average SNRs for the six recording systems are provided in Table 2. These SNRs have been calculated as the mean of absolute differences between sound intensity levels of experimental vowel recordings and experimental pause recordings.

Experimental voice recordings

The complete sound chain with all the original voice samples, pauses, and boundary markers (ie, the upper oscillogram of 555.39 seconds in Figure 1) was radiated by an Inspire T12 loudspeaker model MF1625 (Creative Technology Ltd., Singapore, Singapore) at an average intensity of 80 dB_C at 10 cm, as measured with a CR:832C integrating averaging class 2 sound level meter (Cirrus Research plc). Average sound intensity of the radiated sound chain was chosen to simulate comfortable vowel [a:] levels surrounding 80 dB_{SPL}.[50] The microphone of RS1 and the microphone outlets of RS2–RS6 were placed equidistantly at 10 cm, as controlled with a ruler, and at a nonzero (ie, 45°) azimuthal angle from the loudspeaker, as recommended by, for example, Maryn and Zarowski.[51] This method was hypothesized to maximally mimic real clinical voice recordings. Subsequently, the complete sound chain was radiated 30 times (ie, 6 recording systems × 5 background noise conditions) by the loudspeaker, and captured and stored in WAV format by the mobile recording devices.

Afterward, these 30 complete sound chains were transferred to a personal computer for extraction of 50 sustained vowel and 50 continuous speech parts. To extract a vowel of 2.90 seconds, (1) cursor in Praat's "`View & Edit`" screen (ie, zoomed to view approximately 10 seconds) was positioned manually at its following boundary marker, (2) vowel front selection was set at cursor time −3.95 seconds, (3) vowel back selection was set at cursor time −1.05 seconds, and (4) this selection was saved with the "`Save selected sound as WAV file. . .`" procedure. Duration of 2.90 seconds was chosen to avoid coextraction of pause parts surrounding the vowel due to cursor misplacement. This was iterated for all the 1500 vowel segments. A similar method was applied to extract continuous speech parts: (1) front cursor in the same screen was placed manually at its preceding boundary marker, (2) back cursor in the same screen was situated manually at its following boundary marker, (3) speech front selection was set at front cursor time plus 1.05 seconds, (4) speech back selection was set at back cursor time minus 6.05 seconds, and (5) this selection was saved with the "`Save selected sound as WAV file. . .`" procedure. This was also repeated for all the 1500 continuous speech segments.

Acoustic measures

The following 9 acoustic markers were derived from the 50 experimental [a:] recordings with the program *Praat*:

- median f0;
- measures of f0 perturbation: JL and jitter rap (RAP);
- measures of amplitude perturbation: SL, SLdB, and APQ3;
- HNR;
- GNE following van As-Brooks et al;[38]
- CPPS following Maryn and Weenink.[30]

The AVQI was determined on concatenations of voiced segments of continuous speech with sustained [a:], according to the method described by Maryn and Weenink.[30]

TABLE 2.
Signal-to-noise Ratio (SNR) of the Experimental Voice Recordings: Mean (Standard Deviation).

Audiometric Noise Level	RS1	RS2	RS3	RS4	RS5	RS6
20.5 dB_A	43.9(3.9)	48.1(4.3)	33.5(4.1)	39.4(4.3)	28.4(4.4)	41.9(3.5)
27.6 dB_A	44.0(3.9)	46.4(4.3)	26.7(4.0)	37.1(4.3)	28.1(4.4)	42.9(3.4)
36.9 dB_A	43.8(3.9)	43.3(4.2)	25.7(4.0)	30.4(4.3)	26.1(4.4)	39.8(3.5)
47.0 dB_A	40.9(3.9)	34.6(4.2)	25.3(4.0)	20.8(4.2)	18.9(4.3)	30.0(3.5)
56.7 dB_A	32.7(3.9)	24.6(4.2)	20.6(4.0)	11.3(3.8)	9.9(3.8)	16.2(3.5)

Notes: Gray shades indicate SNR ≤ 30 dB (ie, voice samples with insufficient recording quality).

In the following text, every data set is annotated as following: acronym of acoustic marker in normal font, recording system in subscript, and noise level in subscript. For example, data regarding CPPS recorded with iPhone 5S in 60-dB environmental noise are indicated as data set "$CPPS_{RS4-60}$." In total, there were 300 data sets (ie, 10 acoustic measures × 6 recording systems × 5 noise levels) covering 15,000 data points (ie, 300 data sets × 50 subjects).

Statistical analyses

All statistical analyses in the present study were completed using *SPSS* version 22.0 (SPSS Inc., Chicago, IL) on iOS Lion 10.7.5. Because analysis with the one-sample Kolmogorov-Smirnov test revealed that only 77 out of 300 data sets (ie, a minority of 25.7%) were normally distributed,[b] nonparametric inferential statistical methods were applied. First, for every acoustic marker, nonparametric analysis of variance with Friedman repeated-measures test for five related and dependent samples was used to compare reference RS1 recordings in 20.5 dB_A with RS1 recordings in four levels of environmental noise (ie, 27.6 dB_A, 36.9 dB_A, 47.0 dB_A, and 56.7 dB_A) on acoustic measures of RS1. If this overall statistic yielded a significant result, posthoc Wilcoxon signed-rank tests were used to juxtapose pairs of related samples after Bonferroni correction (ie, results were considered statistically significant at $P \leq 0.013$). These methods were administered to answer the question: Are the 10 acoustic metrics influenced by increasing levels of environmental noise? Second, for every acoustic marker, Friedman repeated-measures test for six related samples was applied to contrast the recordings of the MCDs (ie, RS2, RS3, RS4, RS5, and RS6) in 20.5 dB_A with the reference RS1 recordings also in 20.5 dB_A. If this omnibus statistic resulted in a significant outcome, posthoc Wilcoxon signed-rank tests were used to contrast pairs of related samples after Bonferroni correction (ie, outcomes were considered statistically significant at $P \leq 0.010$). These methods served to answer the question: Are the 10 acoustic metrics affected by MCD? Third, to test for interaction effects of both MCD and environmental noise level, Wilcoxon signed-rank tests were conducted between pairs of RS1 in 20.5 dB_A and RS2–RS6 in 27.6–56.7 dB_A after Bonferroni correction (ie, results were considered statistically significant at $P \leq 0.010$). This method intended to solve the question: Is it clinically acceptable to compare voice-related data from audio samples from MCDs in increased surrounding noise?

RESULTS

Descriptive data

Table 3 summarizes the descriptive data (ie, mean values, standard deviations, and number of undefined values) of all 300 data sets of the present study. Interestingly, the number of undefined values was 0 for RS1, RS2, and RS6 (ie, clinical reference system, iPad and Nokia, respectively). This means that all variables could be determined for all voice recordings, regardless of environmental noise level and dysphonia severity. For RS4 (ie, iPhone), f0 nor the perturbation markers could be calculated for subject 32 with unilateral vocal fold paralysis, but only in the worst noise condition. RS5 (ie, Samsung) could calculate f0, perturbation measures, nor AVQI for subject 32 in all noise conditions, and also not APQ3 for subject 23 after radiation treatment for laryngeal carcinoma in the worst noise condition. Finally, RS3 (ie, Nexus) could not define f0, perturbation values, and AVQI in both subjects 32 and 23, and therefore its recordings of severely dysphonic voices seem to be most susceptible for distortion due to recording circumstances.

Signal-to-noise ratio

From the perspective of SNR (Table 2), and with SNR > 30 dB indicating acceptable recording fidelity,[8] RS1 performed best because it withstood the impact of increased noise of even 56.7 dB_A, followed by RS2 (ie, iPad; accuracy cannot be assured in noise levels above 47.0 dB_A), and by RS4 and RS6 (ie, iPhone and Nokia, respectively; precision cannot be assured in noise levels above 36.9 dB_A). RS3s (ie, Nexus) SNR dropped below 30 dB as soon as external noise was added, whereas RS5s (ie, Samsung) SNR never even exceeded the threshold. On the other hand, in the two most favorable noise conditions (ie, 20.5 dB_A and 27.6 dB_A), the iPad's mean SNR was the only one to outperform the mean SNR of the reference system. When used in ideal circumstances with minimal ambient noise, all recording systems except RS5 yielded highly acceptable (RS1 and RS2) to reasonable (RS3, RS4, and RS6) SNRs. Depending on the volume of surrounding noise, some MCDs may lose their measurement fidelity.

Influence of environmental noise

Table 4 summarizes the inferential findings regarding the impact of environmental noise on acoustic markers from reference experimental recordings. No significant influence was detected for f0. For JL and RAP, omnibus Friedman test indicated a significant influence of environmental noise (with $P = 0.015$ and $P = 0.002$, respectively). However, posthoc Wilcoxon tests for all pairs of noise conditions revealed absence of such influence. It can therefore be concluded that f0, JL, and RAP from RS1 are not influenced by noise levels up to 56.7 dB_A (ie, audiometric speech noise at 60 dB_{SPL}). For SL, SLdB, APQ3, HNR,

[b]The following 77 data sets were normally distributed (ie, *P* values are asymptotic two tailed): $f0_{RS1-0}$: Z = 0.106, *P* = 0.200; GNE_{RS1-0}: Z = 0.124, *P* = 0.054; $f0_{RS1-30}$: Z = 0.106, *P* = 0.200; $f0_{RS1-40}$: Z = 0.106, *P* = 0.200; $f0_{RS1-50}$: Z = 0.106, *P* = 0.200; $f0_{RS1-60}$: Z = 0.106, *P* = 0.200; $CPPS_{RS1-60}$: Z = 0.095, *P* = 0.200; $f0_{RS2-0}$: Z = 0.104, *P* = 0.200; GNE_{RS2-0}: Z = 0.067, *P* = 0.200; $f0_{RS2-30}$: Z = 0.104, *P* = 0.200; GNE_{RS2-30}: Z = 0.079, *P* = 0.200; $f0_{RS2-40}$: Z = 0.104, *P* = 0.200; GNE_{RS2-40}: Z = 0.078, *P* = 0.200; $f0_{RS2-50}$: Z = 0.104, *P* = 0.200; GNE_{RS2-50}: Z = 0.071, *P* = 0.200; $f0_{RS2-60}$: Z = 0.104, *P* = 0.200; GNE_{RS2-60}: Z = 0.069, *P* = 0.200; $f0_{RS3-0}$: Z = 0.108, *P* = 0.200; $f0_{RSS-30}$: Z = 0.107, *P* = 0.200; $AVQI_{RS3-30}$: Z = 0.121, *P* = 0.070; $f0_{RS3-40}$: Z = 0.108, *P* = 0.200; $f0_{RS3-50}$: Z = 0.111, *P* = 0.182; $f0_{RS6-60}$: Z = 0.106, *P* = 0.200; $CPPS_{RS3-60}$: Z = 0.085, *P* = 0.200; $AVQI_{RS3-60}$: Z = 0.091, *P* = 0.200; $f0_{RS4-0}$: Z = 0.104, *P* = 0.200; $f0_{RS4-30}$: Z = 0.104, *P* = 0.200; $f0_{RS4-40}$: Z = 0.104, *P* = 0.200; $f0_{RS4-50}$: Z = 0.103, *P* = 0.200; GNE_{RS4-50}: Z = 0.115, *P* = 0.094; $CPPS_{RS4-50}$: Z = 0.072, *P* = 0.200; $AVQI_{RS4-50}$: Z = 0.121, *P* = 0.066; GNE_{RS4-60}: Z = 0.094, *P* = 0.200; SL_{RS4-60}: Z = 0.097, *P* = 0.200; $SLdB_{RS4-60}$: Z = 0.085, *P* = 0.200; $APQ3_{RS4-60}$: Z = 0.103, *P* = 0.200; HNR_{RS4-60}: Z = 0.095, *P* = 0.200; $CPPS_{RS4-60}$: Z = 0.086, *P* = 0.200; $AVQI_{RS4-60}$: Z = 0.106, *P* = 0.200; $f0_{RS5-0}$: Z = 0.106, *P* = 0.200; $AVQI_{RS5-0}$: Z = 0.107, *P* = 0.200; $f0_{RS5-30}$: Z = 0.106, *P* = 0.200; $f0_{RS5-40}$: Z = 0.106, *P* = 0.200; $CPPS_{RS5-40}$: Z = 0.106, *P* = 0.200; $f0_{RS5-50}$: Z = 0.106, *P* = 0.200; GNE_{RS5-50}: Z = 0.124, *P* = 0.053; $SLdB_{RS5-50}$: Z = 0.119, *P* = 0.082; $APQ3_{RS5-50}$: Z = 0.120, *P* = 0.081; $CPPS_{RS5-50}$: Z = 0.065, *P* = 0.200; $AVQI_{RS5-50}$: Z = 0.090, *P* = 0.200; $f0_{RS5-60}$: Z = 0.125, *P* = 0.057; GNE_{RS5-60}: Z = 0.103, *P* = 0.200; SL_{RS5-60}: Z = 0.116, *P* = 0.118; $SLdB_{RS5-60}$: Z = 0.123, *P* = 0.065; $APQ3_{RS5-60}$: Z = 0.094, *P* = 0.200; HNR_{RS5-60}: Z = 0.097, *P* = 0.200; $CPPS_{RS5-60}$: Z = 0.090, *P* = 0.200; $AVQI_{RS5-60}$: Z = 0.099, *P* = 0.200; $f0_{RS6-0}$: Z = 0.107, *P* = 0.200; GNE_{RS6-0}: Z = 0.105, *P* = 0.200; $f0_{RS6-30}$: Z = 0.105, *P* = 0.200; GNE_{RS6-30}: Z = 0.095, *P* = 0.200; $f0_{RS6-40}$: Z = 0.106, *P* = 0.200; GNE_{RS6-40}: Z = 0.096, *P* = 0.200; $CPPS_{RS6-40}$: Z = 0.096, *P* = 0.200; $f0_{RS6-50}$: Z = 0.106, *P* = 0.200; GNE_{RS6-50}: Z = 0.108, *P* = 0.200; $CPPS_{RS6-0}$: Z = 0.061, *P* = 0.200; $f0_{RS6-60}$: Z = 0.117, *P* = 0.086; GNE_{RS6-60}: Z = 0.107, *P* = 0.200; $APQ3_{RS6-60}$: Z = 0.118, *P* = 0.081; $CPPS_{RS6-60}$: Z = 0.105, *P* = 0.200; $AVQI_{RS6-60}$: Z = 0.090, *P* = 0.200.

TABLE 3.
Mean (M), Standard Deviation (SD), and Number of Undefined Values (NUV) of the 10 Acoustic Measures *per* RS and *per* Environmental Noise Level

	20.5 dB_A	27.6 dB_A	36.9 dB_A	47.0 dB_A	56.7 dB_A
	M(SD);NUV	M(SD);NUV	M(SD);NUV	M(SD);NUV	M(SD);NUV
RS1					
f0	163.9(54.0);0	163.9(54.0);0	163.8(54.0);0	163.9(54.0);0	163.8(54.0);0
JL	1.2(2.1);0	1.2(2.1);0	1.2(2.1);0	1.2(2.1);0	1.2(2.1);0
RAP	0.7(1.2);0	0.7(1.2);0	0.7(1.2);0	0.7(1.2);0	0.7(1.2);0
SL	4.5(5.4);0	4.5(5.3);0	4.6(5.6);0	4.6(5.4);0	4.8(5.6);0
SLdB	0.4(0.5);0	0.4(0.5);0	0.4(0.5);0	0.4(0.5);0	0.4(0.5);0
APQ3	2.3(2.7);0	2.3(2.6);0	2.4(2.8);0	2.3(2.7);0	2.5(2.9);0
HNR	20.3(7.1);0	20.4(7.1);0	20.3(7.1);0	20.3(7.0);0	19.6(6.7);0
GNE	0.8(0.2);0	0.8(0.2);0	0.8(0.2);0	0.8(0.2);0	0.7(0.2);0
CPPS	11.1(3.5);0	11.1(3.5);0	11.1(3.4);0	10.9(3.4);0	10.1(3.1);0
AVQI	4.1(2.2);0	4.2(2.3);0	4.2(2.1);0	4.2(2.2);0	4.5(2.0);0
RS2					
f0	164.2(54.1);0	164.1(54.1);0	164.2(54.0);0	164.1(54.1);0	164.3(53.9);0
JL	1.2(2.0);0	1.2(2.1);0	1.2(1.9);0	1.2(2.0);0	1.3(2.0);0
RAP	0.7(1.1);0	0.7(1.2);0	0.7(1.1);0	0.7(1.2);0	0.7(1.1);0
SL	5.4(5.6);0	5.3(5.2);0	5.4(5.5);0	5.4(5.1);0	6.3(5.3);0
SLdB	0.5(0.5);0	0.5(0.5);0	0.5(0.5);0	0.5(0.5);0	0.6(0.5);0
APQ3	2.8(2.7);0	2.8(2.7);0	2.8(2.7);0	2.8(2.6);0	3.3(2.5);0
HNR	17.9(6.8);0	17.7(6.8);0	17.7(6.7);0	17.4(6.6);0	15.9(5.9);0
GNE	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0
CPPS	11.4(3.4);0	11.3(3.4);0	11.3(3.4);0	11.0(3.2);0	10.1(2.9);0
AVQI	4.0(2.2);0	4.1(2.2);0	4.1(2.2);0	4.2(2.1);0	4.7(1.8);0
RS3					
f0	164.2(54.0);1	164.2(54.1);1	164.3(54.0);1	164.4(53.8);1	156.0(55.4);2
JL	1.2(1.9);2	1.2(2.1);2	1.2(2.0);2	1.3(2.1);2	1.7(2.0);2
RAP	0.7(1.0);2	0.7(1.3);2	0.7(1.0);2	0.7(1.1);2	0.9(1.1);2
SL	5.4(5.6);2	5.3(5.5);2	5.5(5.3);2	7.2(5.2);2	12.9(4.6);2
SLdB	0.5(0.5);2	0.5(0.5);2	0.5(0.5);2	0.6(0.5);2	1.1(0.4);2
APQ3	2.9(3.2);2	2.8(2.8);2	2.9(2.7);2	3.8(2.6);2	7.1(2.2);2
HNR	17.4(6.7);0	17.4(6.7);0	16.9(6.4);0	14.4(5.4);0	8.8(3.8);1
GNE	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.5(0.2);0
CPPS	11.3(3.5);0	11.2(3.5);0	10.7(3.3);0	9.4(2.9);0	7.1(2.1);0
AVQI	4.2(2.1);1	4.2(2.2);1	4.4(2.2);1	5.2(1.8);1	6.7(1.1);1
RS4					
f0	165.3(53.3);0	165.3(53.3);0	165.4(52.9);0	165.2(52.6);0	167.5(51.8);1
JL	1.2(1.8);0	1.1(1.7);0	1.1(1.6);0	1.2(1.7);0	1.2(1.7);1
RAP	0.6(1.0);0	0.6(1.0);0	0.6(0.9);0	0.6(0.9);0	0.7(1.0);1
SL	5.5(4.7);0	5.2(4.4);0	5.5(4.6);0	5.5(4.7);0	6.8(4.5);1
SLdB	0.5(0.4);0	0.5(0.4);0	0.5(0.4);0	0.5(0.4);0	0.6(0.4);1
APQ3	2.9(2.4);0	2.8(2.3);0	2.9(2.5);0	2.9(2.4);0	3.7(2.3);1
HNR	12.9(5.3);0	12.7(5.1);0	12.7(5.1);0	12.6(5.1);0	11.9(4.4);0
GNE	0.8(0.2);0	0.8(0.2);0	0.8(0.2);0	0.7(0.2);0	0.7(0.2);0
CPPS	11.6(3.6);0	11.7(3.6);0	11.6(3.6);0	11.2(3.4);0	10.1(3.1);0
AVQI	5.5(2.2);0	5.6(2.1);0	5.6(2.2);0	5.8(2.1);0	6.2(1.8);0
RS5					
f0	163.9(53.9);1	164.7(52.9);1	164.6(53.0);1	164.3(53.3);1	159.8(52.8);2
JL	1.2(2.1);1	1.3(2.2);1	1.2(1.9);1	1.3(2.1);1	1.5(1.9);2
RAP	0.7(1.2);1	0.7(1.2);1	0.6(1.0);1	0.7(1.2);1	0.9(1.1);2
SL	5.0(5.2);1	4.8(5.2);1	4.9(5.1);1	5.7(4.9);1	9.7(4.4);2
SLdB	0.4(0.5);1	0.4(0.5);1	0.4(0.5);1	0.5(0.5);1	0.9(0.4);2
APQ3	2.6(2.7);2	2.7(3.3);2	2.5(2.5);2	3.2(2.9);2	5.2(2.2);2
HNR	18.8(6.8);0	18.9(6.8);0	18.6(6.6);0	16.9(5.8);0	12.2(4.0);0
GNE	0.8(0.2);0	0.8(0.2);0	0.7(0.2);0	0.7(0.2);0	0.6(0.2);0

(continued on next page)

TABLE 3.
(*continued*)

	20.5 dB_A	27.6 dB_A	36.9 dB_A	47.0 dB_A	56.7 dB_A
	M(SD);NUV	M(SD);NUV	M(SD);NUV	M(SD);NUV	M(SD);NUV
CPPS	10.9(3.4);0	10.9(3.4);0	10.5(3.3);0	9.4(3.0);0	7.1(2.4);0
AVQI	4.0(2.2);1	3.9(2.2);1	4.1(2.0);1	4.7(1.9);1	6.2(1.4);1
RS6					
f0	165.8(53.0);0	165.9(52.9);0	165.6(53.3);0	165.6(53.3);0	159.0(53.1);0
JL	1.2(1.9);0	1.1(1.8);0	1.2(1.9);0	1.2(1.7);0	1.8(1.8);0
RAP	0.6(1.1);0	0.6(1.0);0	0.6(1.1);0	0.7(1.0);0	1.0(1.0);0
SL	5.7(4.8);0	5.9(4.9);0	6.0(4.5);0	7.6(4.5);0	13.7(4.7);0
SLdB	0.5(0.4);0	0.5(0.4);0	0.5(0.4);0	0.7(0.4);0	1.2(0.4);0
APQ3	3.0(2.6);0	3.1(2.6);0	3.2(2.3);0	4.2(2.3);0	7.6(2.7);0
HNR	13.8(5.0);0	13.5(5.1);0	13.3(5.0);0	11.7(4.5);0	7.0(3.2);0
GNE	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.7(0.2);0	0.5(0.2);0
CPPS	10.9(3.2);0	10.8(3.3);0	10.5(3.2);0	9.4(2.9);0	7.0(2.1);0
AVQI	5.1(1.8);0	5.2(2.0);0	5.3(1.9);0	6.0(1.8);0	7.4(1.2);0

and AVQI, overall Friedman tests indicated a significant effect of environmental noise, and for these five measures, posthoc Wilcoxon tests revealed a significant difference only when 20.5 dB_A and 56.7 dB_A conditions were compared (with $P = 0.000$ for these five measures). No significant differences were found for the other conditions with less surrounding noise. Significant differences also emerged for CPPS and GNE between the reference condition and the 40.7 dB_A and 56.7 dB_A conditions (with $P = 0.000$ and $P = 0.009$, respectively). These two measures thus appeared most vulnerable to environmental noise.

Influence of MCD

Table 5 encapsulates the differences in acoustic measures between reference recordings and MCD recordings at 20.5 dB_A (ie, when no noise was artificially added in the audiometric booth). No significant influence was found for f0. For all other measures, however, significant differences surfaced when RS1 recordings were compared with recordings from at least three of the five MCDs. Exceptions were as following: JL for RS1–RS2, RAP for RS1–RS2, GNE for RS1–RS3 and RS1–RS6, CPPS for RS1–RS4, and AVQI for RS1–RS2 and RS1–RS4. Except for f0 and these seven nonsignificant differences, and taken minimal environmental noise into account, it is clear that MCD voice samples have higher acoustic voice quality values than voices recorded with clinical reference equipment.

Combined influence: noise and MCD

Table 6 provides for every RS an overview of surrounding noise levels/thresholds from which significant differences in acoustic measures can be expected (when compared to voice samples recorded with reference equipment in 20.5 dB_A). f0 remained unaffected regardless MCD or external noise condition. JL and RAP on RS2 samples were influenced by 56.7 dBA noise, whereas the condition without additional noise (ie, 20.5 dB_A) already influenced JL and RAP off on RS3, RS4, RS5, and RS6. The measures SL, SLdB, APQ3, and HNR were already impacted by noise from 20.5 dB_A on all MCD samples. GNE was affected by surrounding noise from 47.0 dB_A on RS5, from 36.9 dB_A on RS4,

TABLE 4.
The Influence of Ambient Noise on Various Acoustic Voice Measures ($\alpha = 0.05/4 = 0.013$ for Bonferroni Correction; Gray Shades Indicate Significant Differences)

	Overall Effect (Friedman)		20.5 dB_A–27.6 dB_A (Wilcoxon)		20.5 dB_A–36.9 dB_A (Wilcoxon)		20.5 dB_A–47.0 dB_A (Wilcoxon)		20.5 dB_A–56.7 dB_A (Wilcoxon)	
	χ^2	*P*	Z	*P*	Z	*P*	Z	*P*	Z	*P*
f0	5.83	0.212	NA	NA	NA	NA	NA	NA	NA	NA
JL	12.41	0.015	–1.13	0.259	–0.83	0.407	–1.41	0.160	–0.96	0.339
RAP	17.13	0.002	–1.43	0.153	–0.29	0.775	–1.60	0.110	–0.41	0.683
SL	32.23	0.000	–0.33	0.739	–1.62	0.104	–1.77	0.077	–4.47	0.000
SLdB	42.18	0.000	–0.55	0.583	–1.61	0.107	–1.67	0.096	–4.67	0.000
APQ3	39.33	0.000	–0.82	0.413	–1.26	0.208	–0.65	0.515	–4.36	0.000
HNR	72.78	0.000	–1.81	0.071	–0.55	0.585	–1.22	0.224	–5.68	0.000
GNE	83.98	0.000	–0.78	0.433	–1.28	0.201	–2.60	0.009	–5.18	0.000
CPPS	123.01	0.000	–0.64	0.525	–2.41	0.016	–5.12	0.000	–6.13	0.000
AVQI	57.88	0.000	–0.40	0.692	–0.49	0.622	–1.70	0.089	–5.17	0.000

TABLE 5.
The Influence of Recording System on Various Acoustic Voice Measures (α = 0.05/5 = 0.01 for Bonferroni Correction; Gray Shades Indicate Significant Differences)

	Overall effect (Friedman)		RS1–RS2 (Wilcoxon)		RS1–RS3 (Wilcoxon)		RS1–RS4 (Wilcoxon)		RS1–RS5 (Wilcoxon)		RS1–RS6 (Wilcoxon)	
	χ^2	*P*	Z	*P*	Z	*P*	Z	*P*	Z	*P*	Z	*P*
f0	1.53	0.910	–1.12	0.264	–1.01	0.309	–1.26	0.208	–1.22	0.222	–1.22	0.222
JL	76.98	0.000	–1.37	0.172	–3.62	0.000	–2.79	0.005	–4.11	0.000	–3.60	0.000
RAP	71.70	0.000	–1.40	0.161	–3.79	0.000	–2.96	0.003	–4.36	0.000	–3.74	0.000
SL	61.55	0.000	–3.98	0.000	–4.65	0.000	–3.89	0.000	–5.07	0.000	–4.79	0.000
SLdB	70.33	0.000	–3.68	0.000	–5.16	0.000	–3.55	0.000	–5.11	0.000	–4.44	0.000
APQ3	79.87	0.000	–3.65	0.000	–5.08	0.000	–4.27	0.000	–4.51	0.000	–5.65	0.000
HNR	157.89	0.000	–6.14	0.000	–6.11	0.000	–6.15	0.000	–5.79	0.000	–6.06	0.000
GNE	59.51	0.000	–5.38	0.000	–2.14	0.032	–3.31	0.000	–3.21	0.001	–1.26	0.207
CPPS	114.46	0.000	–3.68	0.000	–5.33	0.000	–2.35	0.019	–3.37	0.000	–5.02	0.000
AVQI	130.49	0.000	–1.85	0.064	–6.09	0.000	–0.60	0.551	–5.51	0.000	–3.78	0.000

and already from 20.5 dB_A on RS2, RS3, and RS6. CPPS was influenced by external noise from 36.9 dB_A on RS3, and already from 20.5 dB_A on all other MCDs. Finally, AVQI was influenced by surrounding noise from 56.7 dB_A on RS2, from 36.9 dB_A on RS3, and already from 20.5 dB_A on RS4, RS5, and RS6.

In other words, all acoustic voice markers from RS6 (ie, Nokia) recordings are already influenced by anechoic booth sound conditions without added noise (ie, 20.5 dB_A condition). Except for HNR and GNE, the same counts for RS4 (ie, iPhone) and RS5 (ie, Samsung) recordings; and excepting CPPS and AVQI, similar results emerged for RS3 (ie, Nexus) recordings. Finally, for RS2 (ie, iPad) recordings, the same findings occurred with the exception of JL, RAP, and AVQI. As a comparison, none of the acoustic measures on reference RS1 recordings were that easily influenced by surrounding noise (cfr. Table 4): it took 47.0 dB_A to affect GNE and CPPS; it took 56.9 dB_A to impact

TABLE 6.
The Mixed Influence of Recording System and Environmental Noise on Various Acoustic Voice Measures (α = 0.05/5 = 0.01 for Bonferroni Correction).

	RS1–RS2 (Wilcoxon)		RS1–RS3 (Wilcoxon)		RS1–RS4 (Wilcoxon)		RS1–RS5 (Wilcoxon)		RS1–RS6 (Wilcoxon)	
	Level		Level		Level		Level		Level	
	Z	*P*	Z	*P*	Z	*P*	Z	*P*	Z	*P*
f0	NS		NS		NS		NS		NS	
	—	—	—	—	—	—	—	—	—	—
JL	56.7 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–3.18	.001	–2.79	.005	–3.62	.000	–3.60	.000	–4.11	.000
RAP	56.7 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–2.96	.003	–2.96	.003	–3.79	.000	–3.74	.000	–4.36	.000
SL	20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–3.98	.000	–3.89	.000	–4.65	.000	–4.79	.000	–5.07	.000
SLdB	20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–3.68	.000	–3.55	.000	–5.16	.000	–4.44	.000	–5.11	.000
APQ3	20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–3.65	.000	–4.27	.000	–5.08	.000	–4.51	.000	–5.65	.000
HNR	20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–6.14	.000	–6.15	.000	–6.11	.000	–6.06	.000	–5.79	.000
GNE	20.5 dB_A		20.5 dB_A		36.9 dB_A		47.0 dB_A		20.5 dB_A	
	–5.38	.000	–3.31	.000	–2.91	.004	–5.31	.000	–3.21	.001
CPPS	20.5 dB_A		36.9 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–3.68	.000	–4.61	.000	–5.33	.000	–5.02	.000	–3.37	.000
AVQI	56.7 dB_A		36.9 dB_A		20.5 dB_A		20.5 dB_A		20.5 dB_A	
	–4.76	.000	–4.23	.000	–6.09	.000	–3.78	.000	–5.51	.000

Notes: For reasons of practicality, only the ambient noise level from which difference between RS1 and MCD started to be significant is mentioned in this table.
Abbreviation: NS, no statistical differences found.

SL, SLdB, APQ3, and AVQI; and it definitely takes higher noise levels to significantly influence f0, JL, and RAP.

DISCUSSION

The present study investigated the influence of environmental noise, MCDs, and the combination of these two factors on a set of 10 diverse acoustic measures that are commonly administered in voice clinics. With contemporary mobile communication apparatuses hypothesized to become increasingly used in voice clinics, their potential as clinical audio recording device and basis for acoustic voice measures required investigation. Furthermore, their mobility may instigate variation in microphone-to-mouth distance and usage in noise-loaded surroundings, especially needing to explore their performance in different signal-to-noise conditions. Sustained vowel and continuous speech recordings of 50 subjects with varying dysphonia severity levels (ie, on a continuum between normal and clear voice, and most severe dysphonia) where therefore radiated and recorded by a reference recording system and five MCDs in five environmental noise conditions.

Review of extensive literature on variability in acoustic data—ie, mostly f0 and perturbation measures—already pointed out significant differences between various recording systems and analysis programs.[7,10–21] It also indicated an important influence of environmental noise on these acoustic voice measures.[5,6,8] However, exploring the combination of recording device (both clinical equipment and mobile appliances) and surrounding noise in these and other acoustic voice measures in a single sample set was required to increase the generalizability of the findings. The present study was designed to fit that purpose, and generally, it can be stated that the results corroborate with earlier findings, at least when recordings of the present study's reference system (ie, RS1) are taken as criterion.

f0 is robust against recording system, environmental noise, and their combination. This corroborates with earlier findings.[7,10–12,14,16] All other measures, however, were infected by both RS and noise condition, and especially by their combination. Furthermore, already in the reference and baseline condition of 20.5 dB_A ambient noise, the majority of the acoustic markers on MCD samples differ significantly from reference RS1 samples, even with highly acceptable SNR level. This contrasts with the finding that acoustic measures on RS1 samples encountered significant impacts only from much higher surrounding noise levels (ie, volumes unlikely to come upon in clinical assessment rooms). Based on these findings, the use of MCD for clinical voice assessment purpose warrants caution, especially for comparison of samples recorded by different RSs in, for example, treatment outcomes studies.

Table 4 revealed GNE and CPPS to be more easily infected by surrounding noise than the other markers. This came as no surprise because they both have been developed to be extra sensitive to noise (ie, glottal noise or perceived breathiness). In Hillenbrand et al,[25] and Hillenbrand and Houde,[26] CPPS outperformed other acoustic metrics in measuring breathiness severity, whereas in Fröhlich et al,[52] GNE emerged as most associated with breathiness ratings. Adding certain amounts of environmental noise triggered these two noise-responsive measures to react when other measures did not, yielding significant differences at 47.0 dB_A instead of 56.7 dB_A.

Even between clinical recording systems that are commercially available, commonly used and regarded as standard clinical tools, there is considerable variability in acoustic voice markers (eg, Karnell et al,[10] Bielamowicz et al,[11] and Smits et al[12]). Expecting MCDs to disagree less than such standard RSs with a criterion RS would be unfair, and therefore, they should not *a priori* be regarded as less proficient recording devices, as long as their SNR is acceptable for clinical assessment purposes and they are used with rigorous consideration for invariant microphone-to-mouth distance and placement, and ambient noise conditions.

Caveats and limitations

The strengths of the present study are following: First, multiple MCDs were examined, containing both tablet computers and smartphones, and covering different operating systems (ie, Android, Apple iOS, and Windows). Therefore, the results of the present study are considered representative for many of the MCDs in the current digital devices market. Second, all levels of dysphonia (ie, from absent to most severe) as well as varying age and both organic and nonorganic vocal and laryngeal pathology were represented in the 50 subjects. The study group therefore was regarded as sufficiently resembling a clinical population. Third, the performance of all recording systems was investigated in different situations with ambient noise levels ranging between almost absent to clearly audible, and thus this method is considered to appropriately mimic real environs. Fourth, this study focused not only on f0 and perturbation measures, but concentrated also on measures like GNE, CPPS, and AVQI and is therefore regarded to sufficiently reflect currently used protocols for acoustic voice assessment methods. Fifth, all experimental samples were recorded while controlling for microphone-to-loudspeaker distance and 45° off-axis placement, minimizing any variability in acoustic data due to microphone or device location.

However, several study weaknesses and caveats regarding the findings are noteworthy. First, the present study did not investigate any of the filtering properties of the recording devices (ie, how voice signals are treated by the recording systems), and therefore, it is impossible to explain why there were differences between them. It would thus be beneficial for future research to inspect which distinct recording properties and filter features cause acoustic voice measures to differ between (mobile and stationary) recordings systems. Second, because not all dysphonia severity gradations were equally represented across the continuum from absent to most severe—that is, the study group predominantly consisted of subjects with very slight to rather moderate degrees of dysphonia—present findings pertain to voices with these particular dysphonia grades. Possible nonlinearity of influence of recording-related noise across all dysphonia levels may therefore be investigated in future research. Third, we used clinically uncommon sound recording equipment (ie, AKG C544L, MPA V L, Focusrite Forte and MacBook Air) as reference systems. Nevertheless, these systems showed to record voice and speech with sufficient SNR outside clinical rooms in an earlier study[51] and were therefore retained as criterion systems in the present study. Fourth, although a diverse set of MCDs was investigated and results of the present study warrant caution in general regarding their clinical implementation, the influence of other MCDs on acoustic voice measures remains unknown until investigated. Fifth, regarding software, there are other mobile device applications available aside from the ones installed on the MCDs of the present study. Furthermore, only the acoustic metrics from the program *Praat* were investigated. Generalizing the findings to other recording apps and analysis software is difficult in the absence of data. Given the trend in conclusions from the present and earlier studies on interprogram as well as intersystem variability (eg, Carson et al,[7] Maryn et al,[13] and Hanschmann et al[17]), it is reasonable to proceed only with uncompressed samples with sufficient SNR. Sixth, next to the (combination of) time-, frequency- and quefrency-domain measures in this study, there are other appealing acoustic markers in the field of clinical voice assessment. Nonlinear dynamics measures, for instance, have often been studied (eg, Zhang et al[53,54]) but were not at our disposal during this study. Their susceptibility to effects of ambient noise and recording equipment also necessitates further investigation. Seventh, the present study only took statistical difference into account. Uloza et al[21] also found significant differences between RSs, but these differences appeared to be clinically irrelevant (ie, did not alter the acoustic markers' accuracy in differentiating between dysphonia and normophonia). Manfredi et al[22] also found strong agreement between recording two smartphones and a reference RS. Therefore, they minimized the clinical relevance of the differences. However, when comparing for example pretherapy and posttherapy data in a treatment outcomes investigation, it is recommended to rule out recording-related influences as much as possible and therefore to adhere to the same equipment. For motives like having voice care recipients recording themselves for extramural analysis,[22] or assisting and motivating patients to achieve therapeutic goals,[55] however, MCDs are particularly convenient and may bridge between clinic and real life.

CONCLUSION

With increased recognition and use of objective techniques in daily clinical practice, together with common availability of MCDs (eg, smartphones and tablet computers), it is expected that MCDs progressively will be used as clinical voice and speech acquisition instruments. However, due to combined differences in hardware, software, and ambient sound condition, acoustic voice quality measures differ between recording systems. The present authors therefore advice voice and speech clinicians to employ MCDs with critical caution, especially when comparing data between systems, and when using them for diagnostic purposes and tracking treatment outcomes.

REFERENCES

1. Behrman A. Common practices of voice therapists in the evaluation of patients. *J Voice*. 2005;19:454–469.
2. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013;22:212–226.
3. Parsa V, Jamieson DG. Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res*. 2001;44:327–339.
4. Buder EH. Acoustic analysis of voice quality: a tabulation of algorithms 1902–1990. In: Kent RD, Ball MJ, eds. *Voice Quality Measurement*. San Diego, CA: Singular Publishing Group; 2000:119–244.
5. Ingrisano DR-S, Perry CK, Jepson KR. Environmental noise: a threat to automatic voice analysis. *Am J Speech Lang Pathol*. 1998;7:91–96.
6. Perry CK, Ingrisano DR, Palmer MA, et al. Effects of environmental noise on computer-derived voice estimates from female speakers. *J Voice*. 2000;14:146–153.
7. Carson CP, Ingrisano DRS, 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.
8. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice*. 2005;19:15–28.
9. Vanhaelewyn B, Pauwels G, De Wolf P, et al. *Measuring Digital Media Trends in Flanders. iMinds Digimeter August–September 2015*. Ghent, Belgium: iMinds; 2015.
10. Karnell MP, Hall KD, Landahl KL. Comparison of fundamental frequency and perturbation measurements among three analysis systems. *J Voice*. 1995;9:383–393.
11. Bielamowicz S, Kreiman J, Gerratt BR, et al. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res*. 1993;39:126–134.
12. Smits I, Ceuppens P, De Bodt MS. Comparative study of acoustic voice measurements by means of Dr. Speech and Computerized Speech Lab. *J Voice*. 2005;19:187–196.
13. Maryn Y, Corthals P, De Bodt M, et al. Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat. *Folia Phoniatr Logop*. 2009;61:217–226.
14. Amir O, Wolf M, Amir N. A clinical comparison between two acoustic analysis softwares: MDVP and Praat. *Biomed Signal Process Control*. 2009;4:202–205.
15. Oğuz H, Kiliç MA, Şafak MA. Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turk J Med Sci*. 2011;41:835–841.
16. Mat Baki M, Wood G, Alston M, et al. Reliability of OperaVOX against Multidimensional Voice Program (MDVP). *Clin Otolaryngol*. 2014;40:22–28.
17. Hanschmann H, Gärtner S, Berger R. Comparability of computer-supported concurrent voice analysis. *Folia Phoniatr Logop*. 2015;67:8–14.
18. Lin E, Hornibrook J, Ormond T. Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatr Logop*. 2012;64:122–130.
19. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications. *J Acoust Soc Am*. 2014;135:EL186–EL192.
20. Vogel AP, Rosen KM, Morgan A, et al. Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. *Folia Phoniatr Logop*. 2014;66:244–250.
21. Uloza V, Padervinskis E, Vegiene A, et al. Exploring the feasibility of smart phone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Otorhinolaryngol*. 2015;272:3391–3399.
22. Manfredi C, Lebacq J, Cantarella G, et al. Smartphones offer new opportunities in clinical voice research. *J Voice*. 2016;doi:10.1016/j.jvoice.2015.12.020; Apr 7, [Epub ahead of print].
23. Titze IR, Winholtz WS. Effect of microphone type and placement on voice perturbation measurements. *J Speech Hear Res*. 1993;36:1177–1190.
24. Svec JG, Granqvist S. Guidelines for selecting microphones for human voice production research. *Am J Speech Lang Pathol*. 2010;19:356–368.
25. Hillenbrand J, Cleveland RA, Erickson RL. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*. 1994;37:769–778.
26. Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Hear Res*. 1996;39:311–321.
27. Michaelis D, Gramms T, Strube HW. Glottal-to-noise excitation ratio: a new measure for describing pathological voices. *Acta Acust*. 1997;83:700–706.
28. Michaelis D, Fröhlich M, Strube HW. Selection and combination of acoustic features for the description of pathologic voices. *J Acoust Soc Am*. 1998;103:1628–1639.

29. Maryn Y, Corthals P, Van Cauwenberge P, et al. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010;24:540–555.
30. 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.
31. Heman-Ackah YD, Heuer RJ, Michael DD, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324–333.
32. Godino-Llorente JI, Osma-Ruiz V, Sáenz-Lechón N, et al. The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders. *J Voice*. 2010;24:47–56.
33. Maryn Y, De Bodt M, Barsties B, et al. The value of the Acoustic Voice Quality Index as a measure of dysphonia severity in subjects speaking different languages. *Eur Arch Otorhinolaryngol*. 2014;271:1609–1619.
34. Heman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. *J Voice*. 2002;16:20–27.
35. Heman-Ackah YD, Sataloff RT, Laureyns G, et al. Quantifying the cepstral peak prominence, a measure of dysphonia. *J Voice*. 2014;28:783–788.
36. Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL J Otorhinolaryngol Relat Spec*. 2004;66:70–73.
37. Maryn Y, Roy N, De Bodt M, et al. Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am*. 2009;126:2619–2634.
38. van As-Brooks CJ, Koopmans-van Beinum FJ, Pols LC, et al. Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech. *J Voice*. 2006;20:355–368.
39. Hirano M. Psycho-acoustic evaluation of voice. In: Arnold G, Winckel F, Wyke B, eds. *Disorders of Human Communication 5. Clinical Examination of Voice*. Vienna, Austria: Springer Verlag; 1981:81–84.
40. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
41. Maryn Y, De Bodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord*. 2010;43:161–174.
42. Reynolds V, Buckland A, Bailey J, et al. Objective assessment of pediatric voice disorders with the Acoustic Voice Quality Index. *J Voice*. 2012;26:672, e1–7.
43. Barsties B, Maryn Y. [The Acoustic Voice Quality Index. Toward expanded measurement of dysphonia severity in German subjects]. *HNO*. 2012;60:715–720.
44. Maryn Y, Kim H-T, Kim J. Auditory-perceptual and acoustic methods in measuring dysphonia severity of Korean speech. *J Voice*. 2015;doi:10.1016/j.jvoice.2015.06.011; Aug 24, [Epub ahead of print].
45. Kankare E, Barsties B, Maryn Y, et al. A preliminary study of the Acoustic Voice Quality Index in Finnish speaking population. In: *Pan-European Voice Conference PEVOC*. 2015.
46. Heylen L, Mertens F. De klinische relevantie van het fonetogram: een onderzoek bij kinderen en leerkrachten. [The clinical relevance of the phonetogram: a study with children and teachers]. *Logopedie*. 1997;11:28–39.
47. Coleman R, Mabis J, Hinson J. Fundamental frequency-sound pressure level profiles of adult male and female voices. *J Speech Hear Res*. 1977;20:197–204.
48. Bosman A. *Speech perception by the hearing impaired* [doctoral dissertation]. Utrecht, The Netherlands: University of Utrecht; 1989.
49. Bosman A, Wouters J, Damman W. Realisatie van een CD voor spraakaudiometrie in Vlaanderen [Realization of a CD for speech audiometry in Flanders]. *Tijdschr Logop Audiol*. 1995;9:218–225.
50. Colton R. Vocal intensity in the modal and falsetto registers. *Folia Phoniatr (Basel)*. 1973;25:62–70.
51. Maryn Y, Zarowski A. Calibration of clinical audio recording and analysis systems for sound intensity measurement. *Am J Speech Lang Pathol*. 2015;24:608–618.
52. Frôhlich M, Michaelis D, Strube HW. Acoustic "breathiness measures" in the description of pathologic voices. In: *International Conference on Acoustics, Speech and Signal Processing ICASSP*. 1998.
53. Zhang Y, McGilligan C, Zhou L, et al. Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps. *J Acoust Soc Am*. 2004;115(5 pt 1):2270–2277.
54. Zhang Y, Jiang JJ, Biazzo L, et al. Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis. *J Voice*. 2005;19:519–528.
55. van Leer E, Pfister RC, Zhou X. An iOS-based cepstral peak prominence application: feasibility for patient practice of resonant voice. *J Voice*. 2016;doi:10.1016/j.jvoice.2015.11.022; Feb 2, [Epub ahead of print].
56. AKG Acoustics. *C544 User Instructions*. Vienna, Austria: AKG Acoustics; 2012.
57. Focusrite Audio Engineering. *Forte User Guide*. Bucks, UK: Focusrite Audio Engineering; 2012.