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*,