

Smartphones Offer New Opportunities in Clinical Voice Research

*C. Manfredi, †J. Lebacqz, ‡G. Cantarella, §J. Schoentgen, *S. Orlandi, *A. Bandini, and ¶P.H. DeJonckere, *Firenze, Italy, †Brussels, Belgium, ‡Milano, Italy, and §Belgium

Summary: Smartphone technology provides new opportunities for recording standardized voice samples of patients and sending the files by e-mail to the voice laboratory. This drastically improves the collection of baseline data, as used in research on efficiency of voice treatments. However, the basic requirement is the suitability of smartphones for recording and digitizing pathologic voices (mainly characterized by period perturbations and noise) without significant distortion. In this experiment, two smartphones (a very inexpensive one and a high-level one) were tested and compared with direct microphone recordings in a soundproof room. The voice stimuli consisted in synthesized deviant voice samples (median of fundamental frequency: 120 and 200 Hz) with three levels of jitter and three levels of added noise. All voice samples were analyzed using PRAAT software. The results show high correlations between jitter, shimmer, and noise-to-harmonics ratio measured on the recordings *via* both smartphones, the microphone, and measured directly on the sound files from the synthesizer. Smartphones thus appear adequate for reliable recording and digitizing of pathologic voices.

Key Words: smartphone–dysphonia–baseline design–voice assessment–synthetic voices.

INTRODUCTION

In recent years, the use of smartphones for clinical applications has gained increasing scientific interest thanks to advancements in digital technology, making these portable devices suitable for recording acoustic signals and transmitting the digitized audio files *via* e-mail. In the field of voice, digital technology enables a decisive improvement in audio quality compared with telephone transmission. For example, Uloza et al¹ explored the potential role of smartphone (Samsung Galaxy Note 3, Samsung, Daegu, South Korea) recordings in screening for laryngeal diseases and for subsequent referral of selected individuals for medical examination and visualization of the larynx, thus improving early diagnosis of laryngeal diseases. Guidi et al² showed that the quality of audio acquisitions from Samsung I9300 Galaxy S III smartphones was adequate for the investigation of fundamental frequency (Fo) features of running speech (using an *ad hoc* Android application) in subjects with bipolar mood disorders. Also, Lin et al³ found that iPhone (Apple A1303, Apple Inc., Cupertino, CA, USA) recordings are suitable for acoustic measurements of voice quality.

If the validity of voice analyses achieved on smartphone recordings (including the most inexpensive ones) sent *via* e-mail is confirmed in dysphonic patients (with a wide range of deviances), new opportunities are opened for clinical voice research: repeated measurements over time become possible without multiple visits to the voice laboratory. The patient can record his/her own voice (according to a standardized protocol) with his/her smartphone and send the audio file to the voice laboratory

by e-mail. Obviously, the basic requirement is the suitability of smartphones for recording and digitizing pathologic voices (that are mainly characterized by period perturbations and noise) without significant distortion. Thus, the investigation about the reliability of acoustic voice parameters obtained using smartphone microphones is of particular clinical interest.¹

Repeated measurements are the basis of single-subject and baseline designs. Multiple baseline designs are widely recognized in many areas of research as easily implemented, highly sensitive, and internally valid. Many areas of research in which randomized-group designs with blinding are disqualified by practical or ethical considerations are easily investigated using at least one of the variants of the multiple-baseline design.⁴ This applies in particular to clinical research in the field of voice pathology, all the more so as it has been shown that short-time variability of acoustic parameters of voice quality is far from negligible.⁵ Therefore, measuring the dependent variable at single time points (pre- and posttreatment measures) may provide a biased estimate.⁶ In a baseline design, voice quality parameters of a single subject are measured repeatedly and plotted as a function of time to establish a baseline (Figure 1). At a given time, a treatment is initiated while measurements are continued. If a change occurs, the posttreatment measurements determine a new baseline, the level of which will differ from the previous baseline. The term “multiple” refers to the fact that each patient included in the study has a different starting baseline, depending on the degree of severity of his/her dysphonia. Figure 1 illustrates this approach: in patient 1, after a set of eight consecutive measurements (showing spontaneous random variations) of a given parameter (eg, jitter % in a sustained /a:/), the treatment (eg, phonosurgery) occurs (fat arrow). An efficient treatment is expected to show a downward shift of the baseline (less jitter). Patient 2, included in the study 1 month later than patient 1, presents with a less deviant voice than patient 1. After five pretreatment measurements, the treatment is given, and again an improvement is noticed in the posttreatment baseline. In the case of a treatment of long duration (eg, voice therapy or antireflux medication), one may expect—in case of success—a downward slope of the regression line determined by the

Accepted for publication December 30, 2015.

From the *Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy; †Neurosciences Institute, University of Louvain, B-1200 Brussels, Belgium; ‡Otolaryngology Department, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy; §Department of Signals, Images and Acoustics, Faculty of Applied Sciences, Université Libre de Bruxelles, Belgium; and the ¶Neurosciences, University of Leuven & Federal Institute for Occupational Diseases, Brussels, Belgium.

Address correspondence and reprint requests to Philippe H. DeJonckere, Neurosciences, University of Leuven & Federal Institute for Occupational Diseases, Brussels, Belgium. E-mail: philippe.dejonckere@med.kuleuven.be

Journal of Voice, Vol. 31, No. 1, pp. 111.e1–111.e7
0892-1997

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
<http://dx.doi.org/10.1016/j.jvoice.2015.12.020>

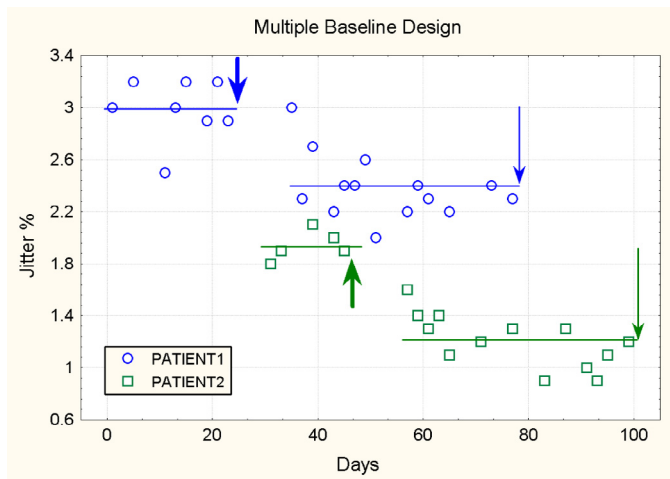


FIGURE 1. Schematic graph with data of two patients consecutively included in a multiple baseline design. In patient 1, eight measurements of a given voice quality parameter (here jitter %) precede the treatment (eg, phonosurgery; *fat arrow*) and define the “pre baseline.” After treatment, 14 additional measurements define the “post baseline”, which is shifted downward (*tiny arrow*), indicating a reduction of the jitter %. Notice that some “pre” measurement points indicate lower jitter values than some “post” measurement points. Patient 2 is included in the study 1 month after patient 1 and has a less deviant voice than patient 1. Here, treatment occurs after five measurements, and there are 13 “post” measurements.

posttreatment measurement points. Adequate statistical approaches are available to deal with such data.⁷

In this study, two smartphones (a very basic, inexpensive one and a high-level one) are tested and compared with direct microphone recordings in a soundproof room. The comparison concerns the main basic acoustical parameters of clinical interest: cycle-to-cycle perturbation (jitter and shimmer) and noise.⁵ Of prime importance is also the voice material. In this experiment, we used synthesized deviant voices that have the advantage of an exact calibration of period perturbation parameters as well as of noise. Such samples have been used in checking the adequacy of voice analysis programs^{8–12} and are used here to test the reliability of the two smartphones.

MATERIALS AND METHODS

Synthesizer

The synthesizer uses a model of the glottal area based on a polynomial distortion function that transforms two excitatory harmonic functions into the desired waveform.^{13,14} The polynomial coefficients are obtained by constant, linear, and invertible transforms of the Fourier series coefficients of the Klatt template cycle that is asymmetric and skewed to the right.¹⁴ This waveform is in fact typical for the glottal area cycle, allowing a maximal glottal area of 0.2 cm². The discrete phase of the harmonic excitation functions changes from iteration to iteration with a step defined by the inverse of the sampling frequency. The sampling frequency is set at 200 kHz to simulate voices, the frequency modulation of which is of the order of 1% of the Fo, thus requiring high temporal resolution. The harmonic excitation functions

are low-pass filtered and down sampled to 50 kHz before their transformation by the distortion function. To simulate voice perturbations as jitter, phase, and/or amplitude fluctuations, disturbances of the harmonic excitation functions are introduced. Specifically, jitter is simulated with a model based on low-pass filtered white noise of adjustable size. The noise signal is obtained by adding pulsatile or aspiration noise to the clean flow rate. Pulsatile noise simulates additive noise due to turbulent airflow in the vicinity of the glottis and its size evolves proportionally to the glottal volume velocity. It is obtained by low-pass filtering white Gaussian noise, the samples of which are multiplied by the clean glottal volume velocity. Low-pass filtering is performed with linear second order filters. Additive noise is measured as the noise-to-harmonics ratio (NHR) of the clean volume velocity signal at the glottis relative to the noise. The synthesizer also generates varying levels of shimmer that automatically increases when jitter increases. Indeed, jitter and shimmer are physiologically linked with each other. Once the glottal area has been obtained, the flow rate is simulated taking into account a model of the glottal impedance and tract load.¹⁵ Each formant is modeled with a second-order bandpass filter. The vocal tract transfer function is obtained by cascading several second-order filters including the nasal and tracheal formants, the frequencies and bandwidths of which are fixed.¹⁶ The bandwidths of the vocal tract formants are calculated *via* the formant frequencies.¹⁷ The radiation at the lips is also taken into account *via* a high-pass filter. The signal is then normalized, dithered, quantized, converted into “.wav” format, and stored on the computer hard disk.

Synthetic voices

The synthesized deviant voice samples consist of sustained /a:/ utterances at a median Fo = 120 and 200 Hz, of 2 seconds of duration with a slight falling and rising intonation, and with three levels of jitter: 0.9%, 2.8%, and 4.5%. For each level of jitter, three levels of additive noise are considered, with a flow rate to aspiration noise of respectively 97.5 dB, 23.8 dB, and 17.6 dB. They perceptually correspond to usual dysphonic patients’ voices, from slightly to severely deviant, rough, as well as breathy. Shimmer increases from about 7% up to 23% with increasing jitter.

Smartphones

The devices were selected at the extremes of the commercial price range. The more expensive one is an HTC One (named hereafter Smart1), the basic, inexpensive one is a Wiko model CINK SLIM2 (named Smart2). Relevant specifications are given in Table 1. Price ratio is 1/15.

Microphone

The microphone is a Sennheiser model MD421U (Wedemark, Germany) (frequency response 30–17,000 Hz) commonly used in the voice laboratory to make recordings of voice patients.

Amplifier and loudspeaker

A Bowers & Wilkins model CM1 (Worthing, UK) loudspeaker was used. Its frequency response is flat ± 1.5 dB between 50 Hz and 20 kHz. It was driven by a Yamaha UK YHT-380 amplifier (Hamamatsu, Japan). The frequency response of the amplifier is

TABLE 1.
Technical Specifications of the Two Smartphones Used in this Experiment

Specification	HTC One (Smart 1)	Wiko (Smart 2)
Operating system	Android 5.0.2	Android 4.2
CPU type	Quad-core	MediaTek MT6572 2-core
CPU freq.	1.7 GHz	1.3 GHz

Abbreviation: CPU, Central Processing Unit.

flat ± 0.5 dB between 20 Hz and 20 kHz, 0.06% total harmonic distortion (THD). The auxiliary (AUX) input of the amplifier was used to reproduce sounds from the synthesizer. The smartphones were fixed on a stand at a 10-cm distance from the center of the loudspeaker. The sound intensity of the loudspeaker was adjusted to 60 dB to correspond to the loudness of human voice.

Soundproof booth

All recordings were made in an IAC Mini 350 (International Acoustics Company, Winchester, UK) soundproof booth certified according to the European Norms, International Organization for Standardization (EN ISO) 9001/14,001 norm.

Analysis program

All data were analyzed with *PRAAT*, a software tool freely available online that enables analysis, synthesis, and manipulation of voice signals (www.praat.org). *PRAAT* has been exhaustively tested with synthetic deviant voices.⁸⁻¹²

RESULTS

Figures 2 and 3 show the jitter values obtained with the two smartphones (Smart1 and Smart2) and the Sennheiser microphone (Micro) as a function of the jitter directly measured on

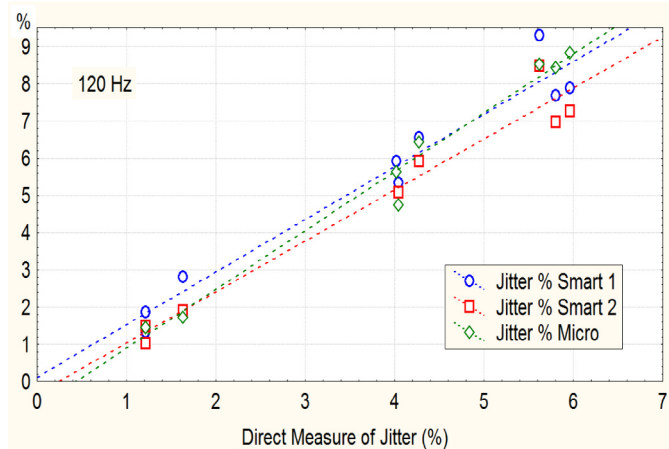


FIGURE 2. Jitter measurements (for the 120 Hz samples) as obtained via smartphone 1, smartphone 2, and Sennheiser microphone plotted against the direct measurement from the synthesized signal. For each device (smartphones and microphone), three jitter levels are measured, and for each level there are three noise levels.

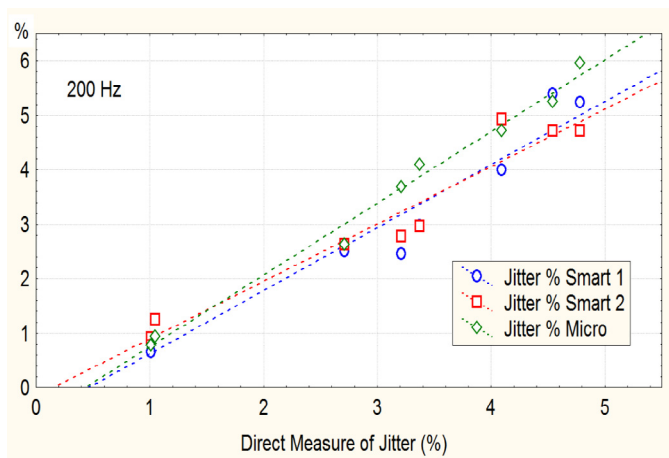


FIGURE 3. Jitter measurements (for the 200 Hz samples) as obtained via smartphone 1, smartphone 2, and Sennheiser microphone plotted against the direct measurement from the synthesized signal. For each device (smartphones and microphone), three jitter levels are measured, and for each level there are three noise levels.

the synthesized signal, for 120 and 200 Hz, respectively. For each jitter level (0.9%, 2.8%, and 4.5%) there are three levels of additive noise (0.2 dB, 0.6 dB, and 0.8 dB). Therefore, three measures are obtained for each device. The figure shows a systematic moderate overestimation of the higher values of jitter. The results obtained with the three devices for the three jitter levels are nevertheless highly correlated. Figures 4 and 5 show—for 120 and 200 Hz respectively—NHR values obtained in the same way with Smart1, Smart2, and the Sennheiser microphone as a function of the direct measurement of additive noise by the software program *PRAAT* on the signal generated by the synthesizer (for each noise level there are three levels of jitter). There is a slight quasi-systematic underestimation of the amount of noise, but again, the results of the three conditions are highly correlated.

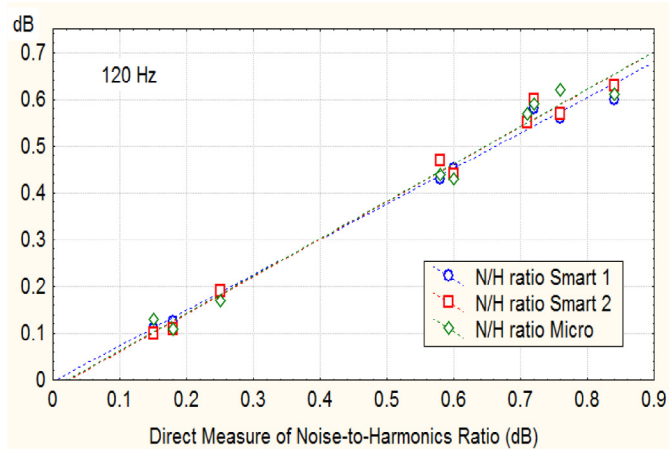


FIGURE 4. Noise-to-harmonics ratio (N/H) ratio measurements (for the 120 Hz samples) as obtained via smartphone 1, smartphone 2, and Sennheiser microphone plotted against the direct measurement from the synthesized signal. For each device (smartphones and microphone), three noise levels are measured, and for each level there are three jitter levels.

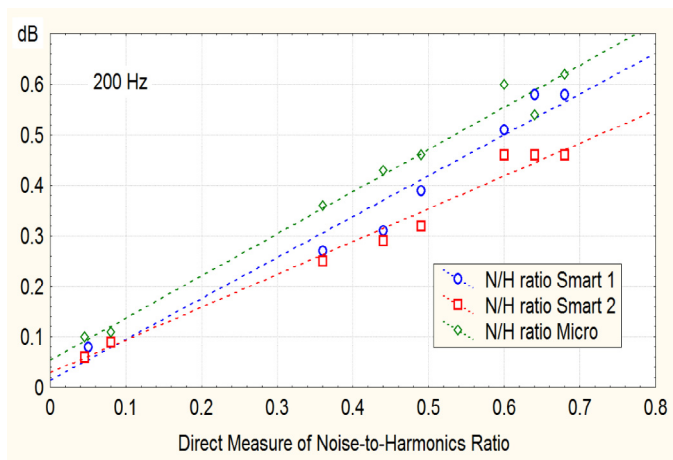


FIGURE 5. Noise-to-harmonics ratio (N/H) ratio measurements (for the 200 Hz samples) as obtained *via* smartphone 1, smartphone 2, and Sennheiser microphone plotted against the direct measurement from the synthesized signal. For each device (smartphones and microphone), three noise levels are measured, and for each level there are three jitter levels.

Table 2 (120 Hz) and Table 3 (200 Hz) show the correlation matrices among measures: Jit k, Shim k, and NHR k, k = 1,2,3 and 4 respectively corresponding to Smartphone 1, Smartphone 2, Sennheiser microphone, and the direct measurement by the software program on the signal generated by the synthesizer. All relevant correlations are very strong ($P < 0.01$).

The intraclass correlation coefficients (ICC) have also been computed.¹⁸ An ICC assesses the reliability of measurement tools by comparing the variability of different measures of the same subject with the total variation across all measures and all subjects. ICC has advantages over correlation coefficient, as more than two measurement tools can be included, and as it is adjusted for the effects of the scale of measurements. An ICC of

0–0.2 indicates poor agreement, 0.3–0.4 indicates fair agreement, 0.5–0.6 indicates moderate agreement, 0.7–0.8 indicates strong agreement, and >0.8 indicates almost perfect agreement. Here, the lowest ICC is 0.81 (Table 4).

The difference plot or Bland-Altman plot¹⁹ is recognized as a relevant method to show agreement between two measurement techniques. The difference between data pairs is graphically presented against their average, and the limits of agreement (mean \pm 1.96 standard deviation) are defined on the plots. In general, the Bland-Altman plot is useful to reveal a relationship between the differences and the magnitude of measurements, to look for any systematic bias, and to identify possible outliers. Figures 6–13 provide the Bland-Altman plots for the jitter percentages and NHR measurements computed by PRAAT for both smartphones, each of them compared with the Sennheiser microphone (120 and 200 Hz). The plots show good agreements, and none of the graphs demonstrates a significant bias, heteroscedasticity, or trend.

DISCUSSION

This paper presents an innovative approach to test the reliability of commercial smartphones in assessing voice quality in clinical applications. Two smartphones selected at the extremes of the commercial price range available today on the market are compared with a high-quality microphone as far as jitter and noise estimation are concerned. Measurements are made on synthetic voice signals and the results are compared with the direct measurement by the software program on the signal generated by the synthesizer. The results show that additive noise has an almost negligible effect on jitter estimation. Similarly, there is almost no effect of jitter on noise estimation. Both smartphones tested—hence also a basic one—record and digitize without significant distortion pathologic voices in a wide range of deviance, as well with respect to perturbation as to additive noise. The chosen median Fo values further let us suppose that there is no difference between male and female voices.

TABLE 2. Correlation Matrix: Measures of Jitter %, Shimmer %, and N/H Ratio Obtained Respectively *via* Smartphone 1 (1), Smartphone 2 (2), Sennheiser Microphone (3), and by Direct Measurement on the Signal Generated by the Synthesizer (4) (120 Hz), for the 120 Hz Samples

	JIT 1	SHIM 1	N/H 1	JIT 2	SHIM 2	N/H 2	JIT 3	SHIM 3	N/H 3	JIT 4	SHIM 4	N/H 4
JIT 1	1.00	.87	.97	1.00	.91	.97	.98	.88	.97	.97	.89	.95
SHIM 1	.87	1.00	.84	.85	.90	.84	.82	.91	.84	.82	.96	.81
N/H 1	.97	.84	1.00	.97	.92	1.00	.97	.93	.99	.97	.89	.99
JIT 2	1.00	.85	.97	1.00	.89	.97	.98	.87	.97	.98	.88	.96
SHIM 2	.91	.90	.92	.89	1.00	.93	.85	.94	.90	.86	.87	.90
N/H 2	.97	.84	1.00	.97	.93	1.00	.97	.93	.99	.98	.89	.99
JIT 3	.98	.82	.97	.98	.85	.97	1.00	.87	.98	.99	.88	.97
SHIM 3	.88	.91	.93	.87	.94	.93	.87	1.00	.91	.89	.95	.93
N/H 3	.97	.84	.99	.97	.90	.99	.98	.91	1.00	.98	.89	.99
JIT 4	.97	.82	.97	.98	.86	.98	.99	.89	.98	1.00	.89	.98
SHIM 4	.89	.96	.89	.88	.87	.89	.88	.95	.89	.89	1.00	.88
N/H 4	.95	.81	.99	.96	.90	.99	.97	.93	.99	.98	.88	1.00

Notes: JIT 1, JIT 2, JIT 3, and JIT 4 are the jitter % obtained with Smartphone 1, Smartphone 2, Sennheiser microphone and by direct measurement respectively. SHIM 1, SHIM 2, SHIM 3, and SHIM 4 are the shimmer % obtained with Smartphone 1, Smartphone 2, Sennheiser microphone, and by direct measurement respectively. N/H 1, N/H 2, N/H 3, and N/H 4 are the N/H ratio obtained with Smartphone 1, Smartphone 2, Sennheiser microphone, and by direct measurement respectively. All relevant correlation coefficients exceed 0.9 ($P < 0.01$).

TABLE 3.

Correlation Matrix: Measures of Jitter %, Shimmer %, and N/H Ratio Obtained Respectively via Smartphone 1 (1), Smartphone 2 (2), Sennheiser Microphone (3), and by Direct Measurement on the Signal Generated by the Synthesizer (4) (120 Hz), for the 120 Hz Samples

	JIT 1	SHIM 1	N/H 1	JIT 2	SHIM 2	N/H 2	JIT 3	SHIM 3	N/H 3	JIT 4	SHIM 4	N/H 4
JIT 1	1.00	.88	.99	.97	.86	.97	.97	.86	.94	.97	.91	.88
SHIM 1	.88	1.00	.93	.91	.97	.94	.92	.98	.94	.93	.97	.71
N/H 1	.99	.93	1.00	.98	.92	.99	.99	.90	.97	.99	.92	.86
JIT 2	.97	.91	.98	1.00	.92	.99	.96	.87	.97	.97	.90	.90
SHIM 2	.86	.97	.92	.92	1.00	.95	.91	.95	.95	.93	.94	.78
N/H 2	.97	.94	.99	.99	.95	1.00	.98	.91	.99	.99	.92	.88
JIT 3	.97	.92	.99	.96	.91	.98	1.00	.92	.98	1.00	.93	.84
SHIM 3	.86	.98	.90	.87	.95	.91	.92	1.00	.93	.92	.96	.68
N/H 3	.94	.94	.97	.97	.95	.99	.98	.93	1.00	.99	.92	.85
JIT 4	.97	.93	.99	.97	.93	.99	1.00	.92	.99	1.00	.93	.87
SHIM 4	.91	.97	.92	.90	.94	.92	.93	.96	.92	.93	1.00	.78
N/H 4	.88	.71	.86	.90	.78	.88	.84	.68	.85	.87	.78	1.00

Notes: JIT 1, JIT 2, JIT 3, and JIT 4 are the jitter % obtained with Smartphone 1, Smartphone 2, Sennheiser microphone and by direct measurement respectively. SHIM 1, SHIM 2, SHIM 3, and SHIM 4 are the shimmer % obtained with Smartphone 1, Smartphone 2, Sennheiser microphone, and by direct measurement respectively. N/H 1, N/H 2, N/H 3, and N/H 4 are the N/H ratio obtained with Smartphone 1, Smartphone 2, Sennheiser microphone, and by direct measurement respectively. All relevant correlation coefficients exceed 0.85 ($P < 0.01$).

Collecting voice samples of dysphonic patients by using a patient’s own smartphone offers a convenient way of increasing the sample size in single-case designs and multiple-baseline designs. This is of particular relevance when investigating ef-

ficacy of treatments. As a matter of fact, by multiplying the sample size by a factor x , the variance of the estimation is typically going down by a factor $1/Vx$. Hence, the reliability of levels and/or trends will be increased, and consequently, so does the statistical sensitivity. Smartphone recordings can also be useful in voice monitoring and in investigating fleeting (reversible) effects on voice of loading and fatigue, which are absent when the patient visits the voice clinic, while being relevant in context of prevention of, for example, occupational voice disorders. For example, a kindergarten teacher with voice complaints pertaining to his/her occupational voice use can be asked to make a short standardized recording before starting a working day and just after a morning of vocal activity. Finally, smartphone recordings can also be used for sequential measurements of other parameters included in standard guidelines for voice

TABLE 4.

Intraclass Correlation Coefficients Among Measures Obtained via Smartphone 1, Smartphone 2, Sennheiser Microphone, and by Direct Measurement on the Signal Generated by the Synthesizer, Respectively for the 120 and 200 Hz Voice Samples

ICC	Jitter	Shimmer	N/H Ratio
120 Hz	0.89	0.81	0.91
200 Hz	0.95	0.87	0.92

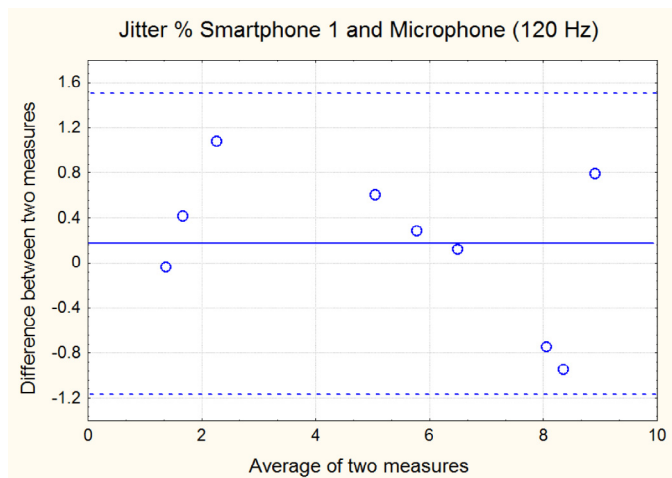


FIGURE 6. Bland-Altman plot for a comparison of jitter % measurements obtained via smartphone 1 and Sennheiser microphone (120 Hz).

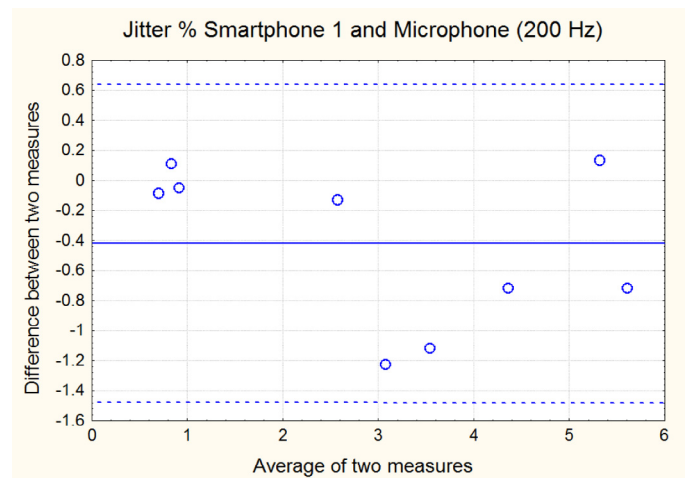


FIGURE 7. Bland-Altman plot for a comparison of jitter % measurements obtained via smartphone 1 and Sennheiser microphone (200 Hz).

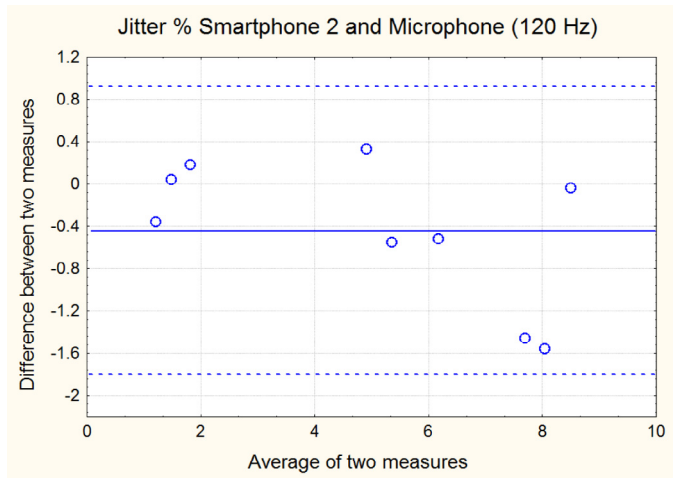


FIGURE 8. Bland-Altman plot for a comparison of jitter % measurements obtained *via* smartphone 2 and Sennheiser microphone (120 Hz).

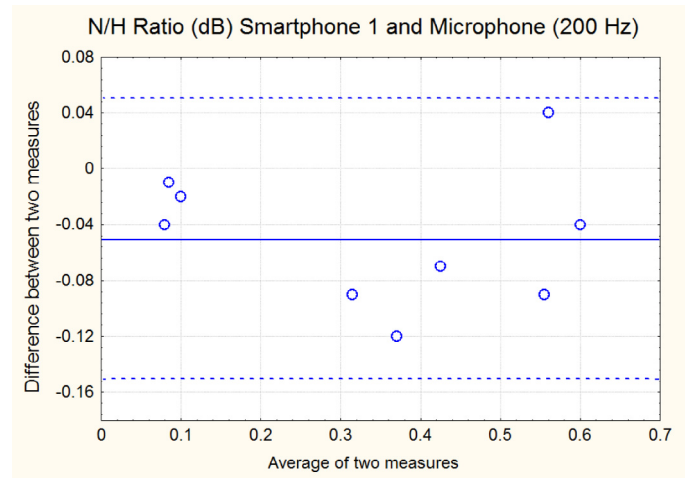


FIGURE 11. Bland-Altman plot for a comparison of Noise-to-harmonics ratio (N/H) measurements obtained *via* smartphone 1 and Sennheiser microphone (200 Hz).

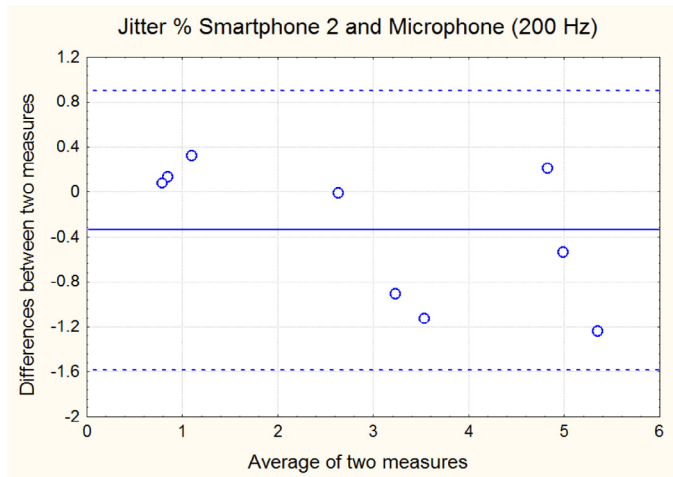


FIGURE 9. Bland-Altman plot for a comparison of jitter % measurements obtained *via* smartphone 2 and Sennheiser microphone (200 Hz).

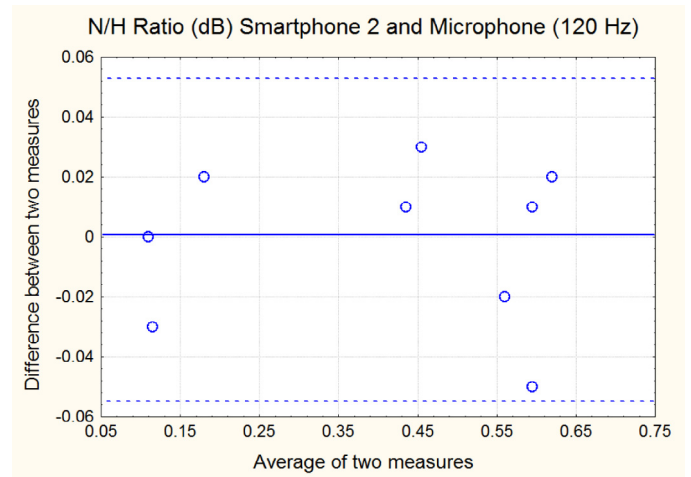


FIGURE 12. Bland-Altman plot for a comparison of Noise-to-harmonics ratio (N/H) measurements obtained *via* smartphone 2 and Sennheiser microphone (120 Hz).

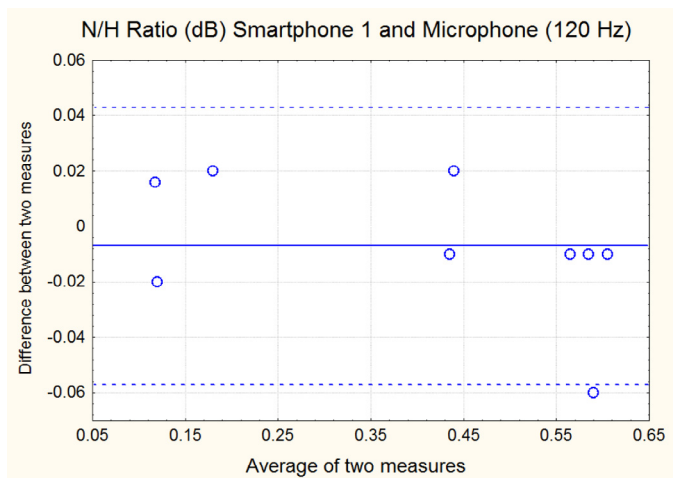


FIGURE 10. Bland-Altman plot for a comparison of Noise-to-harmonics ratio (N/H) measurements obtained *via* smartphone 1 and Sennheiser microphone (120 Hz).

assessment:^{20,21} perceptual rating, maximal phonation time, lowest and highest Fo, and others.

Out of the scope of this article, but of obvious importance, is the protocol for recording the voice samples, which needs to be standardized, as well as the voice material and the environmental conditions. Particularly, ambient noise needs to be controlled, but this is made possible by current smartphone technology, as there exist applications providing the ambient Sound Pressure Level (SPL) level.

CONCLUSIONS

The use of smartphones could provide considerable advantages to clinical voice researchers, mainly by increasing the number of voice samples from a given patient in an easy and inexpensive way, which is essential for baseline designs and for voice monitoring. In this work, we demonstrate the reliability of smartphones with regard to quality of the recordings across a wide range of degrees of deviance (perturbation and additive

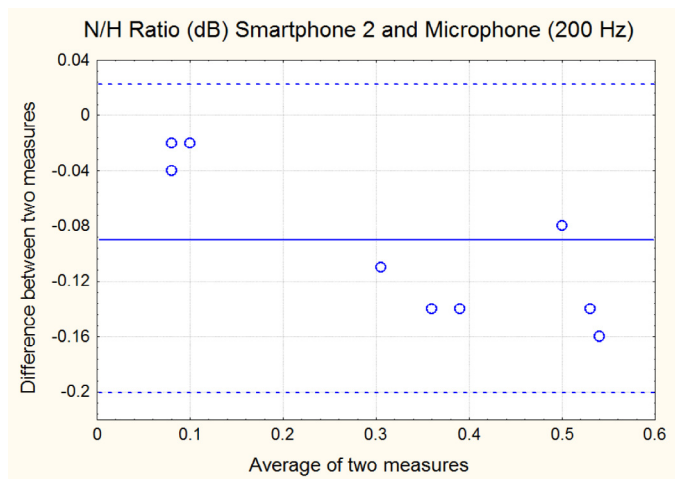


FIGURE 13. Bland-Altman plot for a comparison of N/H measurements obtained via smartphone 2 and Sennheiser microphone (200 Hz).

noise) and in the male/female ranges of F_0 values. The comparison is carried out using synthesized voice signals, which guarantees exact knowledge of reference values for voice quality parameters. Reliability is in fact the basic requirement for the use of smartphones in the transmission of audio signals from the patient to the voice laboratory for the analysis of voice quality. Furthermore, it may be assumed that every type of smartphone is adequate, as we selected two smartphones at the extremes of the commercial price range available today on the market.

Acknowledgment

The authors thank Professor Thomas Bruss (Free University Brussels, Belgium) for his valuable advice.

REFERENCES

1. Uloza V, Padervinskis E, Vegiene A, et al. Exploring the feasibility of smartphone microphone for measurement of acoustic voice parameters and voice pathology screening. *Eur Arch Otorhinolaryngol*. 2015;272:3391–3399.
2. Guidi A, Salvi S, Ottaviano M, et al. Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: system performance evaluation and case study. *Sensors (Basel)*. 2015;15:28070–28087.
3. Lin E, Hornibrook J, Ormond T. Evaluating iPhone recordings for acoustic voice assessment. *Folia Phoniatr Logop*. 2012;64:122–130.
4. Huitema BE. *The Analysis of Covariance and Alternatives: Statistical Methods for Experiments, Quasi-experiments, and Single-Case Studies*. 2nd ed. Hoboken: John Wiley & Sons; 2011.
5. Speyer R, Wieneke GH, DeJonckere PH. The use of acoustic parameters for the evaluation of voice therapy for dysphonic patients. *Acta Acust United Ac*. 2004;90:520–527.
6. Cente DB, Leach RA. The multiple baseline across subjects design: proposed use in research. *J Manipulative Physiol Ther*. 1984;7:231–236.
7. Rhoda DA, Murray DM, Andridge RR, et al. Studies with staggered starts: multiple baseline designs and group-randomized trials. *Am J Public Health*. 2011;101:2164–2169.
8. DeJonckere PH, Schoentgen J, Giordano A, et al. Validity of jitter measures in nonquasi-periodic voices. Part I. Perceptual and computer performances in cycle pattern recognition. *Logoped Phoniatr Vocol*. 2011;36:70–77.
9. Manfredi C, Giordano A, Schoentgen J, et al. Validity of jitter measures in non-quasi-periodic voices. Part II. The effect of noise. *Logoped Phoniatr Vocol*. 2011;36:78–89.
10. DeJonckere PH, Giordano A, Schoentgen J, et al. To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Biomed Signal Process Control*. 2012;7:37–42.
11. Manfredi C, Giordano A, Schoentgen J, et al. Perturbation measurements in highly irregular voice signals: performances/validity of analysis software tools. *Biomed Signal Process Control*. 2012;7:409–416.
12. Rruqja N, DeJonckere PH, Cantarella G, et al. Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia. *Biomed Signal Process Control*. 2014;13:71–78.
13. Schoentgen J. Non-linear signal representation and its application to the modeling of the glottal waveform. *Speech Commun*. 1990;9:189–201.
14. Schoentgen J. Shaping function models of the phonatory excitation signal. *J Acoust Soc Am*. 2003;114:2906–2912.
15. Ananthapadmanaba T, Fant G. Calculation of true glottal flow and its components. *Speech Commun*. 1982;1:167–184.
16. Klatt D. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am*. 1980;67:971–995.
17. Hawks JW, Miller JD. A formant bandwidth estimation procedure for vowel synthesis. *J Acoust Soc Am*. 1995;97:1343–1344.
18. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications and Practice*. Norwalk: Appleton & Lange; 1993:509–516 ISBN 0-8385-1065-5.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
20. DeJonckere PH, Bradley P, Clemente P, 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). *Eur Arch Otorhinolaryngol*. 2001;258:77–78.
21. American Speech-Language-Hearing Association. 2013. Available at: <http://www.asha.org/uploadedFiles/slp/healthcare/AATVoiceEvaluation.pdf>.