Maximal Ambient Noise Levels and Type of Voice Material Required for Valid Use of Smartphones in Clinical Voice Research

***Jean Lebacq, †Jean Schoentgen, ‡Giovanna Cantarella, §Franz Thomas Bruss, ¶Claudia Manfredi, and #Philippe DeJonckere,** **†§#Brussels, Belgium, and ‡Milano and ¶Firenze, Italy*

Summary: Purpose. Smartphone technology provides new opportunities for recording standardized voice samples of patients and transmitting the audio files to the voice laboratory. This drastically improves the achievement of baseline designs, 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 a previous article, this was tested using realistic synthesized deviant voice samples (/a:/) with three precisely known levels of jitter and of noise in all combinations. High correlations were found between jitter and noise to harmonics ratio measured in (1) recordings via smartphones, (2) direct microphone recordings, and (3) sound files generated by the synthesizer. In the present work, similar experiments were performed (1) in the presence of increasing levels of ambient noise and (2) using synthetic deviant voice samples (/a:/) as well as synthetic voice material simulating a deviant short voiced utterance (/aiuaiuaiu/).

Results. Ambient noise levels up to 50 dB_A are acceptable. However, signal processing occurs in some smartphones, and this significantly affects estimates of jitter and noise to harmonics ratio when formant changes are introduced in analogy with running speech. The conclusion is that voice material must provisionally be limited to a sustained /a/.

Key Words: Smartphone–Dysphonia–Recording–Noise–Acoustics.

INTRODUCTION

In recent years, the use of smartphones and web-based systems for clinical applications has gained increasing scientific interest, thanks to developments in digital technology, making these devices suitable for recording acoustic signals and transmitting the digitized audio files.[1,2] Specifically, as far as voice is concerned, digital technology enables a decisive improvement in audio quality compared with telephone transmission. Smartphones are pocket-sized highly mobile computers; they contain the required interfaces for easy voice recording at home or on site. Transmission can be web-based and is no longer restricted by bandwidth limitations of the telephonic pathway.

For general information about potential use of smartphones in pathologic voice research, particularly to help in carrying out single-case designs and multiple baseline designs, the reader is referred to our previous paper.[1]

In a first experiment, we demonstrated the reliability of smartphones with regard to quality of recordings over a wide range of degrees of deviance (perturbation and additive noise) and in the male and female ranges of fundamental frequency (F0) values. The comparison was carried out using realistic synthesized voice signals (sustained /a:/ altered by three levels of jitter and three levels of noise, the two basic acoustic voice quality parameters), which guarantee exact knowledge of reference values for voice quality parameters. The absence of significant distortion by the smartphone (during recording or data processing) is the basic requirement for the use of such devices in the transmission of audio signals from the patient to the voice laboratory for analysis of deviant voice quality. Furthermore, it was assumed that all types of smartphones were likely to be adequate, and we selected two smartphones at the extremes of the commercially available price range. However, our experiments were conducted in a laboratory setting, that is, in a soundproof booth. It was mentioned that for clinical purposes, the sound pressure level of the ambient noise should be controlled while recording voice samples. This is made possible by current smartphone technology (sound measurement applications or "apps"), although so far only some "apps" are really accurate.[3] As regards noise, very recently, Maryn et al[4] found that chains of dysphonic sustained vowels and continuous speech recorded by means of mobile communication devices (two tablet computers and three smartphones) were significantly impacted by ambient noise. They concluded that, due to combined differences in hardware, software, and ambient sound conditions, acoustic voice quality measures may differ between recording systems.

For a voice laboratory setting with direct recording, Deliyski et al[5] found that a level of noise in the acoustic environment of <46 dB was to be recommended, and <58 dB was acceptable. However, practical limits of tolerable ambient noise intensity values for the application considered here—that is, a patient recording his or her voice at home or at the workplace for sending to the voice clinic—are not accurately known. Furthermore, in our first paper,[1] only sustained /a:/ was used as voice material. It is thus worthwhile to assess the extent to which smartphones possibly distort synthesized samples comparable with natural voice

Accepted for publication February 24, 2017.

From the *Neurosciences Institute, University of Louvain, Brussels, Belgium; †B.E.A.M.S. Department, Faculty of Applied Sciences, Université Libre de Bruxelles, Brussels, Belgium; ‡Otolaryngology Department, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy; §Department of Mathematics, Université Libre de Bruxelles, Brussels, Belgium; ¶Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy; and the #Neurosciences, University of Leuven (KULeuven) and FEDRIS (Federal Agency for Occupational Risks), Brussels, Belgium.

Address correspondence and reprint requests to Philippe DeJonckere, Neurosciences, University of Leuven and FEDRIS (Federal Agency for Occupational Risks), Brussels, Belgium. E-mail: philippe.dejonckere@kuleuven.be

Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.

http://dx.doi.org/10.1016/j.jvoice.2017.02.017

productions, for which different reference values for F0 perturbation and noise to harmonics (N:H) ratio are exactly known.

In the first part of this work, the same synthetic voice signals as those used previously (sustained /a:/) were recorded simultaneously by two smartphones in the presence of stepwise increasing levels of ambient noise. The smartphone audio files were sent by e-mail and analyzed using *Praat*. The results of jitter % and N:H ratio could then be compared with those of the direct recording (microphone without added ambient noise) and with those of the direct analysis of the original signal (from computer to computer). In the second part of the present experiments, the sustained /a:/ was replaced by relatively realistic synthetic test signals consisting of a sequence of three repetitions of the fragment /aiu/ with intonation and formant changes, simulating a short voiced utterance, as frequently used in clinical practice.[6,7] Again, three levels of jitter and three levels of noise were introduced in the signals, and the same comparisons were made: the results of jitter % and N:H ratio in the audio files of the smartphones were compared with those of the direct recording (microphone without added ambient noise) and with those of the direct analysis of the original signal (from computer to computer).

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.[8,9] The polynomial coefficients are obtained by constant, linear, and invertible transforms of the Fourier series coefficients of Klatt's template cycle that is asymmetric and skewed to the right.[9] This waveform is in fact typical for the glottal area cycle, allowing a maximal glottal area of 0.2 cm^2. The discrete phase increment of the harmonic excitation functions evolves proportionally to the instantaneous vocal frequency F0. The sampling frequency is set at 200 kHz to simulate voices, the frequency modulation of which is of the order of 1% of the fundamental frequency F0, 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, 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 noisy 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 N:H ratio of the clean volume velocity signal at the glottis relative to the noise. The synthesizer also generates varying levels of shimmer via modulation distortion by the vocal tract transfer function, which automatically increases when jitter increases. Indeed, jitter and shimmer are acoustically linked to each other. Once the glottal area has been obtained, the glottal flow rate is calculated numerically via the interactive voice source model proposed by Rothenberg, which takes into account the glottal impedance and tract load.[10] 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.[11] The bandwidths of the vocal tract formants are calculated via the formant frequencies.[12] The first three formant frequency values have been equal to 640 Hz, 1212 Hz, and 2254 Hz for [a], 230 Hz, 2000 Hz, and 3000 Hz for [i], as well as 298 Hz, 730 Hz, and 2172 Hz for [u].The radiation at the lips is simulated via a high-pass filter. The signals are then normalized, dithered, quantized, converted into ".wav" format, and stored on the computer hard disk.

Synthetic voices

The synthesized deviant voice samples consisted of sustained /a:/ samples at a median F0 of 120 Hz 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 added noise were considered: the lowest level corresponding to a volume velocity to noise ratio at the glottis equal to 17 dB, the intermediate level equal to 23 dB, and the highest level equal to 90 dB. These true levels correspond to numeric N:H ratios obtained via *Praat* equal to 0.2, 0.6, and 0.8, respectively. Perceptually, they correspond to common dysphonic patients' voices, from slightly to severely deviant, rough as well as breathy.

In the second part of the present work, the sustained /a:/ was replaced by realistic synthetic test signals consisting of a sequence of three repetitions of the fragment /aiu/ with slight intonation and formant changes from one stimulus to the next, simulating a short voiced utterance of 4.4 seconds.[13] The jitter and noise levels were similar to those for the /a:/. An example of the spectrogram of an /aiu/ utterance obtained with the *Praat* program is given in Figure 1 (F0 median: 120 Hz; jitter: 3.35%; N:H ratio: 0.28).

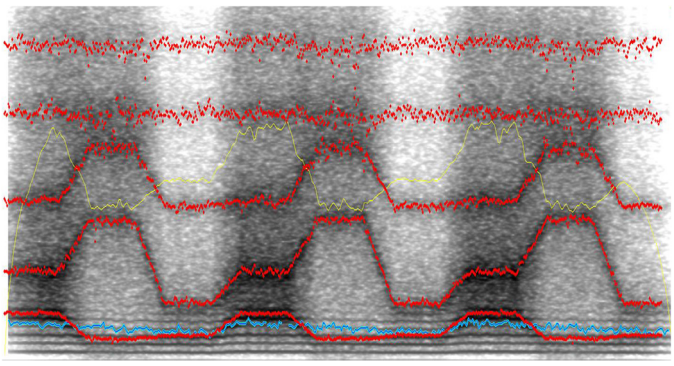

FIGURE 1. Example of spectrogram (0–5 KHz) of a synthetic voice sample (3×/aiu/). Duration: 4.4 seconds; average F0: 120 Hz; jitter %: 3.35%; noise to harmonics ratio: 0.28. *Blue dots*: F0 (total scale = 0–500 Hz linear; slight intonation). *Red dots*: formant locations (total scale = 0–5000 Hz linear). *Yellow line*: intensity (total scale 50–100 dB linear). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Perceptual validation

The realism of the synthetic /aiu/ utterances was verified by means of a perceptual test with 14 bachelor students in speech-language therapy. Ten pairs of audio samples were created, each with two different utterances (/aiuaiuaiu/) selected out of the 18 (9 of 120 Hz and 9 of 200 Hz) used in the present experiments. After adequate information (research on pathologic voice synthesis) and an audio example, the 10 pairs were presented to the students in a small lecture room using hi-fi audio material. After listening to each pair (sample A and sample B), the students were asked to choose, on a prepared form, among four possibilities: (1) both A and B came from human patients; (2) both A and B were synthesized by the computer; (3) A came from a human patient and B from the synthesizer; and (4) A came from the synthesizer and B from a human patient. In the null hypothesis (choosing at random), all possible choices are equally represented: 25% with a confidence interval (±2 standard deviation) of ±7.32% in this case. Considering that all samples were synthetic, response (2) would be expected more than 44 times out of 154 (ie, 32.32%) if the students were able to recognize synthetic voices. The test was performed twice: the observed percentages of response (2) were 29.28 and 26.43. Hence, we may conclude that the students were not able to recognize the synthetic voices.

Smartphones

The two devices were selected at the extremes of the commercially available price range (a price ratio of 13:1 at the time of purchase). The more expensive one was the HTC One (hereafter named Smart1, Taoynan, Taïwan), and the cheaper one was a WIKO Cink Slim2 model (named Smart2, Marseille, France). Relevant technical specifications are given in Manfredi et al.[1]

Microphone

The microphone was a Sennheiser (Wedemark, Germany) MD421U model (frequency response: 30–17,000 Hz) commonly used in the voice laboratory for recording voice patients.

Amplifier and loudspeaker

A Bowers & Wilkins (Worthing, UK) CM1 model loudspeaker was used. Its frequency response is flat ±1.5 dB between 50 Hz and 20 kHz. It was driven by a Yamaha (Hamamatsu, Japan) YHT-380 model amplifier. The frequency response of the amplifier is flat ±0.5 dB between 20 Hz and 20 kHz, with 0.06% Total Harmonic Distortion (THD). The AUX input of the amplifier was used to reproduce sounds from the synthesizer. To avoid directivity effects of the loudspeaker, the smartphones and the microphone were carefully positioned on the axis of the loudspeaker (as there is no aerodynamic noise), fixed on a stand at a 4-cm distance from the center of the loudspeaker. The sound intensity of the loudspeaker was adjusted to correspond to the loudness of a normal human voice, that is, about 75 dB at 4 cm.

Soundproof booth

All recordings were made in an IAC (Winchester, UK) Mini 350 soundproof booth certified according to the EN ISO 9001/14001 norm.

External added noise

The external ambient added noise was a shopping mall ambience noise collected from http://www.google%20soundbible.com Shopping Mall Ambiance (accessed March 13, 2017).

It was provided in the booth by a separate amplifier and loudspeaker, placed at 60 cm from the smartphones, and carefully calibrated using a Wärtsilä 7178 precision integrating sound level meter positioned at the same place as the smartphones. For the experiments, the sound pressure levels of added ambient noise—measured at the level of the smartphone—were 37, 40, 43, 46, 49, 52, 55, 58, and 61 dB_A. In the absence of added noise, the background noise into the booth was 29.7 dB_A.

Analysis program

All data were analyzed with the *Praat* program, 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.[14–17] The two smartphones record audio files in different formats: ".ADTS" for Smart1 and ".OGG" for Smart2. The files were converted into WAV files for analysis using *Praat*.

Statistics

The intraclass correlation coefficient (ICC) is a general measure of agreement or consensus. The coefficient represents agreements between two or more evaluation methods on the same set of data. ICC has advantages over the correlation coefficient: it is adjusted for the effects of the scale of measurements, and it represents agreements from more than two measuring methods.[18] ICCs were calculated using a freely accessible program of the Chinese University of Hong Kong.[19]

RESULTS

Sustained /a:/ 120 Hz

Figure 2 shows the plot of ICCs between the results of jitter and N:H ratio, respectively, versus ambient noise intensity level. "No added noise" corresponds to the background noise of the booth (29.7 dB_A). The average frequency of the /a:/ is 120 Hz. Each ICC was calculated from a table with nine rows (the three jitter levels combined with the three noise levels) and four columns (Smart1 and Smart2, microphone, and direct measurement). For jitter, the ICC remains very high (>.9) up to 49 dB_A ambient noise, and declines beyond 52 dB_A. For N:H ratio, the ICC is very high up to 52 dB_A and drops beyond 55 dB_A.

Sustained /a:/ 200 Hz

Figure 3 similarly shows the ICCs between the results of jitter and N:H ratio, respectively, plotted against ambient noise intensity level. "No added noise" corresponds to the background noise of the booth (29.7 dB_A). The average frequency of the /a:/ is 200 Hz. Each ICC was calculated from a table with nine rows (the three jitter levels combined with the three noise levels) and four columns (Smart1 and Smart2, microphone, and direct measurement). For jitter, the ICC remains very high up to 61 dB_A ambient noise. For N:H ratio, the ICC is very high up to 55 dB_A and drops from 58 dB_A on.

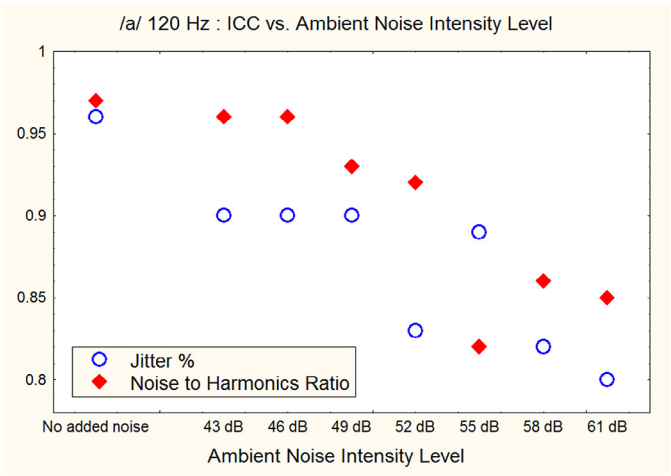

FIGURE 2. ICCs between the results of jitter and N:H ratio plotted against ambient noise intensity level. "No added noise" corresponds to the background noise level of the booth (29.7 dB_A). Voice material consists of /a:/ at an average frequency of 120 Hz. For jitter %, the ICC remains excellent up to 49 dB_A ambient noise, and drops from 52 dB_A on. For the N:H ratio, the ICC is high up to 52 dB_A and drops from 55 dB_A on. ICC, intraclass correlation coefficient; N:H, noise to harmonics.

3× /aiu/ at 120 Hz

Figure 4 shows the ICCs between the results of jitter and N:H ratio, respectively, plotted against ambient noise intensity level. "No added noise" corresponds to the background noise of the booth (29.7 dB_A). Voice material consists of 3× /aiu/ at an average frequency of 120 Hz. Each ICC was calculated from a table with nine rows (the three jitter levels combined with the three noise levels) and four columns (Smart1 and Smart2, microphone, and direct measurement). ICCs are weak, particularly for jitter, and

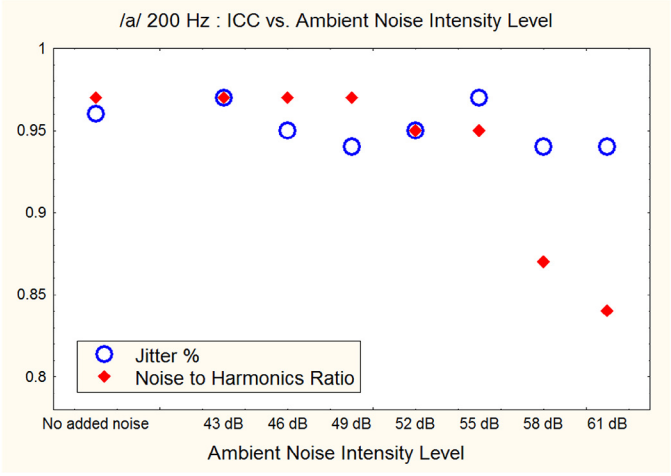

FIGURE 3. ICCs between the results of jitter and N:H ratio plotted against the level of ambient noise intensity. "No added noise" corresponds to the background noise of the booth (29.7 dB_A). Voice material consists of /a:/ at an average frequency of 200 Hz. For jitter %, the ICC remains high up to 61 dB_A ambient noise. For the N:H ratio, the ICC is excellent up to 55 dB_A and drops from 58 dB_A on. ICC, intraclass correlation coefficient; N:H, noise to harmonics.

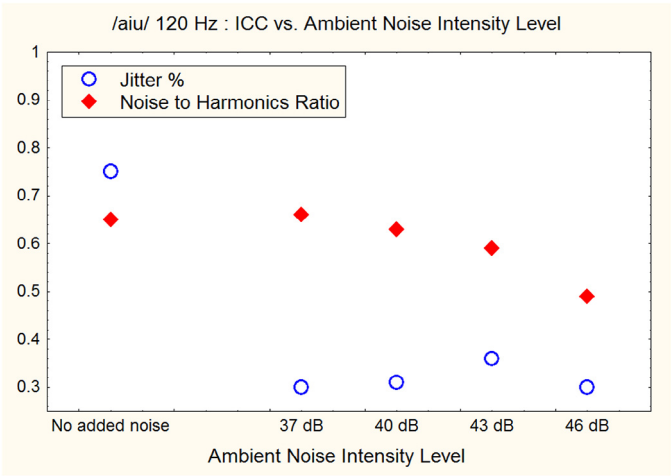

FIGURE 4. ICCs between the results of jitter and N:H ratio plotted against ambient noise intensity level. "No added noise" corresponds to the background noise level of the booth (29.7 dB_A). Voice material consists of 3× /aiu/ at an average frequency of 120 Hz. Considering that values obtained via the smartphones may differ by more than 100% from those obtained via direct recording (the smartphone also uses a microphone), correlations are weak, particularly for jitter, and are insufficient for clinical purposes. ICC, intraclass correlation coefficient; N:H, noise to harmonics.

are systematically insufficient for clinical purposes: in several conditions, the values obtained via one or both smartphones differ by more than 100–150% from those obtained via the microphone.

3× /aiu/ at 200 Hz

Figure 5 similarly shows the ICCs between the results of jitter and N:H ratio, respectively, plotted against ambient noise intensity level. "No added noise" corresponds to the background noise of the booth (29.7 dB_A). Voice material consists of 3× /aiu/ at an average frequency of 200 Hz. Each ICC was calculated from a table with nine rows (the three jitter levels combined with the three noise levels) and four columns (Smart1 and Smart2, microphone, and direct measurement). ICCs are again systematically insufficient for clinical purposes, even without any added noise: an examination of the data shows that, in several conditions, the values obtained via the smartphones differ by more than 100–150% from those obtained via the microphone.

Extent of differences

Table 1 presents the average values of the six differences (Smart1–Smart2–microphone–direct) in jitter % and in N:H ratio for the four types of signals (120–200 Hz; /a:/–/aiuaiuaiuaiu/). The average differences are clearly larger for /aiuaiuaiuaiu/ than for /a:/.

Comparison between the two smartphones

To specifically compare the two smartphones in the cases of low ICC (/aiuaiuaiuaiu/), ICCs were calculated for the data of each smartphone separately with the microphone and the direct measurements (three columns, nine rows for each F0). The results are shown in Table 2.

TABLE 1.
Average Values of the Six Differences (Smartphone 1–Smartphone 2–Microphone–Direct) in Jitter % and in Noise to Harmonics Ratio for the Four Types of Signals (120–200 Hz; /a:/–/aiuaiuaiuaiu/)

	120 Hz				200 Hz			
	/a:/		/aiuaiuaiuaiu/		/a:/		/aiuaiuaiuaiu/	
Ambient Noise	Jitter %	Noise to Harmonics Ratio	Jitter %	Noise to Harmonics Ratio	Jitter %	Noise to Harmonics Ratio	Jitter %	Noise to Harmonics Ratio
29.7 dB_A	0.60	0.04	0.94	0.17	0.12	0.05	0.52	0.13
37 dB_A			2.31	0.16			1.38	0.10
40 dB_A			2.69	0.15			1.60	0.11
43 dB_A	0.62	0.03	2.31	0.18	0.21	0.05	0.71	0.13
46 dB_A	0.67	0.01	2.58	0.19	0.31	0.05	1.48	0.13
49 dB_A	0.84	0.04			0.26	0.05		
52 dB_A	0.62	0.05			0.44	0.09		
55 dB_A	0.65	0.08			0.32	0.06		
58 dB_A	0.49	0.07			0.39	0.14		
61 dB_A	0.74	0.10			0.38	0.14		

Although remaining insufficiently reliable with respect to our criterion (ICC ≥0.9), the cheaper Smart2 systematically records and transmits the /aiuaiuaiu/ signal with better fidelity than the more expensive Smart1.

Figure 6 illustrates, for example, the differences between the two smartphones. Jitter measurements (3× /aiu/, 200 Hz) obtained via Smart1, Smart2, and the Sennheiser microphone are plotted against the direct measurement of the synthesized signal. For each device (smartphones and microphone), three jitter levels are measured, and for each level of jitter there are three noise levels. Values obtained via Smart2 and via the microphone are strongly correlated with each other and with the direct measurements. The correlation is much poorer for Smart1, and individual differences of the order of magnitude of 300% for jitter % are observed between Smart1 and the microphone.

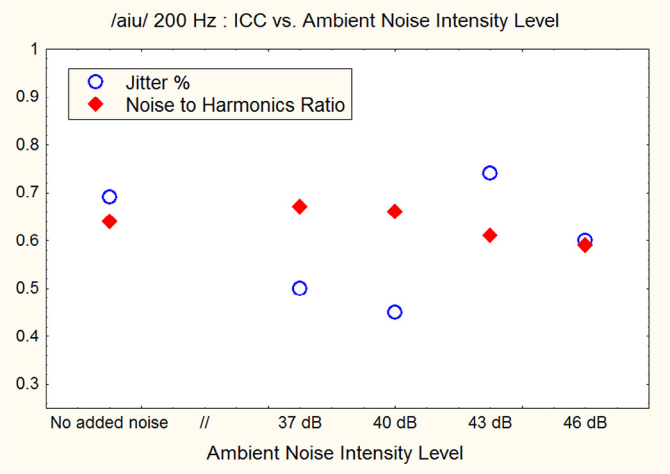

FIGURE 5. ICCs between the results of jitter and N:H ratio plotted against ambient noise intensity level. "No added noise" corresponds to the background noise level of the booth (29.7 dB_A). Voice material consists of 3× /aiu/ at an average frequency of 200 Hz. ICCs are insufficient for clinical purposes, even in the absence of added ambient noise. ICC, intraclass correlation coefficient; N:H, noise to harmonics.

DISCUSSION

The ICC and the level of agreement required for clinical use

The ICC is well suited for the purpose of this research. An important advantage over a Pearson correlation is that the Pearson correlation is invariant to application of separate linear transformations to the two variables being compared. If one is correlating X and Y, where, for example, $Y = 2X + 1$, the Pearson correlation between X and Y is 1, indicating perfect correlation. This simple fact is in the nature of a dimension-free measure as is the correlation coefficient, and could be a serious source of bias in the current study. This is avoided by using the ICC. Furthermore, the ICC directly provides the level of agreement among four categories of measurements. Cicchetti[20] considered an ICC of 0.75 or more as indicating "excellent" agreement. However, for the clinical applications considered in the scope

TABLE 2.
Comparison Between the ICC Values Obtained With the Two Smartphones (Smart1: Smartphone 1; Smart2: Smartphone 2) for Jitter % and N:H Ratio

ICC	Jitter (%)	N:H ratio
Smart1 120 Hz	0.70	0.57
Smart2 120 Hz	0.89	0.74
Smart1 200 Hz	0.68	0.54
Smart2 200 Hz	0.89	0.60

ICCs were calculated for (1) the data of each smartphone separately (2) with those of the direct recording via the Sennheiser microphone and (3) with direct measurements: computer to computer (three columns, nine rows for each F0).
Abbreviations: ICC, intraclass correlation coefficient; N:H, noise to harmonics.

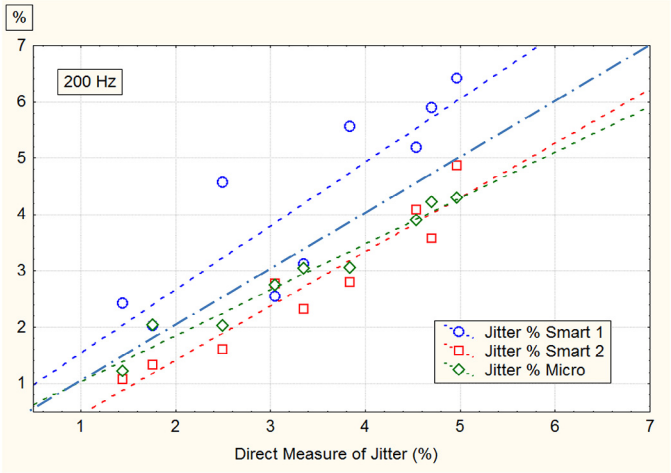

FIGURE 6. Jitter measurements (3× a /aiu/, 200 Hz) obtained via smartphone 1, smartphone 2, and Sennheiser microphone plotted against the direct measurement of the synthesized signal. For each device (smartphones and microphone), three jitter levels are measured, and for each level of jitter there are three noise levels. Values obtained via smartphone 2 and via the Sennheiser microphone are strongly correlated with each other and with the direct measurements. The correlation is much weaker for smartphone 1.

of this work, a more severe criterion is required. Particularly for slightly deviant voices, the difference between acoustic results obtained with the high-quality microphone of the voice laboratory and those obtained via a smartphone may not exceed the short-term spontaneous variability of voice acoustic parameters, which amounts to approximately 25%.[21] In our data set, this corresponds to an ICC of 0.9, which is considered as our threshold limit value.

The sustained /a:/

As shown in Figures 2 and 3, the signal transmission via both smartphones may be considered reliable for clinical use in an external ambient noise level up to 49 dB_A. This is true for both 120 Hz and 200 Hz voices, and for both jitter and N:H ratio. At higher external noise levels, the reliability slightly decreases, but it remains good. This result is qualitatively similar to that obtained by Maryn et al,[4] who used records of human patients with voice disorders. But in their work, these authors analyzed separately neither male and female voices, nor sustained vowels and continuous speech, nor the effects of different levels of voice alterations with regard to jitter and noise. In our work, synthetic voices allowed us to draw separate conclusions for sustained /a:/ and (simulated) connected speech utterances, both altered by known combinations of jitter and noise. Moreover, separate analysis of 120 Hz and 200 Hz voices revealed the more reliable results of 200 Hz voices in the presence of additional ambient noise.

The 3× /aiu/

Figures 4 and 5 show that with /aiuaiuaiu/ utterances, the smartphones become considerably less reliable for recording and transmitting deviant voices, and even totally unreliable in some conditions. Actually, even in the absence of any additional external noise, that is, in the background noise of the booth (29.7 dB_A), unsatisfactory ICC values are obtained. The problem is seemingly due to the smartphones themselves, as correlations between direct measurements and measurements via the microphone are strong (>0.9). As technical data provided by manufacturers are very sparse, it is hard to speculate about the specific element at issue. It seems that, when formants shift, signal processing by the smartphones appears to induce changes that disturb the analysis of F0 perturbations as well as the N:H ratio. The signal processing differs from smartphone to smartphone, and in this work the resulting changes seem more substantial in the more sophisticated and more expensive device. It is known that, to achieve the standard bitrate of 64 kbit/s with the worldwide used 0–4 kHz frequency band at 8 kHz of sampling frequency, telephone digital data are commonly encoded on 8 bits. However, thanks to more efficient data compression techniques, higher audio quality can be achieved on smartphones with higher sampling rates and coding on 16, 24, and even 32 bits.

This is again in line with the results of Maryn et al,[4] who emphasized the differences between devices, although these authors did not test cheap devices. Figure 7 illustrates the point with oscillograms of the same 3× /aiu/ utterance (200 Hz; N:H ratio = 0.46; jitter = 4.01%): the upper trace is the display of the direct signal by *Praat* (from computer to computer); the middle and bottom traces are the same signal recorded and transmitted by Smart1 and Smart2, respectively. A slight distortion is seen in the signal recorded by Smart2, but large distortion is observed in the Smart1 record.

A suggestion for further research is to compare the results obtained via the built-in microphone with those obtained via an external microphone (plugged into the audio jack) with a larger diaphragm and known or testable technical characteristics. Also, the possible interest in this scope of several applications (apps) specifically intended for using a smartphone as a recording device could be investigated. An example of such a dedicated application was recently published in the context of preventive cardiology.[22]

Our findings currently limit the use of smartphones by voice patients to sustained /a:/ for repeatedly transmitting voice samples to the voice laboratory, for example, in the context of single subject studies, follow-up, or baseline designs. The method is not reliable for running speech, at least with some types of smartphones.

CONCLUSION

Following a previous study that demonstrated the reliability of smartphones for recording and transmitting deviant voices of patients (sustained /a:/) to the voice clinic, the present work confirms that a good reliability for F0 perturbation and for N:H ratio measures is guaranteed in the presence of ambient noise levels of 50 dB_A and below. This level of ambient noise can easily be controlled by the patient using a reliable ad hoc application program that incorporates a sound level meter in the smartphone.

However, an important new finding is that this does not apply (so far) to voice samples with formant shifts, present in connected speech. Smartphones apply signal processing that may

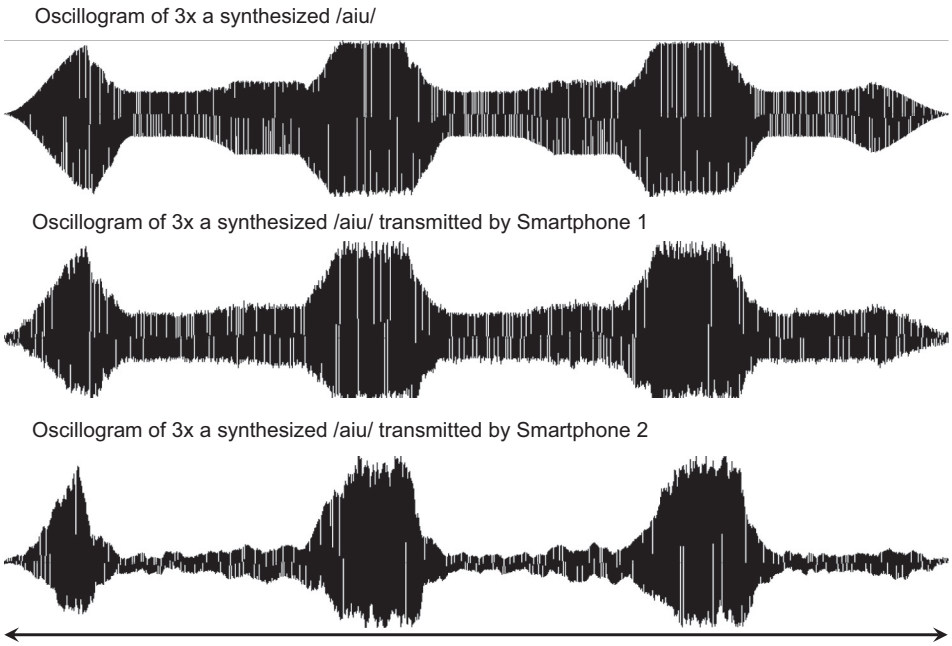

FIGURE 7. Oscillograms of the synthetic voice (3× a /aiu/). *Upper trace*: direct visualization of the signal by *Praat* (from computer to computer). *Middle and bottom traces*: the same signal recorded and transmitted by smartphone 1 and smartphone 2. Minor distortions are seen in the signal of smartphone 2, and much larger distortions in the smartphone 1 signal.

substantially affect the results of the analysis of F0 perturbations and N:H ratio, independent of the level of ambient noise. Provisionally, the use of smartphones for transmitting pathologic voices for acoustic analysis thus needs to be limited to sustained /a:/.

REFERENCES

1. Manfredi C, Lebacq J, Cantarella G, et al. Smartphones offer new opportunities in clinical voice research. *J Voice*. In press;http://dx.doi.org/10.1016/j.jvoice.2015.12.020.
2. Amato F, Cannataro M, Cosentino C, et al. Early detection of voice diseases via a web-based system. *Biomed Signal Process Control*. 2009;4:206–211.
3. Kardous CA, Shaw PB. Evaluation of smartphone sound measurement applications. *J Acoust Soc Am*. 2014;135:EL186–EL192.
4. Maryn Y, Ysenbaert F, Zarowski A, et al. Mobile communication devices, ambient noise, and acoustic voice measures. *J Voice*. 2016;doi:10.1016/j.jvoice.2016.07.023.
5. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice*. 2005;19:15–28.
6. DeJonckere PH. Voice evaluation and respiratory voice assessment. In: Anniko M, Bernal-Sprekelsen M, Bonkowsky V, et al., eds. *Otorhinolaryngology, Head & Neck Surgery*. Heidelberg Berlin: Springer; 2010:563–574.
7. Bandini A, Giovanelli F, Orlandi S, et al. Automatic identification of dysprosody in idiopathic Parkinson's disease. *Biomed Signal Process Control*. 2015;17:47–54.
8. Schoentgen J. Non-linear signal representation and its application to the modeling of the glottal waveform. *Speech Commun*. 1990;9:189–201.
9. Schoentgen J. Shaping function models of the phonatory excitation signal. *J Acoust Soc Am*. 2003;114:2906–2912.
10. Rothenberg M. An interactive model for the voice source. *KTH-QPSR*. 1981;22:1–17.
11. Klatt D. Software for a cascade/parallel formant synthesizer. *J Acoust Soc Am*. 1980;67:971–995.
12. Hawks JW, Miller JD. A formant bandwidth estimation procedure for vowel synthesis. *J Acoust Soc Am*. 1995;97:1343–1344.
13. 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.
14. DeJonckere PH, Schoentgen J, Giordano A, et al. Validity of jitter measures in non-quasi-periodic voices. Part I. Perceptual and computer performances in cycle pattern recognition. *Logoped Phoniatr Vocol*. 2011;36:70–77.
15. 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.
16. 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.
17. 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.
18. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications and Practice*. Norwalk, CT: Appleton & Lange; 1993:509–516. ISBN 0-8385-1065-5.
19. Chang A, Sahota D. Statistics Toolkit (STATTOOLS). Available at: http://www.obg.cuhk.edu.hk/ResearchSupport/StatTools/IntraclassCorrelation_Pgm.php. Accessed December 13, 2016.
20. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6:284–290. doi:10.1037/1040-3590.6.4.284.
21. Speyer R, Wieneke GH, DeJonckere PH. The use of acoustic parameters for the evaluation of voice therapy for dysphonic patients. *Acta Acust United Acust*. 2004;90:520–527.
22. Lotan Y, Snir M, Dranitzki-Elhalel R, et al. Novel mobile application for assessment of extravascular lung water by acoustic analysis of vocal recording. *Eur J Prev Cardiol*. 2016;23(2S):48–49.