Mobile Digital Recording: Adequacy of the iRig and iOS Device for Acoustic and Perceptual Analysis of Normal Voice

*Gisele Oliveira, †Gaetano Fava, ‡Melody Baglione, and ‡Michael Pimpinella, *Brooklyn and †‡New York, New York

Summary: Objective. To determine whether the iRig and iOS device recording system is comparable with a standard computer recording system for digital voice recording.

Methods. Thirty-seven vocally healthy adults, between ages 20 and 62, with a mean age of 33.9 years, 13 males and 24 females, were recruited. Recordings were simultaneously digitalized in an iPad and iPhone using a unidirectional condenser microphone for smartphones/tablets (iRig Mic, IK Multimedia) and in a computer laptop (Dell-Inspiron) using a unidirectional condenser microphone (Samson-CL5) connected to a preamplifier with phantom power. Both microphones were lined up at an equal fixed distance from the subject's mouth. Speech tasks consisted of a sustained vowel "ah" at comfortable pitch/loudness, counting from 1 to 10, and a glissando "ah" from a low to a high note. The samples captured on the iOS devices were transferred via *SoundCloud* in WAV format, and analyzed using the *Praat* software. The acoustic parameters measured were mean, min, and max F0, SD F0, jitter local, jitter rap, jitter ppq5, jitter ddp, shimmer local, shimmer local-dB, shimmer apq3, shimmer apq5, shimmer apq11, shimmer dda, NHR, and HNR.

Results. There were no statistically significant differences for any parameter and speech task analyzed for both iOS devices as compared with the gold standard computer/preamp system (all P values > 0.050). In addition, there were no statistical differences in the perceptual identification of the recordings among devices ($P < 0.001$).

Conclusion. In the present study, the iRig and iOS device may provide reliable digital recording of normal voices.

Key Words: voice–dysphonia–recording–microphone–technology.

INTRODUCTION

Voice recording is a common practice in the field of speech-language pathology.[1–3] Its use is manifold in both the clinical and the research settings. Voice recording can occur at the time of initial evaluation, during treatment, and posttherapy. It can serve as an educational tool for feedback monitoring, help measure progress, and allow for objective capturing of a client's or patient's voice signal for perceptual and acoustic analysis.[1,2,4]

Recording technology has evolved throughout the years. Analog tape recording was popular up to the mid 1990s, followed by a shift to digital and direct-to-computer recording toward the end of the 20th century. Digital recording allowed for more precise and accurate capturing of the audio signal, especially for perturbation voice analysis.[2,3,5,6] During that time, digital recording became the new standard for voice recording, according to a consensus reached among voice professionals.[7] Despite this consensus, however, some clinicians continued to use analog devices due to lack of computer availability in some clinical settings.[2]

Today, wide varieties of recording systems are available to clinicians. They range from low-cost devices, such as portable digital recorders and personal computer recording software, to advanced recording hardware/software packages, such as the *Computerized Speech Lab* (CSL, Kay Elemetrics Corp., Model 4500, NJ). The clinicians' choice of recording device will usually depend on the availability at their facility as well as on their own level of expertise and budget. The literature suggests guidelines for recording devices, techniques, and ambient noise control.[8,9] Ideally, the recorded voice is acoustically identical to the captured voice signal. In reality, a perfect replication is difficult to achieve, but possible to approach if certain conditions are controlled, in particular, environmental noise and microphone selection.[8,9]

The environment can play a great role for voice recording and may affect it according to the environment's dimension, design, and level of background noise.[9] Generally, the average clinic is not soundproof or sound is hard to control (eg, noisy hallway, waiting area, and shared rooms). The choice of microphone will be an important factor in overcoming environmental problems.[2,3,9]

One of the most crucial criterions for recording and analyzing the human voice is microphone selection.[8,9] The most salient characteristics of a microphone for voice-recording purposes are directionality, transduction, frequency response, and mouth-to-microphone distance.[9] Directionality refers to the direction from which the microphone picks up a signal. In regard to directionality, microphones can be omnidirectional or unidirectional. Omnidirectional microphones pick up signals from different directions, while unidirectional microphones, as the name implies, pick up signals coming from one direction, suppressing other surrounding noises and sounds.[9–12] Transduction refers to how the microphone converts sound signal into an electrical signal. Microphones by transducer type can be classified into dynamic, electret, and condenser type.[9] For the purpose of this paper, we shall simply state that dynamic microphones do not require power to operate and have a poorer frequency response, whereas the

Accepted for publication May 31, 2016.
Research previously presented at: The Voice Foundation's 43rd Annual Symposium: Care of the Professional Voice, Philadelphia, USA, May 29, 2014.
From the *Touro College, Brooklyn, New York; †Columbia University Medical Center, New York Presbyterian Hospital, New York, New York; and the ‡The Cooper Union for the Advancement of Science and Art, New York, New York.
Address correspondence and reprint requests to Gaetano Fava, Columbia University Medical Center, New York Presbyterian Hospital, 622 West 168th Street, VC 10, Rm 1001, New York, NY 10032. E-mail: gfava@icloud.com
Journal of Voice, Vol. 31, No. 2, pp. 236–242
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.05.023

electret and condenser types require a power source and have a better frequency response.[9] Frequency response refers to how the microphone captures the signal, according to its design. For example, live singing performance microphones are designed to enhance certain frequencies to produce a better perceptual sound.[9] For voice-recording purposes in a clinic or laboratory, the best frequency response is flat because the signal captured remains closest to the original signal; condenser microphones are considered the standard for recording and the best microphones to achieve flat frequency response.[9,13]

Based on the aforementioned features discussed thus far, the optimal microphone for voice recording in a clinical or laboratory setting would be a unidirectional, condenser microphone because it reduces environmental noises in the background and allows for a flat frequency response. However, a flat frequency response can only be obtained at a certain, established mouth-to-microphone distance to avoid what is known as proximity effect, which is an enhancement of lower frequencies when the microphone is too close to the sound source.[9,10,12] Microphone manufacturers typically do not list the proximity effect distance of their products.[12] As such, this proximity effect dilemma is a crucial problem for optimal microphone selection. The current recommendation is to use either a head-mounted omnidirectional microphone or a handheld unidirectional microphone to reduce the impact of environmental noise, which as discussed above, may be an issue as the average clinic is not soundproof.[14]

Voice perturbation analysis remains a challenging task because of the variations among systems and procedures (eg, different algorithms by acoustic analysis software, recording systems, environments, mouth-to-microphone distances, and microphone types).[1,3,4,9,15,16]

Many computer programs are available for perturbation voice analysis. Two known programs in speech-language pathology are *Praat*[17] and the *Multidimensional Voice Program* (MDVP, Model 5105, Kay Elemetrics, Pine Brook, NJ).[4] It is well documented in the literature that acoustic measurements obtained by different programs do not produce similar results due to algorithmic differences between the programs, thus rendering the measures incomparable.[3,4,6,15,18]

The advancement of science and technology over the last half decade made it possible to achieve sufficient signal approximation in recording systems suitable to clinical and academic purposes.[8,9] The 21st century continues to unravel smaller and more powerful products. Smartphones and tablets are providing the convenience of portability and multifunctionality within one device, replacing previous hardware.[19,20] The growing number of apps is allowing for endless possibilities in terms of daily, personal, and professional needs.[21–24] Some of the advantages in using these devices and applications in the clinical setting are portability, timeliness, and efficiency.[25] With the advent of smartphones and tablets, the use of apps has become popular and many speech-language pathologists are using them in their practices.[11,16,19,25] Testimony to this app-boom in the field of speech-language pathology are many convention sessions devoted to apps, such as at the 2013 and 2014 American Speech-language-hearing Association Conventions,[26–28] the Voice Foundation's 41st and 43rd Annual Symposia,[29–32] and the 29th World Congress of the International Association of Logopedics and Phoniatrics.[31,33] Smartphones and tablets are also providing the opportunity for digital audio recording and transmitting these data via the Internet.[11] Previous studies have attempted to demonstrate the use of smartphones for digital recording.[11,16] The choice of microphone used in previous studies was the internal microphone of the device, which did not follow the microphone selection guidelines for recording set forth by Svec and Granqvist.[9] Moreover, the recordings obtained on the iPhone were captured in M4A, which is a compressed format of the audio signal used by Apple, diminishing overall the quality of the recording, even when converted afterward to a higher audio quality WAV format.[34]

The purpose of the present study was to determine whether the iRig and iOS device recording system is comparable with a standard computer recording system for digital recording of normal voices. The goal was to allow for portability while preserving the technical microphone specifications described in the literature.

METHODS

Perceptual analysis

This study has been approved by the institutional review board of Touro College (Protocol # IRB1335).

Participants

The subjects in this study were 37 vocally healthy adults, aged between 20 and 62 years, with a mean age of 33.9 years, 13 males and 24 females. Subjects were considered vocally healthy as none of the individuals had a voice complaint or a diagnosis of dysphonia. The subjects were recruited among students, faculty members, and acquaintances of the first author at Touro College, Brooklyn, NY. The subjects signed a consent form, participated voluntarily, and were not compensated. A total of 18 subjects were recorded with the iPhone, and 19 subjects with the iPad. The participants were asked to produce a sustained vowel "ah" for 5 seconds at a comfortable pitch and loudness, count from 1 to 10 at a comfortable pitch and loudness, and produce a glissando "ah" from a low note to a high note.

Procedures

The devices used were one iPhone 5 (Model MD634LL/A, iOS 7), one iPad 3rd generation (Model MC706LL/A, iOS 7), and one Dell-Inspiron laptop computer (Dell Inc., Round Rock, TX, USA).

Voice samples were simultaneously recorded in an iPad or an iPhone, and in a computer laptop (Dell-Inspiron). The microphone used for both iPad and iPhone recordings was a unidirectional condenser microphone for smartphones and tablets (iRig Mic, IK Multimedia, Sunrise, FL, USA). The microphone used for the laptop recordings was a unidirectional condenser microphone (Samson-CL5, Samson Technologies Corp., Hauppauge, NY, USA) connected to a preamplifier with phantom power (DigiDesign MBox 2 USB audio interface with 48V phantom power, Avid Technology Inc., Burlington, MA, USA). Both microphones were placed on stands and lined up at an equal fixed distance of 30 cm from the subject's mouth.