

# Accuracy of the quantities measured by four vocal dosimeters and its uncertainty

Pasquale Bottalico<sup>a)</sup>

*Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, Illinois 61820, USA*

Ivano Ipsaro Passione

*Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824, USA*

Arianna Astolfi

*Department of Energy, Politecnico di Torino, Torino, Italy*

Alessio Carullo

*Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy*

Eric J. Hunter

*Voice Biomechanics and Acoustics Laboratory, Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan 48824, USA*

(Received 6 May 2016; revised 17 January 2018; accepted 2 March 2018; published online 22 March 2018)

Although vocal dosimeters are often used for long-term voice monitoring, the uncertainty of the quantities measured by these devices is not always stated. In this study, two common vocal dosimetry quantities, mean vocal sound pressure level and mean vocal fundamental frequency, were measured by four vocal dosimeters (VocaLog2, VoxLog, Voice Care, and APM3200). The expanded uncertainty of the mean error in the estimation of these two quantities as measured by the four dosimeters was performed by simultaneously comparing signals acquired through a reference microphone and the devices themselves. Dosimeters, assigned in random order, were worn by the participants (22 vocally healthy adults), along with a head-mounted microphone, which acted as a reference. For each device, participants produced a sustained /a/ vowel four times and then read a text with three different vocal efforts (relaxed, normal, and raised). The measurement uncertainty was obtained by comparing data from the microphone and the dosimeters. The mean vocal sound pressure level was captured the most accurately by the Voice Care and the VoxLog while the APM3200 was the least accurate. The most accurate mean vocal fundamental frequency was estimated by the Voice Care and the APM3200, while the VoxLog was the least accurate.

© 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5027816>

[LK]

Pages: 1591–1602

## I. INTRODUCTION

Voice disorders can be difficult to assess accurately in the clinical setting since patients' voices in such a setting may not adequately represent their voices in natural settings (e.g., Hunter, 2009). Therefore, long-term ambulatory voice monitoring might more accurately capture how individuals engage in their typical daily activities (Ghassemi *et al.*, 2014; Hunter *et al.*, 2012; Cheyne, 2003; Bottalico and Astolfi, 2012). Such long-term monitoring is often conducted using a vocal dosimeter, a small portable device consisting of a transducer (contact microphone and/or accelerometer) connected to a battery-operated digital recorder and/or analyzer that is attached to the skin of the neck at the jugular notch, directly below the thyroid prominence (with adhesive or a collar). Dosimeters primarily provide sound pressure level (SPL) and fundamental frequency ( $f_0$ ).

One of the early precursors to the concept of real world vocal monitoring was the voice accumulator, which could detect and record phonation time and level over several hours (Buekers *et al.*, 1995). The next generation of dosimeters used an accelerometer attached to the skin of the neck to capture a person's voice. The National Center for Voice and Speech (NCVS) dosimeter (Švec *et al.*, 2004) was developed for research purposes and derived all the necessary values from the skin vibration of the neck. Useful additions to the previous technology were (a) a tool that prompted the wearer to respond to voice use questions to monitor self-report vocal health and (b) its ability to collect voice data over several weeks at a time (Hunter and Titze, 2010). The first commercially available device was the Ambulatory Phonation Monitor 3200 (PENTAX Medical, Lincoln Park, NJ, now discontinued), which attempted to make these devices clinic ready.

The stated bandwidth of the NCVS dosimeter and APM3200 is in the range of 2 Hz to 3 kHz, with a flatness of  $\pm 1.5$  dB in the frequency range 50–1000 Hz. Because uncertainty specifications are not available for the considered

<sup>a)</sup>Electronic mail: pb81@illinois.edu

devices, only preliminary estimations are provided as taken from published material. For the estimation of vocal SPL, the calibration uncertainty of the NCVS dosimeter is 5 dB at a 95% confidence level, while average errors of 3.2 dB with a standard deviation of almost 6 dB have been estimated for the APM3200 (Carullo *et al.*, 2013). Švec *et al.* (2005) demonstrated that the accelerometer BU-7135 (Knowles Electronics, Itasca, IL) could indicate the mean SPL of soft, comfortable or loud voices with an uncertainty of  $\pm 2.8$  dB in 95% of cases if an individual calibration was performed before the monitoring. Hillman *et al.* (2006) found that the average error in the estimation of the SPLs from the acceleration signal was about  $0.2 \pm 2.1$  dB, over a sample of six normal speakers. While these two devices are still in some laboratories, they are no longer available commercially. However, a current accelerometer-based dosimeter (Voice Health Monitor or VHM) is in development and has been reported on Mehta *et al.* (2012) but is not yet commercially available. This new device uses a smartphone to sample and process the raw accelerometer signal.

A second group of vocal dosimeters, which includes the VocaLog2 (Griffin Laboratories, Temecula, CA) and the Voice Care (PR.O. VOICE SRL, Torino, Italy), uses an external contact microphone, while the device VoxLog (Sonvox AB, Umeå, Sweden) uses both a microphone and an accelerometer. According to Carullo *et al.* (2015), after a proper calibration procedure is implemented the Voice Care estimates the SPL parameter with a mean error equal to  $-1.6$  dB with respect to a reference microphone and a standard deviation equal to 2.5 dB, while the fundamental frequency is obtained with a measurement uncertainty of  $\pm 3$  Hz. This calibration procedure allows estimation of a talker's voicing SPL based on a reference microphone (Behringer ECM8000) at a fixed distance of 16 cm in front of the mouth with the talker repeating the vowel /a/ at increasing levels of intensity. The producers of the VocaLog2 and the VoxLog have provided no comparable information.

While dosimeters can provide crucial insights into how subjects or patients use their voices, their importance is tempered by the inability to quantitatively compare the results captured by the various devices. For example, Van Stan *et al.* (2014) compared the results of APM3200, VocaLog (a previous version of VocaLog2), and VoxLog using the VHM dosimeter as the reference signal. All the devices were used on the same participant. For the APM3200 they found a difference of 1.89 dB in the evaluation of the  $SPL_{\text{mean}}$  and a difference of 1.76 Hz for the  $f0_{\text{mean}}$ . A difference of 0.93 dB in the evaluation of the  $SPL_{\text{mean}}$  was found for the VocaLog while the VoxLog showed a difference of 4.29 dB in the evaluation of the  $SPL_{\text{mean}}$  and  $-0.56$  Hz for the  $f0_{\text{mean}}$ . Nevertheless, while these differences appear significant, such differences could be partially due to study methodology (e.g., differences in SPL values could be influenced by the fact that the authors did not refer the SPL values measured by the different devices to the same distance).

Another study, comparing the  $SPL_{\text{mean}}$  values acquired by the VocaLog dosimeter with a sound level meter (SLM), found that the values acquired by the VocaLog were 1.3–1.9 dB higher than the values acquired by the SLM during

the production of a sustained /i/ vowel, while 1.5–2.4 dB higher during a text reading (Searl and Dietsch, 2014).

With the ultimate goal of better understanding the speakers' vocal behavior in real life situations, there has been a steady stream of research over the last several decades focusing on ambulatory voice monitoring. The devices developed have begun to lay the groundwork of data for an increased understanding common voice use as well as the relationship between voice disorders and vocal load in terms of loudness, phonation time and fundamental frequency. However, previous research has not yet adequately associated the measurement uncertainty with the results provided by the dosimeters, a necessary step in order to avoid misinterpretation. Since not all the manufacturers specify the measurement uncertainty of the estimated parameters, the goal of this study was to assess the accuracy of  $SPL_{\text{mean}}$  and  $f0_{\text{mean}}$  estimations of the APM3200, the VoxLog, the VocaLog2, and the Voice Care dosimeters, as well as uncertainty values for each of them.

## II. EXPERIMENTAL METHOD

The evaluation of the accuracy of any device has a relatively standard approach: closeness of agreement between two signals (or calculated metrics from a signal) routing through the device of interest and a reference device (ISO/IEC Guide 98-3, 2008). In this case, the accuracy evaluation used a speech signal which was routed through available vocal dosimeters and a reference microphone, resulting in estimated mean error of  $SPL_{\text{mean}}$  and  $f0_{\text{mean}}$  from all devices. Using these devices, the summary statistics were calculated in order to evaluate the uncertainty of the mean error of the parameters estimated by the devices.

In order to minimize the room effect on the  $SPL_{\text{mean}}$ , a head mounted microphone as close as possible to the mouth was used. Considering that the head mounted microphone cannot be calibrated, a second mic was placed in the room at a fixed distance of 16 cm, which can be calibrated with a Class 1 sound calibrator. The head mounted microphone was calibrated by comparison to the second microphone.

Speech samples were collected and assessed from 22 vocally healthy English speaking participants (11 female, 11 male; 18–59 yr old, mean 24.2 yr old). The participants were not informed of the purpose of the experiment. The participants self-reported normal speech and hearing. Ethical approval to conduct the research was obtained from the Michigan State University's Human Research Protection Program.

### A. Data collection procedures

The participants wore a M80 (Glottal Enterprise, Syracuse, NY) omnidirectional head-mounted microphone placed at 5 cm from their mouth, while simultaneously wearing one of the dosimeters. The participants repeated the speech tasks four times in a random order per each of the four dosimeters. After the device specific calibration procedure, if required by the device, participants were seated in a sound isolation booth in front of a ECM8000 (Behringer, Willich, Germany) ultra-linear measurement condenser microphone placed at 50 cm from the mouth of the

participants. Both the head-mounted and the measurement microphones were connected to a PC via a Scarlett 2i4 (Focusrite, High Wycombe, UK) soundboard, with Audacity 2.0.6 (SourceForge, La Jolla, CA) used as the recording software. The double walled sound isolation booth (2.5 × 2.75 m and h = 2 m) had a mid-frequency reverberation time (RT20) of 0.05 s and the trend over the octave band (125–8000 Hz) was almost flat ( $\mu = 0.062$ ,  $\sigma = 0.011$  s). The RT20 was measured following the standard ISO 3382-2 (2008). The background noise in the room was 25 dB(A). With the chair against the back wall, the participants were asked to keep their head touching the wall behind them in order to keep the distance from the reference microphone as stable as possible for the whole recording.

The reading tasks (a combination of steady vowels and speech) and styles (three vocal efforts) were chosen with the uncertainty analysis in mind, with the intent to create different amplitude (style) and fluctuation (task) from the participants. Moreover, while speech is the primary goal in vocalization, steady phonations are common in communicative exchanges (uh huh, um, etc.) and are still used in clinical environments and in assessments.

The tasks were briefly explained to the participants before recording began. During the recording, instructions and reading materials were presented via a monitor to cue the following: (1) Production of a sustained vowel /a/ for 5 s at a relaxed vocal effort; (2) reading task for 2 min at a relaxed vocal effort; (3) production of a sustained vowel /a/ for 5 s at a normal vocal effort; (4) reading task for 2 min at a normal vocal effort; (5) production of a sustained vowel /a/ for 5 s at a raised vocal effort; (6) reading task for 2 min at a raised vocal effort. The reading task text chosen was “Goldilocks and the Three Bears.” Participants were instructed that the relaxed speech level should be softer than the normal and that the raised level should be louder than the normal speech level, as indicated in the instructions of ISO 9921 (2002).

## B. Device specifications of the four vocal dosimeters

The devices analyzed in the present study were the VocaLog2, VoxLog, Voice Care and APM3200. The main

differences among the devices are the type of transducer (accelerometer and/or contact microphone) and the individual calibration procedures. Table I summarizes the main characteristics of each device which corresponds to a short synopsis of each device.

The VocaLog2 Vocal Activity Monitor was designed to help individuals modify their vocal loudness level through monitoring and feedback. The device consists of a neckband monitor and a USB microphone used during calibration. The laryngeal sensor is a contact microphone. It is placed in the neck band unit and it detects the patients’ vocal activity. The sensor should be located just above the sternal notch, and must be flush against the skin with enough pressure to ensure constant contact. The VocaLog2 unit’s feedback mechanism can vibrate when undesired vocal loudness (both above and below a given threshold) is detected. The VocaLog 2™ desktop application is the software provided with the device for the data acquisition, analysis and calibration. The calibration microphone is placed 30 cm from the individual while he or she completes a series of speech tasks (e.g., reading, producing sustained vowels at minimum and normal voice effort). The VocaLog2 only provides estimates of SPL and registers the presence of phonation once per second.

The VoxLog (SonVox) voice monitoring system is another tool for long-term measurement of voice use. It registers the SPL of the ambient noise, the f0 and SPL of the voice, as well as the percentage of phonation time over the total monitored time. The device consists of a measurement collar for the throat, where an accelerometer and a factory calibrated air microphone are located. The accelerometer collects the data from the acceleration of the skin induced by the vibration of the vocal folds. A fast Fourier transform (FFT) algorithm is applied to the vibrational signal for the evaluation of the fundamental frequency. When the accelerometer detects phonation (the activation level is unknown), the microphone signal is interpreted as the participant’s voice; otherwise, the microphone signal is interpreted as ambient sound (Schalling et al. 2013). The collar is connected to the device where the data are saved and analyzed. VoxLog Discovery is the software provided with the device for the data acquisition and analysis. It shows the SPL, the  $f_{0\text{mean}}$  and the percentage of phonation time over the total

TABLE I. Main characteristics of the four devices compared in this study. SPL represents the sound pressure level, f0 represents the fundamental frequency, and Dt represents the phonation time percentage.

Device	Type of transducer	Frame length	Calibration procedure for SPL	Estimated parameters	Algorithm for f0 evaluation	Uncertainty reported in literature
VocaLog2	Contact microphone	1 s	Yes	SPL	NA	SPL mean error of 1.3–1.9 dB in a vowel task and of 1.5–2.4 dB in a reading task
VoxLog	Accelerometer and calibrated air microphone	0.1 s, 1 s, 5 s, 30 s, 60 s, 180 s and 300 s	No	A-weighted SPL for voice and noise, f0, Dt	Fast Fourier Transform- based	SPL mean error 4.3 dB, f0 mean error of –0.56 Hz
Voice Care	Contact microphone	0.03 s	Yes	SPL, f0, Dt	Autocorrelation- based	SPL mean error –1.6 dB and standard deviation 2.5 dB, f0 absolute error ≤ 3 Hz
APM3200	Accelerometer	0.05 s	Yes	SPL, f0, Dt and other parameters derived from the previous ones	Autocorrelation- based	SPL mean error 3.2 dB with a standard deviation of 6 dB, f0 error 1 Hz

monitored time. Another function in VoxLog Discovery is the possibility of creating analyses that include data from many different recordings, which can be put together as numerical values and charts. During the setup of the VoxLog, it is possible to choose the time sampling for the data acquisition among 0.1, 1, 5, 30, 60, 180, and 300 s. No individual calibration is required because the microphone measuring airborne SPL is factory calibrated.

The Voice Care is essentially an MIAE38 (Midland, Reggio Emilia, Italy) contact microphone and a small data-processing unit (Carullo *et al.* 2012). Like many of the other devices, the microphone is in contact with the jugular notch in order to detect skin acceleration level due to the vibration of the vocal folds. The microphone output is conditioned through a custom circuitry and then sent to an inexpensive micro-controller-based board, which stores the raw samples onto a micro secure digital (SD) card. The off-line processing provides an estimation of SPL,  $f_0$  and phonation time. The Voice Care has a specific calibration procedure to allow for individualized estimation of dB SPL. The calibration is done using a reference microphone (ECM8000, the same type of microphone used for the recordings in the sound booth) directly connected to the device. The participant is placed in front of the reference microphone at a distance of 16 cm and is asked to produce /a/ vowels at different intensities for at least 1 min. Once the calibration has ended, the data stored into the SD card is processed in order to estimate the relationship between the signal at the output of the microphone and the SPL. The  $f_0$  values are estimated on the basis of an autocorrelation-based algorithm. The Voice Care estimates the parameters of interest every 30 ms.

The APM3200 (designed for monitoring a voice throughout a day) measures the amount of time a participant has phonated, identifies when phonations have occurred, and estimates vocal intensity (dB SPL) and fundamental frequency (Hz) during all phonation activity. Further, it purports (albeit untested in this study) to also provide immediate, real-time vibrotactile feedback to the patient during daily activities (based on pre-determined settings entered by the clinician prior to usage). As a voice sensor, a miniaturized accelerometer is mounted on a silicone pad and attached to the neck at the jugular notch using surgical adhesive. Hillman *et al.* (2006) demonstrated that the accelerometer could supply data (i.e., SPL,  $f_0$  and phonation time) for individuals with normal voice quality, as well as those with mild and severe dysphonia. According to its specification, the bandwidth is from 2 Hz to 3 kHz with a flatness of 1.5 dB in the frequency range 50–1000 Hz, while the  $f_0$  is evaluated by an autocorrelation algorithm with measurement errors not greater than  $\pm 1$  Hz. APM3200 device calibration, data handling, and resulting metrics were handled by the provided software. The wearer is placed in front of the provided calibration microphone at a distance of 15 cm. Instructions outline that participants should take a deep breath and sustain the /a/ vowel, starting with a soft voice and steadily increasing volume until the loudest voice is reached. As soon as the subject starts to phonate, the APM software will show on the computer the calibration data points and a linear fit line, which represents the linear correlation between the SPL

recorded by the microphone and the amplitude of the signal captured by the accelerometer on the neck of the patient. While the software creates a calibration curve after at least seven data points have appeared, this does not mean that the calibration has been performed well (Nacci *et al.*, 2013) because the calibration curve may have a low coefficient of determination. Only preliminary estimations of uncertainty specifications have been reported; for the parameter SPL, after a proper calibration the mean error was of 3.2 dB with a standard deviation of almost 6 dB (Carullo *et al.*, 2013). The APM3200 estimated the parameters of interest every 50 ms.

### C. Processing

As described above, participants produced the speech tasks four times, once for each device. For each of the four productions, two simultaneous streams of data were acquired during the measurements performed using each of the four devices: [dev] the data measured by the device and [mic] the data measured by the head mounted microphone, which was calibrated to the reference microphone.

A preliminary calibration procedure of the reference microphone ECM8000 was performed using a Class 1 sound calibrator NC-74 (Rion, Japan) with automatic atmospheric pressure compensation (ref: 94 dB  $\pm$  0.3 dB at 1 kHz  $\pm$  2%). Next, the signal from the head mounted microphone (which could not be directly calibrated with a sound calibrator) was calibrated by comparison to the calibrated signal of the ECM8000 in a fashion patterned after Švec *et al.* (2003). In order to minimize the calibration error due to possible variation of the distance between participant's mouth and microphone, the calibration by comparison was performed by averaging the 12 repetitions of the /a/ vowels (three voice styles per four devices).

Each device, with the exception of VoxLog, was calibrated following the specifications for that device. This procedure included the specified microphone to use for the calibration, distance to place the microphone from the mouth, and tasks to perform during the calibration. The microphones were placed at 30 cm for the VocaLog2, 15 cm for the APM3200, and 16 cm for the Voice Care. Using the SPL calibrated head mounted microphone signal [mic] and the device data stream [dev], four pairs of signals [mic]-[dev], one per each device, were obtained. After this process, all the SPL values were adjusted to estimate vocal SPL at 50 cm.

The signals were processed with the software Matlab R2015b (MathWorks, Natick, MA). For each device, a time history was calculated for  $SPL_{\text{mean}}$  and  $f_{0\text{mean}}$ . The data sampling rates of each of the four dosimeters (signal [dev]) are 1, 0.1, 0.03, and 0.05 s for the VocaLog2, VoxLog, Voice Care, and APM3200, respectively. Among the data sampling rates offered by the VoxLog, 0.1 s was chosen because it is the smallest.

The processing sampling rate of the signal [mic] was chosen considering a trade-off between a good accuracy and the manageability of the database, with the restriction to be at least equal to the sampling rate of the corresponding device. Therefore, the processing sampling rate chosen to be

for the signal [mic] was 0.03 s for the Voice Care comparison and 0.05 s for the other devices.

The dosimeters estimate output values only for the voiced frames. An unvoiced frame is treated as silence with a given output of zero. However, during the unvoiced frames, the reference microphone recorded the background noise, which in the sound booth used for the experiment was equal to 25 dBA. Therefore, a threshold was calculated as the complement of the percentile level of the phonation time in percentage (calculated subdividing the number of the voiced frames found using Praat 5.4.17. with the default setting by the total time of the signal). For example, if the phonation time in percentage was 40%, the threshold was set to percentile level  $L_{60}$ . This threshold was used to discriminate the unvoiced frames in the signal [mic]. Those frames were set to zero.

Regarding the signal [dev], only some devices specify the details about the calibration procedure, such as the distance from the speakers' mouth to which the SPL is referred. Without this information, the SPL can only be relative at best. Therefore, a relative calibration constant for each participant was calculated. This constant was estimated as the difference of the mean value of the /a/ vowel in the normal style between the signals [dev] and [mic]. Thus, it was added to the SPL time history values of the signal [dev] in order to report the results of the two signals at the same distance (50 cm) and to compare them. By adding this constant, possible room acoustics effects on the signal [mic] are taken into account.

The values of the parameter  $f_0$  for the signal [mic] were calculated using Praat with the same time steps chosen for the parameter SPL. The algorithm performed acoustic periodicity detection on the basis of an autocorrelation method. While there are several methods to choose from, the Praat autocorrelation method is more accurate, noise-resistant, and robust than methods based on the cepstrum or comb, as well as the original autocorrelation method (Boersma, 1993). An ad-hoc comparison to using Praat's cross-correlation method yielded a 0.09 Hz difference for the vowel recordings and  $-0.19$  Hz for the speech tasks, which is smaller than the mean errors associated with the devices. It is important to acknowledge that while the results of the analysis performed by Praat are based on estimation algorithms, it represents the most available and used software for speech acoustics analysis. For this reason, it was considered as a reference.

### III. EVALUATION OF THE UNCERTAINTY OF MEASUREMENTS

Because every measurement is prone to error, a measurement result is complete only when accompanied by a quantitative statement of its uncertainty. Where human participants or operators are involved, uncertainty can be minimized by procedures. Within equipment or analysis routines, uncertainty is not based on human variability but on such considerations as time-windowing, thresholds, and hardware size and fit. Measurement uncertainty is quantified by characterizing the distribution of the values attributed to a

measured quantity. In this report, the uncertainty was calculated according to the document [ISO/IEC Guide 98-3 \(2008\)](#). The uncertainty in the result of a measurement is the combination of those components that could contribute to the experimentally observed variability of the output values. Generically those components can be grouped into (1) type A evaluations, estimated by statistical methods, and (2) type B evaluations, estimated by other means. Both of these types of evaluation are based on probability distributions, and the uncertainty components resulting from either type are quantified by variances or standard deviations.

A summary of the uncertainty quantities associated with the two components is given here. The "estimated variance"  $u^2$  characterizing an uncertainty component obtained from a type A evaluation is calculated from series of repeated observations and is the familiar statistically estimated variance  $s^2$ . The "estimated standard deviation"  $u$ , the positive square root of  $u^2$ , is hence  $u = s$  and for convenience is sometimes called a "type A standard uncertainty," "experimental standard deviation of the mean," or "standard error." This quantity provides an estimate of the precision of the mean and is used when one wants to make inferences about data from a sample to some relevant population, differently from the standard deviation that represents the dispersion of the data. (Clark-Carter, 2005).

For an uncertainty component obtained from a type B evaluation, the estimated variance  $u^2$  is evaluated using available knowledge. These include such items as (1) previous measurement data, (2) experience with or general knowledge of the behavior and properties of relevant materials and instruments, (3) manufacturer's specifications, (4) data provided in calibration and other certificates, and (5) uncertainties assigned to reference data taken from handbooks.

In this study, the measurand  $y$  was the Mean Error (ME) for both  $SPL_{\text{mean}}$  and  $f_{0,\text{mean}}$ , both of which were considered separately. The uncertainty contributions due to reproducibility (i.e., the closeness of the agreement between the results of measurements of the same measurand carried out under changed conditions of measurement, [ISO/IEC Guide 98-3, 2008](#)) were considered for the different (a) participants, (b) tasks, and (c) styles.

The type A and type B uncertainties of the ME were evaluated and then combined. The type A uncertainties were obtained from the evaluation of the propagation of the uncertainty among participants over the six combinations of speech tasks and styles. The tasks and styles were chosen in order to create different amplitude (style) and fluctuation (task) of the signals, with the ultimate goal of minimizing the correlations among repeated within-subject measurements. The type B standard uncertainty (evaluated using available knowledge from preliminary analysis of the data) was obtained by considering the uncertainties of the inputs of the ME, i.e., the uncertainties pertaining to the mean values of the time histories from both signal [dev] and signal [mic]. The uncertainty contributions were evaluated according to the following steps:

(Step 0) For signals [dev] and [mic], the mean value  $x_i$  and the experimental standard deviation  $u(x_i)$  for each combination of task and style were evaluated for each participant

from their time histories. A total of 132 pairs of values (mean and experimental standard deviation) were obtained for both [dev] and [mic] (6 task-style combinations  $\times$  22 participants), respectively.

(Step 1) The mean error (ME) of the variables related to the signal [dev] was calculated as the average of the difference between the mean values of [dev] and [mic] obtained in (Step 0), as follows:

$$\text{ME} = \frac{1}{6N} \sum_{i=1}^{6N} (x_{\text{dev},i} - x_{\text{mic},i}) = \frac{1}{6N} \sum_{i=1}^{6N} (\Delta_{xi}), \quad (1)$$

where  $N$  is the number of participants, 6 is the number of task-style combinations,  $x_{\text{dev},i}$  is the mean value of the variables  $y$  pertaining the signal [dev] obtained in (Step 0),  $x_{\text{mic},i}$  is the mean value of the variables  $y$  pertaining the signal

[mic] obtained in (Step 0), and  $\Delta_{xi}$  is the difference between  $x_{\text{dev},i}$  and  $x_{\text{mic},i}$ .

(Step 2) The repeatability contribution of the ME, is estimated as a type A standard uncertainty,  $u_A(\text{ME})$ :

$$\begin{aligned} u_A(\text{ME}) &= s(\text{ME}) \\ &= \left( \frac{1}{\sqrt{6N}} \right) \sqrt{\frac{1}{6N-1} \sum_{i=1}^{6N} (\Delta_{xi} - \text{ME})^2}. \end{aligned} \quad (2)$$

(Step 3) The repeatability contribution of each time history, which represents the intrinsic contribution of each participant, is estimated by propagating the experimental standard deviations  $u(x_{i,\text{mic}})$  and  $u(x_{i,\text{dev}})$  on the ME model (2), thus obtaining a type B standard uncertainty  $u_B(\text{ME})$ :

$$\begin{aligned} u_B(\text{ME}) &= \sqrt{\sum_{i=1}^{6N} \left( \frac{\partial \text{ME}}{\partial x_{i,\text{mic}}} \right)^2 u^2(x_{i,\text{mic}}) + \sum_{i=1}^{6N} \left( \frac{\partial \text{ME}}{\partial x_{i,\text{dev}}} \right)^2 u^2(x_{i,\text{dev}})} \\ &= \sqrt{\frac{1}{(6N)^2} \sum_{i=1}^{6N} u^2(x_{i,\text{mic}}) + \frac{1}{(6N)^2} \sum_{i=1}^{6N} u^2(x_{i,\text{dev}})} = \frac{1}{6N} \sqrt{\sum_{i=1}^{6N} u^2(x_{i,\text{mic}}) + \sum_{i=1}^{6N} u^2(x_{i,\text{dev}})}, \end{aligned} \quad (3)$$

where  $x_{i,\text{mic}}$ ,  $x_{i,\text{dev}}$ ,  $u(x_{i,\text{mic}})$ , and  $u(x_{i,\text{dev}})$  are the mean values and the experimental standard deviations from Step (0).

(Step 4) The final uncertainty of the ME is the *combined* standard uncertainty,  $u_c(\text{ME})$ , which was calculated combining the uncertainty contributions evaluated in Steps 2 and 3, as follows:

$$u_c(\text{ME}) = \sqrt{u_A^2(\text{ME}) + u_B^2(\text{ME})}. \quad (4)$$

The expanded uncertainty  $U_{\text{ME}}$  was obtained by multiplying the combined standard uncertainty  $u_c(\text{ME})$  by a coverage factor  $k$ , as follows:

$$U_{\text{ME}} = k \cdot u_c(\text{ME}). \quad (5)$$

Because the probability distribution characterized  $u_c(\text{ME})$  is approximately normal and the effective degrees of freedom of  $u_c(\text{ME})$  are of significant size, the coverage factor can be assumed equal to 2, that is associated to a risk error  $\alpha$  equal to 5% (ISO/IEC Guide 98-3, 2008). In summary,  $U_{\text{ME}}$  is the final uncertainty while the other two components related to type A and B errors.

## IV. RESULTS

All the devices show some limitations. Because VocaLog2, Voice Care, and APM3200 estimate SPL from the skin vibration levels at the neck using the regression curve performed during the calibration procedures, a poor-fitting regression curve can produce an overestimation or an

underestimation of the SPL parameter. A limitation of the VocaLog2, Voice Care, and APM3200 is related to uncertainties introduced during the calibration of the devices. The participant specific calibration of these devices is the best-fit curve between the signals acquired from the transducer and from an in-air reference microphone and could vary from calibration to calibration. Another limitation is that VocaLog2 has a saturation threshold at 85 dB. VoxLog does not require a calibration performed by the user, but some issues were present during the relaxed vocal style. In this case, when the vibrations of the vocal folds were below the accelerometer signal's preset factory threshold, the VoxLog microphone signals were marked as background noise instead of voice SPL. Because of this mislabeling, the device did not automatically compute the values of the fundamental frequency during these segments.

The summary statistic and the combined uncertainty of the  $\text{SPL}_{\text{mean}}$  were calculated for each device by taking into account only the voiced frames (the values equal to zero were not considered). Since possible differences between tasks were not the focus of the study, task was not a factor in the statistical analysis. Any small task difference would be taken into account by the associated uncertainty.

### A. Sound pressure level

In Table II, the average of  $\text{SPL}_{\text{mean}}$  and the combined standard uncertainty values of [dev] and [mic] for each combination of task and style related to the four devices are reported. In Fig. 1, the mean differences of the SPL between

TABLE II. Averages of  $SPL_{mean}$  and combined uncertainty ( $u_c$ ), in dB, of the signals [dev] and [mic] for each task and style of the four dosimeters.

Task	Style	Signal	VocaLog2		VoxLog		Voice Care		APM3200	
			$SPL_{mean}/dB$	$u_c(SPL_{mean})/dB$	$SPL_{mean}/dB$	$u_c(SPL_{mean})/dB$	$SPL_{mean}/dB$	$u_c(SPL_{mean})/dB$	$SPL_{mean}/dB$	$u_c(SPL_{mean})/dB$
/a/	relaxed	dev	68.7	0.28	67.0	0.10	70.8	0.05	62.9	0.12
		mic	68.6	0.05	68.7	0.05	69.1	0.03	62.6	0.05
	normal	dev	73.9	0.38	73.3	0.10	74.7	0.05	66.8	0.15
		mic	74.1	0.05	73.5	0.05	74.8	0.04	66.7	0.06
	raised	dev	79.7	0.49	84.8	0.07	81.4	0.05	77.9	0.13
		mic	83.3	0.04	84.2	0.04	84.4	0.03	76.6	0.04
reading	relaxed	dev	68.1	0.08	70.2	0.03	68.1	0.03	60.8	0.06
		mic	68.8	0.02	68.3	0.02	68.1	0.01	60.5	0.02
	normal	dev	73.4	0.09	74.4	0.03	72.5	0.03	64.7	0.06
		mic	72.6	0.02	72.2	0.02	72.2	0.02	63.6	0.02
	raised	dev	79.0	0.08	79.1	0.04	77.1	0.03	74.4	0.06
		mic	78.0	0.02	77.5	0.02	77.4	0.02	70.0	0.03

the signal [dev] and [mic] per dosimeter for each task and voice style are shown.

As far as the VocaLog2 is concerned, the only differences higher than 1 dB were found in the /a/ raised task where the signal [dev] resulted in a value of  $-3.6$  dB. As reported in Table V and shown in Fig. 4, the mean error in the  $SPL_{mean}$  estimation by the VocaLog2 was  $-0.40$  dB and the expanded uncertainty was 0.77 dB.

As far as the VoxLog device is concerned, the results for the relaxed style (and, thus, the summary statistic) were calculated without including the values that the device wrongly labeled as noise. In the /a/ relaxed task, the difference between the two signals was higher than 1 dB, probably because of the misrecognition of the voiced frames, while the signal [dev] resulted in an average value higher than 1.9 dB in the reading task. Overall, the mean error in the

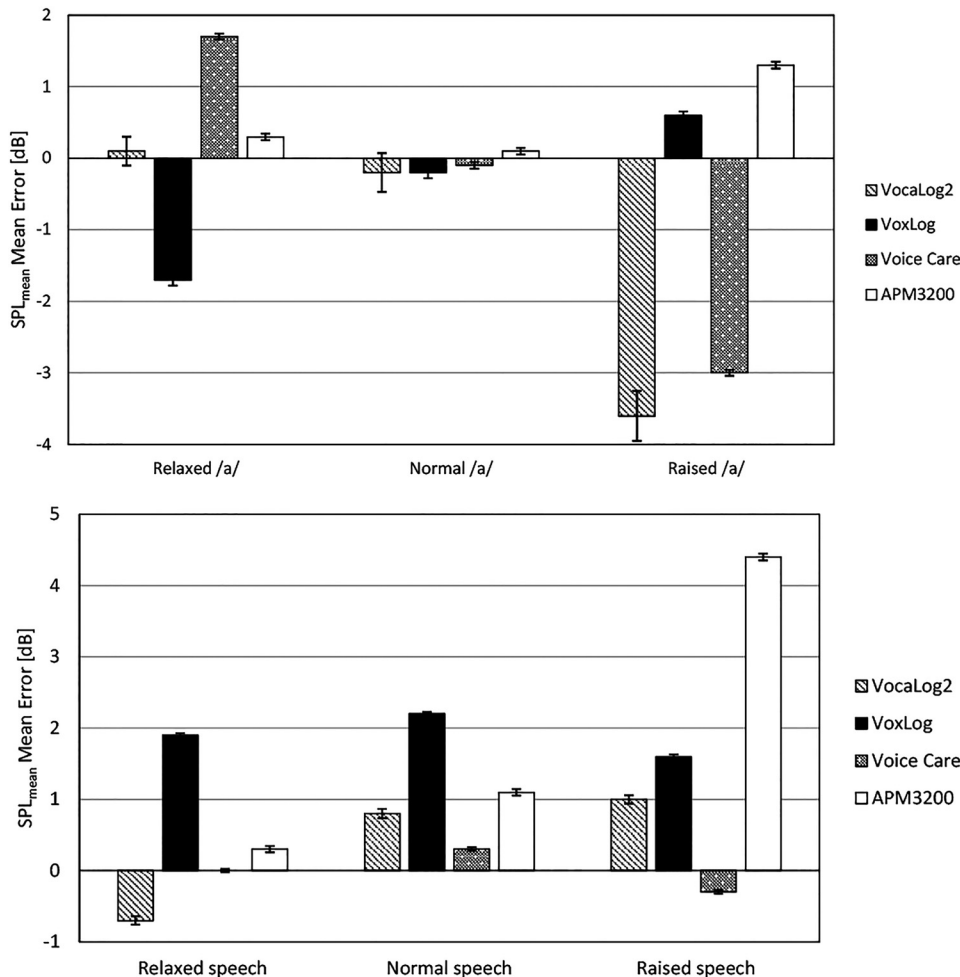


FIG. 1. Mean errors of the  $SPL_{mean}$  between the signal [dev] and [mic] per dosimeter for each task and voice style.

TABLE III. Averages of  $f0_{\text{mean}}$  and combined uncertainty ( $u_c$ ), in Hz, of the signals [dev] and [mic], for each task and style of the four dosimeters for male participants.

Task	Style	Signal	VoxLog		Voice Care		APM3200	
			$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$	$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$	$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$
/a/	relaxed	dev	107	0.64	113	0.13	111	0.10
		mic	106	0.33	112	0.15	106	0.52
	normal	dev	123	2.68	112	0.08	111	0.32
		mic	111	0.12	112	0.12	110	0.10
	raised	dev	149	1.29	138	0.20	150	0.25
		mic	142	0.11	140	0.07	151	0.11
reading	relaxed	dev	119	0.90	104	0.11	104	0.21
		mic	103	0.27	107	0.36	98	0.31
	normal	dev	105	0.51	108	0.13	113	0.23
		mic	109	0.19	110	0.20	109	0.29
	raised	dev	122	0.42	130	0.16	135	0.27
		mic	127	0.20	134	0.21	136	0.28

SPL<sub>mean</sub> estimation by the VoxLog was 0.80 dB and the expanded uncertainty was 0.54 dB.

The Voice Care device showed the smallest differences in the reading task between the two signals compared to other devices. In contrast, in the /a/ task, the signal [dev] showed an average value higher than 1.7 dB in the relaxed style and equal to 3.0 dB in the raised style, compared to the [mic] signal. The behavior of the Voice Care device in estimating the SPL between the relaxed style (SPL approximately in the range of 59–84 dB) and the raised style (SPL approximately in the range of 67–97 dB) is mainly due to an incomplete overlap of SPL ranges covered during the calibration procedure and the monitored task of each participant, as explained in Carullo *et al.* (2015). Finally, the mean error was -0.25 dB and the expanded uncertainty was 0.72 dB.

The APM3200 in both tasks overestimated the values of the voice SPL<sub>mean</sub>. In particular, the average value of the signal [dev] in the raised reading task was higher than 4 dB. Finally, the mean error was 1.15 dB and the expanded uncertainty was 1.01 dB.

## B. Fundamental frequency

In Tables III and IV, the average of  $f0_{\text{mean}}$  and the combined standard uncertainty values of [dev] and [mic] for each combination of task and style related to the four devices are reported for the male and female participants, respectively. For the  $f0_{\text{mean}}$ , the mean values were calculated for each task, style and participant for the VoxLog, Voice Care and APM3200 (the VocaLog2 does not give any values for  $f0$ ). The standard uncertainty was calculated considering an additional variable of gender due to the difference in the value of the  $f0$  among males and females.

The mean differences of the  $f0_{\text{mean}}$  between the signal [dev] and [mic] per dosimeter for each task and voice style are shown, in Fig. 2 and Table III for male participants and in Fig. 3 and Table IV for female participants.

Regarding the VoxLog device, the average value of the signal [dev] in the /a/ task was equal to 12 Hz in the normal style for both males and females and in the relaxed style for females. For the reading task, the devices produced a result equal to 16 and 7 Hz in the relaxed style for males and females, respectively. In the case of normal and raised style

TABLE IV. Averages of  $f0_{\text{mean}}$  and combined uncertainty ( $u_c$ ) in Hz, of the signals [dev] and [mic], for each task and style of the four dosimeters for female participants.

Task	Style	Signal	VoxLog		Voice Care		APM3200	
			$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$	$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$	$f0_{\text{mean}}/\text{Hz}$	$u_c(f0_{\text{mean}})/\text{Hz}$
/a/	relaxed	dev	222	0.82	220	0.10	220	0.12
		mic	210	1.12	221	0.14	212	0.28
	normal	dev	232	5.91	223	0.14	232	0.40
		mic	220	0.67	226	0.09	229	0.43
	raised	dev	249	1.28	253	0.26	255	0.33
		mic	247	0.37	257	0.28	249	1.08
reading	relaxed	dev	198	0.45	199	0.30	192	0.38
		mic	191	0.39	205	0.42	193	0.47
	normal	dev	194	0.46	197	0.32	195	0.41
		mic	193	0.42	205	0.41	196	0.48
	raised	dev	208	0.47	214	0.31	224	0.42
		mic	207	0.39	224	0.38	226	0.46



### Male participants

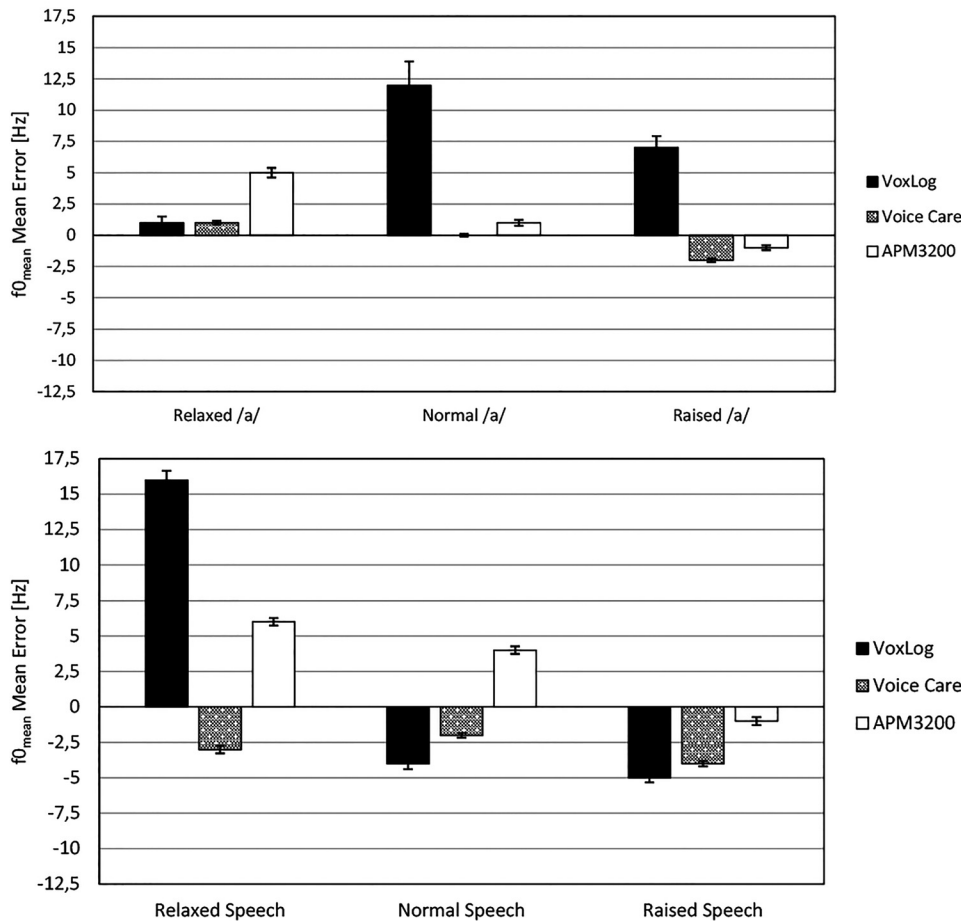


FIG. 2. Mean errors in male participants of the  $f0_{\text{mean}}$  between the signal [dev] and [mic] per dosimeter for each task and voice style.

for males, the device underestimated the  $f0_{\text{mean}}$  of 4 and 5 Hz, respectively. Finally, the mean error was 9.60 Hz and the expanded uncertainty was 3.76 Hz.

For male participants, the Voice Care device obtained differences of  $f0_{\text{mean}}$  from [dev] to [mic] signals lower than 4 Hz for both tasks. Whereas the differences for females were higher mainly for the reading task (-8, -10, and -6 Hz in the normal, raised and relaxed style, respectively). Finally, the mean error was -3.50 Hz and the expanded uncertainty was 0.78 Hz.

Regarding the APM3200 device, the bigger differences for males were found in the relaxed style for both the /a/ and the reading task (5 and 6 Hz, respectively) and in the normal reading (4 Hz). For females, the bigger differences were in the /a/ task (3, 6, and 8 Hz in the normal, raised, and relaxed style, respectively). Finally, the mean error was 2.90 Hz and the expanded uncertainty was 2.45 Hz.

### C. Comparison among devices

Table V lists the values of the uncertainties combined across tasks and styles, the mean errors, the uncertainties of the mean errors, the overall combined uncertainties and the expanded uncertainties for both the signals [dev] and [mic]. For each device, the values of the mean errors and the expanded uncertainties are shown in Fig. 4. The Voice Care

showed the lowest mean error for  $SPL_{\text{mean}}$  (-0.25 dB), while the lowest expanded uncertainty was shown by the VoxLog (0.54 dB). The device with the worst performance was the APM3200 (mean error equal to 1.15 dB and the expanded uncertainty to 1.01 dB). For  $f0_{\text{mean}}$ , the measurements provided by VoxLog had the highest mean error (9.60 Hz) and expanded uncertainty (3.76 Hz). In contrast, the measurements provided by APM3200 showed the lowest mean error (2.90 Hz), while the lowest expanded uncertainty was found for the measurements provided by Voice Care (0.78 Hz).

### V. DISCUSSION

The use of vocal dosimeters in research and medical environments has grown. Nevertheless, their impact is reduced by the lack of uncertainty specifications. This study compares the measured uncertainty during speech tasks of four devices.

For the estimation of the SPL, the VocaLog2 and the Voice Care used a contact microphone as a transducer, while the VoxLog used an air microphone. Regarding the estimation of the  $f0$ , the Voice Care used a contact microphone while VoxLog used an accelerometer. The VocaLog2 does not provide the  $f0$ . In this study, the devices that used a contact microphone for the estimation of the SPL were more

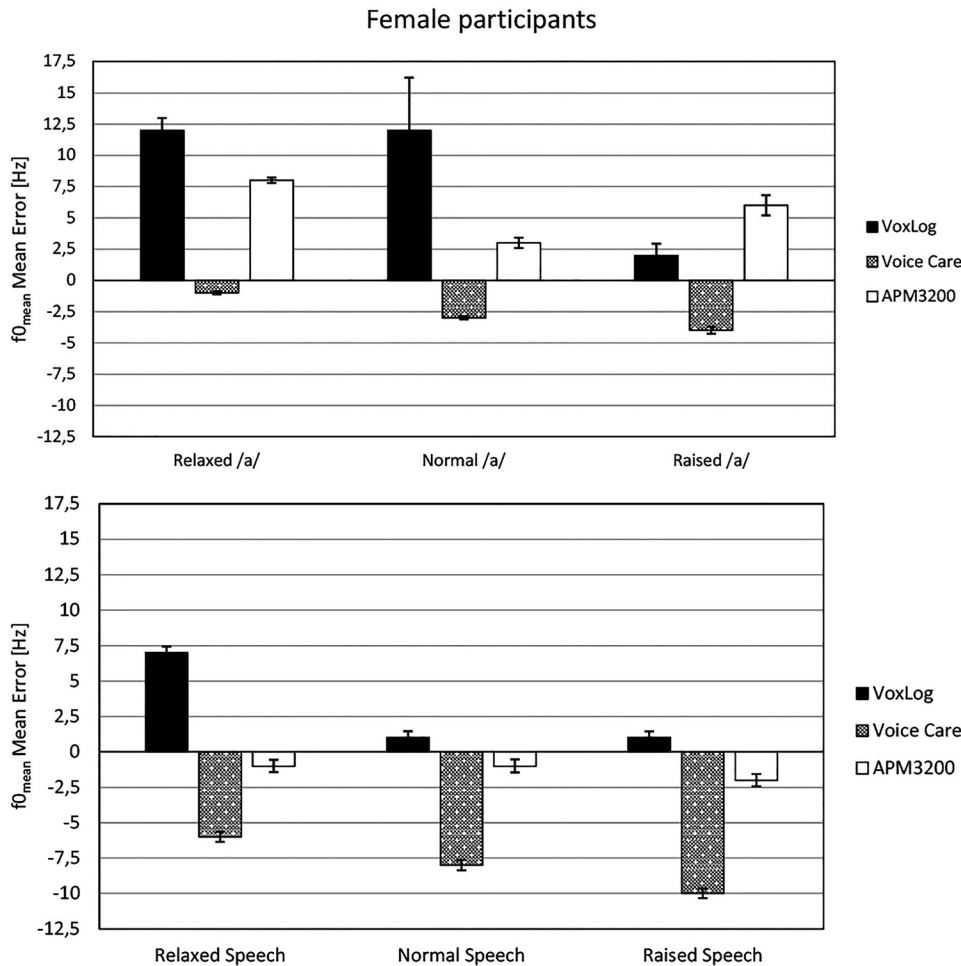


FIG. 3. Mean errors in female participants of the  $f_{0,mean}$  between the signal [dev] and [mic] per dosimeter for each task and voice style.

accurate compared to the APM3200 that used an accelerometer.

The mean error and the expanded uncertainty associated with the measurements provided by VocaLog2 were  $-0.40$  and  $0.77$  dB, respectively. Different from previous studies, the mean errors for the vowels and reading task in this study were  $-1.3 \pm 0.4$  dB and  $0.36 \pm 0.08$  dB, respectively, while Van Stan *et al.* (2014) found a mean error of  $1.3$ – $1.9$  dB in a vowel task and of  $1.5$ – $2.4$  dB in a reading task.

The mean error and expanded uncertainty of the  $SPL_{mean}$  measured by VoxLog were  $0.80 \pm 0.54$  dB. The difference between the present results and the results of Van Stan *et al.* (2014), who found a mean error of  $4.3$  dB, may be attributed to the fact that they referred the SPL

values of the VoxLog and the reference device to different distances. Regarding the  $f_{0,mean}$ , Van Stan *et al.* (2014) found a mean error of  $-0.56$  Hz, while in this experiment a mean error of  $9.60 \pm 3.76$  Hz was found. This difference in the results could have been caused by errors in the identification of the frames as voiced or unvoiced. The device did not automatically compute the values of the fundamental frequency.

The mean errors and the expanded uncertainties measured with Voice Care were equal to  $-0.25 \pm 0.72$  dB for SPL and  $-3.50 \pm 0.78$  Hz for  $f_{0,mean}$ . Carullo *et al.* (2015) found that after following their calibration procedure (explained in the paper), the  $SPL_{mean}$  parameter is estimated with a mean error equal to  $-1.6$  dB and a standard deviation

TABLE V. Mean error, Type A ( $u_A$ ) and Type B ( $u_B$ ) standard uncertainties, combined standard uncertainties ( $u_c$ ) and expanded uncertainties (U) for the  $SPL_{mean}$  in dB and  $f_{0,mean}$  in Hz pertaining to the four dosimeters.

	VocaLog2	VoxLog		Voice Care		APM3200	
	$SPL_{mean}/dB$	$SPL_{mean}/dB$	$f_{0,mean}/Hz$	$SPL_{mean}/dB$	$f_{0,mean}/Hz$	$SPL_{mean}/dB$	$f_{0,mean}/Hz$
ME	-0.40	0.80	9.60	-0.25	-3.50	1.15	2.90
$u_A(ME)$	0.36	0.27	1.78	0.36	0.38	0.50	1.21
$u_B(ME)$	0.12	0.03	0.61	0.02	0.10	0.05	0.17
$u_c,ME$	0.38	0.27	1.88	0.36	0.39	0.51	1.22
$U_{ME}$	0.77	0.54	3.76	0.72	0.78	1.01	2.45

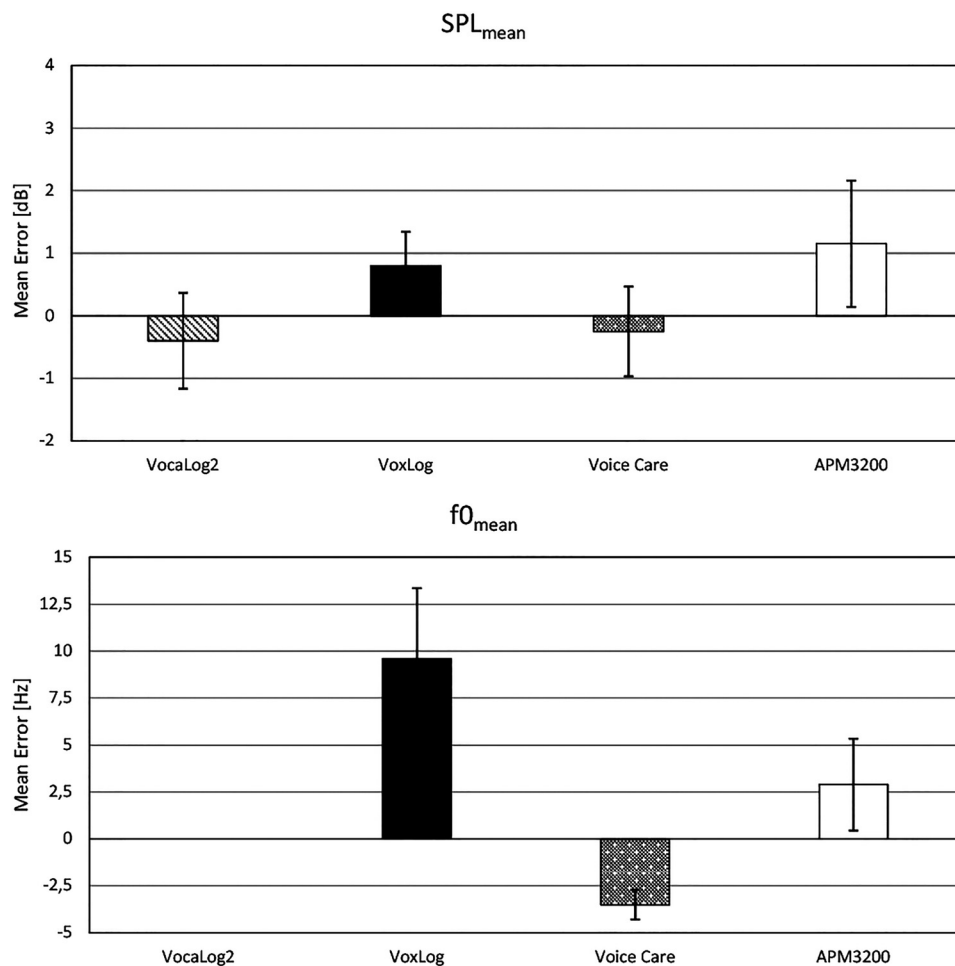


FIG. 4. Mean errors and expanded uncertainties of the SPL<sub>mean</sub> (upper) and f<sub>0</sub><sub>mean</sub> (lower) of the dosimeters and the microphone.

equal to 2.5 dB, while the fundamental frequency was obtained with measurement uncertainty of  $\pm 3$  Hz.

The APM3200 was the only device among the four considered for this study which used only an accelerometer as a transducer. For this device, the results showed the tendency to overestimate the calculation of both SPL<sub>mean</sub> and f<sub>0</sub><sub>mean</sub> ( $1.15 \pm 1.01$  dB and  $2.90 \pm 2.45$  Hz). The present results are comparable with prior studies. For example, Švec *et al.* (2005) reported an accuracy of 2.8 dB, and Hillman *et al.* (2006) found a mean error of  $0.2 \pm 2.1$  dB, while Van Stan *et al.* (2014) found a difference of 1.89 dB in the evaluation of the SPL<sub>mean</sub> and a difference of 1.76 Hz for the f<sub>0</sub><sub>mean</sub>.

As practical advice on the use of dosimeters, if the same device is used in “between-subject” and “within-subject” analyses, the obtained differences are meaningful if they are larger than the expanded uncertainty  $U_{ME}$ ; in the case of different devices used to monitor participants in a “between-subject” analysis, the differences are meaningful if they are larger than the combined effect of the ME and  $U_{ME}$  of each involved device.

## VI. CONCLUSIONS

Vocal dosimeters have become fundamental tools in long-term monitoring of voice use and in the assessment of vocal behaviors that can lead to disorders. The aim of this

study was to evaluate the accuracy and its uncertainty of the quantities mean SPL and mean fundamental frequency measured by four dosimeters and to provide reliable specification of their application in research and medical environments. The dosimeter with the highest mean error in the evaluation of SPL<sub>mean</sub> was the APM3200 (1.15 dB), followed by the VoxLog (0.80 dB), the VocaLog2 (−0.40 dB), and the Voice Care (−0.25 dB), while the dosimeter with the highest expanded uncertainty in the evaluation of SPL<sub>mean</sub> was the APM3200 (1.01 dB), followed by the VocaLog2 (0.77 dB), the Voice Care (0.72 dB) and the VoxLog (0.54 dB). The VoxLog showed problems in the recognition of the voiced frames in the relaxed style, but was more accurate in the raised style. The VocaLog2 and the Voice Care were more accurate in the relaxed style, but underestimated the results in the raised style. The APM3200 showed overestimation in both the relaxed and raised styles.

The dosimeter with the highest mean error in the evaluation of f<sub>0</sub><sub>mean</sub> was the VoxLog (9.60 Hz), followed by the Voice Care (−3.50 Hz) and the APM3200 (2.90 Hz), while the dosimeter with the highest expanded uncertainty in the evaluation of f<sub>0</sub><sub>mean</sub> was the VoxLog (3.76 Hz), followed by the APM3200 (2.45 Hz), and the Voice Care (0.78 Hz). The VocaLog2 was not designed for the computation of the fundamental frequency. The VoxLog did not compute the f<sub>0</sub>

values when it recognized the voiced frames as noise, which happened often in the relaxed style and sometimes in the normal style.

This study focused on the mean vocal behavior of self-reported normal speech talkers over different tasks, however future research is planned with the goal of evaluating the uncertainty and the mean error of instantaneous values, synchronizing the devices with the reference microphone. Moreover, considering that pathological voices are corrupted by noise, fundamental frequency estimation methods are more likely to have higher uncertainty. For this reason, future study will be conducted including participants with common voice disorders.

The assumption of this study was that the values obtained from the microphone represented the true values with negligible uncertainty and that the independency of the results between subjects and within subjects. Regarding the SPL, the measurements performed with the calibrated microphone represented a direct measurement of the sound pressure, while 3 out of 4 dosimeters (Voice Care, APM3200, and VocaLog2) use an individual calibration to estimate the SPL from the skin acceleration or skin pressure levels. The differences in these calibration procedures are substantial contributors to the uncertainties in SPL. The device VoxLog uses an air microphone placed on the neck that can be affected by the lack of direct path between mouth-microphone. Regarding the  $f_0$ , from the signals acquired by both the microphone and the devices, an estimation of the  $f_0$  without a direct measurement was performed. However, the analysis performed by Praat from the microphone signal was considered as reference because it represents the most available and used software for speech acoustics analysis. This represents a limitation of the results and future studies will be performed with direct measurement of  $f_0$  using electroglottography.

## ACKNOWLEDGMENTS

The authors would like to thank for their assistance the members of the Voice Biomechanics and Acoustics Laboratory, Michigan State University. Additionally, they would like to express their gratitude to the participants involved in the experiment, and to the companies developing these devices. We acknowledge Voice Care and Griffin for loaning their devices without cost and thank SonVox for being available to discuss unpublished details about the VoxLog. Additionally, thanks to Kristin Tanner of Brigham Young University for the use of her laboratory's PENTAX Medical APM3200. This research was funded by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under Award No. R01DC012315. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. P.B., A.A., and A.C. have a financial interest in the Voice Care based on a contractual agreement with PR.O. VOICE Srl.

Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proc. of the Institute of Phonetic Sciences*, Vol. 17, No. 1193.

- Bottalico, P., and Astolfi, A. (2012). "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms," *J. Acoust. Soc. Am.* **131**(4), 2817–2827.
- Buekers, R., Bierens, E., Kingma, H., and Marres, E. H. M. A. (1995). "Vocal load as measured by the voice accumulator," *Folia Phoniatr. Logop.* **47**(5), 252–261.
- Carullo, A., Penna, A., Vallan, A., Astolfi, A., and Bottalico, P. (2012). "A portable analyzer for vocal signal monitoring," in *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, pp. 2206–2211.
- Carullo, A., Vallan, A., and Astolfi, A. (2013). "Design issues for a portable vocal analyzer," *IEEE Trans. Instrum. Meas.* **62**(5), 1084–1093.
- Carullo, A., Vallan, A., Astolfi, A., Pavese, L., and Puglisi, G. E. (2015). "Validation of calibration procedures and uncertainty estimation of contact-microphone based vocal analyzers," *Measurement* **74**, 130–142.
- Cheyne, H. A. (2003). "Development and testing of a portable vocal accumulator," *J. Speech Lang. Hear. Res.* **46**, 1457–1467.
- Clark-Carter D. (2005). "Standard error," in *Encyclopedia of Statistics in Behavioral Science*, edited by B. S. Everitt and D. Howell (Wiley, Hoboken, NJ), pp. 1891–1892.
- Ghassemi, M., Van Stan, J. H., Mehta, D. D., Zañartu, M., Cheyne, H. A., Hillman, R. E., and Gutttag, J. V. (2014). "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules," *IEEE Trans. Biomed. Eng.* **61**(6), 1668–1675.
- Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., and Cheyne, H. A. (2006). "Ambulatory monitoring of disordered voices," *Ann. Otol. Rhinol. Laryngol.* **115**(11), 795–801.
- Hunter, E. J. (2009). "A comparison of a child's fundamental frequencies in structured elicited vocalizations versus unstructured natural vocalizations: A case study," *Int. J. Pediatr. Otorhinolaryngol.* **73**(4), 561–571.
- Hunter, E. J., Halpern, A. E., and Spielman, J. L. (2012). "Impact of four nonclinical speaking environments on a child's fundamental frequency and voice level: A preliminary case study," *Lang. Speech Hear. Serv. Sch.* **43**(3), 253–263.
- Hunter, E. J., and Titze, I. R. (2010). "Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings," *J. Speech Lang. Hear. Res.* **53**(4), 862–875.
- ISO 3382-2:2008(E) (2008). "Acoustics—Measurement of Room Acoustic Parameters, Part 2: Reverberation Time in Ordinary Rooms" (International Organization for Standardization, Geneva, Switzerland).
- ISO 9921: 2002(E) (2002). "Ergonomics assessment of speech communication" (International Organization for Standardization, Geneva, Switzerland).
- ISO/IEC Guide 98-3:2008 (2008). "Uncertainty of measurement—Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)" (International Organization for Standardization, Geneva, Switzerland).
- Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne, H. A., and Hillman, R. E. (2012). "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," *IEEE Trans. Biomed. Eng.* **59**(11), 3090–3096.
- Nacci, A., Fattori, B., Mancini, V., Panicucci, E., Ursino, F., Cartaino, F. M., and Berretini, S. (2013). "The use and role of the Ambulatory Phonation Monitor (APM3200) in voice assessment," *Acta Otorhinolaryngol. Ital.* **33**(1), 49–55.
- Schalling, E., Gustafsson, J., Ternström, S., Wilén, F. B., and Södersten, M. (2013). "Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease," *J. Voice* **27**(6), 729–737.
- Searl, J., and Dietsch, A. (2014). "Testing of the VocaLog2 vocal monitor," *J. Voice* **28**(4), 523.e27–523.e37.
- Švec, J. G., Hunter, E. J., Popolo, P. S., Rogge-Miller, K., and Titze, I. R. (2004). "The calibration and setup of the NCVS dosimeter," NCVS Online Technical Memo (No. 2, pp. 1–52).
- Švec, J. G., Popolo, P. S., and Titze, I. R. (2003). "Measurement of vocal doses in speech: Experimental procedure and signal processing," *Logoped. Phoniatr. Vocol.* **28**, 181–192.
- Švec, J. G., Titze, I. R., and Popolo, P. S. (2005). "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," *J. Acoust. Soc. Am.* **117**(3), 1386–1394.
- Van Stan, J. H., Gustafsson, J., Schalling, E., and Hillman, R. E. (2014). "Direct comparison of three commercially available devices for voice ambulatory monitoring and biofeedback," *SIG 3 Perspect. Voice Voice Disord.* **24**(2), 80–86.