# Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings



(Article begins on next page)

# Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings

Antonella Castellana, Alessio Carullo, Arianna Astolfi, Giuseppina Emma Puglisi, and Umberto Fugiglando

Citation: The Journal of the Acoustical Society of America **141**, 2353 (2017); doi: 10.1121/1.4979115
View online: https://doi.org/10.1121/1.4979115
View Table of Contents: https://asa.scitation.org/toc/jas/141/4
Published by the Acoustical Society of America

# Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings

Antonella Castellana,[1,a)] Alessio Carullo,[1] Arianna Astolfi,[2] Giuseppina Emma Puglisi,[2] and Umberto Fugiglando[2]
[1]*Politecnico di Torino, Department of Electronics and Telecommunications, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy*
[2]*Politecnico di Torino, Department of Energy, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy*

(Received 5 December 2016; revised 9 February 2017; accepted 13 March 2017; published online 4 April 2017)

The intra- and inter-speaker variability of speech sound pressure level (SPL) has been investigated under repeatability conditions in this work. In a semi-anechoic chamber, speech from 17 individuals was recorded with a sound level meter, a headworn microphone, and a vocal monitoring device. The subjects were asked to read twice and in sequence two phonetically balanced passages. The speech variability has been investigated for mean, equivalent, and mode SPL from each reading and device. The intra-speaker variability has been evaluated by means of the average among individual standard deviations in the four readings and it reached the maximum of 2 dB for mode SPL. For the inter-speaker variability, the experimental standard deviation of individual averaged SPL parameters among the four repeated measures has been calculated, obtaining the highest value of 5.3 dB for mode SPL. Changes in SPL variability have been evaluated with different logging intervals for each device. The influence of speech material has been investigated by the Wilcoxon test on paired lists of descriptive statistics for SPL distribution and equivalent SPL in the repeated readings. The data reported in this study may be considered as a preliminary reference for the investigation of changes in speech SPL over subjects.
 [http://dx.doi.org/10.1121/1.4979115]

[JFL] Pages: 2353–2363

## I. INTRODUCTION

In different fields that are related to speech, the vocal intensity has been evaluated in terms of statistics or other descriptors of the sound pressure level (SPL), e.g., the mean, mode, or equivalent SPL.

Increased average vocal intensity has been associated to the growth of vocal fold lesions[1] and the deterioration of vocal fold epithelium.[2] In the existing literature, some researchers investigated mean SPL in clinic recordings of a group of patients with vocal nodules and of a control group;[3,4] others used mean SPL as a descriptor of the effects on vocal function of voice therapy.[5] Recently, the development of devices for long-term vocal monitorings[6–8] has allowed a subject's typical vocal behavior to be better characterized by means of descriptive statistics of SPL distributions.[9,10]

In-field voice monitorings have also been used to evaluate the vocal effort of voice professionals, which is a physiological quantity related to voice production that has been quantified in terms of SPL.[11] In particular, among all the professional categories that make a sustained and prolonged voice use, the category of teachers is one of the most affected by voice abuse,[12–14] therefore vocal effort has been widely monitored on them with several methodologies and aims. The mean SPL, mode SPL, and the equivalent SPL, which is the time-weighted average of SPLs, are the most used SPL parameters for the investigation of occupational vocal risk.[15–18]

Moreover, speech SPL has been also estimated in laboratory investigations on speech modifications due to different room acoustics[19] or noise conditions.[20]

The reported researches do not usually take into account the uncertainty contribution due to the repeatability of the SPL estimation related to the subjects involved in the experiments. With the aim of obtaining preliminary results on the spread of repeated measures of SPL of the same subject and of a group of subjects, this study investigates intra- and inter-speaker variability of SPL in continuous speech across repeated readings. The preliminary outcomes provided in this work may be useful to assess the reliability of SPL differences that have been found by researchers in the above-mentioned studies or that will be obtained in future works on speech SPL.

Many studies investigated the variability of SPL parameters, focusing on the possible causes that generate speech modifications. As summarized by Cooke *et al.*,[21] the characteristics related to the addressed listener and to the environment are the dominant factors influencing speech production, as well as the type of speech task.

On the one side, the effect of the acoustical environment on SPL produced by talkers at different communication distances has been examined in several studies (see Pelegrín-Garcia *et al.*[20] for an overview). Among the communication situations where background noise is present, i.e., when realistic communication scenarios are considered, a global increase of speech intensity occurs, leading to the Lombard

a)Electronic mail: antonella.castellana@polito.it

 0001-4966/2017/141(4)/2353/11/$30.00  

speech.[22] However, speech level increases are also observed in the case of talker-to-listener distance increase in absence of noise, perhaps as a form of compensation for perceived listener difficulties.

On the other side, the objective (e.g., vowel, syllable, read-speech, or spontaneous speech) and style (e.g., clear speech or conversational speech) of a speech can significantly affect the voice SPL.[21] "Clear speech" designates any kind of hyper-articulated speech that aims at improving speech intelligibility more than the ordinary and normally articulated conversational speech, which is also uttered at higher speaking rates than the clear speech.[23] An earlier study[24] investigated intra-speaker variation of SPL in 6 subjects related to a reference group data of 15 females and 15 males, while repeating the syllable /pæ/ three times and at three different levels of vocal effort. Corthals[25] investigated the vocal intensity of running speeches collected from 400 subjects with time-weighted SPL estimates, namely equivalent continuous sound levels and percentile levels, which are adequate descriptors for fluctuating sounds.

Other reports investigated the variability of the "comfortable effort level" in different speech materials across experimental sessions. Brown *et al.*[26] recorded 16 subjects during five successive days while producing three times a series of vowels and phrases. Results were reported on within day variation, day to day variation, and subject to subject variation of vocal intensity, which are three different aspects of inter-speaker variability of speech SPL. In a successive work, Brown *et al.*[27] assessed the degree of inter-speaker variability of 50 untrained speakers subdivided into three age groups in a week across utterance types (vowel, reading, and speaking) and recording sessions. Garret and Healey[28] determined the inter-speaker variation of vocal amplitude in three repetitions of connected speech samples acquired from 20 subjects during three time intervals of one day. Sihvo *et al.*[29,30] studied the repeatability and reproducibility of sound level measurement of the softest and loudest possible phonations at five given pitches and between 45-min long readings. The results of the reported studies highlighted that vocal intensity varies from one experimental session to the next when subjects were asked to speak in a comfortable manner. The above review reveals that speech SPL and other related speech parameters[24,31,32] vary within and across speakers, owing to all the aforementioned causes of speech modifications.

The present study is focused on the analysis of the variability of SPL when subjects speak at a comfortable level. Due to the great influence of several factors on speech production, a suitable experimental design has been planned. Environmental effects have not been taken into account as well as the health status of the subjects, since experiments were performed in a semi-anechoic chamber by young talkers who did not report light nor severe voice problems. A proper speech material has been uttered with a normally articulated conversational speech, thus erasing the effects of speech tasks and speaking styles.

Three devices have been employed: a calibrated sound level meter, a headworn microphone, and the vocal analyzer Voice Care. The first one acts as a microphone in air that requires the subjects to remain at a fixed distance during the speech production; the second one is another microphone in air that does not impair the subject from slight movements; the third one is based on a contact microphone that senses the vocal-fold vibrations at the base of the neck.

The purposes of this work can be summarized in the following questions:

(1) How much SPL estimates vary within one speaker in readings?
(2) How much SPL estimates vary in a group of speakers in readings?

These quantities have been assessed for each device and they were named as intra- and inter-speaker variability of SPL, respectively.

Further investigations are related to the influence of speech material on SPL estimates and to the effect of logging intervals on SPL variability.

SPL has been separately computed on readings acquired with each device, thus allowing us to provide preliminary normative data for the assessment of results on SPL obtained in the vast majority of the study in the speech field.

## II. METHOD

Experiments were performed in a semi-anechoic room, where the A-weighted equivalent background noise level was 24.5 dB (33.7 dB unweighted). The mid-frequency reverberation time (from 0.5 kHz to 2 kHz) was 0.11 s. Each participant performed a reading task in the same day and individual measurements were taken subsequently, in a 15-min time interval, thus assuring repeatability conditions.

### A. Subjects

Seventeen native Italian students from Politecnico di Torino (8 males, 9 females) took part in this study (age range 19–26 yrs, mean age 23 yrs). In the semi-anechoic chamber, participants were first asked to perform an audiometric screening test according to the procedure suggested by the iPad-based application titled uHear,[33,34] which provides a hearing sensitivity evaluation per frequency band (from 0.5 kHz to 6 kHz) and with a level-based rating. They obtained results within the normal hearing level. None of them had a history of speech and language disorders, based on self-report, and none of them had professional singing or speaking training.

### B. Speech material

Participants were asked to read aloud two passages twice and in sequence, thus obtaining four repetitions for each subject. The speech material consisted of two standardized phonetically balanced passages (P1 and P2), which were selected being widely used for articulation drills, speech recognition testing, and language studies, because they provide a broad selection of Italian-language sounds.[35] The two passages had different structures and lengths: P1 was a short tale of 300 words and took an average reading time of about 2 min, while P2 was a more expressive text of 124 words and lasted about 1 min.

The choice of readings as experimental tasks was needed to have a various speech material that would have been the same for each participant in the experiment.

Subjects were instructed not to whisper in a soft voice nor shout in a loud voice, but they were advised to choose comfortable levels of loudness and pitch for a normally articulated conversational speech. The texts of P1 and P2 were printed on sheets and laid over a sound absorbing panel hung on a music stand, in front of the speaker's eyes, at a distance of 1 m.

### C. Measurement setup and procedure

The reading uttered by each subject was recorded simultaneously by means of three measurement chains, namely:

(1) A calibrated sound level meter (XL2, NTi Audio, Schaan, Liechtenstein), with a class 1 omnidirectional measurement microphone M2210 by NTi Audio. For the entire period of the test, each subject was asked to stand in front of the microphone, on axis, at the fixed distance of 16 cm as provided by a thin spacer. The recommended mouth-to-microphone distance for this kind of measurements is 30 cm[36,37] and with this suggested distance, when the background noise level is lower than 25 dBA, the low-intensity voice levels can be obtained with a signal-to-noise ratio (SNR) of at least 10 dB.[38] The authors reduced this distance to 16 cm in order to increase the SNR, since the microphone of the sound level meter is an omnidirectional one and it is not affected by the proximity effect,[37] i.e., the increase of the low-frequency boost in the frequency response of a directional microphone as the mouth-to-microphone distance drops;

(2) An omnidirectional headworn microphone Mipro MU-55HN, which was placed at a distance of about 2.5 cm from the lips of the talkers, slightly to the side of the mouth, at about 20°–45° horizontally, depending on the subjects' face shape. The microphone, which exhibits a flatness of $\pm 3$ dB in the range from 40 Hz to 20 kHz, was connected to a bodypack transmitter ACT-30 T, which transmits to a wireless system Mipro ACT 311. The output signal of this system was recorded with a handy recorder ZOOM H1 (Zoom Corp., Tokyo, Japan) that uses a sample rate of 44.1 kHz and 16 bit of resolution. The authors checked SPL linearity and the absence of SPL compression effect using this system, setting the transmitter without the automatic gain control;

(3) A portable vocal analyzer, namely the Voice Care device (PR.O.VOICE, Turin, Italy), which was recently developed at Politecnico di Torino.[7] It consists of a data-logger connected to an Electret Condenser Microphone [ECM AE38 (Alan Electronics GmbH, Dreieich, Germany)], which is fixed at the jugular notch by means of a surgical band, thus sensing the skin vibrations induced by the vocal-fold activity. The output signal of the ECM is suitably conditioned through an analogue circuitry in order to match its characteristics (amplitude and frequency content) to the analogue-to-digital converter internal to a micro-controller based board. The raw samples are stored on an internal memory device (SD card). In order to estimate the speech SPL of the speaker at a fixed distance of 16 cm in front of the mouth, each subject had to perform a preliminary calibration, repeating the vowel /a/ at increasing levels in front of a microphone in air (Behringer ECM8000), used as a reference. Such procedure, which is needed before starting the experiment, is designed to identify the function that relates the voltage signal at the output of the ECM chain to the reference SPLs at the fixed distance from the mouth of the subject under monitoring.[39]

Figure 1 shows a female subject who performed the experiment and who was equipped with all the measurement devices.

Before reading each passage, subjects simultaneously repeated the vowel /a/ and tapped twice the ECM with their hands in order to produce sharp peaks on the speech signals acquired by the two microphones in air and by the ECM, respectively. These peaks were considered as reference points to select signals to be analyzed in the post-processing.

Among all the collected recordings (336 min in total), some of them were discarded in the data processing due to the failure of the preliminary calibration procedure of the Voice Care[39] and/or for incorrect execution of the experiment, (e.g., one subject moved his lips far away from the thin spacer of the sound level meter during the test). Three females performed the experiment only wearing the Voice Care. Therefore, a different number of subjects were taken into account for the three devices: 13 subjects (7 males, 6 females) were considered for the sound level meter, 14 subjects (8 males, 6 females) for the headworn microphone, and 12 subjects (7 males, 5 females) for the Voice Care device. The results were separately analyzed, since a comparison among the different devices is not the goal of the experiment.

### D. Data processing

The stored data obtained for each participant and device were transferred to a Personal Computer and subdivided into

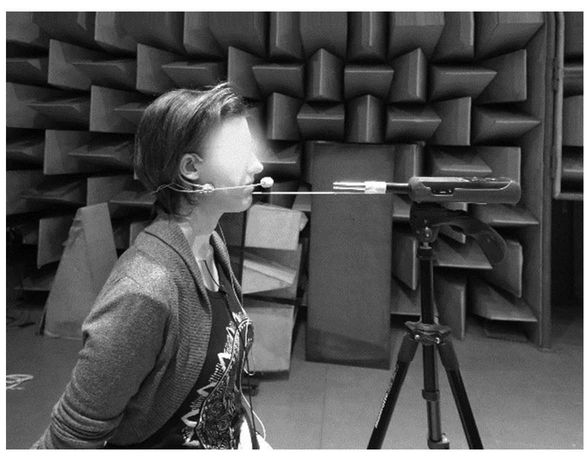

FIG. 1. Female subject while standing in front of the sound level meter XL2 by NTi Audio and wearing the headworn microphone Mipro MU-55HN and the ECM AE38 of Voice Care during the experiment.

different files, using the sharp peaks at the beginning of each reading as the starting time instant for each file. This procedure was done using the software Adobe Audition (version 3.0) for the WAV audio files recorded by the sound level meter and the headworn microphone. A specific MATLAB (R2014b, version 8.4) script was implemented for data stored in the Voice Care device.

Then, each repetition of the two passages collected per each device and subject was post-processed with specific MATLAB scripts for the estimation of speech SPL occurrences, obtaining histograms with a bin resolution of 1 dB. A speech SPL distribution was thus obtained per each reading, based on the logging interval of each device without using any windowing. SPLs were estimated with a logging interval of 1 s for signals acquired from the sound level meter, since it is the most common interval that is set in class 1 sound level meters. The same logging interval was used for signals acquired from the headworn microphone. The samples acquired with the Voice Care were grouped into frames of 30 ms and only *voiced frames* have been processed.[39] The choice of such interval arose from the evidence that the minimum duration of pauses in Italian readings is equal to 60 ms,[40] but pause lengths of 30 ms can also occur in storytelling style speech,[41] so that a 30 ms interval guaranteed an effective discrimination between *voiced* and *unvoiced* frames. This frame duration is also used in other dosimeters that are equipped with contact sensors.[42,43] Suitable root-mean-square voltage threshold was identified in order to distinguish voiced and unvoiced frames per each file, according to a procedure described by Carullo *et al.*[7] SPL values for voiced frames at a fixed distance of 16 cm from the speaker's mouth was then obtained, thanks to the calibration function estimated for each subject.

A calibration sine-wave file at a level of 94 dB, which was registered by coupling the sound level meter to a pressure calibrator B&K 4230, was used as a reference value in the analysis of WAV signals acquired with the sound level meter. A different reference value was used in the analysis of data recorded by the headworn microphone and it was estimated by means of a comparative calibration procedure between the headworn microphone and the sound level meter, used as a reference device (see details in the Appendix).

### E. Speech SPL parameters and analysis

SPL occurrences of each reading constituted SPL distributions that characterized each individual speech sample. The mean, mode, and equivalent SPLs ($SPL_{mean}$, $SPL_{mode}$, and $SPL_{eq}$, respectively), which are the most representative descriptive parameters for the intensity of speech production, were obtained for each reading and subject. The estimation of $SPL_{eq}$ from data acquired with the Voice Care device was performed implementing the same equation proposed by Ŝvec *et al.*,[44] as follows:

$$\mathrm{SPL_{eq}} = 10\log_{10}\left(\frac{1}{N}\sum_{n=1}^{N}\left[k_v(n)10^{\mathrm{SPL}(n)/10}\right]\right), \tag{1}$$

where $n$ is the frame index, $N$ is the total number of frames in the analyzed segment of speech, $k_v$ is the voicing unit step function (1 for voiced and 0 for unvoiced frame), and $\mathrm{SPL}(n)$ is the SPL value within the frame $n$.

To estimate the SPL variability, the type A method proposed in the Guide to the Expression of Uncertainty in Measurement (GUM) (Ref. 45) has been followed both in the experimental design and in the result processing. SPL values have been considered as random variables and SPL variability has been estimated as the experimental standard deviation of the available data.

#### 1. Intra-speaker variability of speech SPL

With the purpose of finding the intra-speaker variability of speech SPL that occurred across readings, the experimental standard deviation of the four repeated measures for each $i$th subject, hereafter referred as $s_i$, was calculated for $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$.

Then, for each device-group and SPL parameter, the average of $s_i$ values ($\bar{s}$) and its 95% confidence interval (CI) for the mean based on a $t$ critical value were calculated. This estimate has been considered as the mean descriptive parameter for intra-speaker variability in speech SPL, since it denotes, on average, the variability of vocal intensity referred to a general speaker. The $t$ critical value changed depending on the number of subjects who performed the experiment with each device[45] (i.e., it was calculated as 2.18 for the sound level meter-group, 2.16 for the headworn-group, and 2.20 for the Voice Care-group).

A further investigation of the individual SPL variability has been performed by the estimation of the maximum differences among the four repeated measures ($\Delta$) of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ for each subject and device.

#### 2. Inter-speaker variability of speech SPL

Aiming to quantify the individual variability of speech SPL among speakers, also known as inter-speaker variability, the experimental standard deviation of each device-group, $s(g)$, was calculated for $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$, according to the following expression:

$$s(g) = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(r_j - \bar{r})^2}, \tag{2}$$

where $n$ is the number of subjects, $r_j$ represents the average of the four SPL measures obtained from the four repeated readings for each subject, and $\bar{r}$ is the overall mean among the $r_j$ values for each device.

This quantity denotes the variability of vocal intensity in a group of speakers and it has been considered as the mean descriptive parameter for inter-speaker variability in speech SPL.

The standard deviation of the mean, or standard error, $s_m$, was also obtained as the ratio between $s(g)$ and the root square of $n$, where $n$ is the device-group sizes of the participants in the experiment. This estimate may be a reference for the investigation of changes in speech SPL over groups

TABLE I. Results on speech SPL variability obtained from the readings recorded with the calibrated sound level meter (SLM) XL2 at 16 cm from the speaker's mouth. Intra-speaker variability results: average of the individual standard deviations of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ in the four readings, $\bar{s}$, and 95% CI for the mean based on a $t$ critical value; minimum and maximum differences ($\Delta$) of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ in the four repeated readings among subjects. Inter-speaker variability results: group mean and experimental standard deviation, $s(g)$, of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ obtained from all subjects.

| Variability | $SPL_{eq}$ (dB) | | $SPL_{mean}$ (dB) | | $SPL_{mode}$ (dB) | |
|---|---|---|---|---|---|---|
| Intra-speaker | $\bar{s}$ (CI) | min, max $\Delta$ | $\bar{s}$ (CI) | min, max $\Delta$ | $\bar{s}$ (CI) | min, max $\Delta$ |
| | 0.4 (0.2–0.6) | 0.2, 2.2 | 0.6 (0.4–0.8) | 0.3, 2.6 | 1.0 (0.7–1.3) | 1.0, 4.0 |
| Inter-speaker | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ |
| | 76.3 | 3.9, 1.1 | 74.4 | 3.5, 1.0 | 76.6 | 4.0, 1.1 |

of subjects or for the same group of subjects in different conditions, when a comparison between averaged measures has to be performed. It represents a significant parameter to be used as a reasonable uncertainty contribution for the mean value of the group of data, since it takes into account the group size.

### 3. Influence of reading material on SPL variability

Further analysis on speech SPL distributions has been conducted in order to investigate if the reading of two different passages can affect SPL variability, comparing differences in material P1 and P2 according to the voice intensity produced. For each speech SPL distribution, the following descriptive statistics were calculated: mean ($SPL_{mean}$), median ($SPL_{median}$), and mode ($SPL_{mode}$) as measures of location of the distribution; standard deviation ($SPL_{sd}$) and the interval between the maximum and minimum value ($SPL_{int}$) as measures of its variance, kurtosis ($SPL_{kurt}$), and skewness ($SPL_{skew}$) for the characterization of the distribution shape.

With the purpose of investigating the speech SPL distributions, the two-tailed Wilcoxon signed ranks test has been applied, that is a non-parametric test based on dependent paired samples.[46] All the descriptive statistics of the SPL distribution and $SPL_{eq}$ were calculated for each repetition and subject involved in the study, and two pairs were thus obtained for each subject, one related to the two readings of the first passage (P1a–P1b) and the other related to the two readings of the second passage (P2a–P2b). The average values of each SPL parameter between the two readings of each passage were also calculated for each subject (P1m–P2m). The Wilcoxon signed ranks test has been applied to all the paired lists of descriptive statistics for SPL distributions related to each group device. The adopted statistical test does not require any specific assumptions on the distribution, and the null hypothesis ($H_0$) states that $MD = 0$, where $MD$ is the median of the difference between the descriptors of the paired sample in the two readings of each reading passage. The one-sample Kolmogorov-Smirnov test verified that data in each list did not come from a normal distribution, except for the kurtosis values of the SPL distributions ($SPL_{kurt}$) obtained from Voice Care, thus justifying the use of a non-parametric test for the analysis.

### 4. Influence of logging intervals on SPL variability

Further investigations have been carried out for determining how different logging intervals can affect the speech SPL variabilities. Vocal data acquired with the sound level meter and the headworn microphone has been post-processed with a frame length of 30 ms, 250 ms, and 500 ms. The same analyses described in Secs. II E 1 and II E 2 were then performed.

## III. RESULTS

### A. Speech SPL variability

Table I shows the results of speech SPL variability obtained from the readings that were recorded with the sound level meter at 16 cm from the speaker's mouth. $SPL_{eq}$ shows the minimum variability within one speaker, having the minimum $\bar{s}$, that is 0.4 dB (95%-CI between 0.2 dB and 0.6 dB). Furthermore, $SPL_{eq}$ shows the lowest range between the minimum and maximum $\Delta$, which is equal to 2 dB, while $SPL_{mode}$ has both the maximum intra- and inter-speaker variability, showing $\bar{s}$ equal to 1.0 dB (95%-CI between 0.7 dB and 1.3 dB) and $s(g)$ of 4.0 dB. The intra-speaker variability of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ presents values at least 4 times lower than those of the inter-speaker variability. The standard error $s_m$ is equal to 1.0 dB for $SPL_{mean}$ and 1.1 dB for both $SPL_{eq}$ and $SPL_{mode}$.

Table II shows the results of speech SPL variability obtained from the readings that were recorded with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth. $SPL_{eq}$ shows the minimum variability within one speaker, with both the minimum $\bar{s}$ of 0.5 dB (95%-CI between 0.3 dB and 0.7 dB) and the lowest range between the minimum and maximum $\Delta$ of 2.2 dB,

TABLE II. The same as Table I. Data refers to speech SPL obtained from the readings recorded with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth.

| Variability | $SPL_{eq}$ (dB) | | $SPL_{mean}$ (dB) | | $SPL_{mode}$ (dB) | |
|---|---|---|---|---|---|---|
| Intra-speaker | $\bar{s}$ (CI) | min, max $\Delta$ | $\bar{s}$ (CI) | min, max $\Delta$ | $\bar{s}$ (CI) | min, max $\Delta$ |
| | 0.5 (0.3–0.7) | 0.1, 2.3 | 0.6 (0.5–0.8) | 0.2, 2.4 | 1.1 (0.7–1.5) | 1.0, 5.0 |
| Inter-speaker | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ |
| | 95.1 | 5.0, 1.3 | 93.2 | 4.7, 1.3 | 95.4 | 5.3, 1.4 |

TABLE III. The same as Table I. Data refers to the readings recorded with the Voice Care, which estimates speech SPL at 16 cm from the speaker's mouth.

| Variability | $SPL_{eq}$ (dB) | | $SPL_{mean}$ (dB) | | $SPL_{mode}$ (dB) | |
|---|---|---|---|---|---|---|
| Intra-speaker | $\bar{s}$ (CI) | min, max Δ | $\bar{s}$ (CI) | min, max Δ | $\bar{s}$ (CI) | min, max Δ |
| | 0.8 (0.3–1.0) | 0.3, 5.2 | 0.6 (0.3–0.9) | 0.2, 3.9 | 1.5 (0.8–2.2) | 1.0, 9.0 |
| Inter-speaker | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ | Group mean | $s(g)$, $s_m$ |
| | 77.9 | 2.8, 0.8 | 77.7 | 2.8, 0.8 | 79.4 | 3.0, 0.9 |

while $SPL_{mean}$ shows the minimum variability among speakers, with $s(g)$ equal to 4.7 dB. $SPL_{mode}$ has the maximum values for both the inter- and intra-speech variability, with $\bar{s}$ equal to 1.1 dB (95%-CI between 0.7 dB and 1.3 dB) and $s(g)$ of 5.3 dB. The intra-speaker variability of $SPL_{mode}$ and $SPL_{eq}$ presents values at least 5 and 10 times lower than those of the inter-speaker variability, respectively. The standard error $s_m$ is 1.3 dB for $SPL_{eq}$ and $SPL_{mean}$ and 1.1 dB for $SPL_{mode}$.

Results on SPL variability that have been obtained from readings recorded with the Voice Care, whose data refers to 16 cm from the speaker's mouth, are summarized in Table III. $SPL_{mean}$ shows the lowest intra-speaker variability, with both the minimum $\bar{s}$, that is 0.6 dB (95%-CI between 0.3 dB and 0.9 dB), and the lowest range between the minimum and maximum Δ of 3.7 dB, while $SPL_{mode}$ shows the highest inter-speaker variability with $s(g)$ equal to 3.0 dB. The values of intra-speaker variability are at least 3 times lower than the inter-speaker ones, except for $SPL_{mode}$ that has the variability contributions that differ less than 1 dB. The standard error $s_m$ is 0.8 dB for $SPL_{eq}$ and $SPL_{mean}$ and 0.9 dB for $SPL_{mode}$.

In the present study, the absolute values of the estimated SPL parameters have not been mentioned, because they are not directly included in the questions under investigation. However, Tables I, II, and III report the group mean of each SPL parameter as complementary data for the inter-speaker variability and Fig. 2 shows some details about the speech levels, since it summarizes for each SPL parameter and device the individual mean of the four repeated measures with the respective standard deviations ($s$) and the overall mean value with the relative experimental standard deviations, $s(g)$.

### B. Influence of reading material on SPL parameters

Table IV shows the $p$-values obtained for each group-device, and for each paired list of SPL parameters related to P1a–P1b, P2a–P2b, and P1m–P2m. None of the paired lists of quantities present significant differences across the two readings of the same passage among subjects ($p$-values >0.05), with the exception of $SPL_{kurt}$ for the readings of the second passage acquired with the Voice Care. A main result of this analysis is that, generally, each subject in each device-group repeated the reading of the same passage with similar speech levels. On the other hand, from the analysis of the paired lists of P1m and P2m significant differences

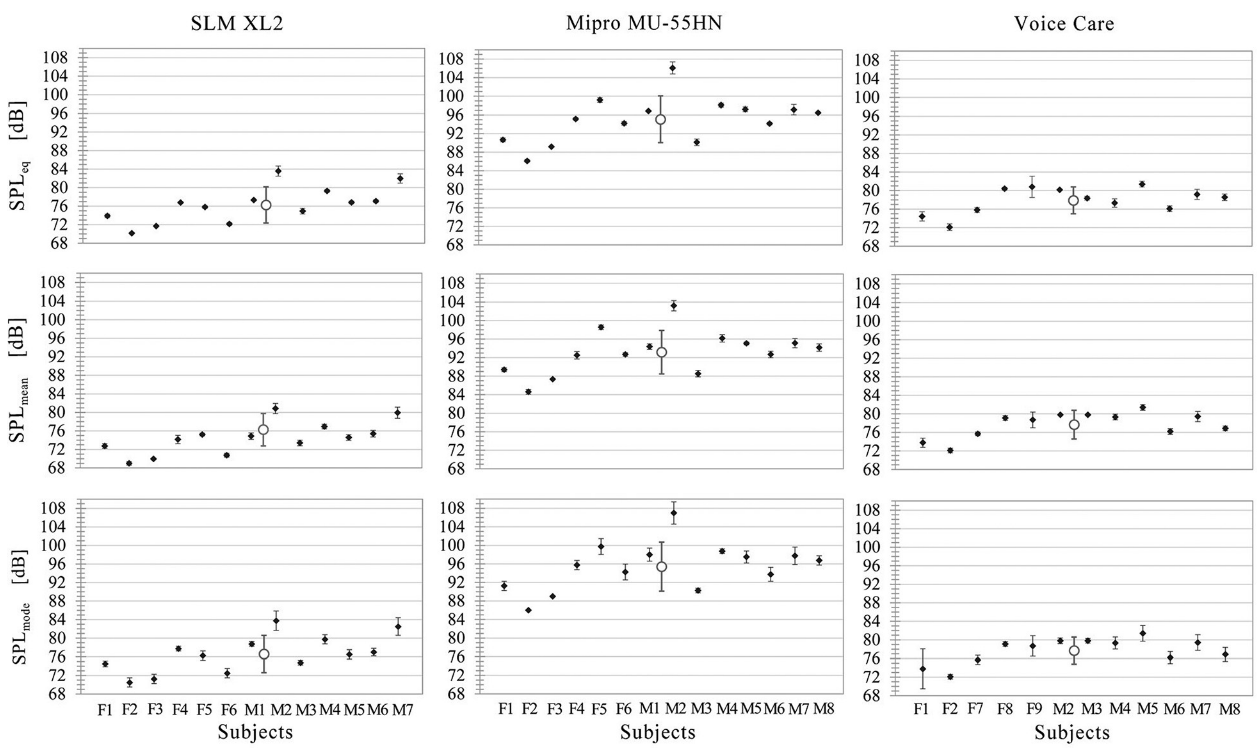


FIG. 2. Averaged values of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ in the two readings of the two passages for each subject, four total repetitions (diamond points); bars indicate the experimental standard deviation, $s$, for each subject. Overall mean values among subjects are indicated as circle points; bars indicate the experimental standard deviations, $s(g)$, of averaged values.

TABLE IV. *P*-values of the two-tailed Wilcoxon signed ranks test of the paired lists of descriptive statistics for the SPL distributions and $SPL_{eq}$, related to the repetitions of the first passage (P1a, P1b) and of the second passage (P2a, P2b). *P*-values refer also to pooled data from the two readings (P1m, P2m). Values lower than a significance level of 0.05 are in bold and indicate the rejection of the null hypothesis H0: MD = 0, where MD is the median of the population of the differences between the paired sample data.

| Device | Speech materials | $SPL_{mean}$ (dB) | $SPL_{sd}$ (dB) | $SPL_{median}$ (dB) | $SPL_{mode}$ (dB) | $SPL_{kurt}$ (-) | $SPL_{skew}$ (-) | $SPL_{int}$ (dB) | $SPL_{eq}$ (dB) |
|---|---|---|---|---|---|---|---|---|---|
| SLM XL2 | P1a–P1b | *0.556* | *0.464* | *0.125* | *0.828* | *0.588* | *0.984* | *0.840* | *0.151* |
| | P2a–P2b | *0.576* | *0.852* | *0.625* | *0.305* | *0.305* | *0.210* | *0.721* | *0.490* |
| | P1m–P2m | *0.924* | ***0.040*** | *0.250* | *0.117* | ***0.040*** | *0.124* | ***0.034*** | *0.138* |
| Mipro MU-55HN | P1a–P1b | *0.571* | *0.424* | *0.188* | *0.766* | *0.658* | *0.886* | *0.307* | *0.140* |
| | P2a–P2b | *0.690* | *0.572* | *1.000* | *0.090* | *0.391* | *0.199* | *0.764* | *0.419* |
| | P1m–P2m | *0.653* | ***0.010*** | ***0.002*** | ***0.035*** | ***0.016*** | *0.092* | ***0.035*** | ***0.009*** |
| Voice Care | P1a–P1b | *0.938* | *0.231* | *1.000* | *0.654* | *0.727* | *1.000* | *0.510* | *0.787* |
| | P2a–P2b | *0.574* | *0.924* | *1.000* | *0.941* | ***0.031*** | *0.176* | *0.488* | *0.639* |
| | P1m–P2m | *0.310* | *0.197* | *0.766* | *0.199* | *0.360* | *1.000* | *0.214* | *0.916* |

have been found. $SPL_{sd}$, $SPL_{kurt}$, and $SPL_{int}$ significantly change in readings of the two passages acquired with the sound level meter. In the case of the headworn microphone, SPL parameters corresponding to P1m that result significantly different from SPL parameters obtained by P2m are $SPL_{sd}$, $SPL_{median}$, $SPL_{mode}$, $SPL_{kurt}$, $SPL_{int}$, and $SPL_{eq}$. None of the SPL parameters significantly changes for the Voice Care. These outcomes reveal that subjects recorded with the sound level meter and the headworn microphone tended to read the two passages with different sound speech levels. Therefore, negligible intra-speaker variability would be expected in the repetition of the same passage, but non-negligible intra-speaker variability could be expected between the readings of the two different passages. These findings validates our choice about the experiment, since two readings of two different passages may guarantee a sufficiently diversified speech material.

### C. Influence of logging intervals on SPL variability

Table V shows results on speech SPL variability obtained by post-processing the readings acquired with the sound level meter and the headworn microphone with different logging intervals. The intra-speaker variability, $\bar{s}$, for $SPL_{eq}$ keeps constant by post-processing the reading samples with logging intervals equal to 1 s, 750 ms, 500 ms, 250 ms, and 30 ms, both for the sound level meter and the headworn microphone. For both the microphones, $SPL_{mean}$ has a deviation of 0.1 dB among $\bar{s}$ values, while $SPL_{mode}$ shows an upward trend of $\bar{s}$ when logging intervals decrease. An extreme result of 6.4 dB can be easily noticed in the $\bar{s}$ values obtained from data acquired with the headworn microphone and post-processed with a logging interval of 30 ms. It is due to the conjunction of two phenomena, that are the use of 30 ms frame length and the internal noise of the measurement chain of the headworn microphone. Such a frame length is short enough to obtain several SPL occurrences of the unvoiced frames, which could have SPL values similar to the background noise in the semi-anechoic chamber. This assumption has been confirmed, since a silent period of 10 s was recorded with the headworn microphone and the equivalent level was equal to 46 dB, which actually is the internal noise of the headworn microphone.

Figure 3 shows two distributions of SPL occurrences, which both exhibit a bimodal shape, obtained using a 30 ms logging-interval of a reading that was simultaneously acquired with the sound level meter and the headworn microphone. The SPL distribution that refers to the sound level meter has the lowest peak-level equal to 34 dB, while the SPL distribution of the headworn microphone has the lowest peak-level equal to 48 dB. As highlighted by Hodgson *et al.*,[16] the lowest-peak level of a long-term speech corresponds to the background noise that occurs during the voice monitoring. The SPL distribution obtained from the sound level meter reflects this finding, since a correspondence between the

TABLE V. Results of speech SPL variability obtained by post-processing the reading voice signals of readings with different logging intervals. Speech samples are recorded with the calibrated sound level meter (SLM) XL2 at 16 cm from the speaker's mouth and with the headworn microphone Mipro MU-55HN at a distance of 2.5 cm from the speaker's mouth. Intra-speaker variability results: average of the individual standard deviations, $\bar{s}$, of $SPL_{eq}$, and the mean $SPL_{mean}$ and mode $SPL_{mode}$ in the four readings. Inter-speaker variability results: experimental standard deviation, $s(g)$, of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ obtained from all subjects.

| SPL parameter (dB) | Logging interval (ms) | SLM XL2 | | Mipro MU-55HN | |
|---|---|---|---|---|---|
| | | $\bar{s}$ | $s(g)$ | $\bar{s}$ | $s(g)$ |
| $SPL_{eq}$ | 1000 | 0.4 | 3.9 | 0.5 | 5.0 |
| | 750 | 0.4 | 3.9 | 0.5 | 5.0 |
| | 500 | 0.4 | 3.9 | 0.5 | 5.0 |
| | 250 | 0.4 | 3.9 | 0.5 | 5.0 |
| | 30 | 0.4 | 3.9 | 0.5 | 5.1 |
| $SPL_{mean}$ | 1000 | 0.6 | 3.5 | 0.7 | 4.7 |
| | 750 | 0.6 | 3.4 | 0.7 | 4.5 |
| | 500 | 0.7 | 3.3 | 0.7 | 4.5 |
| | 250 | 0.7 | 3.1 | 0.8 | 4.3 |
| | 30 | 0.7 | 2.7 | 0.8 | 3.9 |
| $SPL_{mode}$ | 1000 | 0.9 | 4.0 | 1.1 | 5.3 |
| | 750 | 1.0 | 4.1 | 1.1 | 4.9 |
| | 500 | 1.1 | 4.1 | 1.3 | 4.9 |
| | 250 | 1.2 | 4.1 | 1.4 | 5.0 |
| | 30 | 1.3 | 3.9 | **6.4** | **13.9** |

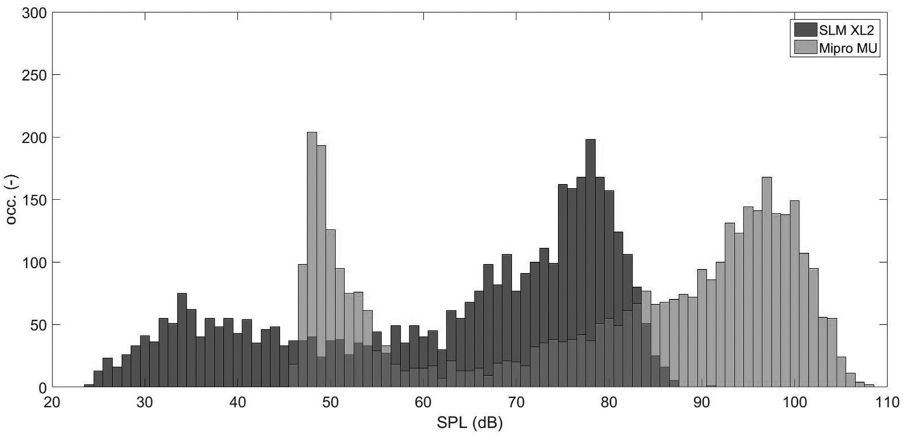

FIG. 3. Two distributions of SPL occurrences obtained from the analysis of a reading that was simultaneously acquired with both the SLM XLS (dark gray bar) and the headworn microphone Mipro MU-55HN (light gray bar). The logging interval used in the post-processing was 30 ms.

lowest peak-level (34 dB) and the background noise that was measured in the semi-anechoic chamber (33.7 dB, as reported in Sec. II) has been found. For the headworn microphone, a difference of 2 dB between the internal noise and the lowest peak-level occurs. However, it seems that occurrences of both low SPLs and internal noise have been accumulated at 48 dB, determining the highest peak-level in correspondence of that value. This phenomenon results in 7 out of 56 SPL distributions with the highest occurrence near to the internal noise level, thus achieving the extreme $\bar{s}$ value of 6.4 dB.

Table V also shows results on the inter-speaker variability of SPL parameters in the two microphones. Despite the varying of logging intervals, $s(g)$ remains the same for $SPL_{eq}$ with a deviation of 0.1 dB for the headworn microphones, while it shows a downward trend when logging intervals decrease both in $SPL_{mean}$ and $SPL_{mode}$, with the exception of the extreme value of $s(g)$ that corresponds to the 30 ms logging interval in $SPL_{mode}$, that is 13.9 dB. This anomalous $s(g)$ behavior can be attributed to the same phenomenon that has been explained above.

## IV. DISCUSSION

The aim of the present study was to examine how speech SPL varies within one subject and in a group of speakers across repeated readings acquired with three devices, with the attempt to provide preliminary data on the intra- and inter-speaker variability in speech SPL, respectively. In a semi-anechoic chamber, 17 individuals with no speech and language disorders were recorded with a class-1 sound level meter, a headworn microphone, and a portable vocal analyzer, while reading out twice and in sequence two phonetically balanced passages. Three parameters related to the speech SPL have been obtained from each reading and device, which are $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$.

### A. Speech SPL variability

The average value of the individual standard deviations among the repeated measures of the two passages, $\bar{s}$, has been considered as the descriptive parameter for the intra-speaker variability of speech SPL. The experimental standard deviation, $s(g)$, of $SPL_{eq}$, $SPL_{mean}$, and $SPL_{mode}$ obtained from all the subjects have been used for characterizing the speech SPL inter-speaker variability.

Due to the different computation algorithm implemented by the Voice Care to estimate SPL distributions, i.e., the vocal analyzer estimates SPLs only on *voiced frames*, the outcomes from the three devices have not been compared. However, some common considerations on the variability of SPL parameters in the three devices can be made. The intra-speaker variability of $SPL_{eq}$ and $SPL_{mean}$ results negligible for the three devices, that is within 1 dB, while it is higher for $SPL_{mode}$, reaching 2 dB. These outcomes are not surprising, since they reflect the type of parameter under analysis: $SPL_{eq}$ and $SPL_{mean}$ express average measures, while $SPL_{mode}$ represents the most frequent observation among SPL occurrences.

The results of this study cannot be compared with most of the outcomes by other researches, because of the difference in the experimental procedure and measurement equipment. Previous works on speech SPL in readings investigated the intra-speaker variability of vocal intensity within days or within times in a day.[26–28] In the present study, instead, the evaluation has been done within successive reading tasks performed in a few minutes, in order to ensure repeatability conditions.

Table I shows that the inter-speaker variability of $SPL_{eq}$ estimated from signals acquired with the sound level meter was 3.9 dB. This outcome is in agreement with the result that Corthals[25] found for the youngest group of participants in his experiment. It should be noted that even if both the groups are made of young people, the age range of young subjects who participated to Corthals's experiment (from 7 to 17 yrs) did not match the one of participants who were involved in this study (from 19 to 26 yrs). Otherwise, the standard deviation of vocal intensity, in relative dB, that Brown *et al.*[26] obtained for the reading task in the young group of people, i.e., 1.9 dB, resulted definitely lower than the inter-speaker variability of SPL parameters that has been found in this work for both the sound level meter and the headworn microphone, which is shown in Tables I and II, respectively.

### B. Considerations on reading material

The *p*-values of the two-tailed Wilcoxon signed rank test of the paired lists of descriptive statistics for SPL distributions and $SPL_{eq}$, related to the repetitions of the first and second passage, reveal additional aspects of speech SPL variability: people tend to read the same passage without variations in SPL, i.e., $p$-value>0.05 for P1a–P1b and P2a–P2b pairs, while they have a tendency to read different speech materials with altered SPL, i.e., some $p$-value<0.05 occurred for P1m–P2m. However, the reading order of P1 and P2 was not counterbalanced in the subjects, so that a time recording effect can happen, that is the two readings of P2 always followed the two readings of P1 and P1m and P2m had different SPLs due to an effect of either speaker fatigue or speaker habituation to the recording environment. An additional aspect is that P2 is a more expressive passage than P1. Therefore, the one-left-tailed Wilcoxon signed rank test has been performed to the SPL paired lists. $SPL_{sd}$ for the sound level meter, $SPL_{sd}$, $SPL_{median}$, $SPL_{mode}$, and $SPL_{eq}$ for the headworn microphone have $p$-value<0.05. In other words, for these two device-groups, values of such SPL parameters in the first passage resulted significantly lower than those in the second passage, thus having the presence of a certain time recording effect or the more expressive nature of the second passage as possible reasons. None of the SPL parameters has a $p$-value<0.05 for the Voice Care.

The results obtained for the Voice Care from both the two- and one-tailed Wilcoxon signed ranks test give an indication that pauses, i.e., *unvoiced frames* that are discarded in the process algorithm, are relevant in the distribution of SPL.

### C. Considerations on logging intervals

It should be noted that SPL distributions from reading samples of 1 min using logging intervals of 1 s have only 60 points, thus their descriptive statistics could have some random variation linked to the low number of data points, especially for $SPL_{mode}$. With shorter frame durations, instead, a greater number of points are included in the histogram of the measured data, thus allowing an underlying theoretical random distribution to be estimated, so that $SPL_{mode}$ becomes a good estimator of the peak of the distribution. Despite this, the intra-speaker variability of $SPL_{mode}$ increases as logging intervals decrease, while a not clear trend has been identified for the inter-speaker variability of $SPL_{mode}$. In addition, anomalous values of $SPL_{mode}$ variability have been found for the readings acquired with the headworn microphone that were post-processed with a 30 ms frame duration, due to the reasons explained in the Sec. III C.

$SPL_{mean}$ reveals slight changes in the intra-speaker variability and a decrease of the inter-speaker variability as logging intervals become shorter. As highlighted by Švec *et al.*,[44] there is evidence that different $SPL_{mean}$ values can be obtained for the same voice signal when different logging intervals are used in the analysis, so that modifications in $SPL_{mean}$ variabilities can be expected.

Eventually, both the intra- and inter-speaker variability of $SPL_{eq}$ keep quite constant as logging intervals change, according to its definition of time-weighted average of SPLs.

### D. Limitations and application of the results

The results reported in the present study may be affected by the lower reproducibility due to the relative position between the subject and the devices during the experiment. For the sound level meter, subjects could have slightly moved their lips from the thin spacer during the readings. The arch of the headworn microphone is crucial for two main reasons: it could have slightly changed the distance from the lips and the microphone during the experiment because of its thinness and it has a fixed length that caused a different horizontal angle from the mouth, depending on the subjects' shape of the face. Therefore, the microphone could be placed in the airflow area for some subjects, thus acquiring unwanted artefacts despite the use of the windscreen. Further precautions are needed in future research. In addition, the estimation of SPL from the wearable vocal analyzer is affected by the sensitivity of the ECM with respect to body activity, the so-called tissue-borne effects, which could occur during voice monitoring. It provides an additional contribution of uncertainty in the estimation of speech SPL.[39]

The outcomes in the present study are preliminary, mainly because of the limited number of subjects who took part in the experiment. Further researches should involve more subjects and it could be useful to ask the speakers to repeat more than four readings, in order to obtain more reliable values from individual standard deviations.

It is also important to consider the application of these preliminary types of normative data. The results of the intra-speaker variability may be particularly useful in studies that investigate individual differences in speech SPL, which can be measured in two different periods or conditions. The outcomes of the inter-speaker variability may be a reference for the investigation of changes in speech SPL over groups of subjects. When a comparison between averaged measures among groups of subjects have to be performed, researchers may refer to $s(g)$ values given in this study and calculate the standard deviation of the mean ($s_m$), or standard error, which can be obtained as a ratio between $s(g)$ and the root square of $n$, where $n$ is the group size of the participants in the experiment. It is important to underline that the use of values given in this paper is limited to situations in which the equipment and experimental setup are the same as those in the present study.

Researchers often make comparisons between different situations, e.g., states of health and room acoustic conditions, and evaluate SPL trends in a subject or among groups of subjects in long-term monitorings. As a general rule, when differences are greater than $\bar{s}$ and $s(g)$, it can be assumed that the new aspect that changes the previous situation has a significant influence on the intensity in speech production in a single subject or in groups of subjects, respectively.

## V. CONCLUSIONS

The present paper deals with the variability of speech SPL within a speaker, i.e., intra-speaker variability, and in a group of people, i.e., inter-speaker variability, in successive readings that have been recorded with a sound level meter, a headworn microphone, and a portable vocal analyzer, which is named Voice Care.

The main conclusions are as follows:

(a) For each device, the intra-speaker variability results within 1 dB for $SPL_{eq}$ and $SPL_{mean}$, while it reaches

2 dB for $SPL_{mode}$. In addition, it was always lower than the respective inter-speaker variability.

(b) The inter-speaker variability of the three devices ranges from 2.8 to 5.3 dB, having always the highest values for $SPL_{mode}$.

(c) For the sound level meter and the headworn microphone, negligible changes of descriptive statistics for SPL distributions and $SPL_{eq}$ have been obtained in the repetition of the same passage, while significant differences have been found in readings of different passages. Fewer modifications have been highlighted for the Voice Care, probably due to the different post-processing that it implements for the SPL estimation.

(d) Both the intra- and inter-speaker variability of $SPL_{eq}$ remain constant as logging interval changes, while $SPL_{mean}$ and $SPL_{mode}$ show moderate to high sensitivity with respect to the logging interval used in the post-processing.

## ACKNOWLEDGMENTS

The kind cooperation of the subjects who voluntary participated in the experiment made this study possible. Particular thanks are extended to the I.N.Ri.M. researchers who made their laboratories available for the experiments.

## APPENDIX: CALIBRATION PROCEDURE FOR THE HEADWORN MICROPHONE

The characterization was performed in the anechoic chamber of Politecnico di Torino, where the measured A-weighted equivalent background noise level was 26.2 dB. Initially, the sound level meter was calibrated by coupling it to a pressure calibrator B&K 4230, which provides a nominal pressure of 1 Pa @ 1 kHz, and a calibration sine-wave file at a level of 94 dB was recorded. Then, both the sound level meter and the headworn microphone were placed at a distance of 2.5 cm from the mouth of a B&K type 4128 Head and Torso Simulator, HATS (B&K, Nærum, Denmark), on-axis. The HATS was connected through the amplifier ALPINE MRP T222 (Alpine Electronics, Inc., Tokyo, Japan) and the audio device TASCAM US-144 (TEAC America, Inc., Montebello, CA) to a notebook PC. The software DIRAC 5 was run to generate different SPLs of ICRA noise[47] in the usual range observed in professional voice users (from 55 to 72 dB @ 1 m).[13] ICRA noise was preferred to standard signals like white or pink noise due to its speech-like spectral and temporal properties. For each SPL, the output signals of the headworn microphone and the reference device were simultaneously acquired and post-processed by means of MATLAB scripts that estimated $SPL_{eq}$, using the calibration wave file of the sound level meter as a reference. The difference between $SPL_{eq}$ values estimated from data recorded by the two devices was added to the headworn microphone levels in order to obtain calibrated values.

[1]T. Masuda, Y. Ikeda, H. Manako, and S. Komiyama, "Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers," Acta Otolaryngol. **113**(4), 547–552 (1993).

[2]E. E. Levendoski, C. Leydon, and S. L. Thibeault, "Vocal fold epithelial barrier in health and injury: A research review," J. Speech Lang. Hear. Res. **57**(5), 1679–1691 (2014).

[3]E. B. Holmberg, P. Doyle, J. S. Perkell, B. Hammarberg, and R. E. Hillman, "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," J. Voice **17**(3), 269–282 (2003).

[4]M. Hirano, S. Tanaka, M. Fujita, and R. Terasawa, "Fundamental frequency and sound pressure level of phonation in pathological states," J. Voice **5**(2), 120–127 (1991).

[5]E. B. Holmberg, R. E. Hillman, B. Hammarberg, and M. Södersten, "Efficacy of a behaviorally based voice therapy protocol for vocal nodules," J. Voice **15**(3), 395–412 (2001).

[6]Online official website, http://www.sonvox.com/index.html (Last viewed November 23, 2016).

[7]A. Carullo, A. Vallan, and A. Astolfi, "Design issues for a portable vocal analyzer," IEEE Trans. Instrumen. Meas. **62**(5), 1084–1093 (2013).

[8]D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," IEEE Trans. Biomed. Eng. **59**(11), Pt. 2, 3090–3096 (2012).

[9]J. H. Van Stan, D. D. Mehta, S. M. Zeitels, J. A. Burns, A. M. Barbu, and R. E. Hillman, "Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects," Ann. Otol., Rhinol. Laryngol. **124**(11), 864–874 (2015).

[10]M. Ghassemi, J. H. Van Stan, D. D. Mehta, M. Zanartu, H. A. Cheyne, R. E. Hillman, and J. V. Guttag, "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 (2014).

[11]H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**(6), 3438–3451 (2000).

[12]E. J. Hunter and I. R. Titze, "Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus nonoccupational settings," J. Speech Lang. Hear. Res. **53**, 862–875 (2010).

[13]P. Bottalico and A. Astolfi, "Investigations into vocal doses and parameters pertaining to primary school teachers in classrooms," J. Acoust. Soc. Am. **131**(4), 2817–2827 (2012).

[14]A. Astolfi, P. Bottalico, A. Accornero, M. Garzaro, J. Nadalin, and C. Giordano, "Relationship between vocal doses and voice disorders on primary school teachers," in *Proceedings of the 9th European Conference on Noise Control* (2012).

[15]N. Durup, B. Shield, S. Dance, and R. Sullivan, "An investigation into relationships between classroom acoustic measurements and voice parameters of teachers," Build. Acoust. **22**, 225–241 (2015).

[16]M. R. Hodgson, R. Rempel, and S. Kennedy, "Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures," J. Acoust. Soc. Am. **105**, 226–233 (1999).

[17]L. C. Cantor Cutiva, G. E. Puglisi, A. Astolfi, and A. Carullo, "Four-day follow-up study on the self-reported voice condition and noise condition of teachers: Relationship between vocal parameters and classroom acoustics," J. Voice **31**(1), 120.e1–120.e8 (2017).

[18]G. E. Puglisi, A. Astolfi, L. C. Cantor Cutiva, and A. Carullo, "Four-day-follow-up study on the voice monitoring of primary school teachers: Relationships with conversational task and classroom acoustics," J. Acoust. Soc. Am. **141**(1), 441–452 (2017).

[19]A. Astolfi, A. Carullo, L. Pavese, and G. E. Puglisi, "Duration of voicing and silence periods of continuous speech in different acoustic environments," J. Acoust. Soc. Am. **137**(2), 565–579 (2015).

[20]D. Pelegrín-García, B. Smits, J. Brunskog, and C. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments," J. Acoust. Soc. Am. **129**(4), 1981–1990 (2011).

[21]M. Cooke, S. King, M. Garnier, and V. Aubanel, "The listening talker: A review of human and algorithmic context-induced modifications of speech," Comput. Speech Lang. **28**, 543–571 (2014).

[22]H. L. Lane, B. Tranel, and C. Sisson, "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**, 618–624 (1970).

[23]M. A. Picheny, N. I. Durlach, and L. D. Braida, "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446 (1986).

[24]E. B. Holmberg, R. E. Hillman, J. S. Perkell, and C. Gress, "Relationships between intra-speaker variation in aerodynamic measures of voice

production and variation in SPL across repeated recordings," J. Speech Hear. Res. **37**(3), 484–495 (1994).
[25]P. Corthals, "Sound pressure level of running speech: Percentile level statistics and equivalent continuous sound level," Folia Phoniatr. Logop. **56**, 170–181 (2004).
[26]W. S. Brown, T. Murry, and D. Hughes, "Comfortable effort level: An experimental variable," J. Acoust. Soc. Am. **60**, 696–700 (1976).
[27]W. S. Brown, Jr., R. J. Morris, and T. Murry, "Comfortable effort level revisited," J. Voice **10**(3), 299–305 (1996).
[28]K. L. Garrett and E. C. Healey, "An acoustic analysis of fluctuations in the voices of normal adult speakers across three times a day," J. Acoust. Soc. Am. **82**, 58–62 (1987).
[29]M. Sihvo and E. Sala, "Sound level variation findings for pianissimo and fortissimo phonations in repeated measurements," J. Voice **10**(3), 262–268 (1996).
[30]M. Sihvo, P. Laippala, and E. Sala, "A study of repeated measures of softest and loudest phonetions," J. Voice **14**(2), 161–169 (2000).
[31]E. B. Holmberg, J. S. Perkell, R. E. Hillman, and C. Gress, "Individual variation in measures of voice," Phonetica **51**, 30–37 (1994).
[32]I. R. Titze, "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906 (1989).
[33]J. Szudek, A. Ostevik, P. Dziegielewski, J. Robinson-Anagor, N. Gomaa, B. Hodgetts, and A. Ho, "Can uHear me now? Validation of an iPod-based hearing loss screening test," J. Otolaryngol.–Head Neck Surg. **41**(S1), S78–S84 (2012).
[34]J. C. Wang, S. Zupancic, C. Ray, J. Cordero, and J. C. Demke, "Hearing test app useful for initial screening, original research shows," Hear. J. **67**, 32–35 (2014).
[35]I. Vernero, M. Gambino, R. Stefanin, and O. Schindler, *Cartella Logopedica. Età Evolutiva* (*Logopedic Collection. Childhood*) (Edizioni Omega, Torino, 1998).
[36]H. K. Schutte and W. Seidner, "Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography," Folia Phoniatrica et Logopaedica **35**, 286–288 (1983).
[37]J. G. Švec and S. Granqvist, "Guidelines for selecting microphones for human voice production research," Am. J. Speech-Lang. Pathol. **19**, 356–368 (2010).
[38]H. Sramkova, S. Granqvist, C. T. Herbst, and J. G. Švec, "The softest sound level of the human voice in normal subjects," J. Acoust. Soc. Am. **137**(1), 407–418 (2015).
[39]A. Carullo, A. Vallan, A. Astolfi, L. Pavese, and G. E. Puglisi, "Validation of calibration procedures and uncertainty estimation of contact-microphone based vocal analyzers," Measurement **74**, 130–142 (2015).
[40]E. Campione and J. Veronis, "A large-scale multilingual study of silent pause duration," in *Proceedings of the Speech Prosody Conference*, edited by Bernard Bel and Isabelle Marlien, Aix-en-Provence, Laboratoire Parole et Langage (2002), pp, 199–202.
[41]P. Sarkar and K. S. Rao, "Modeling pauses for synthesis of storytelling style speech using unsupervised word features," Procedia Comp. Sci. **58**, 42–49 (2015).
[42]P. S. Popolo, M. K. Rogge, J. G. Švec, and I. R. Titze, "Technical considerations in the design of a wearable voice dosimeter," The National Center for Voice and Speech Online Technical Memo 5, version 1.1 (2002). Available at http://www.ncvs.org/ncvs/library/tech/NCVSOnlineTechnicalMemo05.pdf (Last viewed November 23, 2016).
[43]P. S. Popolo, J. G. Švec, and I. R. Titze, "Adaptation of a pocket PC for use as a wearable voice dosimeter," J. Speech Lang. Hear. Res. **48**, 780–791 (2005).
[44]J. G. Švec, I. R. Titze, and P. S. Popolo, "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," J. Acoust. Soc. Am. **117**(3), 1386–1394 (2005).
[45]JCGM 100, Evaluation of measurement data—Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology (2008), http://www.bipm.org/en/publications/guides (Last viewed November 23, 2016).
[46]J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference* (Taylor & Francis, London, 2003), pp. 215–223.
[47]W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann, "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," Audiology **40**(3), 148–157 (2001).