AJSLP

Research Article

Calibration of Clinical Audio Recording and Analysis Systems for Sound Intensity Measurement

Youri Maryn[a,b,c,d] and Andrzej Zarowski[a]

Purpose: Sound intensity is an important acoustic feature of voice/speech signals. Yet recordings are performed with different microphone, amplifier, and computer configurations, and it is therefore crucial to calibrate sound intensity measures of clinical audio recording and analysis systems on the basis of output of a sound-level meter. This study was designed to evaluate feasibility, validity, and accuracy of calibration methods, including audiometric speech noise signals and human voice signals under typical speech conditions.
Method: Calibration consisted of 3 comparisons between data from 29 measurement microphone-and-computer systems and data from the sound-level meter: signal-specific comparison with audiometric speech noise at 5 levels, signal-specific comparison with natural voice at 3 levels, and cross-signal comparison with natural voice at 3 levels. Intensity measures from recording systems were then linearly converted into calibrated data on the basis of these comparisons, and validity and accuracy of calibrated sound intensity were investigated.
Results: Very strong correlations and quasisimilarity were found between calibrated data and sound-level meter data across calibration methods and recording systems.
Conclusions: Calibration of clinical sound intensity measures according to this method is feasible, valid, accurate, and representative for a heterogeneous set of microphones and data acquisition systems in real-life circumstances with distinct noise contexts.

The sound pressure level (SPL) is a measure of the sound intensity of the vocal output, which is routinely measured during various phonatory tasks in the clinical assessment of voice and its disorders. Because speech sound intensity is related to various physiological events across the lower and upper airways, it can be regarded as an important diagnostic marker in the clinical voice assessment. For example, patients with unilateral vocal fold paralysis often lack the ability to speak/phonate sufficiently loud. Most of these patients are not able to adequately adduct the vocal folds and therefore cannot build up sufficient subglottal pressure. Furthermore, the paralyzed vocal fold exhibits a loss of tonicity and mass, resulting in weakness and bowing, and preventing vocal power (Miller, 2004). On the other hand, patients with muscle tension dysphonia and vocally abusive behaviors are often found to speak excessively loud (Altman, Atkinson, & Lazarus, 2005; Van Houtte, Van Lierde, & Claeys, 2009) due to increased phonatory effort.

Acoustic analysis of vocal intensity typically consists of determining one or more of the following descriptive statistics on intensity data of voice recordings during sustained vowel, continuous speech, and/or other tasks such as singing, crescendo phonation, and decrescendo phonation: mean/median/mode, root-mean-square, standard deviation, coefficient of variation, minimum, maximum, range, specific percentiles and interquartile range, short-term perturbation, long-term perturbation, and so forth (Buder, 2000). Sound intensity is often metered to quantify the habitual/modal loudness of connected speech within specified speaking environments (e.g., Gelfer & Young, 1997; Healey, Jones, & Berky, 1997). Also the phonetogram (i.e., voice range profile, voice area, phonatory range profile, or phonational profile), for example, is a common clinical method to measure the minimum and maximum intensity of voice productions at various fundamental frequencies across the physiological phonatory range (Baken & Orlikoff, 2000; Heylen, Wuyts, Mertens, & Pattyn, 1996; Ma, 2011). Furthermore, the

[a]Sint-Augustinus General Hospital, European Institute of ORL, Antwerp, Belgium
[b]University College Ghent, Belgium
[c]University of Antwerp, Belgium
[d]Vlaamse Vereniging voor Logopedisten, Belsele, Belgium
Correspondence to Youri Maryn: youri.maryn@gza.be

Editor: Krista Wilkinson
Associate Editor: Julie Barkmeier-Kraemer
Received June 14, 2014
Revision received December 13, 2014
Accepted May 28, 2015
DOI: 10.1044/2015_AJSLP-14-0082

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

intensity contour of voice recordings can be investigated to assess the prosodic syllable accentuation or the presence of intensity modulation in cases with essential vocal tremor (Lester, Barkmeier-Kraemer, & Story, 2013).

Measuring sound intensity of voice signals is a physiologically and diagnostically relevant method that is commonly administered in clinical speech and voice assessment protocols. It is therefore crucial to measure it as correctly/reliably as possible, especially when comparing sound recordings within and across speakers and clinics. However, there are various pitfalls with the measurement of vocal intensity. Healthy adult human vocal tracts can produce sounds from approximately 30 dB to approximately 120 dB with the microphone 30 cm from the mouth, which corresponds with about 45 dB and about 135 dB with 5 cm mouth–microphone distance (Švec & Granqvist, 2010). This range of dB-data physically corresponds with a very wide range of sound pressure levels between approximately 0.004 Pascal and approximately 20.000 Pascal. Both microphone and sound card of the digital sound acquisition system thus need to handle loud voice productions as well as voice signals that are 5,000 times less loud without distortion and with linear output (i.e., without favoring nor hindering certain intensity levels). During phonetography, for example, participants are asked to phonate as softly and loudly as possible and at different intermediate levels. To value the intensities of these phonations appropriately, it is essential that they are weighted equally by the data acquisition system. Furthermore, differences in sound sensitivity between microphones and in hardware as well as software input levels across recording systems logically induce undesired variation in measured sound intensities, impeding accurate data collection and analysis, standardization of methods, and comparability of vocal intensities of two or more recording systems between and within voice clinics.

To enable reliable measurement, the Professional Services Board of the American Speech-Language-Hearing Association (ASHA), for example, demands regular calibration of clinical/audiological equipment (ASHA, 1984). It is therefore crucial to calibrate the clinical intensity measurement before recording and analyzing sound signals (Ma, 2011). However, research on comparability and calibration of intensity measures across clinical recording and analysis systems is sparse. To our knowledge, only Winholtz and Titze (1997) performed an experiment to convert microphone-and-computer-based intensity measures to sound-level meter–based data and to assess the validity/accuracy of this conversion method. They worked with three types of signals in an acoustically treated sound booth: (a) a 400-Hz tone presented at 60 dB, 70 dB, 80 dB, 90 dB, and 100 dB by an artificial sound source; (b) a sustained vowel [a:] produced by two female and two male participants; and (c) the Rainbow passage text read aloud by a woman and a man. All input stimuli were simultaneously recorded with a sound-level meter at 30 cm of the sound source and with a head-mounted condenser microphone at 8 cm. The C-filtered dB_{SPL} output (i.e., dB_C) of the sound-level meter was adjusted to compensate the difference in distance, and the computer data were processed to establish a conversion value on the basis of the difference between the known reference value of the sound source and the processed value of the calibration tone. Comparison between both outputs yielded a favorable overall accuracy of less than 1 dB (Winholtz & Titze, 1997). A number of issues, however, emerge from this study. First, the source-to-sound-level meter distance was almost four times longer than the source-to-microphone distance, necessitating a compensation of the SPL data before comparing the two sets of output data. Second, a Fortran computer program code had to be requested from the authors and implemented for the conversion of the microphone-and-computer signal. Third, next to the sound-level meter (i.e., an indispensable item in any calibration method), an extra sound source had to be obtained when calibrating with constant tones. Fourth, the validity of this calibration method was investigated in only one recording system, and therefore the generalizability of the outcome to other data acquisition systems is questionable. The present study was undertaken to investigate feasibility, validity, and accuracy of three straightforward comparison methods with the output of a sound-level meter as criterion for intensity calibration in multiple systems for clinical voice and speech recording and analysis.

Method

Recording Systems

To obtain reimbursement for voice therapy from the Belgian health care system, Belgian voice clinicians were recently urged to objectively quantify their findings. This has triggered many speech-language therapists from inside as well as outside voice clinics to administer acoustic measures of fundamental frequency, sound intensity, and so forth, during routine voice assessment. Given the crucial role of sound intensity measurement in the clinical voice assessment, the Vlaamse Vereniging voor Logopedisten (i.e., VVL, Flemish Association for Speech-Language Therapists) organized three sound intensity calibration sessions in an attempt to increase reliability of and comparability across audio recording and analysis systems. Members of the VVL encountering problems with or lacking confidence in calibration of their own sound acquisition and analysis system, were advised to enroll in one of the calibration sessions under the following conditions: bring along all required equipment (i.e., laptop computer, microphone, external sound card, and cables), have the program Praat (Paul Boersma & David Weenink; Institute of Phonetic Sciences, University of Amsterdam, the Netherlands; http://www.praat.org/) installed on the computer, and be able to make a recording with this system. During these sessions, the sound intensity measurement of 31 recording systems was calibrated according to the comparison method. Gross technical specifications of these data acquisition systems are summarized in Table 1. All systems represented a variety of laptop/notebook computers equipped with either a head-mounted condenser microphone plus external sound card

Table 1. Summary of several characteristics of the data acquisition systems.

Number	Type of computer	Operating system	External sound card/preamplifier	Microphone	Software (version)	Room	SNR
1	MacBook Air	OS X 10.8.5	Focusrite Forte	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	44.8 dB
2	Fujitsu Siemens	MS Windows Vista	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	33.7 dB
3	Hewlett Packard ProBook	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.61)	Hotel room	38.5 dB
4	Hewlett Packard ENVY	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.39)	Hotel room	34.9 dB
5	MacBook Pro	OS X 10.9.1	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	47.4 dB
6	Lenovo T500	MS Windows XP	None	Samson C01U	Praat (5.3.61)	Hotel room	21.6 dB
7	ASUS X535	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	50.7 dB
8	MacBook Air	OS X 10.7.5	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	33.7 dB
9	Sony Vaio	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Hotel room	37.2 dB
10	MacBook Air	OS X 10.9.1	None	Samson C01U	Praat (5.3.53)	Hotel room	19.2 dB
11	Sony Vaio	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.61)	Hotel room	34.0 dB
12	Hewlett Packard ProBook	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.61)	Hotel room	36.9 dB
13	Compaq	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Secretariat	38.2 dB
14	MacBook Pro	OS X 10.7.5	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Secretariat	50.6 dB
15	Acer Aspire	MS Windows 7	None	Samson C01U	Praat (5.3.29)	Secretariat	40.1 dB
16	Sony Vaio	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Secretariat	51.6 dB
17	Dell Vostro	MS Windows 7	None	Samson C01U	Praat (5.3.61)	Secretariat	40.5 dB
18	Samsung R610	MS Windows Vista	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.61)	Secretariat	41.6 dB
19	HP	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Secretariat	46.8 dB
20	HP ENVY	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	40.1 dB
21	Samsung	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	45.4 dB
22	Samsung	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.59)	Secretariat	45.7 dB
23	Dell Latitude	MS Windows 7	None	Samson C01U	Praat (5.3.53)	Secretariat	30.5 dB
24	MacBook Air	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.53)	Secretariat	42.9 dB
25	HP Pavilion	MS Windows 7	None	Logitech USB desktop	Praat (5.3.59)	Secretariat	40.0 dB
26	Sony Vaio	MS Windows 7	None	Samson C01U	Praat (5.3.64)	Secretariat	34.6 dB
27	HP Pavilion	MS Windows Vista	None	Samson Meteor	Praat (5.3.64)	Secretariat	44.3 dB
28	Acer Aspire	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	44.1 dB
29	Toshiba Tecra	MS Windows 8	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	42.3 dB
30	Sony Vaio	MS Windows 7	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	45.6 dB
31	MacBook Air	OS X 10.9.1	Focusrite iTrack Solo	AKG C544L + MPA V L	Praat (5.3.64)	Secretariat	51.9 dB

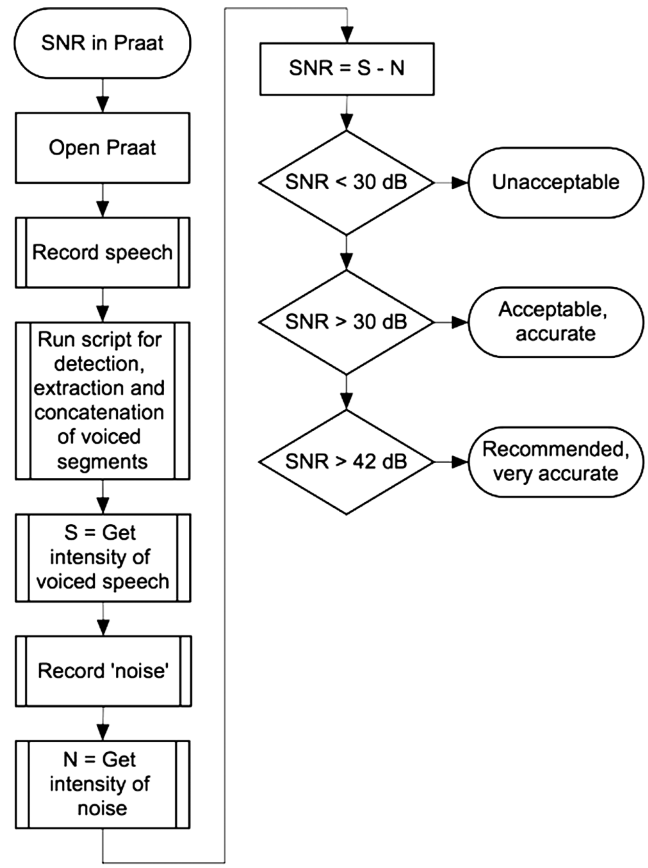
Note. SNR = signal-to-noise ratio.

or a hand-held/standing condenser microphone with USB connection. The program Praat operated in either Microsoft Windows or Apple Macintosh environment. The first calibration session was organized in a hotel room in the city Sint-Niklaas, and the second and third calibration sessions took place at the secretariat of the VVL in the city Belsele. Both venues were chosen for practical reasons: their central location in the region of Flanders and the ability to calibrate in a relatively quiet environment. Ambient noise level were 42.2 dB_C in the hotel room and 37.8 dB_C in the secretariat.

Signal-to-Noise Ratio

To control the signal-to-noise ratio (SNR) of the data acquisition systems, pronunciation of the two sentences, "*Papa en Marloes staan op het station. Ze wachten op de trein,*" at comfortable loudness and pitch by the owners was recorded together with several seconds of speechless/mute/relatively silent sound. Because only phonation parts are of interest during clinical voice assessment, the voiced segments of the two sentences were detected, extracted, and concatenated according to Parsa and Jamieson (2001) and with the customized Praat-script of Maryn, Corthals, Van Cauwenberge, Roy, and De Bodt (2010; Appendix 1). The resulting chained waveform was considered *signal*, whereas the waveform during relative silence was considered *noise* generated by erroneous sound sources. The SNR was calculated as signal intensity minus noise intensity. Studies of Deliyski and colleagues (Deliyski, Shaw, & Evans, 2005; Deliyski, Shaw, Evans, & Vesselinov, 2006) show that a SNR ≥ 30 dB is warranted for voice measures to be valid and reliable and that a SNR ≥ 42 dB is required to obtain 99% accuracy in acoustic voice quality measures. A SNR-threshold of 30 dB was therefore chosen in the present study to differentiate between acceptable (i.e., SNR ≥ 30 dB) and unacceptable (i.e., SNR < 30 dB) recording condition. Systems 6 and 10 yielded SNR of 21.6 dB and 19.2 dB, respectively, and were therefore excluded from further participation in the present study. The operators of these two systems were given advice on how to increase the SNR (e.g., by using another microphone, reducing the noise in the room, or by implementing an external sound card). A procedural outline with all steps to establish the SNR of an audio recording system is provided in Figure 1.

Figure 1. Flowchart illustrating the procedure for determining the signal-to-noise ratio (SNR) of a data acquisition system using the program Praat.



Intensity Calibration

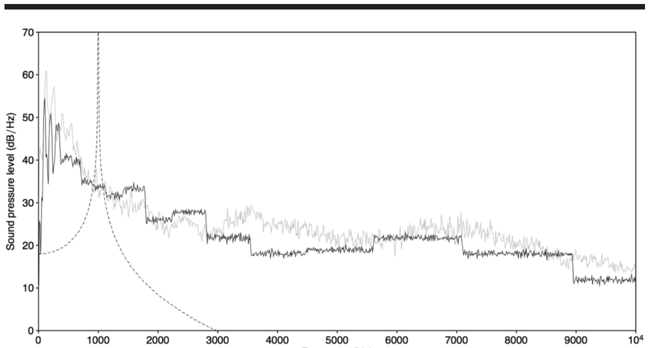
To calibrate estimates of vocal sound intensities across the phonatory range, the sound intensity measurements in the program Praat (i.e., the measured sound intensity [SIM]) in dB were compared with the output of a CR:832C integrating averaging Class 2 sound-level meter (Cirrus Research plc, Hunmanby, North Yorkshire, UK; i.e., the expected sound intensity [SIE]) in C-weighted sound pressure level (i.e., dB_C) and fast-time weighting mode with a 125-ms averaging window. The dB_C scale was chosen because of its relatively flat response curve across phonatory relevant frequencies: only −0.8, −0.2, 0.0, 0.0, 0.0, −0.2, and −0.8 dB for the frequencies 60, 125, 250, 500, 1000, 2000, and 4000 Hz, respectively (ASA, 2001). Winholtz and Titze (1997, p. 419) also used the C-filter response. The sound-level meter's equivalent continuous C-weighted sound pressure level (i.e., $L_{Ceq,T}$) was taken as SIE. This statistic refers to the dB_C level of a continuous steady sound that has the same C-weighted sound energy as the actual sound within a specified time interval T (Maekawa & Lord, 1994). SIM was compared with SIE according to the following steps.

1. An acceptable intensity switch setting of the recording equipment was sought by (a) shifting the microphone gain of the sound card device driver interface to its maximum position, (b) decreasing the gain (or input level) of the external amplifier so that the loudest [a:] productions were not clipped, and (c) increasing the input level of the external amplifier so that the softest [a:] sounds were still recorded and could be adequately differentiated from background noise. All following sounds were recorded without any additional adjustments of the switches in the computer configuration software and on the external sound card/amplifier. Because these settings may shift over time (e.g., involuntarily due to children playing with them and

use of equipment for other purposes) and to enable restoration of the calibration settings when this occurs, participants were advised to save a printscreen of the computer's audio settings window and to mark accurately the position of the external controller with a small sticker or a painted dot.

2. Audiometric speech noise (i.e., acoustic noise specifically created from speech material to mask the spoken stimuli during speech audiometry), derived from all monosyllabic words from the audiometry lists of the Dutch Association for Audiology (Bosman, 1989; Bosman et al., 1995), was used as calibration signal. Other artificial signals have been advocated for the calibration of an audio recording system, such as a 1-kHz tone presented at 94 dB (e.g., Howard & Murphy, 2010). However, audiometric speech noise can be presented as long as needed at fixed intensities, and as demonstrated by the spectra in Figure 2, it is much more similar to natural speech than a tone. To proceed with the calibration protocol, the audiometric speech noise was radiated by an Inspire T12 loudspeaker model MF1625 (Creative Technology Ltd., Singapore, Singapore) at five different intensities between SIE = 57.1 dB_C and SIE = 101.6 dB_C and captured by the microphones of the clinical recording systems and the sound-level meter. The intensity level of the radiated signal was regulated with the turning knob on the loudspeakers. The two microphones were placed equidistantly and adjacent to each other at 10 cm from the loudspeaker, as controlled with a ruler. The microphone-to-loudspeaker angle was 45°. This microphone placement was chosen as a trade-off between not too far to reduce the influence of room acoustics (i.e., ventilation noise, computer noise, reverberation, etc.) and not too close and at a nonzero (i.e., 45°) azimuthal angle from the mouth to reduce proximity and p-popping effects (i.e., signal artifacts due to plosive noise of [p]). This step yielded for every recording system five datasets for each of the two variables: Praat's SIM_{noise} in dB and the sound-level meter's SIE_{noise} in dB_C.

Figure 2. Illustration of the spectral configuration of recorded normophonic speech (gray contour), audiometric speech noise (black contour, solid line), and a single tone at 1 kHz (black contour, dashed line).



3. To calibrate not only with artificial audiometric speech noise but also with representative/natural voice sounds, all participants (i.e., the recording system owners) were asked to sustain the vowel [a:] at three different intensities: relatively soft, habitual/comfortable, and relatively loud. A method similar to the calibration of synthetic noise signal intensity was applied. The clinical recording microphones and the sound-level meter were located equidistantly and adjacent to each other at 10 cm from the mouth, as controlled with a ruler. The microphone-to-mouth angle was also 45°. Per data acquisition system, this step generated three datasets for each of the two variables: Praat's SIM_{vowel} in dB and the sound-level meter's SIE_{vowel} in dB_C.

4. To convert SIM data into SIE data on the basis of linear regression function, all the data (i.e., five SIM_{noise}, three SIM_{vowel}, five SIE_{noise}, and three SIE_{vowel} for every recording system) were collected in a Microsoft Excel spreadsheet before further data analysis and statistics.

5. To explore the correction of sound intensity of noise signals in a first experiment, linear regression was applied to convert SIM_{noise} into calibrated SIM_{noise} according to the linear function $SIM_{noise\text{-}cal}$ = (a × SIM_{noise}) + b, where a and b are constants indicating the slope and the ordinate intercept of the fit line through the coordinates, respectively. This was accomplished for every recording system separately. For example, for the third recording system in this study, comparison of the SIM_{noise} data, 27.3, 41.1, 49.2, 57.5, and 62.1 dB, with the corresponding SIE_{noise} data, 59.7, 70.8, 79.1, 87.5, and 92.0 dB_C, yielded a correlation coefficient of r = .998, and resulted in the $SIM_{noise\text{-}cal}$ data, 58.8, 71.7, 79.2, 87.0, and 91.3 dB, according to the linear function $SIM_{noise\text{-}cal}$ = (0.9372 × SIM_{noise}) + 33.361, respectively. Subtraction of the original SIE_{noise} from the corresponding $SIM_{noise\text{-}cal}$ provided absolute differences (i.e., errors) of 0.9, 0.9, 0.2, 0.5, and 0.7 dB, respectively, and an average error of 0.6 dB.

6. In a similar way, to examine the correction of sound intensity of vowel signals in a second experiment, SIM_{vowel} was transformed into calibrated SIM_{vowel} with the linear function $SIM_{vowel\text{-}cal}$ = (c × SIM_{vowel}) + d, where c and d are constants indicating the slope and the ordinate intercept of the fit line through the coordinates, respectively. This was undertaken for every recording system singly. By way of illustration, comparing the SIM_{vowel} data, 46.0, 56.2, and 78.1 dB, with the SIE_{vowel} data, 75.5, 88.5, and 110.6 dB_C of the third recording system, the linear function $SIM_{vowel\text{-}cal}$ = (1.0801 × SIE_{vowel}) + 26.621 was generated and yielded the $SIM_{vowel\text{-}cal}$ data, 76.3, 87.3, and 111.0 dB. This function was accompanied by r = .998 and resulted in an average error of 0.8 dB (i.e., on the basis of absolute differences of 0.8, 1.2, and 0.4 dB, respectively).

7. To check if it is reasonable to transfer the correction formula from the speech noise signal calibration to vowel signals in a third experiment, SIM_{vowel} was transformed into calibrated $SIM_{vowel\text{-}noise}$ with the linear function $SIM_{vowel\text{-}noise\text{-}cal} = (a \times SIM_{vowel}) + b$, where a and b are constants indicating the slope and the ordinate intercept of the fit line through the coordinates, respectively. This was carried out for every single recording system.
8. Exemplified for the third data acquisition system in the present study, the linear function-based equation $SIM_{vowel\text{-}noise\text{-}cal} = (0.9372 \times SIM_{vowel}) + 33.361$ was applied on the SIM_{vowel} data, 46.0, 56.2, and 78.1 dB, and produced the $SIM_{vowel\text{-}noise\text{-}cal}$ data, 76.5, 86.0, and 106.6 dB. Subtraction of the original SIE_{vowel} from the corresponding $SIM_{vowel\text{-}noise\text{-}cal}$ provided errors of 1.0, 2.5, and 4.0 dB, respectively, and an average error of 2.5 dB.

A procedural outline illustrating input level adjustments and calibration methodology is provided in Figure 3.

Statistics

All statistical analyses were completed using IBM SPSS Statistics Version 22.0 (SPSS Inc., Chicago, IL), and results were considered statistically significant at $p \leq .05$. Because analysis with the one-sample Kolmogorov-Smirnov test revealed that none of the variables were normally distributed (SIE_{noise}: $Z = 0.104$, asymptotic two-tailed $p = .001$; $SIM_{noise\text{-}cal}$: $Z = 0.104$, asymptotic two-tailed $p = .001$; SIE_{vowel}: $Z = 0.109$, asymptotic two-tailed $p = .012$; $SIM_{vowel\text{-}cal}$: $Z = 0.117$, asymptotic two-tailed $p = .005$; and $SIM_{vowel\text{-}noise\text{-}cal}$: $Z = 0.109$, asymptotic two-tailed $p = .012$), nonparametric statistical methods were applied.

To investigate the validity of the three calibration experiments, Spearman rank correlation coefficients (i.e., r_S) were determined between SIE and calibrated (i.e., converted) SIM. For the first calibration method with audiometric speech noise, r_S was calculated between SIE_{noise} and $SIM_{noise\text{-}cal}$. For the second calibration method with natural vowel productions, r_S was tallied between SIE_{vowel} and $SIM_{vowel\text{-}cal}$. For the third method in which vowel intensities were calibrated on the basis of the transformation equation for noise signal, r_S was determined between SIE_{vowel} and $SIM_{vowel\text{-}noise\text{-}cal}$. Furthermore, average error scores as well as Wilcoxon signed rank tests for two related samples were applied to examine the similarity and difference between the three above-mentioned pairs of data (i.e., SIE_{noise} vs. $SIM_{noise\text{-}cal}$, SIE_{vowel} vs. $SIM_{vowel\text{-}cal}$, and SIE_{vowel} vs. $SIM_{vowel\text{-}noise\text{-}cal}$), and thus to assess the accuracy of the three calibration methods.

Results

Signal-to-Noise Ratio

The last column of Table 1 lists the SNR values of the recordings systems in the present study. From the 29 data acquisition systems with acceptable recording quality, however, the SNR data varied from 30.5 to 51.9 dB with mean SNR = 41.8 dB and a standard deviation of 5.8 dB.

Calibration of Audiometric Speech Noise Signal Intensity

Table 2 summarizes the descriptive statistics of SIM_{noise}, SIE_{noise}, and $SIM_{noise\text{-}cal}$. Whereas SIM_{noise} had a mean of 47.4 dB in the program Praat, the sound-level meter's mean SIE_{noise} was 79.6 dB_C, illustrating a mean difference of 32.2 dB between measured and expected sound intensity, respectively. The large discrepancy between SIE_{noise} and SIM_{noise} originated in the relatively low setting of the external amplifier's level. To bridge this difference, SIM_{noise} was transformed into $SIM_{noise\text{-}cal}$ according to recording system-specific linear regression functions, which resulted in an average $SIM_{noise\text{-}cal}$ of 79.6 dB.

Table 2 demonstrates that the descriptive statistics of SIE_{noise} and $SIM_{noise\text{-}cal}$ are nearly identical. This is confirmed by the inferential statistics in Table 3. Absolute errors between these two variables varied between 0.0 and 2.3 dB with a mean error of only 0.4 dB. Statistical testing with Wilcoxon's method confirmed the quasisimilarity ($p = .932$) between SIE_{noise} and $SIM_{noise\text{-}cal}$, as demonstrated by the box-and-whiskerplots in Figure 4. Furthermore, statistical testing of correlation revealed a very strong proportional relationship. The $r_S = 0.996$ is illustrated in the scatterplot of Figure 5. $SIM_{noise\text{-}cal}$ was thus highly comparable to the sound-level meter's SIE_{noise}.

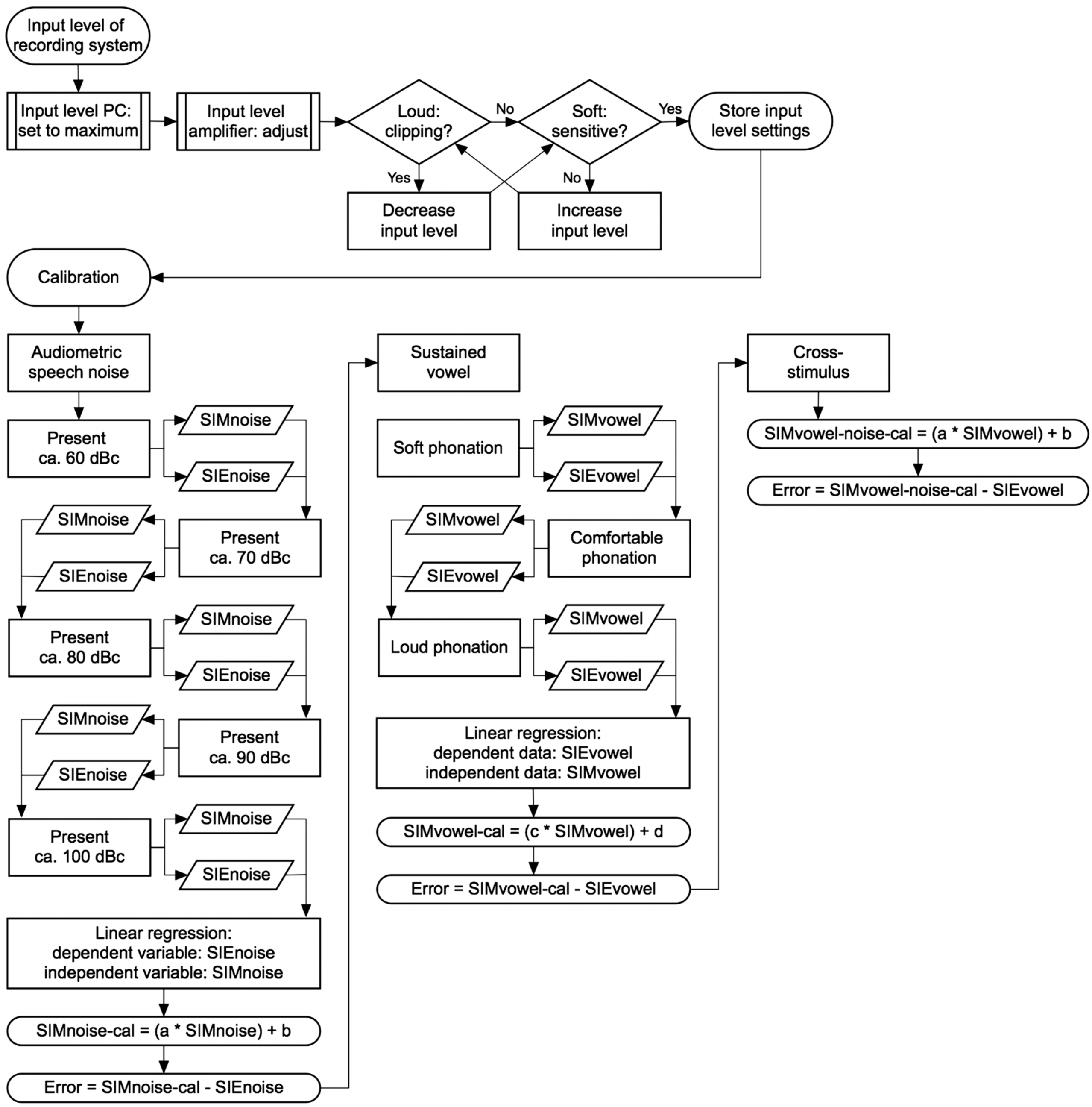
Calibration of Natural Voice Signal Intensity

Table 2 also summarizes the descriptive statistics of SIM_{vowel}, SIE_{vowel}, and $SIM_{vowel\text{-}cal}$. Whereas SIM_{vowel} in the program Praat resulted in a mean of 53.9 dB, mean SIE_{noise} of the sound-level meter was 86.3 dB_C. Similar to the calibration with audiometric speech noise, this represents an average difference of 32.4 dB between measured and expected sound intensity, respectively. As in the calibration with speech noise signals, this substantial dissimilarity between SIE_{vowel} and SIM_{vowel} stemmed from the relatively low setting of the external amplifier's level. To overcome this discrepancy, SIM_{vowel} was converted into $SIM_{vowel\text{-}cal}$ with recording system-specific linear regression functions, which yielded an average $SIM_{vowel\text{-}cal}$ of 86.3 dB.

The descriptive statistics of SIE_{vowel} and $SIM_{vowel\text{-}cal}$ in Table 2 are almost uniform, and Table 3 shows absolute errors varying between 0.0 and 2.2 dB with a mean error of only 0.5 dB. The Wilcoxon signed rank test demonstrated quasisimilarity ($p = .933$) between SIE_{vowel} and $SIM_{vowel\text{-}cal}$, as shown in the box-and-whiskerplots in Figure 4. Furthermore, $r_S = .998$ corresponds with a very strong proportional relationship, as illustrated in the scatterplot of Figure 6. $SIM_{vowel\text{-}cal}$ was thus particularly comparable to the sound-level meter's SIE_{vowel}.

Also, when the transformation functions from the calibration with audiometric speech noise were applied to SIM_{vowel}, there was a mean error of 2.3 dB, with absolute

Figure 3. Flowchart illustrating the procedure for (1) adjusting the input levels of the audio recording and analysis system (i.e., input level of recording system), and (2) determining the calibration formula and corresponding errors according to the three methods in the present study (i.e., calibration). SIM = measured sound intensity; SIE = expected sound intensity.



differences ranging between 0.0 and 8.4 dB. These errors, however, were not significant (p = .246). Furthermore, correlation between SIE_{vowel} and $SIM_{vowel\text{-}noise\text{-}cal}$ yielded r_S = .970. Both data sets were therefore also considered highly comparable, yet less than $SIM_{noise\text{-}cal}$ with SIE_{noise} and $SIM_{vowel\text{-}cal}$ with SIE_{vowel} (i.e., the dots in Figure 7 are more scattered around the regression fit line than in Figures 5 and 6).

Discussion

Measuring sound intensity is routine and especially relevant in clinical speech and voice assessment, from both physiological (i.e., as it is related to subglottal, glottal, and supraglottal phenomena) and diagnostic (e.g., as one of the markers in the voice range profile or vocal tremor analysis) perspectives. However, to measure and standardize sound

Table 2. Descriptive statistics of the data dispersion within the seven data sets from the 29 recording and analysis systems in the present study.

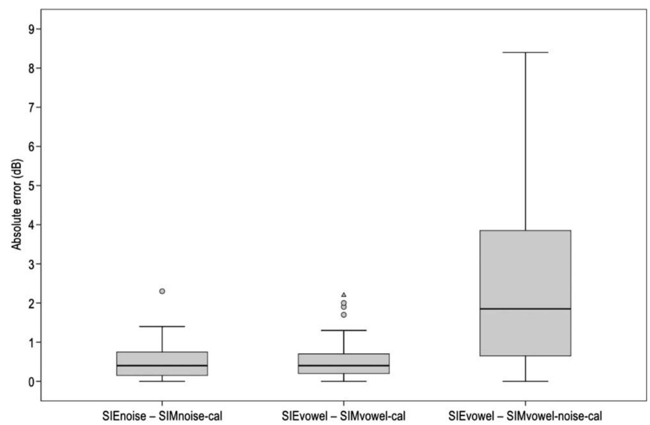
Sound intensity	Min.	Max.	Q1	Q3	IQR	*M*	*SD*
SIM_{noise} (dB)	18.3	84.2	36.5	58.9	22.5	47.4	14.4
SIE_{noise} (dB_C)	57.1	101.6	69.9	90.8	20.9	79.6	13.4
$SIM_{noise\text{-}cal}$ (dB)	57.0	101.6	70.3	90.5	20.2	79.6	13.4
SIM_{vowel} (dB)	29.2	83.9	40.7	67.6	27.0	53.9	14.3
SIE_{vowel} (dB_C)	61.3	111.2	74.7	101.7	27.1	86.3	14.9
$SIM_{vowel\text{-}cal}$ (dB)	61.5	111.5	74.3	101.9	27.6	86.3	14.8
$SIM_{vowel\text{-}noise\text{-}cal}$ (dB)	62.6	112.2	74.4	99.9	25.4	86.2	13.9

Note. Min. = minimum; Max. = maximum; Q1 = first quartile; Q3 = third quartile; IQR = interquartile range; SIM = measured sound intensity; SIE = expected sound intensity.

intensity correctly, reliably, and validly within and across clinical audio recording and analysis systems, it is imperative to minimize variability and therefore to calibrate and correct the sound intensity data of such systems according to the output of a sound-level meter (ASHA, 1984; Ma, 2011). This is similar to the requirements for audiometric equipment (Walker, Dillon, & Byrne, 1984) and oto-acoustic emission probes (Rasetshwane & Neely, 2011; Siegel & Hirohata, 1994). Because information on calibration of sound intensity measurement for the purpose of voice and speech evaluation is scarce, the present study was designed to assess the validity and accuracy of a straightforward calibration method in multiple (i.e., 29) audio recording systems. This method included both audiometric speech noise signals and natural voice signals at different intensity levels, and the sound- level meter's output in dB_C was considered the criterion on which the intensity data of the recording systems were corrected.

The findings of this—especially the inferential statistics in Table 3 and the scatterplots in Figures 5–7—first show that the investigated calibration procedures prove to be valid. This legitimates methods in which the output of the data acquisition system is first compared with and then adjusted on the output of a sound-level meter. All correlations between calibrated intensity measures and criterion data of the sound-level meter were very strong across the 29 recordings systems. Because input levels of the external amplifiers were a priori set close to the minimum setting to prevent sound wave clipping (i.e., without losing the ability

Figure 4. Box-and-whiskerplots illustrating the dispersion of the absolute differences between the intensity levels of sound-level meter (i.e., expected sound intensity [SIE]) and the intensity levels of calibrated data acquisition systems (i.e., measured sound intensity [SIM]).



to record quietest phonations and taking quantization noise in consideration), all signals were recorded relatively attenuated. On average 32.4 and 31.9 dB were therefore added for vowel and speech noise signals, respectively.

Second, with mean absolute errors of merely 0.5 and 0.4 dB for natural voice and artificial noise signals, respectively, it can be stated that SIE and calibrated SIM are nearly equal. This is demonstrated in Figure 4 and strongly confirmed by the nonsignificant difference statistics in Table 3. Winholtz and Titze (1997) found mean absolute errors of 0.8 and 0.6 dB for human sustained vowels and 400 Hz sinewaves, respectively. The outcome of the calibration method in the present study can therefore be considered highly acceptable and highlights its accuracy. Only when a cross-signal method was applied (i.e., when the linear conversion formulae from noise signals were applied in the calibration of vowel signal intensity) was there a greater mean absolute error of 2.3 dB. Discrepancy in accuracy between signal-specific and cross-signal methods may be caused by differences in intensity variation between speech noise (i.e., presented with relatively invariant intensity) and human voice (i.e., with many intensity fluctuations). Furthermore, notwithstanding that spectral contours of audiometric speech noise and human voice resemble each other (as demonstrated in Figure 2), they are not equal.

Table 3. Summary of inferential statistics on association as well as difference between the three pairs of data from the 29 recording systems with sufficient signal-to-noise ratio (SNR) in the present study.

Data pair	r_S (*P*)	Wilcoxon test statistic (*P*)	Absolute error Min.	Absolute error Max.	Absolute error *M*	Absolute error *SD*
SIE_{noise} – $SIM_{noise\text{-}cal}$	.996 (.000)	0.085 (.932)	0.0	2.3	0.4	0.4
SIE_{vowel} – $SIM_{vowel\text{-}cal}$	.998 (.000)	0.084 (.933)	0.0	2.2	0.5	0.5
SIE_{vowel} – $SIM_{vowel\text{-}noise\text{-}cal}$	.970 (.000)	–1.161 (.246)	0.0	8.4	2.3	2.1

Note. Min. = minimum; Max. = maximum; SIE = expected sound intensity; SIM = measured sound intensity.

Figure 5. Scatterplot illustrating the very strong association in audiometric speech noise levels between the sound-level meter and the calibrated recording systems. According to the procedure, the plot clearly shows five batches of dots surrounding approximately 60, 70, 80, 90, and 100 dB_C. The linear regression fit line through the scatters is defined as $SIM_{noise\text{-}cal} = (1 \times SIE_{noise}) + 0.16$. SIM = measured sound intensity; SIE = expected sound intensity.

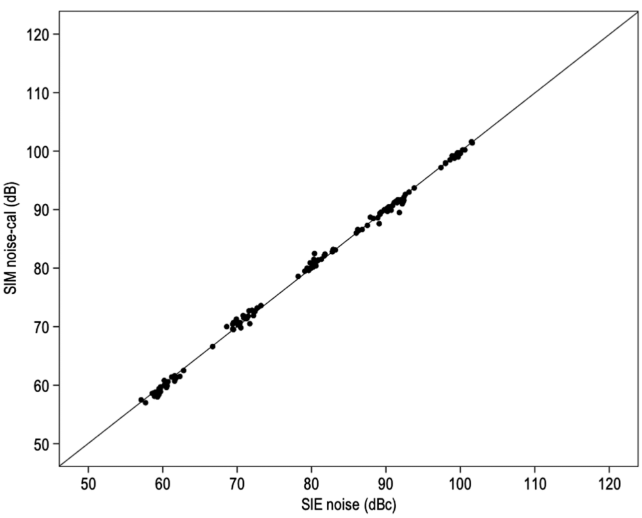
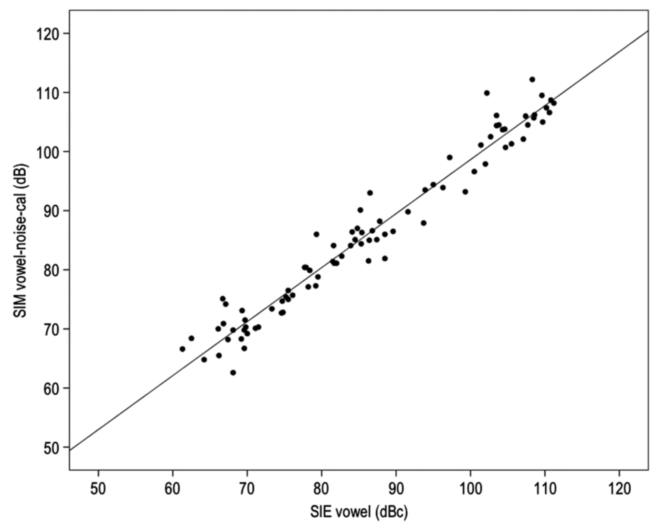


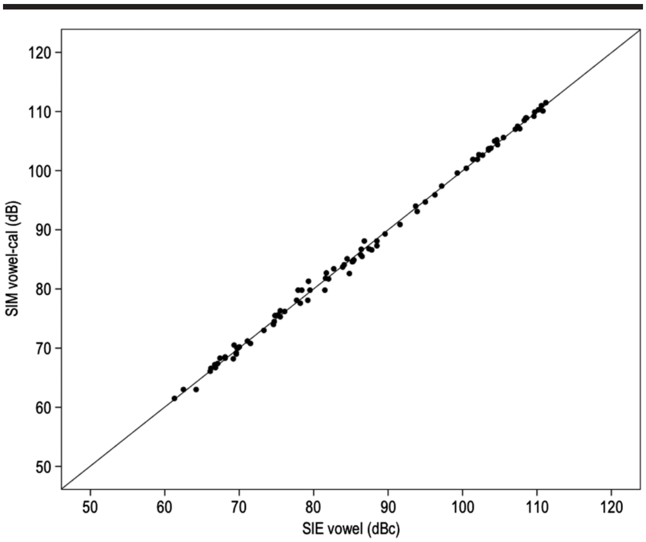
Figure 7. Scatterplot illustrating the very strong association in natural voice levels between the sound-level meter and the recording systems calibrated with the formulae from the method with the audiometric speech noise. The linear regression fit line through the scatters is defined as $SIM_{vowel\text{-}noise\text{-}cal} = (0.91 \times SIE_{vowel}) + 7.33$. SIM = measured sound intensity; SIE = expected sound intensity.



Depending on vocal tract configuration, for example, this may count for the 2–5 kHz zone (as demonstrated in Figure 2 for an adult male participant). This a zone where C-weighting of the sound-level meter slightly affects the sound intensity measurement (i.e., −0.2 dB at 2 kHZ, −0.8 dB at 4 kHz, and −1.3 dB at 5 kHz; ASA, 2001) and, for example, the AKG C544L microphone is entirely flat according to its frequency response curve (AKG Acoustics, 2011). Such dissimilarity may have added mean absolute error in the cross-signal method. Although the differences between SIE_{vowel} and $SIM_{vowel\text{-}noise\text{-}cal}$ were nonsignificant ($p = .246$), such a cross-signal method for calibrating sound intensity is less accurate and valid than the signal-specific methods (i.e., procedures resulting in $SIM_{vowel\text{-}cal}$ and $SIM_{noise\text{-}cal}$). Therefore, the cross-method is considered subservient, and clinicians are encouraged to apply the signal-specific calibration methods. The method with relatively stable human vocalizations is preferred when calibrating with C-weighted sound pressure level, to minimize C-weighting effects by matching as much as possible the spectral contours of calibration signal and clinical analysis signal. Methods with more steady nonhuman stimuli (e.g., audiometric speech noise) can be applied when calibrating with zero-weighted sound pressure level, because then there is no frequency weighting to account for.

Third, the time to calibrate a clinical audio recording system with the two abovementioned stimuli (i.e., audiometric speech noise and human voice signals) never exceeded approximately 25 min, including arrangement of the hardware/software intensity controllers, determination of SNR, and brief explanation regarding the results. This time could easily be reduced to a couple of minutes if only one calibration method would be applied, for example with steadily sustained vowels at three intensity levels. With a program such as Praat, a natural voice source from, for example, the clinician (or at least a vocally normal person who is able

Figure 6. Scatterplot illustrating the very strong association in natural voice levels between the sound-level meter and the calibrated recording systems. The dots can be grouped in three batches between approximately 60–75 dB_C (i.e., soft phonation), 75–90 dB_C (i.e., intermediate or habitual phonation), and 90–115 dB_C (i.e., loud phonation). The linear regression fit line through the scatters is defined as $SIM_{vowel\text{-}cal} = (1 \times SIE_{vowel}) + 0.18$. SIM = measured sound intensity; SIE = expected sound intensity.



to steadily sustain a vowel, such as clinic personnel or staff member) vocalizing at different intensities, or from synthetic signals played at different intensities, and with a spreadsheet with linear regression/fit line function, the present study demonstrated the comparison method for calibration of sound intensity to be exceptionally feasible.

Fourth, what's the required frequency of calibration? The sensitivity of for instance electret condenser microphones has been shown to be exceptionally stable across environment conditions. It varies with only 0.010 dB per degree Celsius, shifts with merely 1 dB if continuously exposed to severe environment of 99% relative humidity and 55 °C for 2 years, and changes with less than −0.2 dB per 10% increase in ambient pressure (GenRad, 1977). Even after 28 years when handled with care in the laboratory, or after outdoor use for 13 years, the accuracy of electret condenser microphones appeared to be almost unaffected (Yasuno & Miura, 2006). Therefore, with microphone-to-mouth distance and hardware and software settings kept constant, with considerate use of the recording equipment, and with invariable clinical environment characteristics (in terms of temperature, atmospheric pressure, and air humidity), verification of the performance of recording system is needed only periodically (e.g., once every half year or even year).

Because this study (a) included audiometric speech noise and human voice signals at various intensity levels, (b) determined SNR in voiced segments of continuous speech, (c) calibrated to comparatively flat C-weighted dB_{SPL} scale, (d) controlled for a constant distance of 10 cm between microphone/sound-level meter and loudspeaker/mouth, (e) across 29 different audio recording system and voices, (f) and took place at two different locations with regular room acoustics, its results are considered representative for many voice and speech analysis systems inside and outside specialized clinics.

Limitations

First, a sound-level meter is indispensable in any calibration method as the criterion for actual sound intensity. However, not all voice and speech clinicians have this instrument at their disposal. In this case, the clinician may ask to borrow a sound-level meter for calibration purposes at a professional association, technical city services, police department, related university department, and so forth. Another option is to purchase a decent quality sound-level meter, which are available at reasonable price.

Second, the calibration sessions did not occur in silent accommodations, as for example in an anechoic audiometry booth, but in rooms with usual ambient noise levels. However, 29 of 31 (i.e., 93.5%) recording systems yielded SNR of at least 30 dB, regardless of the presence of environmental noise (e.g., computer fan and traffic). Furthermore, these room acoustics were considered clinically representative as they approximate the acoustic circumstances in which many clinicians organize voice recording and analysis (i.e., not all voice clinicians work in a sound-treated chamber).

Third, when using microphones in close proximity to the lips (as was the case in 24 of the 31 recording systems in this study [i.e., 77.4%] used head-mounted microphones at about 10 cm from the mouth), clinicians must be aware that the intensity of phonations at maximum loudness may exceed the microphone's maximum SPL, causing nonlinear intensity data. The calibration factors then no longer apply to such loud signals. The loudest phonation in the present study was produced at 111.4 dB_C by the female clinician with the second recording system, including an AKG C544L headset microphone. Because the maximum SPL of this microphone is 126 dB_{SPL} (AKG Acoustics, 2011, p. 20), such loud phonation never prompted intensity inflation or nonlinearity. However, when a person phonates even louder and/or the microphone is placed even closer to the mouth, sound wave clipping and nonlinear effects must be anticipated. To reduce the risk of encountering such nonlinearity, calibration methods including multiple gain settings depending on phonatory task (e.g., more gain during soft phonation, less gain during loud phonation) could be explored.

Conclusion

Sound intensity is an important marker in clinical voice and speech assessment. To be valid and accurate, however, calibration of sound intensity measures is indispensable. The calibration methods in the present study have proven to be feasible and highly valid, accurate, and representative for diverse audio recording and analysis systems in real-life environments with various noise contexts. Next to factors such as SNR, microphone type and placement relative to the source, environmental noise type and level, hardware, software, and so forth (Deliyski et al., 2005, 2006; Howard & Murphy, 2010; Švec & Granqvist, 2010; Vogel & Morgan, 2009), sound intensity calibration should be added to the list of factors to standardize and account for in acoustic methods for voice and speech assessment.

Acknowledgments

The first author did all the calibrations himself within his salaried mandate as member of Executive Committee of the Vlaamse Vereniging voor Logopedisten [Flemish Association of Speech-Language Therapists].

References

AKG Acoustics. (2011). *C544: User instructions*. Vienna, Austria: Author.

Altman, K. W., Atkinson, C., & Lazarus, C. (2005). Current and emerging concepts in muscle tension dysphonia: A 30-month review. *Journal of Voice, 19,* 261–267.

American National Standard Institute. (2001). *Design response of weighting networks for acoustical measurements* (ANSI/ASA S1.42-2001). New York, NY: Author.

American Speech-Language-Hearing Association. (1984). *Professional service board standards for accreditation of professional* *service programs in speech-language pathology and audiology*. Rockville, MD: Author.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice* (2nd ed.). San Diego, CA: Singular.

Bosman, A. J. (1989). *Speech perception by the hearing impaired* (Doctoral thesis). Utrecht, the Netherlands: University of Utrecht.

Bosman, A. J., Wouters, J., & Damman, W. (1995). Realisatie van een CD voor spraakaudiometrie in Vlaanderen [Realization of a CD for speech audiometry in Flanders]. *Logopedie en Audiologie, 9,* 218–225.

Buder, E. H. (2000). Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. In R. D. Kent & M. J. Ball (Eds.), *Voice quality measurement* (pp. 119–244). San Diego, CA: Singular.

Deliyski, D. D., Shaw, H. S., & Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice, 19,* 15–28.

Deliyski, D. D., Shaw, H. S., Evans, M. K., & Vesselinov, R. (2006). Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica, 58,* 274–288.

Gelfer, M. P., & Young, S. R. (1997). Comparisons of intensity measures and their stability in male and female speakers. *Journal of Voice, 11,* 178–186.

GenRad. (1977). *Microphones and accessories: Instruction Manual*. Concord, MA: Author.

Healey, E. C., Jones, R., & Berky, R. (1997). Effects of perceived listeners on speakers' vocal intensity. *Journal of Voice, 11,* 67–73.

Heylen, L. G., Wuyts, F. L., Mertens, F. W., & Pattyn, J. E. (1996). Phonetography in voice diagnoses. *Acta Otorhinolaryngologica Belgica, 50,* 299–308.

Howard, D. M., & Murphy, D. T. (2010). *Voice science, acoustics and recording*. San Diego, CA: Plural.

Lester, R. A., Barkmeier-Kraemer, J., & Story, B. H. (2013). Physiologic and acoustic patterns of essential vocal tremor. *Journal of Voice, 27,* 422–432.

Ma, E. P. M. (2011). Voice range profile: Phog. In E. P. M. Ma & E. M. L. Yiu (Eds.), *Handbook of voice assessments* (pp. 253–267). San Diego, CA: Plural.

Maekawa, Z., & Lord, P. (1994). *Environmental and architectural acoustics* (2nd ed.). Boca Raton, FL: CRC Press.

Maryn, Y., Corthals, P., Van Cauwenberge, P., Roy, N., & De Bodt, M. (2010). Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels. *Journal of Voice, 24,* 540–555.

Miller, S. (2004). Voice therapy for vocal fold paralysis. *Otolaryngologic Clinics of North America, 37,* 105–119.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *Journal of Speech, Language, and Hearing Research, 44,* 327–339.

Rasetshwane, D. M., & Neely, S. T. (2011). Calibration of otoacoustic emission probe microphones. *The Journal of the Acoustical Society of America, 130,* EL238–EL243.

Siegel, J. H., & Hirohata, E. T. (1994). Sound calibration and distortion product otoacoustic emissions at high frequencies. *Hearing Research, 80,* 146–152.

Švec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology, 19,* 356–368.

Van Houtte, E., Van Lierde, K., & Claeys, S. (2009). Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge. *Journal of Voice, 25,* 202–207.

Vogel, A. P., & Morgan, A. T. (2009). Factors affecting the quality of sound recording for speech and voice analysis. *International Journal of Speech-Language Pathology, 11,* 431–437.

Walker, G., Dillon, H., & Byrne, D. (1984). Sound field audiometry: Recommended stimuli and procedures. *Ear and Hearing, 5,* 13–21.

Winholtz, W. S., & Titze, I. R. (1997). Conversion of a head-mounted microphone signal into calibrated SPL units. *Journal of Voice, 11,* 417–421.

Yasuno, Y., & Miura, K. (2006). Sensitivity change with long-term preservation and practical use of electret condenser microphone. *Acoustical Science & Technology, 27,* 302–304.