An in-ear speech database in varying conditions of the audio-phonation loop

Rachel E. Bouserhal, Antoine Bernier, and Jérémie Voix

Citation: The Journal of the Acoustical Society of America **145**, 1069 (2019); doi: 10.1121/1.5091777
View online: https://doi.org/10.1121/1.5091777
View Table of Contents: https://asa.scitation.org/toc/jas/145/2
Published by the Acoustical Society of America

The following article appeared in *The Journal of the Acoustical Society of America* 145, 1069 (2019)
and may be found at https://doi.org/10.1121/1.5091777.

An in-ear speech database in varying conditions of the audio-phonation loop

Rachel E. Bouserhal,[a),b)] Antoine Bernier,[b)] and Jérémie Voix[b)]
École de technologie supérieure, 1100 Rue Notre-Dame O, Montréal, Québec, Canada

(Received 7 August 2018; revised 21 December 2018; accepted 5 February 2019; published online 26 February 2019)

With the rise of hearables and the advantages of using in-ear microphones with intra-aural devices, accessibility to an in-ear speech database in adverse conditions is essential. Speech captured inside the occluded ear is limited in its frequency bandwidth and has an amplified low frequency content. In addition, occluding the ear canal affects speech production, especially in noisy environments. These changes to speech production have a detrimental effect on speech-based algorithms. Yet, to the authors' knowledge, there are no speech databases that account for these changes. This paper presents a speech-in-ear database, of speech captured inside an occluded ear in noise and in quiet. The database is bilingual (in French and in English) and is intended to aid researchers in developing algorithms for intra-aural devices utilizing in-ear microphones.
 https://doi.org/10.1121/1.5091777

[ICB] Pages: 1069–1077

I. INTRODUCTION

With the rapid advancements in data science and digital signal processing along with the diminishing size of sensors, wearable technologies, interconnecting the human body to electronics, have attracted researchers' interest in the last ten years (Pantelopoulos and Bourbakis, 2010). In particular, *hearables* have gained significant research effort because of their widespread use and their potential for numerous applications (Hunn, 2016; Johansen *et al.*, 2017; Voix, 2017). Hearables are intra-aural devices capable of more than just audio playback and can include, among others, health monitoring, enhanced communication, augmented hearing, as well as voice command. Frequently, signals of interest are captured using either an in-ear acoustic microphone placed in an occluded ear canal (Bouserhal *et al.*, 2013; Voix, 2017), or a bone-conduction microphone that presses against the ear canal at its opening (Hunn, 2016). For speech applications, the developed algorithms must be robust to the uncontrolled conditions of everyday life, the changes in speech production caused by blocking the ear canal at its opening, as well as the unconventional placement of the microphones. Still, validation of these algorithms is conducted using existing speech corpora that are not representative of the actual conditions when wearing an intra-aural device. In this paper, an in-ear database of speech captured in adverse conditions while wearing an intra-aural device equipped with in-ear microphones is presented.

Several factors contribute to changes in speech production when wearing an intra-aural device. Namely, speech production is primarily governed by two systems: feedback and feedforward. The *audio-phonation loop*, is the feedback system of how one hears oneself over three different paths: the direct air-conduction path, the bone conduction path, and the indirect air-conduction path (reflections from surfaces in the room) (Bouserhal *et al.*, 2017b; Garnier *et al.*, 2010; v. Békésy, 1949). The feedforward system allows talkers to anticipate the sensory consequences of their speech production and gets more efficient with experience (Tourville and Guenther, 2011). In quiet rooms that are not notably reverberant, with open ears, the direct air-conduction path dominates the audio-phonation loop (Zwislocki, 1957). Disturbing one of the three feedback paths causes changes in speech production. A common manifestation of this feedback system is the Lombard effect (Brumm and Zollinger, 2011; Lombard, 1911). In the presence of background noise, because the audio-phonation loop is disturbed, talkers adjust their vocal effort in an attempt to remain intelligible (Brumm and Zollinger, 2011; Hotchkin and Parks, 2013; Junqua *et al.*, 1999). Consequently, Lombard speech differs from normal speech in that it has an increased overall amplitude, increased vowel intensity and duration, decreased duration for unvoiced consonants, an increased spectral center of gravity, and an increase in both the fundamental frequency (f_0) and first formant (F1) (Bottalico *et al.*, 2017; Garnier and Henrich, 2014; Lane and Tranel, 1971; Summers *et al.*, 1988). These changes in speech production are collectively beneficial for speech intelligibility. Studies have shown that, when tested with additive noise, Lombard speech is more intelligible than speech produced in quiet (Cooke and Lecumberri, 2012; Pittman and Wiley, 2001; Summers *et al.*, 1988).

The presence of noise, however, is not the only way that the audio-phonation loop can be disturbed. Obstructing the ear canal with an intra-aural device also alters speech production. When compared to speech produced with open ears, these variations are particularly pronounced in the presence of ambient noise (Bouserhal *et al.*, 2016; Casali and Horylev, 1987; Navarro, 1996; Tufts and Frank, 2003). The majority of the research done in this area is with the use of

[a)]Electronic mail: rachel.bouserhal@etsmtl.ca
[b)]Also at: Centre for Interdisciplinary Research in Music Media and Technology, Montréal, Québec, Canada.

hearing protection devices (HPD) and hearing aids. Blocking the ear canal at its opening causes it to act like a low pass filter, amplifying the bone and tissue conducted vibrations generated by a talker speaking. This phenomenon, known as the occlusion effect, disrupts the audio-phonation loop and causes the bone conduction path to dominate over the other two. An extensive review of speech production while wearing HPDs has been done by Byrne (2014). Results on speech production in quiet while wearing HPDs have been inconsistent. Some studies showed that talkers wearing HPDs in quiet increased their speech level between 3 and 6 dB (Bouserhal *et al.*, 2016; Casali and Horylev, 1987; Kryter, 1946), while other studies showed no significant difference in speech level (Navarro, 1996; Tufts and Frank, 2003). Tufts and Frank (2003) attribute these discrepancies to level of occlusion and attenuation caused by the earplug. In addition, taking into account the feedforward system, the degree of experience as well as awareness of the participant may also contribute to these inconsistencies. Participants with an efficient feedforward system, may be better at inferring the consequences of ear occlusion, thus, maintaining a stable speech level, even when their direct-path auditory feedback is lowered. Whereas other participants, with less efficient feedforward systems, rely mostly on the altered direct auditory path, causing changes in their speech level when occluded in quiet. In noise, all studies have been consistent in showing that talkers do not adjust as much to the noise when wearing HPDs compared to when they do not (Byrne, 2014; Casali and Horylev, 1987; Hoemann *et al.*, 1984; Howell and Martin, 1975; Tufts and Frank, 2003). In general, Lombard speech was not fully engaged when wearing HPDs in noise; speech levels did not increase as much and the upward shift in the spectral center of gravity was smaller. Consequently, it has been shown that speech produced in noise by talkers wearing HPDs is less intelligible than its open-ear counterpart (Howell and Martin, 1975; Tufts and Frank, 2003). The changes in speech caused by occluding the ear and its interaction with the presence of noise are significant and must be addressed when considering speech-based algorithms developed to be used with intra-aural devices.

In addition to variations in speech production, the placement and type of microphone used as part of the intra-aural device are important to consider. In recent years, the use of in-ear microphones has gained notable research interest (Bouserhal *et al.*, 2017a; Bulbuller *et al.*, 2006; Denby *et al.*, 2010; Kurcan, 2006). This is because in-ear microphones placed in occluded ear canals are less susceptible to the effects of background noise, since they are placed past the passive attenuation of the earplug, and can capture a diverse range of human-produced audio signals such as heartbeats and breathing (Bouserhal *et al.*, 2018; Martin and Voix, 2017). However, due to the properties of bone and tissue conduction, speech captured inside the occluded ear is "boomy," having an amplified low-frequency content and a limited bandwidth of 2 kHz (Bouserhal *et al.*, 2017a). Nonetheless, in-ear microphone speech has a useful amount of mutual information with speech captured in-front of the mouth (Bouserhal *et al.*, 2015). This allows for manageable ways of artificial bandwidth extension of speech captured with an in-ear microphone without the use of an additional air-conduction microphone placed in front of the mouth, which is commonly the case when using bone-conduction microphones (Shin *et al.*, 2012).

Existing speech databases have given rise to extensive development and validation of speech-based algorithms. Typically, clean speech corpora, such as the IEEE Sentences (Rothauser, 1969) and TIMIT (Zue *et al.*, 1990) are composed of phonetically balanced sentences produced by male and female talkers. They are usually recorded in a controlled environment, like an acoustically treated sound room, meeting some criteria in terms of reverberation time and residual background noise level. Beyond clean speech, corpora such as the NTIMIT (Network TIMIT) (Jankowski *et al.*, 1990), Noizeus (Hu and Loizou, 2007), and the AURORA corpus (Hirsch and Pearce, 2000) simulate the behavior of telecommunication terminals by bandpass filtering clean speech or mixing in additive noise to the filtered signals. In addition, corpora like UT-SCOPE (Ikeno *et al.*, 2007) containing speech produced in noise also exist. The advantage of the UT-SCOPE database is access to clean Lombard speech, achieved by recording signals using a microphone placed in front of the talkers' mouth while noise is played through headphones. However, the UT-SCOPE corpus is meant to simulate a noisy open-ear condition and does not consider things like the occlusion effect or the attenuation of the headphones. Le Roux *et al.* (2015) did an extensive review of existing corpora, their cost, size and realism. To the authors' knowledge, no speech corpora that account for the effect of noise as well as occluding the ear exist. Furthermore, even with the advantages offered by using in-ear microphones, there are no databases of in-ear microphone speech signals captured from occluded ears. Still, the problems of communication while wearing intra-aural devices is of significant research interest as it affects both the consumer world of hearables as well as the occupational safety and health world of HPDs.

In this paper, a speech in-ear (SpEAR) database is presented. This database is aimed to deepen the understanding of speech production in noise as well as provide a standardized dataset for researchers working with in-ear microphone speech-based algorithms. SpEAR is a bilingual database of French (Vaillancourt *et al.*, 2005) and English hearing in noise test (HINT) sentences (Nilsson *et al.*, 1994). HINT sentences are phonetically balanced, with uniform length and representation of natural speech, and are used to assess speech intelligibility in noise and in quiet. The lists have been made publicly available in both French and English, allowing for their widespread use. As part of SpEAR, speech is collected from talkers wearing an intra-aural device using three different microphone placements in four different conditions. Using in-ear microphones, outer-ear microphones and a microphone placed in front of the mouth, clean, noisy and clean Lombard speech are collected. In addition, open-ear clean speech is recorded to be used as reference. The methods used, including participant recruitment, apparatus used, and a detailed description of the recording conditions are presented in Sec. II. The data accessible as part of

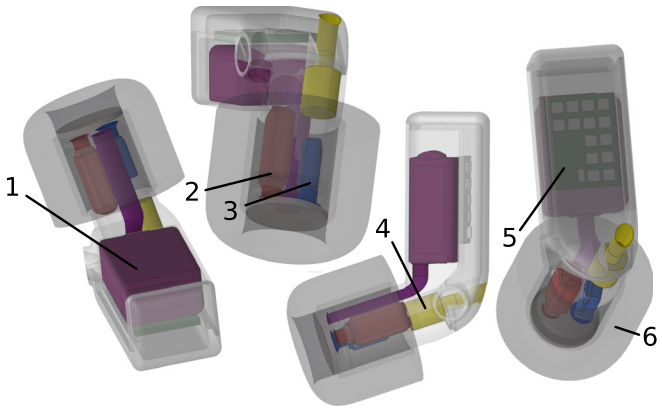


FIG. 1. (Color online) Transparency view of the CAD drawing of the earpiece, highlighting its components and associated sound channels, showing the woofer (1), the tweeter (2), the IEM (3), the OEM (4), the cross-over (5), and the foam tips (6).

SpEAR are presented in Sec. III. A precursory acoustic analysis over the different conditions is performed in Sec. IV to characterize basic changes in speech production and to aid the user of the database in selecting data that fits their needs. Results of the acoustical analysis are presented in Sec. V, followed by conclusions in Sec. VI.

II. METHODS

A. Participants

Participants were recruited via email and word of mouth, following the approval of the Comité d'éthique pour la recherche, the internal review board (IRB) of the École de technologie supérieure (H20170103). In total, 25 people participated in the experiment. Of all the participants, one male French talker was excluded from the study for whispering in all conditions. For the English corpus, 11 participants consisting of six females and five males with a mean age of 34 and 36, respectively, were included in the study. Four of the English talkers, including one female and three males, were not native speakers. No formal assessment of the degree of foreign accent was made, but all were self-identified fluent speakers. For clarity and ease of access, each talker's native language as well as the language spoken for the database are identified in the metadata. For the French corpus, 13 participants consisting of four females and nine males with a mean age of 25 and 30, respectively, were retained for the study. All francophone participants were native French speakers from either France or Québec. All participants had normal hearing with thresholds of 25 dB hearing level (HL) or lower at each octave band frequency from 0.25 to 8 kHz, verified using tonal audiometry for both ears.

B. Apparatus

Recordings were done in a double-wall audiometric booth (Eckel Noise Control Technologies, Morrisburg, Ontario, Canada) using a MacBook Pro laptop (Apple Inc., Cupertino, California), running MATLAB 2015 (MathWorks, Natick, Massachusetts), connected to a Roland OCTA-CAPTURE (Roland Corporation, Hamamatsu, Shizuoka Prefecture, Japan) sound card. For the majority of the recordings, participants wore an intra-aural device equipped with in-ear microphones (IEM), outer-ear microphones (OEM), and miniature loudspeakers, connected to a cross-over, located inside the ear as seen in Fig. 1. The components used in the earpiece are as follows: the IEMs are Sonion 50GE31 (Sonion, Plymouth, Minnesota), the OEMs are Knowles FG-23652, the woofers are Knowles CI-22955, and the tweeters are Knowles WBFK-30095 (Knowles Electronics, Itasca, Illinois). Since occluding the ear canal has an effect on the frequency response of an IEM, the behavior of the IEM in an average ear canal is measured on a GRAS 45CB acoustic test fixture (GRAS Sound & Vibration, Holte, Denmark). Figure 2 shows the estimated transfer function between the IEM and the GRAS RA0045 ear simulator of the acoustic test fixture, measured by playing pink noise through the internal loudspeakers of the earpiece. The IEM itself has two characteristics deviating from a uniform response. First, it exhibits a first order low frequency roll-off at about 250 Hz. This was found to be a desirable characteristic for an IEM, since ear canal deformation resulting from jaw movement when talking caused quasi-static pressure changes inside the closed volume that is the occluded ear canal. On a previous earpiece prototype with flatter response IEM, this was found to cause acoustic overload and clipping of the signal at the transducer. Second, the IEM exhibits a peak at 8 kHz, followed by a second order high frequency roll-off beyond 10 kHz. Anti-resonances can be seen on the IEM response relative to the GRAS microphone response when the in-ear loudspeaker is used as a source. This is attributed to reflections off the end of the coupler and happens at frequencies for which the distance from the IEM to the end of the coupler

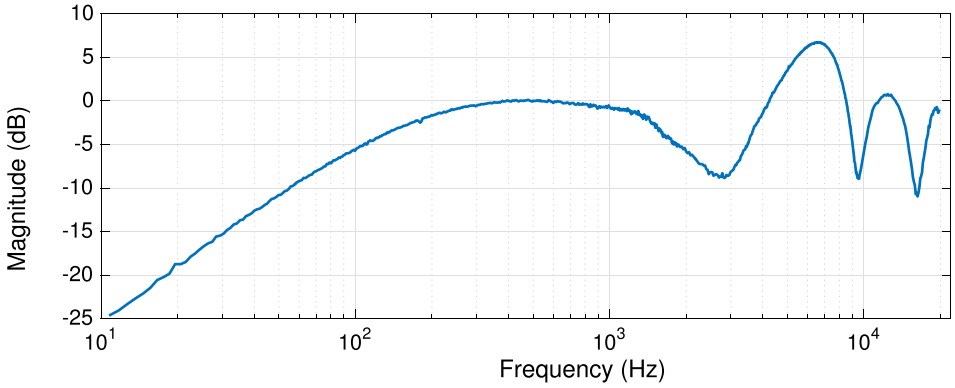
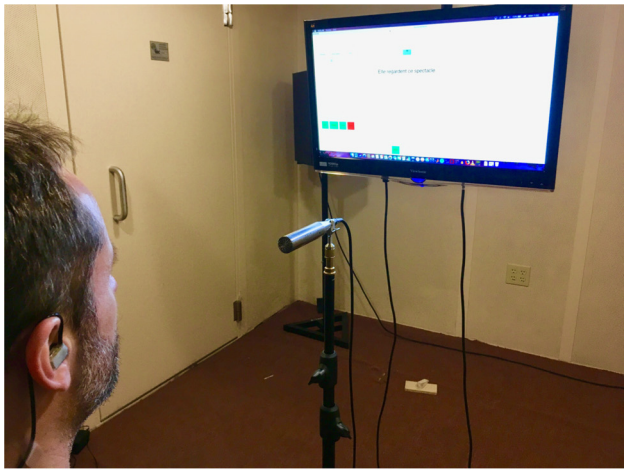


FIG. 2. (Color online) The estimated transfer function between the IEM and the microphone of the ear simulator, estimating the differences between measurement locations and occluded ear acoustics.



(A)

(B)

FIG. 3. (Color online) (A) An example showing the setup during the recording, including the placement of the REF microphone and the screen, with a (B) close up of the earpiece inside the participant's ear.

and back causes the reflected wave to be out of phase with the incident wave (Hiipakka *et al.*, 2010). They can be observed at around 3, 9, and 15 kHz.

To attenuate external noise and to occlude the talkers, Comply^{TM} Tx-200 tips (Hearing Components, Inc., Oakdale, Minnesota) were used with the earpiece. In addition, a GRAS 40HF 1-inch low-noise microphone (GRAS Sound & Vibration, Holte, Denmark) was placed 30 cm from the mouth at a 0° angle of incidence. This microphone is referred to as the REF microphone for the remainder of this paper. HINT sentences were displayed on a screen at 1 m from the participants head and its angle was adjusted to accommodate each participant's comfort. An example of the setup, including a participant equipped with the earpiece is shown in Fig. 3.

C. Procedure and recording conditions

To start, participants were asked to insert the earpieces in their ears to the best of their abilities. Once the earpiece was inserted, to check for a good acoustical seal, pink noise was played at 85 dBA using a loudspeaker placed 30 cm from the participants' head. A well-inserted earpiece was accepted if the attenuation between the OEM signal and the IEM signal was at least 8 dB at 250 Hz and the coherence between the two microphones was at least 0.8 (Voix and Laville, 2009). The foam tip of the earpiece was adjusted in the participants' ears until a good acoustical seal was satisfied. Subsequently, participants were asked to hum in a form of a frequency sweep from the lowest to the highest frequency they could manage. The occlusion effect was estimated as the difference in level between the IEM signal and the OEM signal at 250 and 500 Hz (Bernier and Voix, 2013; Kuk *et al.*, 2005). Next, participants were asked not to talk while factory noise from the NOISEX-92 database (Varga and Steeneken, 1993) was played in diffuse field, within the audiometric booth, at 95 dBA for 3 s and recorded using the IEM. The purpose of this recording was to estimate the residual noise inside the ear for each participant so that a noisy environment could then be regenerated using the internal loudspeakers of the earpiece.

Speech was recorded at 48 kHz sampling rate and 32-bit resolution in four different conditions. Table I lists and describes the noise conditions of the microphones. To allow for the training of denoising algorithms, such as those using adaptive filtering (Bouserhal *et al.*, 2017a), and to simulate a realistic environment, speech was recorded while factory noise from the NOISEX-92 database (Varga and Steeneken, 1993) was played in diffuse field, within the audiometric booth, at 95 dBA. Factory noise was selected to mimic an industrial environment. To better understand the noise, the long-term average spectrum as well as the modulation spectrum are presented in Fig. 4. Noise level measurements were made before the recording using the REF microphone placed at the location of the participant's head. In addition, clean speech was recorded while the earpiece was worn as well as with open ears. This helps assess any changes in speech production caused by simply wearing hearing protection, even in quiet conditions. Finally, to have access to clean Lombard speech that can be conveniently analyzed, noise was played directly in the ears leaving the OEMs and REF microphone free of noise. The noise regenerated inside the ear was filtered to reproduce the spectral characteristics that were measured by the IEM when it was previously played in diffuse field and recorded under the earpiece. This was done for each participant by designing a filter matching the inverse transfer function estimate between the IEM and the input to the miniature loudspeakers inside each ear and applying it to the individually recorded residual noise before playing it back through the miniature loudspeakers of the earpiece. Calculating the inverse transfer function estimate was done by sending pink noise to the miniature loudspeakers and recording it with the IEM as the earpiece was inserted, while looping back the pink noise to the sound card to obtain a time-aligned reference. The coefficients of the filter were those of an adaptive filter with a step-size of 0.01, where the

TABLE I. A list of the four recording conditions and the states of the speech signals picked up by each microphone.

Condition	IEM	OEM	REF
Open-ear quiet	N/A	N/A	Clean
Occluded quiet	Clean	Clean	Clean
Occluded noisy (ambient noise)	Noisy	Noisy	Noisy
Occluded noisy (regenerated in-ear noise)	Noisy	Clean	Clean

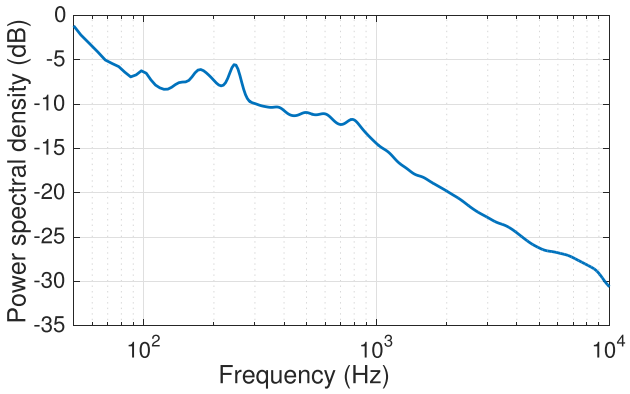


FIG. 4. (Color online) (A) The long-term average spectrum and (B) the modulation spectrum of the factory noise used in the experiment.

input signal $x(n)$ was the IEM signal and the desired signal $d(n)$ was the looped back signal at sound card, as shown in Fig. 5. Due to acoustical constraints on the frequency response of the loudspeaker and its enclosure in the earpiece, the residual noise could not be matched without error at all frequency bands. Nonetheless, the mean deviation between the overall sound pressure level (SPL) of the recorded residual noise and the regenerated residual noise over all participants was 1.4 and 1.8 dB with standard deviation of 1.0 and 1.6 dB, for the left and right ear, respectively. An example of the recorded residual noise and its regenerated counterpart is shown in Fig. 6.

A list of 234 HINT sentences, for both English and French, was divided into three lists of 78 sentences. Each of the three occluded conditions was assigned a list of unique sentences. A fourth list of 78 sentences was created from the first 26 sentences of each of the three unique aforementioned lists, to be used during the open-ear quiet condition. Every list was read in its entirety by every talker for each condition. Therefore, the first 26 sentences of each list were repeated twice by each talker, once in the open-ear condition and once in their respective occluded condition.[1] This was done to ensure that in each condition a reference list of sentences was available so that comparisons could be drawn on the phonetic level.

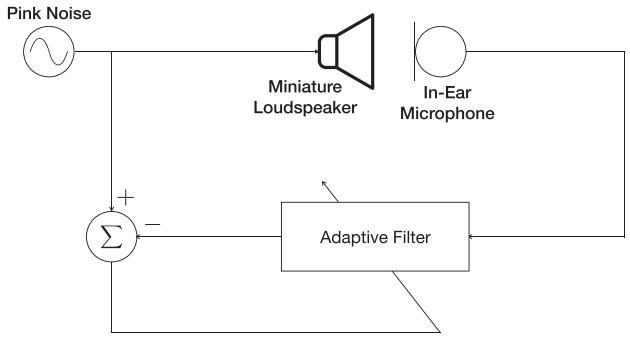


FIG. 5. A schematic of the adaptive filtering technique used to extract the coefficients of a filter representing the inverse transfer function estimate between the IEM and the input to the loudspeakers. The extracted coefficients are used for the regeneration of the in-ear noise signal.

III. THE DATABASE

As part of the collected corpus, data beyond the speech signals is made available to the user. For each participant, the following data are available: age, sex, language spoken, native language, hearing thresholds, attenuation of earplug, occlusion effect, measured residual noise levels during ambient noise, measured regenerated noise levels in the ear, and speech sentences in four different conditions. Attenuation of the earpiece at each ear are 1×9 vectors representing attenuation at each octave band frequency. Occlusion effect estimates are provided as a 1×2 vector with the measured overall SPL values in dB at 250 Hz and 500 Hz in the first and second column, respectively. Sex, native language, and spoken language are strings with one value for each participant. Hearing thresholds at each ear are 1×9 vectors representing monaural octave band thresholds. Overall SPL measurements of the residual noise inside the right and left ear as well as SPL measurement of the regenerated noise in both ears are also provided as 1×2 vectors for each participant. Speech signals are WAV files and $78 \times 4 = 312$ sentences are associated with each participant.[1] Therefore, in total, 7488 sentences are available including 1872 sentences in each condition. This database can be accessed by filling out a request through the research group's website.[1]

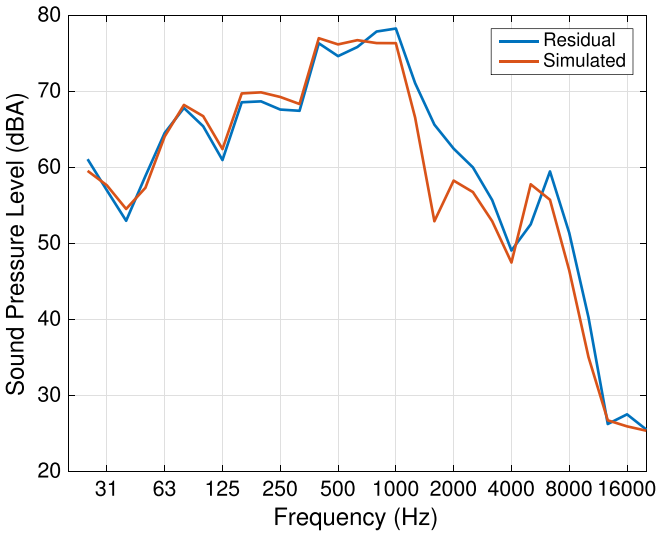


FIG. 6. (Color online) An example of the recorded residual noise in the ear and its regenerated counterpart.

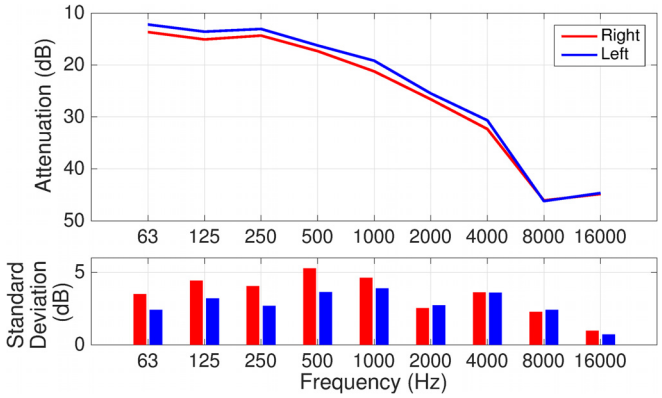


FIG. 7. (Color online) The mean attenuation of the earpiece over all participants for the right and left ear and the standard deviation at each octave band frequency.

IV. PRECURSORY ACOUSTICAL ANALYSIS

A precursory analysis is performed to describe the changes provoked in each noise condition and the spread of the data accessible as part of the database. To better illustrate the available data of the database, mean, standard deviation, maximums and minimums are calculated and presented. The statistical analysis tool, R (Team, 2013) and the *lme4* package (Bates *et al.*, 2014) are used to perform a linear mixed effects analysis of the relationship between the participants' speech levels and the various variables in the study. Participants as well as the sentences read are treated as random effects. The presence of noise (quiet or noisy), as well as whether the earpiece was worn, are treated as fixed effects in the model without any interaction term. Furthermore, the effects of sex (male or female) and language (French or English) are studied by adding these binary variables separately to the model while including interaction terms with both the presence of noise and whether or not the participant is occluded. Likelihood ratio tests between the full model with and without the effect in question are used to compare the goodness of each model. Three conditions are compared, the open-ear quiet condition, the occluded (i.e., the earpiece is worn) quiet condition, and the occluded noisy condition, where noise was regenerated directly inside the ear. The objective of this analysis is to aid future users of the database to be informed about the broad relationships and interactions between the conditions and the variables examined. It aims to facilitate the decision of the user on what data fits their needs.

V. RESULTS OF ACOUSTICAL ANALYSIS

Figure 7 presents the mean attenuation of all participants at each octave band frequency and the respective standard deviation for each ear. Overall, the observed maximum and minimum attenuation are 24.7 and 8.1 dB for the right ear, and 23.2 and 8.5 dB for the left ear. The occlusion effect estimates for all participants in the right and left ear at 250 and 500 Hz are presented in Fig. 8. On average the estimated occlusion effect over all talkers for both ears is at 19.7 dB with a 5.2 dB standard deviation and 16.5 dB with a 4.9 dB standard deviation at 250 and 500 Hz, respectively. Occlusion effect estimates range from a minimum of 5.0 dB to a maximum of 31.2 dB at 250 Hz, and a minimum of 7.1 dB to a maximum of 28.2 dB at 500 Hz. The variations in level of occlusion can be attributed to a combination of the insertion depth of the earpiece and the geometry of the participant's ear canal. Figure 9 compares the measured residual noise levels against the regenerated levels for each participant in each ear. The mean residual noise over all participants is 79.9 and 81.2 dBA with standard deviation of 4.3 and 3.7 dB for the right and left ear, respectively. Similarly, the mean regenerated noise levels are 78.5 and 80.2 dBA with standard deviation of 4.7 and 3.6 dB for the right and left ear, respectively. As discussed in Sec. II C, the mean difference between the residual and regenerated noise inside the ear is 1.4 and 1.8 dB with a standard deviation of 1 and 1.6 dB for the right and left ear, respectively. The maximum difference in level between the residual noise and the regenerated noise is found at 3.8 and 6.2 dB for the right and left ear, respectively.

For the linear mixed effect analysis, visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. The presence of noise and occlusion (i.e., the earpiece is worn) have a significant ($p < 0.001$) effect on the speech level of participants. On average participants speak at 57.9 dBA with 1.1 dB standard

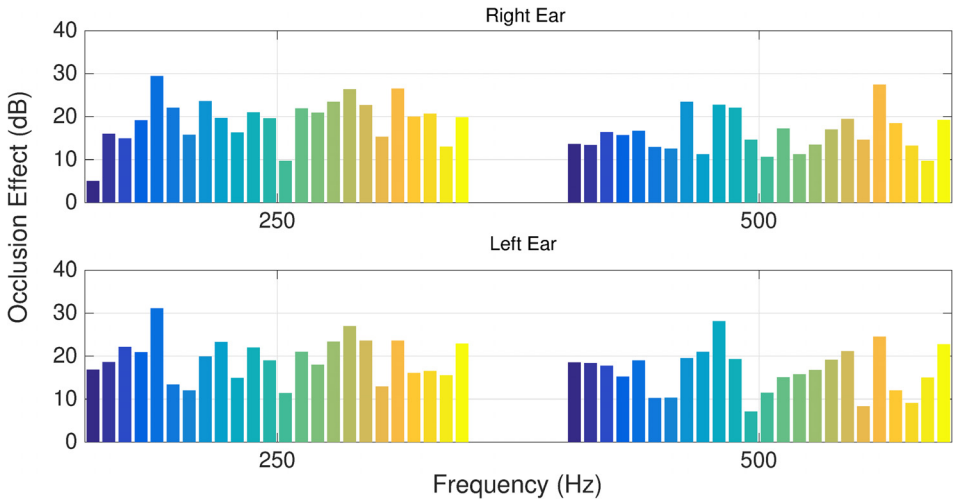


FIG. 8. (Color online) A bar graph representing the estimated occlusion effect at 250 and 500 Hz for the right and left ear, for all participants.

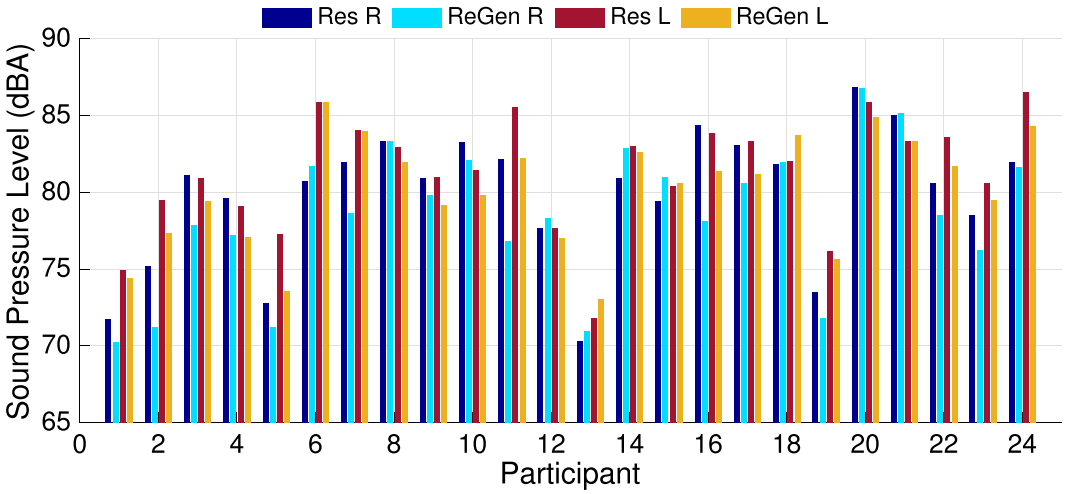


FIG. 9. (Color online) A bar graph representing SPL values of the residual noise (Res) measured inside the ear when ambient noise at 95 dBA is played in the room compared to the SPL values of the regenerated (ReGen) noise played directly inside the ear, for all participants in the right (R) and left (L) ear.

deviation in the quiet open-ear condition. Once occluded and in quiet, on average participants raise their speech level by 2.6 with 0.1 dBA standard deviation compared to the open-ear condition. This is in contrast to Tufts and Frank (2003) and Navarro (1996) who showed no change in speech level when occluded in quiet. However, this could be explained by differences in occlusion effect and attenuation of the ear-plug as well as the feedforward system described in Sec. I.

To visualize the changes in speech level for each participant in each condition, Fig. 10 presents boxplots of speech levels over all sentences in each condition for each participant in the open-ear quiet, occluded quiet and occluded noisy conditions. Compared to the open-ear quiet condition, most participants (excluding participant 21 and 22) raised their speech level on average once occluded in quiet and all participants raised their speech level on average in the presence of noise. Figure 11 presents a box plot of the speech levels at the three conditions. At the introduction of 95 dBA of factory noise, the average speech level increases on average by 6.5 dB with 0.1 dB standard deviation. This is consistent with existing literature that showed speech level to increase between 1.3 and 1.8 dB for every 10 dB increase in noise from 60 dB (Bouserhal *et al.*, 2016; Tufts and Frank, 2003).

Analysis showed that males and females do not speak at different average speech levels. However, upon introducing the interaction of noise and the presence of an earpiece, female participants appear to raise their voice in the presence of noise by 0.8 dB with 0.2 dB standard deviation ($p < 0.001$) and when they are occluded by 0.6 with 0.2 dB standard deviation ($p < 0.001$) more than males. This is consistent with Junqua *et al.* (1999) who showed that men and women do not react the same way to the Lombard effect. It is not understood exactly why females reacted more strongly to the Lombard effect than males. However, Junqua *et al.* (1999) also showed that there was an increase in the 4–5 kHz frequency band for vowels uttered by females under the Lombard effect which was not observed for males. This increase in the high frequencies as well as the use of an A-weighting for level measurements in this study could explain these observed differences. No difference in speech level according to language spoken is found in the data. Anglophones and francophones speak on average at the same levels over all conditions.

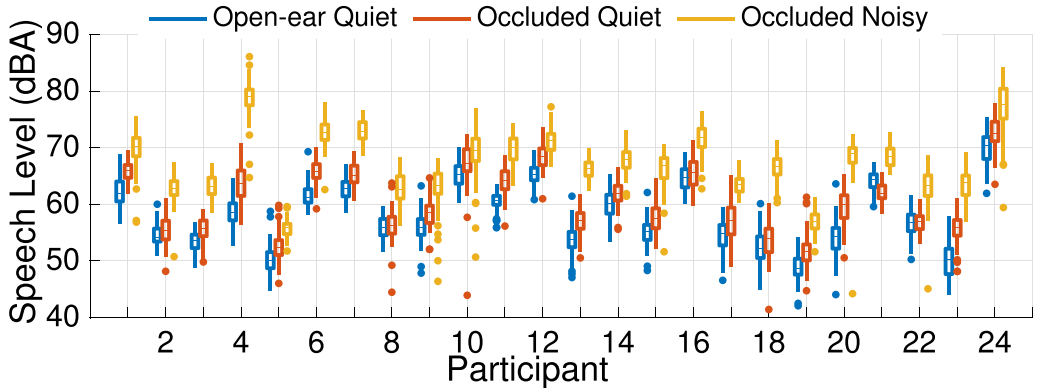


FIG. 10. (Color online) Boxplot of speech levels over all sentences in each condition for each participant in the open-ear quiet, occluded quiet and occluded noisy conditions.

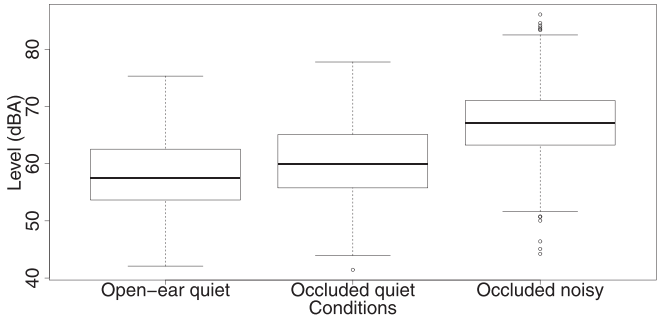


FIG. 11. Boxplot of speech levels as captured by the REF microphone in three conditions: the open-ear quiet condition (mean = 57.9 dBA), the occluded quiet condition (mean = 60.5 dBA), and the occluded noisy condition (mean = 67.0 dBA).

VI. CONCLUSIONS

The in-ear speech database, SpEAR, presented in this paper is meant to respond to a lack of in-ear speech databases in noisy conditions. It comprises of clean and noisy speech collected from occluded talkers, including a reference set of open-ear speech produced in quiet. It is intended to aid in the development, optimization, and validation of speech algorithms for intra-aural devices. Furthermore, it could aid in understanding changes in speech production on the phonetic level caused by being occluded. Having access to more than the speech signals also allows for a deeper understanding of the changes provoked from wearing intra-aural devices in adverse conditions. SpEAR could accelerate research advancements in the consumer market of hearables, as well as the occupational safety and health field of hearing protection devices.

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding received from the Centre for Interdisciplinary Research in Music Media and Technology, the Fonds de recherche du Québec–Nature et technologies, the Natural Sciences and Engineering Research Council of Canada, and the NSERC-EERS Industrial Research Chair in In-Ear Technologies. The authors would also like to thank Stijn Rebry for his help with the statistical analysis and Sanjeev Kumar Singh for his early work on the MATLAB GUI. Thanks to all the participants that volunteered and contributed in making this database.

[1]See supplementary material at https://doi.org/10.1121/1.5091777 for the list of sentences; a sample of the WAV files; and http://critias.etsmtl.ca/spear/.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and s4," R package version **1**(7), 1–23.

Bernier, A., and Voix, J. (**2013**). "An active hearing protection device for musicians," in *Proceedings of Meetings on Acoustics ICA2013*, Vol. 19, p. 040015.

Bottalico, P., Passione, I. I., Graetzer, S., and Hunter, E. J. (**2017**). "Evaluation of the starting point of the Lombard effect," Acta Acust. Acust. **103**(1), 169–172.

Bouserhal, R., Chabot, P., Sarria-Paja, M., Cardinal, P., and Voix, J. (**2018**). "Classification of nonverbal human produced audio events: A pilot study," in *Interspeech*, 2–6 September, Hyderabad, India.

Bouserhal, R. E., Bockstael, A., MacDonald, E., Falk, T. H., and Voix, J. (**2017b**). "Modeling speech level as a function of background noise level and talker-to-listener distance for talkers wearing hearing protection devices," J. Speech, Lang. Hear. Res. **60**(12), 3393–3403.

Bouserhal, R. E., Falk, T. H., and Voix, J. (**2013**). "Integration of a distance sensitive wireless communication protocol to hearing protectors equipped with in-ear microphones," in *Proceedings of Meetings on Acoustics ICA2013*, Vol. 19, p. 040013.

Bouserhal, R. E., Falk, T. H., and Voix, J. (**2015**). "On the potential for artificial bandwidth extension of bone and tissue conducted speech: A mutual information study," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5108–5112.

Bouserhal, R. E., Falk, T. H., and Voix, J. (**2017a**). "In-ear microphone speech quality enhancement via adaptive filtering and artificial bandwidth extension," J. Acoust. Soc. Am. **141**(3), 1321–1331.

Bouserhal, R. E., Macdonald, E. N., Falk, T. H., and Voix, J. (**2016**). "Variations in voice level and fundamental frequency with changing background noise level and talker-to-listener distance while wearing hearing protectors: A pilot study," International journal of audiology **55**(sup1), S13–S20.

Brumm, H., and Zollinger, S. A. (**2011**). "The evolution of the Lombard effect: 100 years of psychoacoustic research," Behaviour **148**(11-13), 1173–1198.

Bulbuller, G., Fargues, M., and Vaidyanathan, R. (**2006**). "In-ear microphone speech data segmentation and recognition using neural networks," in *12th Digital Signal Processing Workshop and 4th Signal Processing Education Workshop*, pp. 262–267.

Byrne, D. (**2014**). "Influence of ear canal occlusion and air-conduction feedback on speech production in noise," Ph.D. thesis, University of Pittsburgh.

Casali, J. G., and Horylev, M. J. (**1987**). "Speech discrimination in noise: The influence of hearing protection," in *Proceedings of the Human Factors Society Annual Meeting*, SAGE Publications Sage CA: Los Angeles, CA, Vol. 31, pp. 1246–1250.

Cooke, M., and Lecumberri, M. L. G. (**2012**). "The intelligibility of Lombard speech for non-native listeners," J. Acoust. Soc. Am. **132**(2), 1120–1129.

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., and Brumberg, J. S. (**2010**). "Silent speech interfaces," Speech Commun. **52**(4), 270–287.

Garnier, M., and Henrich, N. (**2014**). "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?," Comput. Speech Lang. **28**(2), 580–597.

Garnier, M., Henrich, N., and Dubois, D. (**2010**). "Influence of sound immersion and communicative interaction on the Lombard effect," J. Speech, Lang., Hear. Res. **53**(3), 588–608.

Hiipakka, M., Tikander, M., and Karjalainen, M. (**2010**). "Modeling of external ear acoustics for insert headphone usage," J. Audio Eng. Soc. **58**, 269–281.

Hirsch, H.-G., and Pearce, D. (**2000**). "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*.

Hoemann, H., Lazarus-Mainka, G., Schubeius, M., and Lazarus, H. (**1984**). "Effect of noise and the wearing of ear protectors on verbal communication," Noise Control Eng. J. **23**(2), 69–77.

Hotchkin, C., and Parks, S. (**2013**). "The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems," Biol. Rev. **88**(4), 809–824.

Howell, K., and Martin, A. (**1975**). "An investigation of the effects of hearing protectors on vocal communication in noise," J. Sound Vib. **41**(2), 181–196.

Hu, Y., and Loizou, P. (**2007**). "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun. **49**(7), 588–601.

Hunn, N. (**2016**). "The market for hearable devices 2016–2020," Technical Report, http://www.nickhunn.com.

Ikeno, A., Varadarajan, V., Patil, S., and Hansen, J. H. L. (**2007**). "UT-Scope: Speech under Lombard effect and cognitive stress," in *IEEE Aerospace Conference*, pp. 1–7.

Jankowski, C., Kalyanswamy, a., Basson, S., and Spitz, J. (**1990**). "NTIMIT: A phonetically balanced, continuous speech, telephone bandwidth speech database," *International Conference on Acoustics, Speech, and Signal Processing*, pp. 109–112, doi: 10.1109/ICASSP. 1990.115550.

Johansen, B., Flet-Berliac, Y. P. R., Korzepa, M. J., Sandholm, P., Pontoppidan, N. H., Petersen, M. K., Larsen, J. E., and Stephanidis, C. (**2017**). "Hearables in hearing care: Discovering usage patterns through

IoT devices," in *Universal Access in Human–Computer Interaction. Human and Technological Environments* (Springer, New York), pp. 39–49.
Junqua, J.-C., Fincke, S., and Field, K. (**1999**). "The Lombard effect: A reflex to better communicate with others in noise," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 4, pp. 2083–2086.
Kryter, K. D. (**1946**). "Effects of ear protective devices on the intelligibility of speech in noise," J. Acoust. Soc. Am. **18**(2), 413–417.
Kuk, F., Keenan, D., and Lau, C.-c. (**2005**). "Vent configurations on subjective and objective occlusion effect," J. Am. Acad. Audiol. **19**(9), 747–762.
Kurcan, R. S. (**2006**). "Isolated word recognition from in-ear microphone data using hidden Markov models (HMM)," Ph.D. thesis, Naval Postgraduate School, Monterey, California.
Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech, Lang., Hear. Res. **14**(4), 677–709.
Le Roux, J., Vincent, E., Hershey, J. R., and Ellis, D. P. (**2015**). "Micbots: Collecting large realistic datasets for speech and audio research using mobile robots," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5635–5639.
Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of the elevation of the voice"), Ann. Mal. L'Oreille Larynx **37**, 101–119.
Martin, A., and Voix, J. (**2017**). "In-ear audio wearable: Measurement of heart and breathing rates for health and safety monitoring," IEEE Trans. Biomed. Eng. **65**, 1256–1263.
Navarro, R. (**1996**). "Effects of ear canal occlusion and masking on the perception of voice," Percept. Mot. Skills **82**(1), 199–208.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099.
Pantelopoulos, A., and Bourbakis, N. (**2010**). "A survey on wearable sensor-based systems for health monitoring and prognosis," IEEE Trans. Systems, Man, Cybernet., Part C (Applic. Rev.) **40**(1), 1–12, http://ieeexplore.ieee.org/document/5306098/.
Pittman, A. L., and Wiley, T. L. (**2001**). "Recognition of speech produced in noise," J. Speech, Lang. Hear. Res. **44**(3), 487–496.
Rothauser, E. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Shin, H. S., Kang, H.-G., and Fingscheidt, T. (**2012**). "Survey of speech enhancement supported by a bone conduction microphone," in *Proceedings of 10. ITG Symposium Speech Communication*, VDE, pp. 1–4.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**(3), 917–928.
Team, R. C. (**2013**). "R: A language and environment for statistical computing."
Tourville, J. A., and Guenther, F. H. (**2011**). "The diva model: A neural theory of speech acquisition and production," Lang. Cognit. Processes **26**(7), 952–981.
Tufts, J. B., and Frank, T. (**2003**). "Speech production in noise with and without hearing protection," J. Acoust. Soc. Am. **114**(2), 1069–1080.
Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., Soli, S. D., and Giguère, C. (**2005**). "Adaptation of the hint (hearing in noise test) for adult Canadian francophone populations [Adaptación del hint (prueba de audición en ruido) para poblaciones de adultos canadienses francófonos]," Int. J. Audiol. **44**(6), 358–361.
Varga, A., and Steeneken, H. J. (**1993**). "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun. **12**(3), 247–251.
v. Békésy, G. (**1949**). "The structure of the middle ear and the hearing of one's own voice by bone conduction," J. Acoust. Soc. Am. **21**(3), 217–232.
Voix, J. (**2017**). "The ear beyond hearing: From smart earplug to in-ear brain computer interfaces," in *24th International Congress on Sound and Vibration, ICSV24*, London.
Voix, J., and Laville, F. (**2009**). "The objective measurement of individual earplug field performance," J. Acoust. Soc. Am. **125**(6), 3722–3732.
Zue, V., Seneff, S., and Glass, J. (**1990**). "Speech database development at MIT: TIMIT and beyond," Speech Commun. **9**, 351–356.
Zwislocki, J. (**1957**). "In search of the bone-conduction threshold in a free sound field," J. Acoust. Soc. Am. **29**(7), 795–804.