Research Article

Modeling Speech Level as a Function of Background Noise Level and Talker-to-Listener Distance for Talkers Wearing Hearing Protection Devices

Rachel E. Bouserhal, Annelies Bockstael, Ewen MacDonald, Tiago H. Falk, and Jérémie Voix

Purpose: Studying the variations in speech levels with changing background noise level and talker-to-listener distance for talkers wearing hearing protection devices (HPDs) can aid in understanding communication in background noise.

Method: Speech was recorded using an intra-aural HPD from 12 different talkers at 5 different distances in 3 different noise conditions and 2 quiet conditions.

Results: This article proposes models that can predict the difference in speech level as a function of background noise level and talker-to-listener distance for occluded talkers. The proposed model complements the existing model presented by Pelegrín-García, Smits, Brunskog, and Jeong (2011) and expands on it by taking into account the effects of occlusion and background noise level on changes in speech sound level.

Conclusions: Three models of the relationship between vocal effort, background noise level, and talker-to-listener distance for talkers wearing HPDs are presented. The model with the best prediction intervals is a talker-dependent model that requires the users' unoccluded speech level at 10 m as a reference. A model describing the relationship between speech level, talker-to-listener distance, and background noise level for occluded talkers could eventually be incorporated with radio protocols to transmit verbal communication only to an intended set of listeners within a given spatial range—this range being dependent on the changes in speech level and background noise level.

Talkers adjust their vocal effort with varying talker-to-listener distance (Fux, Feng, & Zimpfer, 2011) in the presence of noise (Lane & Tranel, 1971), to express emotion (Schröder, 2001), and in different acoustic environments (Bottalico, 2017; Brunskog, Gade, Bellester, & Calbo, 2009). In an effort to understand strain on voice observed in teachers, changes in vocal effort caused by room acoustics have been well documented (Astolfi, Carullo, Pavese, & Puglisi, 2015; Brunskog et al., 2009; Pelegrín-García, Smits, Brunskog, & Jeong, 2011; Puglisi, Astolfi, Cantor Cutiva, & Carullo, 2017). It has been shown that as reverberation time increases, speech level decreases (Brunskog et al., 2009; Pelegrín-García, Smits, et al., 2011); however, Astolfi et al. (2015) and Puglisi et al. (2017) proved that as reverberation increases in everyday environments such as classrooms, voice level increases due to the Lombard effect generated by the increase in internal noise that is correlated with the increase in reverberation. Furthermore, an increase in the reverberation time measured at a talker’s ears was found to be strongly associated with an increase in the perceived noise condition inside of classrooms, which may lead to a raised voice (Cantor Cutiva, Puglisi, Astolfi, & Carullo, 2017). Bottalico (2017) and Astolfi et al. (2015) confirmed the findings by Pelegrín-García, Smits, et al. (2011) that phonation time increases as reverberation time of the room increases.

Disclosure: The authors have declared that no competing interests existed at the time of publication.
Changes in vocal effort as a function of varying talker-to-listener distance have also been well studied and modeled (Pelegrín-García, Smits, et al., 2011; Traummüller & Eriksson, 2000; Zahorik & Kelly, 2007). The study done by Zahorik and Kelly (2007) showed that talkers adjust their vocal effort according to their acoustical environment as well as to the communication distance. In addition, Pelegrín-García, Smits, et al. (2011) proposed a model of speech levels as a function of the talker-to-listener distance as well as the room acoustics.

The increase in speech levels as a result of the onset of noise is known as the Lombard effect (Bottalico, Passione, Graetzer, & Hunter, 2017; Lombard, 1911; Zollinger & Brumm, 2011). The Lombard effect and the increasing speech levels with changing talker-to-listener distance are done both involuntarily and voluntarily by a talker to enhance speech intelligibility by the listener. Studies have shown that the Lombard effect is manifested differently when talkers are trying to communicate in noise compared to performing a reading task with no targeted listener (Junqua, Fincke, & Field, 1999). Garnier, Dohen, Loevenbruck, Welby, and Bailly (2006) showed that changes in acoustics for Lombard speech are not purely physiological in nature but are rather a controlled enhancement in an effort to increase speech intelligibility.

Another implication of the Lombard effect and changes in vocal effort due to room acoustics is the presence of a feedback mechanism between vocal production and perception, working to adapt speech performance (Brumm & Field, 1999). Garnier, Henrich, & Dubois, (2010). As illustrated in Figure 1, there are three main components that affect the perception of one’s own voice (Lehnert & Giron, 1995; Pörschmann, 2000):

(a) Direct air conduction: Sound travels from the talker’s mouth to the ear through propagation in the open air.

(b) Bone conduction: Sound transmitted through bone and tissue conduction inside the skull. Direct stimulation of the cochlea can occur through vibrations of the skull vibrating the cochlear fluid. In addition, indirect stimulation can occur through the excitation of the air entrapmed in the ear canal vibrating the eardrum, thus stimulating the cochlea as well as the displacement of the stapes relative to its oval window.

(c) Indirect air conduction: Sound travels from the talker’s mouth and then reflects off of surfaces around the talker traveling back to the talker’s ear.

It is therefore reasonable to assume that wearing hearing protection devices (HPDs) affects this feedback path and thus causes deviations in speech production. For open ears, the air-conduction pathways are the primary feedback paths for a talker (Henry & Letowski, 2007). Blauert, Els, and Schroeter (1980) identified that direct transmission of sound from skull vibrations to the cochlea are 40 and 70 dB less effective in the high frequencies and the low frequencies, respectively, making sound transmission through air conduction superior to bone conduction. This also implies that, in the open-ear condition, the bone-conduction pathway may be neglected. When it comes to the perception of one’s own voice, however, the significance of the contribution from each path is debatable. Békésy (1949) concluded that the air- and bone-conduction paths equally contribute to one’s hearing of one’s own voice. However, a more recent study by Pörschmann (2000) observed that, except for the midfrequencies (700–1200 Hz) where bone conduction had a slightly superior contribution, air conduction was the dominant contributor to the hearing of one’s own voice. It is important to note that there are large deviations between talkers in terms of the contribution of bone conduction to self-perceived speech (Maurer & Landis, 1990). In summary, for different frequency ranges, both the bone- and air-conduction paths are significant contributors to the perception of one’s own voice. They act as the main feedback paths that aid talkers in correcting their speech to become more intelligible.

Occluding the ear canal with an HPD reduces the effect of two of the three feedback paths, the direct and indirect air-conduction paths, but amplifies the bone-conduction path. While speaking, the skull vibrates, causing the soft tissue of the ear canal to vibrate as well. For the open-ear condition, these vibrations are not noticeable, as a substantial part of the energy escapes through the opening of the ear canal. However, when the ear canal is blocked, the energy from the soft tissue vibrations in the ear canal builds up, resulting in larger sound waves causing an amplification of the bone-conducted sounds in the ear canal. This phenomenon is called the occlusion effect. The location at which the ear canal is blocked determines the strength of the occlusion effect (Dean & Martin, 2000). Blocking the ear canal right at its opening causes larger occlusion effect than when it is blocked closer to the eardrum. This can be explained by modeling the open and occluded ear canals as open and closed pipes of different lengths. The occlusion effect can also be modeled with electronic circuits where the open-ear

---

**Figure 1.** Illustration of the three paths affecting the perception of one’s own voice: (a) direct air conduction, (b) bone conduction, and (c) indirect air conduction.
canal resembles a high-pass filter and the occluded ear canal is the removal of this high-pass filter (Brummund, Sgard, Petit, & Laville, 2014). Therefore, when the ear canal is blocked and bone conduction is amplified, it can be assumed that the bone-conduction path dominates the audiophonation loop. Consequently, speech production is altered when wearing HPDs in quiet and in noise (Byrne, 2014; Casali, Horylev, & Grenell, 1987; Tufts & Frank, 2003). In general, when wearing HPDs, talkers do not raise their speech levels in noise as much as they do without HPDs. Tufts and Frank (2003) showed that when exposed to the same noise level, there is a 4- to 11-dB decrease in speech level between the open-ear condition and the occluded condition. The changes in speech levels and fundamental frequency caused by occluding the ear in noise have been well studied. A good review of the literature is done by Byrne (2014). However, studies on the effect of occluding the ear canal on variations in speech levels caused by changing talker-to-listener distance are still limited.

In this article, models of the talker-to-listener distance as a function of the background noise level and the talker’s speech levels for the occluded ear are presented. These models build on the model presented by Pelegrín-García, Smits, et al. (2011) and based on the results of a recent study by the authors (Bouserhal, Macdonald, Falk, & Voix, 2016). The models are meant to better represent changes in speech levels for a talker wearing HPDs in noise with changing communication distance. Because a large intersubject variability in speech level is expected (Bouserhal et al., 2016; Pelegrín-Garcia, Fuentes-Mendizábal, Brunskog, & Jeong, 2011), it is of interest to investigate whether including a talker-dependent factor in the model would reduce variability. A talker’s unoccluded speech levels at specific distances contain relevant information on the talker’s adaptation to changes in talker-to-listener distance. Knowledge of this adaptation when unoccluded could be incorporated in a talker-dependent model as a reference and extended to better predict how a talker would adjust his or her speech level to distance and noise when occluded. Reaching a model with small prediction intervals could potentially be integrated with HPDs equipped with radio capabilities to enhance the communication experience for users in noisy environments (Bouserhal, Falk, & Voix, 2013). This could be done by instructing the radio to transmit verbal communication from the talker only to listeners within a specific spatial range. This range will be determined using the model, based on the talker’s changes in speech levels and the level of background noise. Therefore, it is important to consider the practical implementation of such a model. For such a model to be implemented with advanced HPDs, the talker-dependent data required should be kept as limited as possible due to realistic constraints. Consequently, finding one distance where the speech level best characterizes a talker’s adaptation to distance, as well as keeping prediction intervals small, could be implementable and realistically used as a reference in a model predicting speech level as a function of talker-to-listener distance and background noise level for occluded talkers.

Method and Materials

Experimental Setup

The models developed in this work are based on the data collected from a recent study by the authors and supported by the “Comité d’éthique pour la recherche,” the institution’s internal review board. For details on the experimental setup and procedure, the reader is encouraged to refer to that work (Bouserhal et al., 2016). To summarize, the study involved 12 volunteer participants (10 men, two women) ranging in age from 23 to 34 years, with a mean age of 28 years. No audiogram was performed to measure hearing thresholds, but no participants reported any known hearing loss. Participants were only instructed on the procedure prior to participation and had no knowledge of the reason for the study. They were equipped with an intra-aural HPD containing outer-ear microphones, in-ear microphones, and miniature loudspeakers in the ear, as depicted in Figure 2. They stood in a long corridor and were asked to lead a listener through a set of standardized geographical maps (Anderson et al., 1991; Pelegrín-García, Smits, et al., 2011) from start to finish at varying levels of noise and varying talker-to-listener distances. The speech was not interactive because the listener gave no feedback to the talker. The corridor is the same one used during the Pelegrín-García, Smits, et al. (2011) experiments, and the room acoustics as reported are shown in Table 1. Table 2 shows the 25 different experimental conditions performed. Speech was recorded using the outer-ear microphones placed on the outside surface of the HPD. Recordings were made at a sampling frequency of 48 kHz using a Fireface UCX soundcard (RME) and a Windows computer running MATLAB (version 2015b; MathWorks). For each noise condition (i.e., 70, 80, and 90 dB SPL), factory noise from the NOISEX-92 database (Varga & Steeneken, 1993) was recorded once prior to the experiment in an anechoic booth using the outer-ear microphones. Noise levels were measured using a sound level meter and averaged over 15 s ($L_{eq, 15s}$). The choice of noise levels was made similar to those used by Tufts and Frank (2003) for comparison. Background noise levels were unweighted (SPL) values; however, all speech recordings were run through an A-weighting filter for level assessment. The A-weighted SPL value was chosen over an overall SPL value to better match the analysis bandwidth to the speech communication bandwidth and to apply less weight to any extraneous low-frequency parasitic noise that could have been picked up by the microphones, given the ambient background noise (heating/air conditioning, etc.). At the start of each experiment, the fit of the HPD was tested to ensure a good acoustic seal by undertaking the following procedure: A broadband white noise was played over external loudspeakers with the talker’s head facing the two external loudspeakers and recording simultaneously for 2 s using in-ear and outer-ear microphones on each ear and calculating the attenuation of the earplug using a procedure known as field-microphone-in-real-ear and described in Voix and Laville (2009). The schematic of this procedure.
is illustrated in Figure 3. An example of the attenuation curve of a well-fitted earplug is shown in Figure 4. The resulting transfer function from a good acoustical seal was then used to filter the aforementioned noise signals, and noise was played directly in the participant’s ears for each noise condition. Therefore, the residual noise inside the occluded ear canal was known for each participant and differed based on the participant’s individual transfer function. This left the outer-ear microphones noise free and allowed for recording of clean Lombard speech.

**Model Calculation**

To find the most appropriate model to fit the data, the model presented by Pelegrín-García, Smits, et al. (2011) was used as a starting point. The proposed model by Pelegrín-García, Smits, et al. (2011)\(^1\) was as follows:

\[
L_w = a + \alpha + (b + \beta) \times \log_2(D/1.5) + \epsilon, \tag{1}
\]

where \(L_w\) is the speech power level; \(a\) and \(b\) are the coefficients of the fixed effects; \(\alpha\), \(\beta\), and \(\epsilon\) are the coefficients for the random effects; and \(D\) is the talker-to-listener distance in meters. Pelegrín-García, Smits, et al. (2011) use speech power levels, \(L_w\), to represent the strength of speech sounds. However, in this work, unlike Pelegrín-García, Smits, et al. (2011), this magnitude is represented using the on-axis sound pressure level, which is a different yet valid way of representing the strength of speech sounds (Pelegrín-García, Smits, et al., 2011). For the model created from the observations in the corridor, the values for the fixed effects and random effects presented by Pelegrín-García, Smits, et al. (2011) are shown in Table 3. It is hypothesized that:

1. The data collected in the open-ear quiet condition fits in the model in Equation 1.
2. Model 1 fails once the ear is occluded and in the presence of noise.

From previous research, it is known that the noise level has a statistically significant effect on the level of noise when wearing HPDs (Bouserhal et al., 2016; Tufts

### Table 1. Physical volume, reverberation time, and background noise level as reported by Pelegrín-García, Smits, et al. (2011).

<table>
<thead>
<tr>
<th>(V) (m(^3))</th>
<th>(T_{30}) (s)</th>
<th>(L_{N,Aeq}) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>410</td>
<td>2.34</td>
<td>37.7</td>
</tr>
</tbody>
</table>

\(^1\)Reprinted with permission. Copyright © 2011, Acoustical Society of America
In addition, from Bouserhal et al. (2016), it has been shown that there is a significant interaction between distance and noise level. In addition, from Pelegrín-García, Smits, et al. (2011) as well as from Traunmüller and Eriksson (2000), there is a logarithmic relationship between speech level and distance. This knowledge is, thus, accounted for while creating the new models using the lme4 package (Bates, 2010) supplied in the open access R system for statistical computing (Ihaka & Gentleman, 1996). In the model, only statistically significant independent variables and interaction effects are included. Statistical significance is based on \( p \) values calculated using single-term deletion (Bates, 2010): A chi-square test is performed on the difference in Akaike information criterion value of the model with and without the variable or interaction effect under study. Statistical model assumptions are verified by visual inspection of the residual’s Q-Q plot together with Shapiro–Wilk test and Kolmogorov–Smirnov test (Kutner, Nachtsheim, Neter, & Li, 2005). For significance, an alpha of .05 is used. Using the “bootMer” function from the lme4 package, model-based parametric bootstrap is performed to calculate predictions. For each model, 1,000 simulations are run. New random effect and residual error values are generated for each simulation (Bates, 2010). Different models are compared to assess the benefits of adding different main and interaction effects to the model. The least complicated model with the significantly lowest Akaike information criterion value is retained. To visualize the results, the 0.025, 0.500, and 0.975 quantiles are calculated.

Three models are generated to predict speech level. The first model is a talker-independent model, needing no information about the talker. However, as previously discussed, a large intersubject variability is expected. In an effort to reduce the effect of such variability on the prediction intervals, the second model is talker dependent and uses as a reference the talker’s unoccluded speech levels at all the measured distances. Still, a model needing all five distances (going up to 30 m) as reference is not practical due to realistic constraints, such as time and available space. Lastly, the use of one particular distance as a reference to the model is investigated to create a third model that is talker dependent, with small prediction intervals relying on only one distance as a reference. Because the independent variables of the talker-dependent models are measured on very different scales, they were standardized for numerical stability.

### Results

#### Comparison With Existing Model

Here the collected data in the quiet condition, occluded and unoccluded, is compared to Model 1. Because...
the model by Pelegrín-García, Smits, et al. (2011) was based on the level of speech, unweighted speech levels are used for comparison in this section. The levels from the open-ear condition are plotted against Model 1 in Figure 5. It can be seen that for the open-ear condition, except for the 1-m condition, most speech levels fall within the range of 1 SD above and below Model 1. This corroborates the model by Pelegrín-García, Smits, et al. (2011) and extends its validity to distances not tested by that study. However, as can be seen from Figure 6, Model 1 does not properly reflect the changes in speech level for the occluded ear. This demonstrates the need for a model taking into account the occluded ear and the effect of noise.

**Proposed Models**

**Talker-Independent Model**

Without taking into account any prior information about the participants, the new model including the effect of noise as well as the interaction between noise and distance is given as follows:

\[
L_{wA} = a + \alpha + (b + \beta) \times \log_2(D/1.5) + (c + \gamma) + N + (g + \lambda) \times \log_2(D/1.5) \times N + u + \varepsilon,
\]

where \(c\) and \(g\) are the coefficients of the fixed effects due to the distance, noise, and the interaction between the two, respectively; \(\beta, \gamma, and \lambda\) are their respective standard errors; \(N\) is the level of noise outside the HPD in dB SPL; \(u\) is the standard deviation of the random effect (participant); and \(\varepsilon\) is the standard deviation of the residual error. All values for the model in Equation 2 are given in Table 4. Note that now the model takes in speech levels in dB(A) for the reasons discussed in the Experimental Setup section. The 0.025, 0.500, and 0.975 quantiles of Model 2 and their respective predictions for each noise condition are shown in Figure 7. As can be seen, the prediction intervals are 10 dB wide, and there is a substantial overlap between noise conditions. Although this model properly describes the relationship between speech level, noise, and distance for occluded talkers, it is not reliable in practice.

**Talker-Dependent Models**

Including the open-ear (aka unoccluded) levels of the talkers at each distance in the model causes the interaction effect between distance and noise to become statistically insignificant. This is understandable because the level is referenced at each distance to the level in quiet and the main contributor to the change in level becomes the noise. Therefore, the model becomes

\[
L_{wA} = a + \alpha + (b + \beta) \times \log_2(D/1.5) + (c + \gamma) \times N + (z + \varsigma) \times S + (h + \kappa) \times S \times N + u + \varepsilon,
\]

where \(z\) and \(h\) are the coefficients of the fixed effects due to the unoccluded speech level \(S\) at a given distance and its interaction with noise, respectively; \(\varsigma\) and \(\kappa\) are their respective standard errors; and \(S\) is the unoccluded speech level at the respective distance. All values for the model in Equation 3 are given in Table 4. Figure 8 shows the 0.025, 0.500, and 0.975 quantiles of predictions made for \(S = 50, 55, 60, 65,\) and 70 dB(A) for the two extreme noise conditions, quiet and 90 dB SPL. Including the unoccluded levels at all distances reduces the prediction intervals to 5 dB. However, this is at a cost of reducing the effect of distance on speech level. To better demonstrate this, Figure 9 shows...
the prediction intervals for all noise conditions with unoccluded speech level at 60 dB(A) for each distance. As can be seen, the relationship between speech level and distance is smoothed out. This is unsurprising, as the effect of distance is accounted for by the unoccluded speech levels at each distance. Consequently, even though including all distances as a reference to the model decreases the prediction interval, it reduces the relationship between distance and speech level and is not practical.

However, this motivates the idea that including a known speech level at a given distance may reduce the prediction intervals while still retaining the interaction effect between noise and distance and without being too demanding of the user. This was tested for levels for each distance of 1, 5, 10, 20, and 30 m. To assess the benefits of each particular distance in the model, the standard deviation of the random effects are compared between the different models; this is the variation between participants and the residual variation not captured by the model. The standard error associated with the model at each distance is shown in Table 5. The highest standard deviation can be seen at the 1-m and 30-m conditions. The lowest standard deviation is found when using the unoccluded speech level at the 10-m distance. In this model, all two-way interaction effects become statistically significant: the interaction effect between noise level and distance \((p < .01)\), between speech level at 10 m and distance \((p < .001)\), and between noise level and speech level at 10 m \((p < .01)\). The model relying only on the 10-m unoccluded speech level is given as

\[
L_w = a + \alpha (b + \beta) \times \log_2(D/1.5) + (c + \gamma) \times N + (g + \lambda) \times \log_2(D/1.5) \times N + (z + \zeta) \times S_{10} + (h + \kappa) \times S_{10} \times N + (r + \rho) \times \log_2(D/1.5) \\
\times S_{10} + u + \epsilon, \tag{4}
\]

where \(r\) is the coefficient of the fixed factor due to the interaction between the unoccluded speech level at 10 m \(S_{10}\) and distance and \(\rho\) is its respective standard error. All values for the model in Equation 4 are given in Table 4. The 0.025, 0.500, and 0.975 quantiles of predictions made for \(S_{10} = 50, 55, 60, 65, \) and 70 dB(A) for the two extreme noise conditions, quiet and 90 dB SPL, of Model 4 are shown in Figure 10. It can be seen that using only the 10-m unoccluded speech levels reduces the prediction intervals to less than 5 dB. In addition, because using only the occluded speech level at 10 m removes the distance dependence of the unoccluded speech level present in Model 3, the effect of variable distance on the outcome speech level becomes stronger again in Model 4. Figure 11 illustrates this better by showing only \(S_{10} = 60\) dB(A) for all the noise conditions. Not only are overlaps reduced between noise conditions but also within each noise condition, and speech levels correspond to a smaller range of distance. Furthermore, it

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed effects</th>
<th>Standard error</th>
<th>Random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Values for fixed and random effect coefficients and standard errors for Models 2, 3, and 4 relating speech level with distance and noise for the occluded ear.
requires the user to provide only one speech level at 10 m, which is a reasonable effort considering the benefits in the reduction of prediction intervals and the overlap between conditions.

Discussion

The models presented in this work indicate that talkers increase their speech level at the 1-m talker-to-listener distance when the ear is occluded. This can also be seen by comparing Figures 5 and 6. This is contradictory to Tufts and Frank (2003) and Navarro (1996), who found no significant increase in speech level when wearing HPDs in quiet at a 1-m distance. However, it is in accordance with older studies, such as Casali et al. (1987) and Kryter (1946), that reported up to 4 dB increase in speech level after wearing HPDs. As Tufts and Frank (2003) have explained, this could be a consequence of different levels of occlusion, which affect the perception of one’s own voice: High levels of occlusion caused by shallow-fitted HPDs could cause an increase in speech levels compared to the open-ear condition. In Bouserhal et al. (2016), it was discussed that this might have been a consequence of the design of the experiment, because the unoccluded quiet condition was the last condition to be recorded after the 90-dB SPL occluded condition, causing the talkers to greatly reduce their speech levels after removing the HPD. However, after seeing how well the unoccluded levels corroborate the model by Pelegrín-García, Smits, et al. (2011), this hypothesis seems no longer valid. Nonetheless, considering that only one talker was speaking at a time and the corridor was relatively quiet, the decrease in speech level in the unoccluded quiet condition compared to the occluded quiet condition could be caused by the reverberance of the corridor. It would be of interest

Figure 8. Prediction intervals at 0.025, 0.500, and 0.975 of the talker-dependent model in Equation 3 using the unoccluded speech levels at all five distances as reference. Prediction intervals, reduced to about 5 dB, are plotted for unoccluded speech levels at S = 50, 55, 60, 65, and 70 dB(A) for the quiet condition and 90 dB SPL for the noise condition.

Figure 9. Prediction intervals at 0.025, 0.500, and 0.975 of the talker-dependent model in Equation 3 using the unoccluded speech levels at all five distances as reference for S = 60 dB(A). Prediction intervals are flattened, and one speech level within one noise condition can correspond to multiple distances.
to investigate this in future studies and to better understand the nature of this decrease in level.

It may seem counterintuitive that providing only the unoccluded distance at 10 m would result in a better model than including much more detailed information of unoccluded speech levels, that is, measured at different distances. However, looking at the standard deviation in Table 5 showing highest values at the 1-m and 30-m conditions helps to explain this. It is speculated that at 1 m the talkers use their "conversation" voice, which can vary substantially between talkers even when they share the same gender and age (Traunmüller & Eriksson, 2000). This can also be seen in Figure 5 where speech levels at 1 m are the most spread out and speech levels at 10 m are more clustered at the exception of two talkers. However, this would not explain why the 30-m condition has the highest standard deviation, because levels seem to be more clustered at that distance as well. In this case, it is assumed that the cluster is due to a saturation effect in the speech caused by physiological limitations similar to those caused by very high noise levels (Bottalico et al., 2017). This is also expected from the logarithmic relationship between distance and speech level. Therefore, it is suspected that the 10-m condition best characterizes the adaptation of speech level caused by distance.

Even though data from only 12 talkers were used to create the models, the intersubject variability is accounted for in the models and is not considered to be a limiting factor. In addition, because the model with best prediction intervals in Equation 4 is talker dependent, it further reduces the need to have a large number of participants, as it accounts for a baseline speaker-specific speech level.

Figure 10. Prediction intervals at 0.025, 0.500, and 0.975 of the talker-dependent model in Equation 4 using the unoccluded speech levels at 10 m as reference. Prediction intervals reduced to less than 5 dB are plotted for unoccluded speech levels at $S_{10} = 50, 55, 60, 65, \text{and } 70 \text{ dB(A)}$ for the quiet condition and 90 dB SPL for the noise condition.

Figure 11. Prediction intervals at 0.025, 0.500, and 0.975 of the talker-dependent model in Equation 4 using the unoccluded speech levels at 10 m as reference for $S_{10} = 60 \text{ dB(A)}$. Prediction intervals are the best in this model and show lowest ranges of distances corresponding to the same speech level within one noise condition.
Table 5. The standard deviation associated with the models using the unoccluded speech level at only one of the given distances as a reference level.

<table>
<thead>
<tr>
<th>Distance (m)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.062</td>
</tr>
<tr>
<td>5</td>
<td>2.512</td>
</tr>
<tr>
<td>10</td>
<td>2.244</td>
</tr>
<tr>
<td>20</td>
<td>2.576</td>
</tr>
<tr>
<td>30</td>
<td>3.064</td>
</tr>
</tbody>
</table>

Conclusions

There is a clear relationship between speech level, background noise level, and talker-to-listener distance for persons wearing HPDs. Three models were presented describing this relationship. These models are complementary to the model presented by Pelegrín-Garcia, Smits, et al. (2011), and they include the effects of noise and occlusion on the variations in speech levels with changing talker-to-listener distance. A talker-independent model nicely captures the trend but has large prediction intervals. The most practical model with the best prediction intervals is talker dependent and requires an initial reference unoccluded speech level at 10 m. In future work, this model will be investigatated to be used to predict the intended communication distance of a talker wearing HPDs and transmitting information only to a radius of listeners within that spatial range, which will make speech transmission over personal radio in noise while wearing HPDs a more pleasant experience.

Acknowledgment

The authors would like to acknowledge the funding received from the Centre for Interdisciplinary Research in Music Media and Technology (for authors Bouserhal, Falk and Voix), the Fonds de recherche du Québec – Nature et technologies (2015-NC-181280, for Voix), the Natural Sciences and Engineering Research Council of Canada (402126-2012, 402237-2011, for Voix and Falk), the Erasmus Mundus student exchange program in Auditory Neuroscience (for Bouserhal, MacDonald and Voix), and the Natural Sciences and Engineering Research Council of Québec – Nature et technologies (2015-NC-181280, for authors Bouserhal, Falk and Voix), the Fonds de recherche du Québec – Nature et technologies (2015-NC-181280, for authors Bouserhal, Falk and Voix), the Natural Sciences and Engineering Research Council of Canada (402126-2012, 402237-2011, for Voix and Falk), the Erasmus Mundus student exchange program in Auditory Neuroscience (for Bouserhal, MacDonald and Voix), and the NSERC-EERS Industrial Research Chair in In-Ear Technologies (for Bouserhal and Voix).

References


