Modeling Speech Production in Noise for the Assessment of Vocal Effort for Use with Communication Headsets

Rachel E. Bouserhal
École de technologie supérieure, Université de Québec, Montréal, Canada.
Centre for Interdisciplinary Research in Music Media and Technology, Montréal, Canada.

Jérémie Voix*
École de technologie supérieure, Université de Québec, Montréal, Canada.
Centre for Interdisciplinary Research in Music Media and Technology, Montréal, Canada.

Tiago H. Falk
Institut national de la recherche scientifique, Centre EMT, Montréal, Canada.
Centre for Interdisciplinary Research in Music Media and Technology, Montréal, Canada.

Summary
A Radio Acoustical Virtual Environment (RAVE) is being developed to address issues occurring when communicating in noise while wearing Hearing Protection Devices (HPD). RAVE mimics a natural acoustical environment by transmitting the speaker's voice signal only to receivers within a given radius, the distance of which is calculated by considering the speaker's vocal effort and the level of background noise. To create a genuine RAVE, it is necessary to understand and model the speech production process in noise while wearing HPDs. Qualitative open-ear and occluded-ear models of the vocal effort as function of background noise level, exist. However, few take into account the effect of communication distance on the speech production process and none do so for the occluded-ear. To complement these models, quantitative data is used to generate quantitative open-ear and occluded-ear models, representing the relationship between vocal effort, communication distance, background noise level and type of HPD. These models can later be implemented within radio-communication headsets used in the proposed RAVE. Speech production models for occluded-ear accounting for the intended communication distance are presented in qualitative terms.

PACS no. 43.70.+i, 43.72.+q

1. Introduction

Using radio communication in noisy environments is a practical and affordable solution allowing communication between people with Hearing Protection Devices (HPD). Traditionally, one of its weaknesses lies in the lack of designating receivers: all those carrying a personal radio (walkie-talkie, etc.) are subjected to the broadcasted signal regardless of whether or not they are the intended listeners. Receiving irrelevant communication is annoying and contributes to the daily accumulated noise dose [1]. A new concept of a "Radio-Acoustical Virtual Environment" (RAVE) is being developed [2]. RAVE intends to mimic a natural acoustical environment by transmitting a communication signal only to people within a specific spatial range. This range is defined as the intended communication distance of the speaker.

Speakers with normal hearing adjust their vocal effort in the presence of noise [3], when trying to communicate at a distance [4] and to express emotion [5]. These adjustments still occur when wearing HPDs, however, they are altered as a function of the type of HPD worn [6, 7]. These changes in vocal effort as a function of the background noise and the type of HPD have been studied and modelled [8]. Interestingly, none of the studies include the effect of the intended communication distance to the model.

This paper presents a review of the current known work of the speech production process in noise both for the open-ear and the occluded-ear condition in Sections 2 and 3 respectively. We also propose a new model that includes the effect of communication distance to the speech production process in noise with

* Corresponding author: jeremie.voix@etsmtl.ca

Paper presented at
Euronoise 2015, the 10th European Congress and Exposition on Noise Control Engineering
Maastrich, The Netherlands, May 31-June 3, 2015)

HPDs in Section 4. Finally, the conclusions are presented in Section 5.

2. Open-Ear Condition

Naturally, speakers raise their voice when speaking in noise. This is called the Lombard Effect [3]. The Lombard effect has been well studied for the open-ear condition. Studies have shown that speakers raise the level of the their voice by 1-6 dB for every 10 dB of noise increase [9]. Multiple studies have found that Lombard speech, i.e. speech produced in noise, increases the speaker's fundamental frequency, f_0, [10, 11] by 0.6-2.5 semitones [12]. A gender-dependant increase in the spectral center of gravity also occurs during Lomabrd speech [6].

In quiet conditions, in turn, speakers raise their vocal effort to reach farther distances. As the communication distance doubles speakers raise their vocal level between 1.3-6 dB [13, 14, 15]. A study done by Zahorik et. al showed that speakers adjust their vocal effort according to their environment as well as the communication distance [15]. The speakers' f_0 as well as first formant, $F1$, also increase as a function of distance. As the vocal intensity increases, f_0 increases by 5 Hz/dB while $F1$ increases by 3.5 Hz/dB [16]. The changes in f_0, Δf_0, caused by increase in communication distance, and thus vocal intensity, was studied to be unique and telling of an increase in effort as a consequence of the increase in distance [4].

It is evident from previous studies that adjustments in the vocal effort as a consequence of either increase in communication distance or the presence of noise varies from speaker to speaker but follows the same trend across speakers. Vocal intensity, and changes in the speaker's f_0 are good indicators of raising the vocal effort. Zahorik et al. suggested that speakers adequately try to match the degradation in their vocal intensity due to propagation loss over distances [15]. Let us consider the model presented in [14] for the vocal power level as a function of the distance. In this case, we choose the model created for the speech produced in an anechoic room. This is because it eliminates any corrections from reverberation and it is the model that best fits data collected from other studies. The model is as follows:

$$L_w = 59.54 + 2.96 \times \log_2 \left(\frac{d}{1.5} \right) \quad (1)$$

where L_w is the speech power level in decibels (dB) and d is the communication distance in meters. As a function of distance, the vocal power level can be graphed as shown in Fig. 1. Combining the model in Eq. 1 and what we know about communication in noise we can create a model that incorporates the presence of noise as a correction factor to the model. This will lead to a model that relates the vocal effort to the level of background noise and the intended communication distance. If we consider the presence of noise to be anything above 60 dB(SPL) for the and average that a speaker's level will increase by 3 dB for every 10 dB of noise, then the modified model becomes:

$$L_w = 59.54 + 2.96 \times \log_2 \left(\frac{d}{1.5} \right) + n \times [10 + 0.3 \times (N - 60)] \quad (2)$$

where n is 0 in quiet and 1 if the noise is greater than 60 dB and N represents the level of background noise in dB(SPL). The addition of the 10 dB accounts for an initial increase at the onset of noise that can be estimated from [6]. For example, the vocal power of a speaker exposed to 70 dB of noise can be compared to that of a speaker in quiet as shown in Fig. 2. From Eq. 1, a speaker in quiet trying to reach a distance of 50 m will speak at an estimated power level, L_w, of 74.5 dB while, from Eq. 2, a speaker in 70 dB noise trying to reach the same communication distance will speak at 87.5 dB. It is important to keep in mind that the model presented in equation Eq. 2 has not been validated but merely a prediction based on the already available data from previous studies. In the next section we review the effects on wearing hearing protection devices on the speech production process.

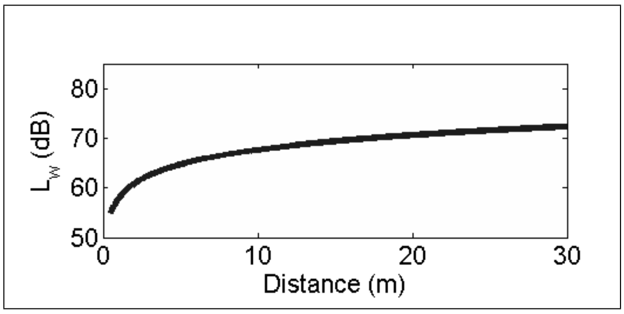

Figure 1. Vocal power level as a function of communication distance as presented in [14].

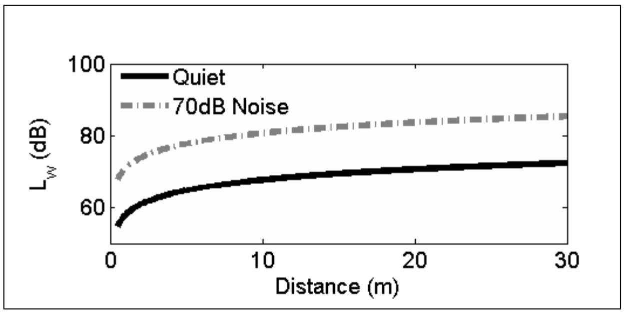

Figure 2. Comparing vocal power level of a speaker in quiet and a speaker in noise as a function of communication distance.

3. Occluded-Ear Condition

Blocking the ear canal path causes a resonance of the bone conducted vibrations caused by speech, causing speakers to hear an amplified 'boomy' version of their voice as they speak. This phenomenon is called the "occlusion effect" [17]. The contributions of the occlusion effect on changes in speech production while wearing HPDs is arguable. In fact, it's one's perception of his/her own voice that greatly affects the speech production process in noise [6]. A speaker's perception of their own voice level compared to the level of noise is the driving factor in the speech production process. Studies have shown that speakers wearing HPDs do not react to increase in noise levels as much as speakers not wearing HPDs. Tufts et al. report a 4-11 dB decrease in the level of speech produced in noise while wearing earplugs compared to speech produced in noise without HPDs. In the presence of 60 dB(SPL) of noise, while wearing earplugs, speakers did not increase their vocal effort from the quiet condition. Also, overall speech level increased by only 5 dB even though the noise increased 40 dB [6]. In other words, while wearing HPDs speakers adjust their vocal effort by only 1.25 dB for every 10 dB increase in noise. In quiet, however, speakers wearing earplugs did not significantly alter their overall speech level [6, 18] from their open-ear level. None of the studies performed on the occluded ear looked at the effects of the communication distance.

If we assume that the model from [14] presented in Eq. 1 still holds for speech production as a function of communication distance and we treat the use of HPDs as a correction factor just as we did in Eq. 2 then the model becomes:

$$L_w = 59.54 + 2.96 \times \log_2\left(\frac{d}{1.5}\right) + n \times 0.125 \times (N - 60) \qquad (3)$$

where again n is 0 in quiet and 1 if the noise is greater than 60 dB and N represents the level of background noise. The three conditions, a speaker in quiet with open ears, a speaker in noise with open ears and a speaker in noise wearing HPDs, are compared in Fig. 3. This model would imply two assumptions:

1. In noise wearing HPDs does not greatly affect the speech production process as a function of the communication distance from the open-ear condition.
2. In quiet wearing HPDs would not affect the speech production process as a function of distance.

Based on the studies of speech production in noise while wearing HPDs the first assumption seems reasonable. However, intuitively, wearing HPDs might still alter the speech production process as a function of the communication distance, making assumption 2 invalid. The effects of communication distance and wearing HPDs in noise on the speech production process must be better studied. In the next section we present an experimental protocol to model the speech production process while wearing HPDs as a function of the background noise level, the speaker's vocal effort and the intended communication distance.

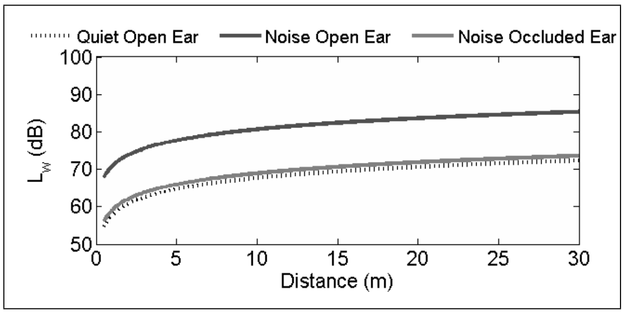

Figure 3. Comparing vocal power level of a speaker in quiet, a speaker in noise, and a speaker in noise wearing HPDs as a function of communication distance.

4. Proposed Experimental Protocol

To model the speech production process while wearing HPDs as a function of the background noise as well as the intended communication distance, an experimental protocol must be designed. Normal hearing individuals will be recruited to perform an instruction task. Each participant will be equipped with the intra-aural communication earpiece shown in Fig. 4. This communication earpiece is chosen for several reasons:

1. It is intra-aural, so it can be fitted into a participant's ear using different tips (roll-down foam plug, rounded flanged tips, malleable silicon wax, custom molded earpiece) and, thus, causing different levels of the occlusion effect.
2. It contains a microphone and miniature loudspeaker inside the ear as well as a microphone outside the ear.
3. It is the earpiece used for the radio-acoustical environment described in Section 1.

Having a miniature loudspeaker inside the ear allows us to play noise inside the ear directly. This leaves the speech signal captured with the outer-ear microphone free of noise and easier to process. The in-ear microphone can capture a noisy speech signal from inside the ear. This allows us to look at the difference in speech level between the outside and the inside of the ear and, in quiet, to measure the occlusion effect.

4.1. Experiment

Participants will be asked to give instructions to a listener in a corridor at 5 different communication distances: 0.3 m, 5 m, 10 m, 15 m, and 20 m. These distances were chosen to cover a wide range of distances and vocal efforts. Since the participants will be wearing HPDs the effects of reverberation can be ignored. At each distance, the speaker will be asked

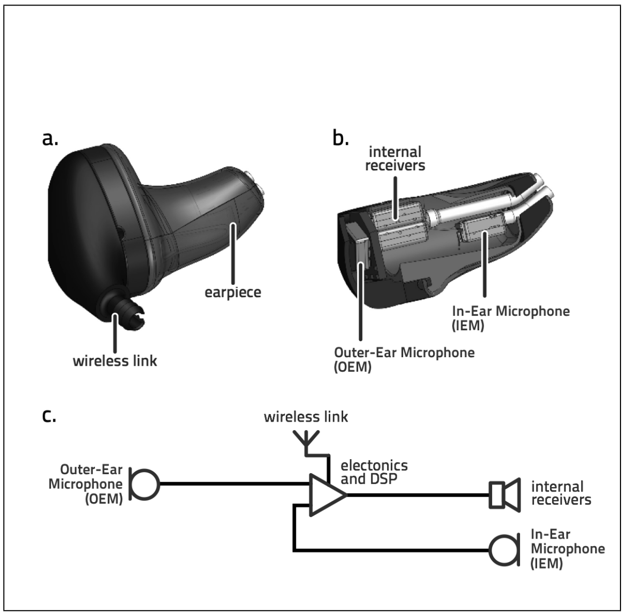

Figure 4. Intra-aural communication earpiece (a), its electroacoustical components (b), and equivalent schematic (c).

to instruct the listener to show him/her a color and a digit, 20 different times. The speaker will have 4 different colors (Red, Green, Blue, Yellow) to choose from and 10 different digits (0-9). The speaker can choose any combination he/she desires and can even repeat combinations. This is done so that the speech is natural and not read, mimicking a realistic situation. This procedure will be repeated for 5 different conditions: in quiet and in pink noise ranging from 60 dB to 90 dB at increments of 10 dB. Since the noise will be played directly inside the ear, only residual noise (after the passive attenuation of the plug) will be played. Once the participant is fitted with the earpiece the transfer function of the earpiece will be measured by playing white noise at a high level (∼85dB(SPL)) using a loudspeaker outside the ear which will be recorded using both the OEM and IEM. After the transfer function is found, the noise is filtered and played inside the ear. During the recordings the level of the speech, as well as the speaker's fundamental frequency, f_0, will be recorded at all conditions. The data is then collected from all the participants and a model will be found to fit it. This will give a relationship between vocal effort, background noise level and intended communication distance while wearing HPDs.

5. Conclusions

Communication is a key part of any workplace. Unfortunately, the use of currently available HPDs tends to affect communication. Modelling speech production while wearing HPDs as a function of the noise level and the intended communication distance can aid in alleviating the communication problem for personal radio systems. In this paper, we review the existing models of speech production in quiet, as a function of distance and in noise with and without the use of hearing protection devices. We also propose an experimental procedure to model speech production in noise while wearing HPDs to include the effects of the communication distance, which is currently not found in the literature.

Acknowledgement

This work was made possible via funding from the Centre for Interdisciplinary Research in Music Media and Technology, the Natural Sciences and Engineering Research Council of Canada, and the Sonomax-ETS Industrial Research Chair in In-Ear Technologies.

References

[1] K. Mazur and J. Voix, "A case-study on the continuous use of an in-ear dosimetric device," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 3274–3274, 2013.

[2] R.E. Bou Serhal, T.H. Falk, and J. Voix, "Integration of a distance sensitive wireless communication protocol to hearing protectors equipped with in-ear microphones.," in *Proceedings of Meetings on Acoustics*. Acoustical Society of America, 2013, vol. 19, p. 040013.

[3] J.C. Junqua, S. Fincke, and K. Field, "The Lombard effect: a reflex to better communicate with others in noise," *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, pp. 2083–2086 vol.4, 1999.

[4] T. Fux, G. Feng, and V. Zimpfer, "Talker-to-listener distance effects on the variations of the intensity and the fundamental frequency of speech," *Cognition*, pp. 4964–4967, 2011.

[5] M. Schröder, "Emotional speech synthesis: A review," *Source*, vol. 1, pp. 2–5, 2001.

[6] J.B. Tufts and T. Frank, "Speech production in noise with and without hearing protection," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, pp. 1069, 2003.

[7] J.G. Casali, M.J. Horylev, and J.F. Grenell, "A pilot study on the effects of hearing protection and ambient noise characteristics on intensity of uttered speech," *Trends in Ergonomics/Human Factors IV, edited by SS Asfour, Elsevier Science Pub.,(North-Holland)*, pp. 303–310, 1987.

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

[9] H. Lane and B. Tranel, "The lombard sign and the role of hearing in speech," *Journal of Speech, Language, and Hearing Research*, vol. 14, no. 4, pp. 677–709, 1971.

[10] J.C. Junqua, "The Lombard reflex and its role on human listeners and automatic speech recognizers.," *The Journal of the Acoustical Society of America*, vol. 93, no. 1, pp. 510–24, Jan. 1993.

[11] M. Garnier and N. Henrich, "Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?," *Computer Speech and Language*, vol. 28, no. 2, pp. 580–597, 2014.

[12] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise.," *The Journal of the Acoustical Society of America*, vol. 124, no. November 2008, pp. 3261–3275, 2008.

[13] H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women, and children.," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, pp. 3438–3451, 2000.

[14] D. Pelegrín-García, B. Smits, J. Brunskog, and C. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments.," *The Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 1981–90, Apr. 2011.

[15] P. Zahorik and J. W. Kelly, "Accurate vocal compensation for sound intensity loss with increasing distance in natural environments.," *The Journal of the Acoustical Society of America*, vol. 122, no. 5, pp. EL143–50, Nov. 2007.

[16] J. S. Liénard and M. G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels.," *The Journal of the Acoustical Society of America*, vol. 106, no. 1, pp. 411–22, July 1999.

[17] A. Bernier and J. Voix, "An active hearing protection device for musicians," in *Proceedings of Meetings on Acoustics*. Acoustical Society of America, 2013, vol. 19, p. 040015.

[18] R. Navarro, "Effects of ear canal occlusion and masking on the perception of voice," *Perceptual and motor skills*, vol. 82, no. 1, pp. 199–208, 1996.