Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Noise
Session 1pNSa: Advanced Hearing Protection and Methods of Measurement II

1pNSa5. Integration of a distance sensitive wireless communication protocol to hearing protectors equipped with in-ear microphones.

Rachel E. Bou Serhal, Tiago H. Falk and Jérémie Voix*

***Corresponding author's address: Ecole de Technologie Superieure, Universite du Quebec, 1100 rue Notre-Dame Ouest, Montréal, H3C 1K3, QC, Canada, jeremie.voix@etsmtl.ca**

Using radio communication in noisy environments is a practical and affordable solution allowing communication between workers wearing Hearing Protection Devices (HPD). However, typical radio communication systems have two main limitations when used in noisy environments: first, the background noise is disturbing the voice signal picked-up and transmitted, and second, that voice signal goes to all listeners on the same radio channel regardless of their physical proximity. A new concept of a so-called " Radio Acoustical Virtual Environment" (RAVE) addressing these two issues is presented. Using an intra-aural instantly custom molded HPD equipped with both an in-ear microphone and miniature loudspeaker, undisturbed speech is captured from inside the ear canal and transmitted over the wireless radio to the remote listener. The transmitted signal will only be received by listeners within a given spatial range, such range depending on the user's vocal effort and background noise level. This paper demonstrates the technological challenges to overcome and the methodology involved in the implementation of RAVE.

Published by the Acoustical Society of America through the American Institute of Physics

Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Hearing protection has been widely discussed and researched. Several Hearing Protection Devices (HPD) have been developed to protect workers' hearing from noisy environments. HPDs come in several different shapes and sizes and can be made from a variety of materials. The two main types of HPDs are intra-aural i.e. earplugs, and supra-aural i.e. earmuffs (Berger, 2003). Depending on the type of HPD worn, as well as, the spectrum of the noise and the wearer's hearing ability, wearing HPDs could limit communication (Berger, 2003). Good communication in a work environment is vital. Unfortunately, workers must make compromises between protecting their hearing and maintaining good communication. There are several different ways that are used to communicate in noise, one could:

a) Remove the HPD: get closer to a listener and adjust vocal effort to communicate

b) Use passively filtered HPD: flat attenuation HPDs could be beneficial for speech communication as they do not attenuate high frequencies as much as other HPDs.

c) Use a hand-held radio device: use of a walkie-talkie allows for distance communication with multiple people while remaining stationary (with HPDs or without).

d) Use of a communication headset: usually an earmuff with a miniature loudspeaker and an external boom microphone. The voice picked up by the boom microphone is transmitted through either a wired or wireless network to a remote listener.

Although these techniques are feasible and commonly utilized, their performance is unsatisfactory. Removing an HPD to communicate is counter-productive, potentially harmful to the worker's hearing and requires the workers to be in close proximity. Passively filtered HPDs do not require the user to remove the HPD for communication, but the speaker must still be in close range for the listeners to understand. As a result of the excessive levels of background noise, the persons communicating will naturally increase their vocal effort to compensate for such conditions in comparison to a quite environment. Using a hand-held radio overcomes the problem of proximity but still requires the removal of the HPD. The best current alternative is the use of HPDs that are equipped with an external microphone called a boom microphone and connected to a personal radio system. Although a step in the right direction, these headsets still present the following inconvenience: the external microphone will not only pick up the user's voice but background noise as well, which dramatically affects intelligibility.

Another issue associated with using any kind of radio transmitter, is that it does not distinguish a receiver and all communication is sent to everyone on the same radio channel. Thus, the users' radio is often flooded with irrelevant conversation that could be annoying and somewhat loud and thus contributing to the noise dose. Clearly there is a need for a device that provides good noise attenuation as well as good communication without compromising the performance of one or the other.

Proposed Approach

We propose a new concept called "Radio Acoustical Virtual Environment" (RAVE) in which workers in noisy environments can achieve intelligible communication without hindering their hearing protection. RAVE uses an advanced intra-aural instantly custom molded HPD, shown in Figure 1, equipped with an In-Ear Microphone (IEM), a miniature loudspeaker, a Digital Signal Processor (DSP), an Outer-Ear Microphone (OEM) and Wireless Radio (WR) capabilities. Such a device can capture a somewhat undisturbed speech signal from inside the ear (referred to as IEM speech). Because the signal captured originates from bone conducted vibrations, it lacks higher frequencies. Thus, the IEM signal must first be enhanced in its high frequency content. Once enhanced, the IEM signal is coded and sent to an appropriate radius of listeners based on the acoustical features of the produced speech and the level of background noise.

This paper introduces the design of RAVE and the methodology involved in realizing such a protocol. The next section discusses different techniques available for the enhancement of the IEM speech signal followed by the concept of vocal effort coding. Then we discuss the envisioned experimental work required to obtain RAVE and the final section presents our conclusions.

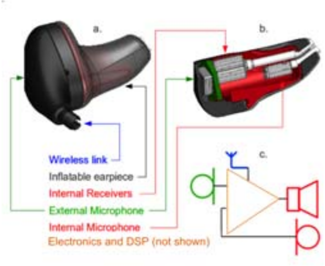


FIGURE 1: Overview of digital custom earpiece (a), its electroacoustical components (b), and equivalent schematic (c).

ENHANCEMENT OF THE IEM SPEECH

When speech is captured conventionally (with a boom microphone), to be sent over a radio network in a noisy environment, it is disturbed and contains the noise picked up by the exposed microphone, even when using a directional microphone. On the other hand, capturing speech from inside the protected ear allows for the transmission of a less-disturbed speech signal that will not require extra de-noising usually achieved by the electronics within the radio. When the ear canal is blocked by an in-ear device, there is a regeneration of the speech inside the ear canal and one experiences what is called the occlusion effect (Berger, 2003). The occlusion effect allows for the capturing of speech inside the ear, which is useful in noisy environments. Because of cranial bone conduction, this signal is "boomy", containing most of its energy in the lower frequencies while missing important high frequency content (Bernier and Voix, 2010). The difference between the frequency content of the IEM speech and the OEM speech (referred to as REF) of the utterance /u/, for a male speaker, is demonstrated in Figure 2. In Figure 2, we notice that above 1.8 kHz, the IEM signal is missing important high frequency content. As a consequence of the IEM signal's limited bandwidth, fricative consonants such as /s/ and /f/, and nasals such as /n/ and /m/ are unintelligible. The IEM signal is thus perceived as having lower quality and intelligibility than "free air speech", or speech that is recorded near the mouth. To solve this, the IEM signal could be expanded using Bandwidth Extension (BWE) of the speech signal as will be reviewed in the next section.

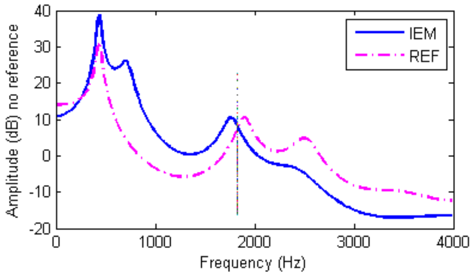


FIGURE 2: IEM vs. REF spectral envelopes of the utterance /u/ from the word 'canoe', showing the increased low frequency content and the missing high frequency content.

Bandwidth Extension (BWE)

In this section, we introduce some BWE techniques commonly utilized in the field of speech signal processing (a good reference of common speech terms can be found in (O'shaughnessy, 2000)). Many different BWE techniques exist, and the proper choice depends on the desired results and available resources. BWE can range from spectral estimation and expansion through excitation signal extension, to Vector Quantization (VQ) and codebook mapping. Iser et al. give a good review of the basics of such techniques (Iser, Bernd *et al.*, 2008). In the past, the need for BWE arose because of the limited bandwidth of the telephone network. The narrow bandwidth of a telephone is about 3.5 kHz leaving some significant parts of human speech unrepresented. In this context, wideband signals refer to signals that can represent the entire vocal range while narrowband signals can only represent a limited part of the vocal range. With access to an IEM and an OEM, BWE can be used for our purposes by treating the IEM signal as the narrowband signal and the free-air speech captured by the OEM as the wideband signal.

One BWE technique is *excitation signal extension*. This technique involves three main procedures: *envelope extraction*, *excitation signal extraction*, and *excitation signal extension* (Iser, Bernd *et al.*, 2008). The envelope extraction technique depends on the Linear Predictive Coding (LPC) analysis of the narrowband signal. The excitation signal extraction and extension could be done using several methods: non-linear characteristics approach, spectral shifting approach and the function generator approach. Another way BWE can be achieved, is by *wideband spectral envelope expansion*. To estimate the wideband spectral envelope, several methods are available, such as neural networks, linear mapping, and codebooks (Iser, Bernd *et al.*, 2008). Statistically based methods also exist, such as the statistical recovery function used by Cheng et al. (Cheng *et al.*, 1994), and Gaussian Mixture Models (GMM) (Park and Kim, 2000). Wideband spectral envelope estimation differs from excitation signal extension in that it requires a training data set. While only the narrowband input is required for excitation signal extension, the estimation of the wideband spectral envelope requires a sufficiently large training data set that contains the desired sampling rate and bandwidth (Iser, Bernd *et al.*, 2008).

With all these available techniques, listed in Figure 3, it is important to assess the resources available to choose a practical and efficient technique with good performance. Some things to consider are the computational complexity and cost of the algorithm, power consumption and whether the algorithm will be speaker dependent or speaker independent. *Excitation signal extension* and *spectral envelope expansion* could be used for speaker independent BWE. Quality may be increased with speaker dependent techniques using spectral envelope expansion at the cost of some practicality. When speaker dependent algorithms are used the user must train the algorithm. Although speaker dependent algorithms may lead to better quality reconstructed speech, they are less robust when compared to speaker independent algorithms. Small variations in speech for instance, caused by a common cold, may lead to undesirable results. This could be palliated by making the algorithm re-trainable. However, this is impractical and may lead users to abandoning the use of the device. It is thus important to evaluate such adverse effects and assure that the BWE algorithm used is practical, efficient, and reliable.

VOCAL EFFORT CODING

In this section we discuss the various vocal modes and their relationship with physical distance between a speaker and a listener. Naturally, human beings adjust their vocal effort to compensate for changes in their environment. One can whisper a confidential message, call out for a meeting or shout out for help. It is important to distinguish "vocal effort" from "vocal level". The latter suggests a change in Sound-Pressure Level (SPL) while vocal effort involves a lot more than just changes in SPL (Traunmüller and Eriksson, 2000). Zhang et al. (2007) classified

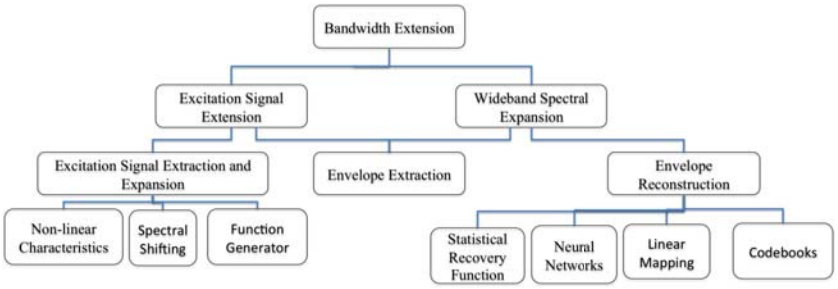

Bandwidth Extension
Excitation Signal Extension
Wideband Spectral Expansion
Excitation Signal Extraction and Expansion
Envelope Extraction
Envelope Reconstruction
Non-linear Characteristics
Spectral Shifting
Function Generator
Statistical Recovery Function
Neural Networks
Linear Mapping
Codebooks

FIGURE 3: Classification of different bandwidth extension techiniques applicable to in-ear microphone signal pickup inside workers' ears.

5 speech modes: (1) whispered, (2) soft, (3) neutral, (4) loud, and (5) shouted. Each of these speech modes is characterized by its deviations from the neutral speaking condition. Many studies have been done to characterize each speech mode as to enhance speaker recognition systems and other applications. In particular, whispered and shouted speech require the most dramatic change in excitation (Zhang and Hansen, 2007) and have thus received a lot of attention. Our interest lies mostly with the shouted speech mode and the changes in acoustical features that occur.
As documented by many, as the vocal effort increases so does the fundamental frequency, *F0*. Another widely accepted change in the formants is the increase of the first formant, *F1* (Liénard and Di Benedetto, 1999) (Elliot, 2000) (Garnier *et al.*, 2008). Liénard and Di Benedetto (1999), however, also claim an increase in the second formant, *F2*, for females but this has not yet been widely accepted. Shouted sentences have increased initial *F0* slope but a decreased final *F0* (Fux *et al.*, 2011) and a decreased spectral slope (Zhang and Hansen, 2007). They are longer in duration which is caused by longer word duration, but have a decreased silence duration (Zhang and Hansen, 2007). Typically, shouted speech is detected based on *F0*, *F1* and the spectral tilt (Nanjo *et al.*, 2009). A summary of these changes can be seen in Table 1.

Traunmüller et al. describe vocal effort as *"the quantity that ordinary speakers vary when they adapt their speech to the demands of an increased or decreased communication distance"* (Traunmüller and Eriksson, 2000). As distance increases so does the vocal effort. In fact, Brungart et al. report that as distance doubles the intensity increases by 8 dB, while Liénard et al. report that *F0* increases at 3.5 Hz/dB (Fux *et al.*, 2011) (Liénard and Di Benedetto, 1999). Distance, however, is not the only time we adjust our vocal effort. When our ability to hear our own voice changes, as a result of background noise for example, our vocal effort changes (Junqua, 1993). This is known as the *Lombard* effect. Although Lombard speech may share some characteristics with shouted speech, it is unique and cannot be treated the same way as shouted speech. Speakers vary their vocal effort based on the spectrotemporal properties of the background noise. In fact, significant differences of adjustments in the presence of white noise and babble noise have been reported (Traunmüller and Eriksson, 2000). A summary of the acoustical changes cause by Lombard speech as found by Junqua (1993) is shown in Table 1 .

When wearing HPDs, the Lombard effect is also a contributing factor in decreased speech intelligibility, from the perspective of both the speaker and the listener. Wearing hearing protection in noise not only affects the way speech is heard, it changes the way speech is produced. At the speaking end, Tufts and Frank (2003) studied the differences in speech acoustics when produced in noise while wearing hearing protection. For people with normal hearing, the level of adjustment in vocal effort as the level of noise increased was less when hearing protection was worn than without. As the level of noise increased from 60 dB to 100 dB SPL, speakers not wearing hearing protection increased their speaking leved by about 40dB, while those wearing hearing protection increased their vocal level by only 3 dB to 15 dB (Tufts and Frank, 2003). At the hearing end, studies by Candido Fernandes (2003) report that in environments with +5 dB Signal-to-Noise Ratio (SNR) and +10 dB SNR , wearing hearing protection decreases the intelligi-

TABLE 1: Summary of acoustical differences between shouted speech and Lombard when compared to neutral speech

Acoustical Feature	Shouted speech	Lombard speech
F0	Increased frequency	Increased frequency (more dominant in male speakers)
F1	Increased frequency	Increased frequency (more dominant in female speakers)
F2	Increased frequency (females only)	Increased frequency (females only)
Sentence Duration	Increased duration	Increased duration
SPL	Increased level	Slightly increased level

bility of speech. However, at -5 dB and -10 dB SNR, wearing hearing protection increases speech intelligibility by up to 10% (Candido Fernandes, 2003). It is also useful to note that studies by Giguère and Dajani (2009) report that persons wearing HPDs prefer an SNR of 13.5 dB when listening to speech in noise. This can be utilized in the experimental procedures as discussed in the next section.

The studied changes that characterize the different vocal efforts could be utilized in our application. However, as can be concluded from the preceding discussion, to correctly make a link between vocal effort and intended communication distance while wearing HPDs in noise, the effects of the Lombard effect along side the occlusion effect must be considered.

ENVISIONED EXPERIMENTAL APPROACH

In order to reach our goal to provide workers with intelligible communication without compromising their hearing protection through our proposed "Radio Acoustical Virtual Environment", we must answer the following research questions:

(1) *What is the relationship between vocal effort and communication distance in noise with and without HPDs? What are the changes in relevant acoustical features between the case where HPDs are worn and when they are not?*

(2) *To what extent does post processing of the IEM speech enhance intelligibility? Are there ways to train a speaker in noise to further enhance intelligibility?*

To answer the questions above, several tests must be performed on a large control group of human subjects for data collection. Below, is a list of tests that we anticipate could be helpful in advancing our research. For the following tests the speaker will be in a quiet environment but exposed to background noise through the in-ear device and asked to communicate at different levels of this background noise. This will leave the OEM free of noise enabling it to pick up a clean speech signal. Since the transfer function between the IEM and REF will always be the same, the respective IEM level can be figured out from the found REF signal. The control group for theses tests will consist of normal hearing people and have an equal number of females and males. There are two main types of tests that we envision carrying out:

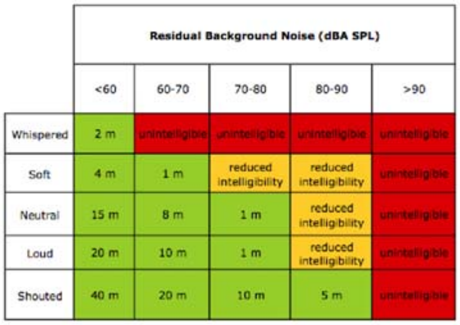
(1) The first test will involve two normal hearing human subjects. One subject will be assigned as the speaker the other subject will be assigned as the listener. The speaker will be asked to relay a set of actions to the listener, for example, *"Pick up the hammer"*. The listener will have to correctly perform the requested action. Once the listener is successful the speech from the speaker is saved for analysis and annotated with the distance between speakers

and level of background noise. The same test will be repeated for multiple speakers and listeners, with and without hearing protection, in silence and different levels of background noise.

(2) The second test will utilize a moving cardboard target (equipped with a measurement microphone) that, again gradually moves farther away. At a fixed background noise level the speaker will be asked to speak so that her/his speech is intelligible to the moving target. Using objective speech intelligibility measures such as the one presented by Giguère *et al.* (2009), the speaker will receive a cue of whether or not the information was understood by analysis of the measurement microphone signal. The speaker's own speech will be played back to them and the speaker will be asked to adjust their speech to make it more intelligible.

Each test will contribute to a certain aspect of our research. The first test will help us collect the necessary data to map vocal effort and variations in relevant acoustical features of speech, with and without the use of HPDs, to intended communication distance. This test will also allow us to assess what acoustic features of speech are robust enough to be used in coding the vocal effort. An interesting consideration is the SPL of the speech. Conventionally SPL is not used to characterize different levels of vocal effort because of the unfixed position of the microphone. However, in our application the microphone location is stationary and could be used along with other features to code the vocal effort. The second test could indicate whether training a speaker how to speak in noise could further increase the intelligibility. It could also provide significant data on how much facial cues and gestures from a human listener are useful to the speaker and the listener. For example, Erber (1969) reports that lip reading in -10 dB SNR can increase speech intelligibility in noise by about 60%.

At the conclusion of these tests we envision producing a relationship as portrayed in Figure 4. The green blocks represent distances where speech is intelligible for the given vocal effort and background noise level. The yellow blocks represent areas of reduced intelligibility or areas where intelligibility is achieved only with reinforcement from facial cues or gestures. Red blocks represent areas where speech is unintelligible. Note, the numbers in this table are strictly for illustrative purposes and do not yet come from research data. Once this table is compiled, the vocal effort of the speaker may be coded and sent to an appropriate radius of intended listeners through an *ad-hoc* radio system such as cognitive radios (Li *et al.*, 2011). Figure 5 demonstrates the anticipated performance of RAVE. If a worker is speaking at 70 dBA SPL in a quiet environment the radio signal will be transmitted to anyone within a 20 m radius. As the level of noise increases and the vocal effort of the speaker remains constant the transmitting distance will decrease. Therefore, in an extremely noisy environment the transmitting distance of the radio will only be 5 m to compensate for such phenomena as the Lombard effect.



	Residual Background Noise (dBA SPL)				
	<60	60-70	70-80	80-90	>90
Whispered	2 m	unintelligible	unintelligible	unintelligible	unintelligible
Soft	4 m	1 m	reduced intelligibility	reduced intelligibility	unintelligible
Neutral	15 m	8 m	1 m	reduced intelligibility	unintelligible
Loud	20 m	10 m	1 m	reduced intelligibility	unintelligible
Shouted	40 m	20 m	10 m	5 m	unintelligible

FIGURE 4: Illustrative table of relationship between vocal effort and communication distance in the presence of background noise while wearing HPDs.

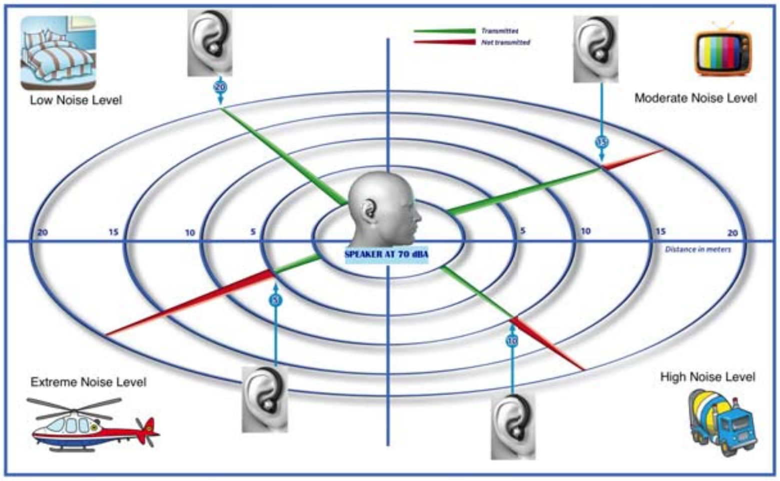


FIGURE 5: Illustration of functionality of RAVE. The green and red lines represent the areas where the signal is transmitted and not transmitted, respectively.

CONCLUSIONS

Good hearing protection is currently achieved at the cost of decreased communication while good communication is achieved at the cost of jeopardizing good hearing protection. Providing workers with satisfactory hearing protection and communication is still difficult and requires the compromise of one or the other. In this paper, we propose a new distance sensitive protocol that provides intelligible speech to workers wearing hearing protection. Using changes in acoustical features of speech the vocal effort will be coded and the speech signal will be sent in a way that mimics a natural acoustical environment. The "Radio Acoustical Virtual Environment" discussed will allow workers to communicate without the need to remove their HPDs and without having to move closer to their listener. Undisturbed speech from inside the ear canal will be captured and transmitted over wireless radio to the remote listener. The transmitted signal will only be received by listeners within a given spatial range, this range depending on the user's vocal effort and background noise level. Providing workers with such a device will enhance their work experience and potentially promote the use of HPDs in noisy work environments.

ACKNOWLEDGMENTS

The authors would like to thank Sonomax Technologies Inc. and its *Sonomax-ÉTS Industrial Research Chair in In-Ear Technologies* (CRITIAS) for their financial support. The authors also acknowledge the financial support from MITACS- Accelerate research internship program.

REFERENCES

Berger, E. (**2003**). *The noise manual* (Aiha).

Bernier, A. and Voix, J. (**2010**). "Signal characterization of occluded ear versus free-air voice pickup on human subjects", Canadian Acoustics **Vol. 38**, pp. 78–79.

Candido Fernandes, J. a. (**2003**). "Effects of hearing protector devices on speech intelligibility", Applied Acoustics **64**, 581–590.

Cheng, Y., O'Shaughnessy, D., and Mermelstein, P. (**1994**). "Statistical recovery of wideband speech from narrowband speech", Speech and Audio Processing, IEEE Transactions on **2**, 544–548.

Elliot, J. (**2000**). "Comparing the Acoustic Properties of Normal and Shouted Speech: A Study in Forensic Phonetics", in *8th Aus. Int. Conf. Speech Sci. & Tech*, 154–159.

Erber, N. (**1969**). "Interaction of audition and vision in the recognition of oral speech stimuli", Journal of Speech, Language and Hearing Research **12**, 423.

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

Garnier, M., Wolfe, J., Henrich, N., and Smith, J. (**2008**). "Interrelationship between vocal effort and vocal tract acoustics : a pilot study Music Acoustics Group , School of Physics , University of New South Wales , Sydney , Australia Département Language and Cognition , GIPSA-Lab , Grenoble , France", **2**, 3–6.

Giguère, C. and Dajani, H. R. (**2009**). "Noise exposure from communication headsets: the effects of external noise, device attenuation and effective listening signal-to-noise ratio", INTER-NOISE . . . .

Giguère, C., Laroche, C., Vaillancourt, V., and Soli, S. D. (**2009**). "A predictive model of speech intelligibility in noise for normal and hearing-impaired listeners wearing hearing protectors", INTER-NOISE . . . .

Iser, Bernd, Schmidt, G., and Minker, W. (**2008**). *Bandwidth Extension of Speech Signals* (ISBN 978-0-387-68898-5).

Junqua, J. C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers.", The Journal of the Acoustical Society of America **93**, 510–24.

Li, J., Zhou, Y., Lamont, L., and Gagnon, F. (**2011**). "A Novel Routing Algorithm in Cognitive Radio Ad Hoc Networks", in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 1–5 (IEEE).

Liénard, J. S. and Di Benedetto, M. G. (**1999**). "Effect of vocal effort on spectral properties of vowels.", The Journal of the Acoustical Society of America **106**, 411–22.

Nanjo, H., Nishiura, T., and Kawano, H. (**2009**). "Acoustic-Based Security System: Towards Robust Understanding of Emergency Shout", 2009 Fifth International Conference on Information Assurance and Security **1**, 725–728.

O'shaughnessy, D. (**2000**). *Speech communications: human and machine* (Universities press).

Park, K.-y. and Kim, H. S. (**2000**). "Narrowband to wideband conversion of speech using GMM based transformation", Spectrum 1843–1846.

Traunmüller, H. and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children.", The Journal of the Acoustical Society of America **107**, 3438–51.

Tufts, J. B. and Frank, T. (**2003**). "Speech production in noise with and without hearing protection", The Journal of the Acoustical Society of America **114**, 1069.

Valin, J.-M. (**2002**). "Extension spectrale d'un signal de parole de la bande téléphonique à la bande AM", Ph.D. thesis, Sherbrooke University.

Zhang, C. and Hansen, J. H. L. (**2007**). "Analysis and Classification of Speech Mode : Whispered through Shouted", in *Proceedings of the Interspeech*, 2289– 2292.