International Journal of Industrial Ergonomics 43 (2013) 526–535

Contents lists available at SciVerse ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Improving speech understanding in communication headsets: Simulation of adaptive subband processing for speech in noise

Eric R. Bernstein [a,*], Anthony J. Brammer [a,b], Gongqiang Yu [a]

[a] Ergonomics Technology Center, University of Connecticut Health Center, 263 Farmington Av., M.C. 2017, Farmington, CT 06030-2017, USA
[b] Envir-O-Health Solutions, 4792 Massey Lane, Ottawa, ON, Canada KIJ 8W9

ARTICLE INFO

Article history:
Received 30 April 2012
Received in revised form
6 October 2012
Accepted 8 October 2012
Available online 10 November 2012

Keywords:
Speech understanding optimization simulation
Delayless subband active noise reduction with speech signal-to-noise ratio improvement
Communication headset simulation

ABSTRACT

Speech communication headsets are necessary for many high-noise environments to maintain interaction between individuals and facilitate safe working conditions. However, current hearing protection devices intended to protect hearing health can impede speech communication or expose persons to sound pressure levels (SPLs) that could lead to excessive noise exposure if a communication signal is presented improperly. This paper explores an adaptive subband communication algorithm, based on a delayless subband active noise reduction architecture, intended to adjust the communication channel gain to provide an appropriate speech signal power in relation to the instantaneous environmental noise power. The method monitors SPLs underneath the ear cup of a communication headset to provide a target speech signal-to-noise ratio without exceeding safe noise exposure thresholds. A series of computer simulations derived from a real-world communication headset model are used to compare the method developed with a traditional passive attenuation headset and a commercial active noise reduction design. The simulations demonstrate the ability of the adaptive subband communication algorithm to adjust automatically the speech signal gain for improved intelligibility while maintaining healthy noise exposure levels.

Relevance to industry: The electro-acoustic performance of an active speech communication headset is explored by simulation. The concept integrates a subband active noise control algorithm with an adaptive gain control structure to improve speech intelligibility in a noisy environment. The concept automatically selects appropriate communication channel gain levels without exceeding hearing damage thresholds or requiring user input, and is directly applicable to a practical device.

1. Introduction

Many occupational environments expose workers to noise levels that are high enough to cause noise-induced hearing loss with prolonged exposure. The National Institute of Occupational Safety and Health found that over 9 million U.S. workers in a variety of aviation, manufacturing, transportation, military, and other job categories are subject to occupational noise exposure levels exceeding 85 dBA (NIOSH, 1998). For these workers, hearing protection is necessary to prevent damage and maintain hearing acuity later in life. In many of these occupations, particularly in aviation and military applications, a hearing protector with an integrated speech communication channel, or communication headset, is needed to maintain communication with coworkers. Cardosi et al. (1998) found that three of the seven non-mechanical related near mid-air collisions reported in the U.S. Aviation Safety and Reporting System from 1991 to 1996 were the result of misunderstanding of aircraft call signs (e.g., mistaking 'United Airline flight 123' for 'American Airline flight 123'). Pilot and controller workload, as well as fatigue were cited as major factors contributing to these communication errors. LaTourette et al. (2003) found that emergency response personnel, particularly fire fighters, face significant problems when trying to communicate with coworkers over the background noise generated by engines, sirens, power tools, and other sources at incident scenes. In these situations, hearing protection is required that not only limits noise

Abbreviations: ANR, active noise reduction; DFT, discrete Fourier transform; DSP, digital signal processor; FFT, fast Fourier transform; FIR, finite impulse response; FXLMS, filtered-X least mean squares; IFFT, inverse fast Fourier transform; LMS, least mean squares; MRT, modified rhyme test; PNR, passive noise reduction; SNR, signal-to-noise ratio; SPL, sound pressure level; STI, speech transmission index; TNR, total noise reduction.

* Corresponding author. 9 Squire Rd., Windsor, CT 06095, USA. Tel.: +1 860 604 4718; fax: +1 860 679 1989.
E-mail addresses: eric.bernstein@uconn.edu (E.R. Bernstein), brammer@uchc.edu (A.J. Brammer), gyu@uchc.edu (G. Yu).

0169-8141/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ergon.2012.10.004

exposure, but also optimizes speech intelligibility to ensure that critical communication is received and understood. More importantly, the optimizations to speech intelligibility should require minimal user interaction so that users can focus on job tasks rather than on adjusting communication channel gain.

While passive hearing protection devices, such as circumaural earmuffs or earplugs, provide good mid-to-high frequency attenuation, they are not suited for every noise environment. Shaw and Thiessen (1962) demonstrated that the attenuation of circumaural ear cups is directly affected by a variety of factors including the stiffness of the cushion between the ear cups and the head, the air volume trapped between the ear cup and the head, and the quality of the seal of the earmuff to the head. As a result, passive earmuffs provide poor attenuation performance at low frequencies. Therefore, an additional control mechanism is required if additional attenuation is needed at these frequencies to provide adequate hearing protection.

1.1. Active noise reduction

Active Noise Reduction (ANR) uses an electronic control system to provide additional noise attenuation at low frequencies. The ANR system uses a miniature loudspeaker within the ear cup to generate an opposite phase signal designed to cancel the environmental noise. By integrating ANR into circumaural ear cups, or earplugs, the effective attenuation range of the device is extended leading to improved hearing protection.

Two fundamentally different ANR control structures have been integrated into previous hearing protection devices. A feedforward ANR hearing protector uses a microphone outside the ear cup to generate the cancellation signal (Brammer et al., 1995). This allows for cancellation of both periodic as well as impulsive environmental sounds, since the incoming noise is detected prior to reaching the point of control at the loudspeaker. A feedback ANR hearing protector, which is used in most commercial ANR hearing protectors, uses a microphone underneath the ear cup to generate the cancellation signal (Nelson and Elliott, 1992; Song et al., 2005). As a result of the intrinsic time delay of the control system, periodic noise is canceled better than transient noise since the latter may pass the point of control before the opposite phase sound can be generated. The primary benefit of the feedback structure is that since the only microphone is located underneath the ear cup, this structure is relatively insensitive to the direction of the environmental noise relative to the user. However, careful microphone placement with a feedforward ANR design can generally overcome the more directional nature of this ANR structure. Therefore, the ANR structures explored in this paper are all feedforward designs due to the expected performance benefits for canceling both periodic and transient low frequency sounds. Combined feedforward and feedback systems exist that attempt to address the limitations of the individual control structures (Ray et al., 2006), but the computational complexities of these algorithms were anticipated to restrict the computational resources needed for communication signal processing.

A block diagram of a typical single input, single output feedforward ANR system using the traditional Filtered-X Least Mean Squares (FXLMS) algorithm can be seen in Fig. 1. This design uses two microphones; a reference microphone outside the ear cup (R) and an error microphone underneath the ear cup (E), with a secondary source loudspeaker underneath the ear cup (S) to generate a cancellation signal. The system works by first detecting environmental noise at the reference microphone. A cancellation signal is immediately generated by the current Finite Impulse Response (FIR) control filter and the signal is output by the secondary source loudspeaker. The ear cup attenuated environmental noise and the secondary noise combine at the error microphone located near the entrance of the ear canal, which monitors the noise cancellation. The error signal and a filtered version of the reference signal are then used as inputs to a Least Mean Squares (LMS) optimization routine that iteratively adjusts the coefficients of the FIR control filter over time (Kuo and Morgan, 1996). At steady state, the LMS-adjusted filter coefficients should minimize the power of the error signal and thus minimize the amount of noise underneath the ear cup. A communication signal can be added to the control signal if speech communication is desired. Since the communication signal is essentially incoherent with the primary noise signal recorded at the reference microphone, it results in minimal perturbation of the ANR control system.

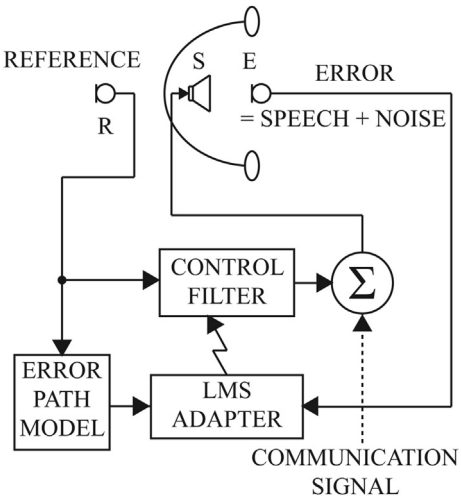

Fig. 1. Block diagram of a typical fullband feedforward FXLMS hearing protector with integrated communication channel.

1.2. Subband ANR

While the FXLMS algorithm is the simplest and most common algorithm for feedforward ANR, it has several disadvantages. The best convergence rates are achieved using a white noise input signal that spreads the signal power equally across all frequencies. Most real world noises have power focused in certain frequencies, such as the rotational frequencies of industrial equipment. The unequal frequency distribution causes longer convergence time and adversely affects the ability of the system to reach an optimal steady state filter. Also, every filter coefficient is normally updated for every input sample. Thus, high order control filters lead to high computational cost and longer processing times to generate a control signal for a given input.

Morgan and Thi (1995) developed the delayless subband adaptive filter architecture to address similar problems in the field of acoustic echo cancellation. Subband processing breaks the frequency range into a series of parallel frequency bands, each with its own adaptive filter. Fig. 2 gives the frequency ranges of each subband filter used in this study. The delayless subband filter uses multirate processing techniques to resample the input signals, which leads to slower update rates for individual filters but also a significant reduction in computational complexity even for high order filters. Several articles have explored similar subband algorithms intended to further reduce computational cost for high order filters (Park et al., 2001; Wu et al., 2008). Additionally, since the subband filters are independent of each other, this architecture is better suited for noise environments with a wide variability of sound pressure level (SPL) over the frequency range. Optimal

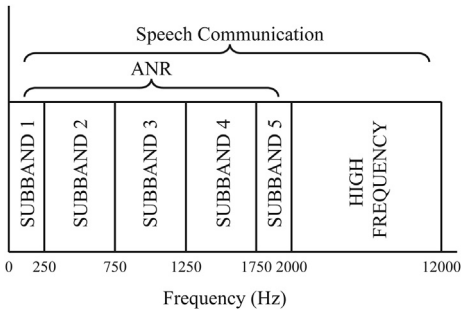

Fig. 2. Frequency division of the ANR control range into subbands compared with the speech communication range.

convergence rates are achieved with a near "white" noise input to each subband, but variations in the power between subbands does not affect the convergence of the total system. While some researchers have explored the computational requirements of implementing subband ANR in modern processors (Qiu et al., 2006), to the authors' knowledge the subband ANR algorithm has never before been applied to ANR hearing protection devices, and the effects of integrating a self-adapting speech communication channel into such a system have not been explored.

1.3. Communication signal processing

Speech intelligibility is governed by a variety of factors. One of the most common methods to improve intelligibility is to increase the speech signal-to-noise ratio (SNR) relating the SPL of the speech sounds with the SPL of the environmental noise. Hearing protection and active control systems attempt to raise the speech SNR of the communication channel by reducing the SPL of the environmental noise. However, if reduction of environmental noise is insufficient, the communication channel SPL must be increased by the user. This poses a problem in environments with varying environmental noise SPLs where a user must regularly adjust communication channel gain to maintain intelligibility. Additionally, since the gain affects the communication SPL over the entire communication signal bandwidth, power is added equally over the total frequency range. If the environmental noise spectrum only contains power in a narrow range, then additional communication power could be added selectively outside this frequency range beyond what is necessary to maintain the desired overall SNR.

Communication signals must also be accounted for when evaluating the noise exposure of the user. In many high noise environments, the user can set the communication channel gain to a level that provides acceptable speech SNR but also exceeds hearing damage thresholds that could eventually lead to noise-induced hearing loss (Crabtree, 2002; Giguère and Dajani, 2009). By automatically controlling gain levels, a reduced gain communication signal could be presented that results in minimal decrease in speech intelligibility but a significant reduction in noise exposure. Reduced sound pressure means users can maintain communication for longer periods of time with decreased risk of hearing damage.

1.4. Communication headset simulation

In this paper, a feedforward subband ANR communication headset algorithm is described and its performance investigated. The primary goal was the development of an adaptive subband communication enhancement algorithm. To this end, it was determined that a computer simulation would facilitate the development goal better than coding the new algorithm directly on a physical implementation. Computer simulation of the subband headset design allows for rapid prototyping of new speech processing algorithms as well as automated optimization of system parameters rather than the extensive testing with human subjects required when developing a physical device (Bockstael et al., 2011). While it is the intent to transfer the developed algorithm to a physical device, this paper focuses on the development of the subband communication enhancement algorithm and verification of the design goals using two different simulations designed to replicate real-world listening scenarios.

To ensure the relevance of the simulation results, a feedforward ANR hearing protector was assembled for modeling the acoustics and electro-acoustics of a real-world device. Simulation of a physical system ensures that the developed algorithm can eventually be implemented to obtain real-world results. The simulation models the transmission of sound through the ear cup, the performance of the loudspeaker (for reproducing ANR and communication signals) and the SPL of sound reaching the ear. The major limitation of the simulation is that it does not model the computational speed of the digital signal processing system needed to implement the algorithm. Therefore, care must be taken to ensure that the computational complexity of the ANR and speech processed signals does not exceed the processing time limit imposed by the requirement to maintain phase cancellation necessary for feedforward active control.

This paper explores two simulations to demonstrate the effectiveness of the developed subband communication enhancement algorithm compared with traditional communication headsets. In the first simulation, the speech transmission index (STI) is used to demonstrate the adaptive capability of the algorithm to adjust the characteristics of the communication channel to optimize speech intelligibility. The second simulation explores the effects on speech intelligibility when the total SPL detected at the ear is limited to desirable noise exposure levels. The subband communication enhancement algorithm developed integrates active control and speech enhancement techniques in order to improve speech intelligibility and limit total noise exposure. More importantly, the algorithm demonstrates the ability to merge the very different bandwidths of active control and speech enhancement with a minimal increase in computational cost thus ensuring that it can eventually be implemented in a real-world system. The methods described in this paper are the subject of a patent application (Brammer and Bernstein, 2011).

2. Methods

2.1. ANR communication headset

A physical device, illustrated in Fig. 3, was obtained by modifying a commercially available ear cup with the custom electronics necessary for feedforward active control, namely the reference and error microphones as well as the secondary source loudspeaker. The ear cup was selected from a series of industrial ear cup designs based on a method for predicting the total attenuation of a feedforward control system (Nelson and Elliott, 1992) discussed in Section 2.1.2. The ear cup was modified to include the loudspeaker element from an HD600 headphone (Sennheiser Electron Corporation, Old Lyme, CT) and two hearing instrument miniature microphones (FG-23629, Knowles Electronics, Itasca, IL). A custom analog electronic front-end provided the signal conditioning necessary to interface the microphone and speaker elements with the analog-to-digital (ADS8361, Texas Instruments, Dallas, TX) and digital-to-analog (DAC8534, Texas Instruments) converters. Calculations were performed using a TMS320C6713B (Texas

Fig. 3. Active hearing protector physical model demonstrating placement of active sensing and control elements.

Instruments) digital signal processor (DSP) integrated into a DSP development board (DSK6713, Spectrum Digital, Stafford, TX) programmed using the Code Composer Studio integrated development environment (Texas Instruments) with a variety of C and assembly programs.

2.1.1. Passive attenuation of the earmuffs

The attenuation of a feedforward ANR communication headset at frequencies above approximately 1 kHz is determined primarily by the passive performance of the earmuff (i.e., ear cup + cushion) rather than the active control system (Bernstein et al., 2010; Shaw and Thiessen, 1962). Evaluation of the passive attenuation performance of the earmuffs used an in-ear probe microphone system to measure SPLs at the eardrum (Brammer et al., 2009). This system allows for repeated placement of the tip of the probe microphone within 5 mm of the eardrum and has been shown to be accurate to within ±2 dB for SPLs at frequencies up to 6 kHz. SPLs without hearing protection were recorded on a dynamic signal analyzer (SR785, Stanford Research Systems, Sunnyvale, CA) with the human subject seated in a white noise environment generated in the University of Connecticut Health Center reverberation room. The subject was then fitted with the unmodified hearing protector and the resulting difference between the two sound pressures was used as an estimate of the passive noise reduction performance, PNR(ω).

All subjects gave their informed consent to participate in the study, which was conducted according to the provisions of the University of Connecticut Health Center's ethics review board.

2.1.2. Predicting the active attenuation of earmuffs

The geometry of the ear cup can have a significant effect on the performance of the active control system. Estimation of the performance boundaries of a proposed active control system prior to implementation can provide valuable information before the time and money are spent adding the electronic components to an ear cup. Nelson and Elliott (1992) demonstrate that the active noise reduction, ANR(ω), performance of a feedforward ANR system is dependent on the coherence function between the input of the reference microphone and the input of the error microphone, $\gamma^2_{re}(\omega)$, according to

$$\mathrm{ANR}(\omega) = 1 - \gamma^2_{re}(\omega). \quad (1)$$

Coherence is a predictability measure that in this case indicates how much of the error signal can be predicted by the reference signal and eventually controlled by ANR. The advantage of this method is that the two microphones can be moved between a series of ear cup designs to provide a performance estimate rather than fitting microphones and loudspeakers into every ear cup under test. The expected total noise reduction, TNR(ω), of the complete system was obtained through summation of the passive and predicted active noise reduction estimates. The summation is valid provided the total noise reduction so estimated does not exceed the threshold for bone-conducted sound.

The coherence measurements were conducted using a white noise sound source in a reverberation room with an average T60 time of 3 s. The diffuse field generated by this environment provides an acoustic worst case scenario since the controller cannot use any directional cues to produce the optimal control filter. The error microphone was positioned just outside the entrance to the ear canal. The reference microphone was mounted on the outside surface of the ear cup in various positions to find an optimal placement that provided the most predicted ANR. The coherence of the two microphone signals was measured by a dynamic signal analyzer (SR785) and converted into dB for comparison with the passive attenuation results.

2.2. Primary and secondary path estimation

A system identification scheme was used to model the primary and secondary paths for the computer simulations used for algorithm development. The primary path transfer function describes the propagation of sound through the passive ear cup from the reference microphone to the error microphone, i.e., from R to E in Fig. 1. The secondary path transfer function describes the reproduction of sound by the secondary source and its propagation to the error microphone, i.e., from S to E in Fig. 1. The primary path transfer function was measured on a human subject wearing the active hearing protector in an anechoic chamber by placing a loudspeaker 1 m away from the hearing protector along the axis formed by a line passing from the reference microphone to the error microphone. The configuration formed a progressive sound field that maximizes the processing time available for calculation of the cancellation signal, thereby creating an ideal environment for algorithm development. Despite the advantages for algorithm development, testing of the final algorithm in additional sound field models or in a physical implementation will be required to ensure performance in a variety of real world, rather than simulated, sound fields. Signals from the reference and error microphone were recorded by the DSP at a sampling rate of 24 kHz. A Recursive Least Squares algorithm (Haykin, 2002) in a system identification configuration was programmed using Matlab (MathWorks, Natick, MA) to adapt the coefficients of an FIR filter until it best models the primary path of the hearing protector. The number of filter coefficients of the filter was adjusted to ensure accurate modeling of the low frequency components of the transfer function.

Identification of the secondary path required a white noise signal generated by the secondary source loudspeaker and the signal received by the error microphone, both again recorded at a sampling rate of 24 kHz. The secondary path identification algorithm used the loudspeaker drive signal as an input and the error microphone signal as the desired signal, rather than the reference microphone signal used as the input for the estimation of the primary path.

These primary and secondary path impulse response models are used in all headset simulations to ensure that the algorithms produced results that were consistent with the physical implementation of an ANR communication headset.

2.3. Passive headset simulation

Simulations involving a passive attenuation communication headset use the primary and secondary path models for calculation of the sound pressure at the ear. A primary noise external to the ear cup is attenuated through the primary path before it is sensed at the error microphone and eardrum. Communication signals are reproduced by the secondary source loudspeaker and are sensed at the error microphone after being "filtered" by the secondary path.

2.4. Fullband ANR headset simulation

The FXLMS algorithm is the most common adaptive algorithm used for feedforward ANR. As already noted, this algorithm adapts the coefficients of the FIR control filter over time to values that best minimize the square of the signal received at the error microphone. Since this signal corresponds to the SPL, the total acoustic noise reaching the eardrum is minimized when the filter reaches a steady state value. For the purposes of this paper, as the coefficients of the single FXLMS adaptive filter cover the entire ANR frequency range, this filter can be described as a fullband filter.

2.5. Delayless subband ANR headset simulation

The delayless subband ANR algorithm, illustrated in Fig. 4, expands upon the fullband FXLMS algorithm by dividing the reference and the error signals into a series of subband signals each representing a small portion of the entire frequency range. Each subband adapts independently from the other filters, which can lead to improved convergence rates. A bank of filters, called analysis filters, are included in the "generate subbands" block that produces multiple subband filter signals from a single input signal from either the reference or error microphones. The analysis filter bank is constructed using a prototype filter for the lowest frequency subband, designed using an algorithm described by Mitra (2006) that is shifted in the frequency domain using a Discrete Fourier Transform (DFT) modulation (Lee et al., 2009).

The subband filters are eventually combined into a single wideband filter that is used to generate the appropriate ANR cancellation signal. The wideband filter includes the frequency range of the fullband FXLMS filter and is needed to generate a single cancellation signal to the secondary source loudspeaker rather than a series of signals from the individual subbands. For a wideband filter with N coefficients in a system with M subbands, each subband has a reduced number of coefficients, $2N/M$, that only needs to be updated every $M/2$ samples. This reduction in the number of coefficients and update rate forms the primary advantage of the subband architecture since it allows for a significant reduction in computational cost for high order control filters when compared with the fullband algorithm.

The coefficients of the wideband filter are determined from the subband filters by first calculating the Fast Fourier Transform (FFT) of each of the subband filter coefficient vectors. The frequency responses of each subband filter are then pieced together (Lee et al., 2009) to form the frequency response of the desired wideband filter. An Inverse Fast Fourier Transform (IFFT) is finally performed to obtain the corresponding wideband filter coefficients. The resulting output control signal is combined with the communication signal and output by the loudspeaker. The 'delayless' reference in this filter design refers to the fact that the output signal is generated without passing through any subband filters. Traditional subband filtering introduces delays that could inhibit active control if causality constraints are violated by processing times that exceed the propagation time of the primary noise from the point of detection (the reference microphone) to the point of control (the secondary source).

Conversion of the subband filters into the necessary wideband filter coefficients adds an additional computation beyond a typical fullband algorithm, but the total cost remains less than the fullband system. For a detailed derivation of the computational cost of the subband ANR system, see Morgan and Thi (1995). In the implementations used in this simulation, the fullband 1024 coefficient filter is estimated to require 2048 multiplies per input sample, while the subband 1024 coefficient filter with eight subbands and a 64 tap subband prototype filter is expected to require 1372 multiplies per input sample, resulting in a 35% improvement in computational cost. For the purposes of this study, the subband ANR system was operated at a sampling rate of 4 kHz with a total of eight subbands, although redundancies require only five of the subband filters to be calculated for proper operation (Morgan and Thi, 1995). While additional subbands can further reduce computational cost, the selection of the number of subbands in this study was driven by compatibility with the communication processing goals rather than computational demands.

The modifications to the delayless subband ANR algorithm explored in this paper are the inclusion of the communication signal and an additional step to estimate the error signal SPLs in each

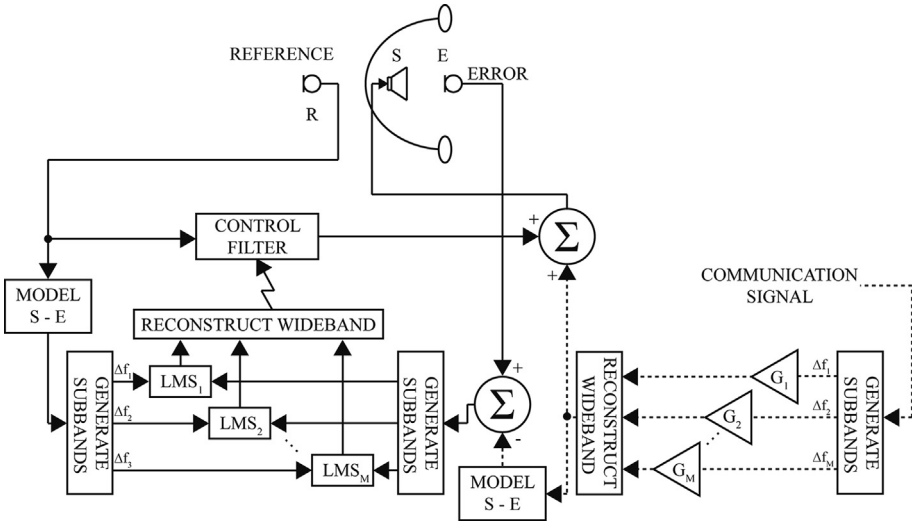

Fig. 4. Block diagram of the delayless subband ANR hearing protector with adaptive communication channel gain.

subband. In order to provide a communication signal at an SNR appropriate to the background noise, the SPL underneath the ear cup is required. However, this information must be obtained with minimal increase in computational cost in order to meet the delay requirements imposed by feedforward control. The subband ANR algorithm allows for estimation of the error signal power within each of the narrow frequency bands of the subband filters rather than just the power over the entire frequency range. Using the subband signals generated from the error microphone, an exponential averaging technique is implemented to estimate the power of the background noise within each subband. This noise estimate accounts for not only the passive attenuation of noise through the primary path, but also the noise levels after active control.

While the communication signal generally will not significantly disturb the performance of the ANR system, it does disrupt the estimation of the background noise levels underneath the ear cup. The resulting error signal contains the background noise levels to be measured and the communication signal reproduced by the secondary source. As a result, the latter must be removed from the error signal recorded. This is accomplished by filtering the communication signal by the estimate of the secondary path transfer function and subtracting it from the recorded error signal.

2.5.1. Adaptive communication channel gain

The ANR and communication systems operate at different sampling rates to accommodate the wide disparity in signal bandwidth associated with ANR versus speech signals. Speech communication can carry important information from the 125 Hz to 8 kHz octave bands, and thus the sampling rate of the communication system was set to 24 kHz in order to capture features throughout the desired frequency range. Conversely, feedforward ANR effectiveness is generally negligible above 1 kHz. Therefore, a downsampling technique was implemented to reduce the sampling rate of the ANR system to 4 kHz in order to optimize the computational efficiency of the merged ANR and communication processing as well as provide a minimization of the propagation delay through the ANR system (Bai et al., 2001; Brammer et al., 1997). Fig. 2 shows the frequency ranges of individual subbands of the subband ANR system as well as the frequency range of the communication channel. Note that due to the method of modulation of the prototype filter, the first and last subbands will always have half the bandwidth of the remaining subbands.

The communication channel is divided into subband signals using identical filters as those used to process the error signal, hence ensuring that noise and speech power data are obtained from identical frequency bands. A similar exponential averaging operation was performed on the speech signal to obtain estimates of the speech signal power in each subband. The communication signal power is calculated for subbands within the frequency range of active control as these are the only frequencies for which the corresponding error signal power is also calculated.

Once instantaneous estimates of the error and speech powers in each of these subbands are known, the speech SNR can be derived for any communication channel gain G according to:

$$\mathrm{SNR} = \frac{G^2 \cdot C(k)}{E(k)}, \tag{2}$$

where $C(k)$ and $E(k)$ represent the instantaneous communication and error signal powers, respectively.

If the desired speech SNR is fixed, this equation can be solved to determine the communication channel gain in each subband necessary for the speech signal to maintain a certain power ratio related to the background noise levels. By continuously adjusting the instantaneous gain in each subband to that ideal for speech communication, the system is able to adapt the current communication signal to a level that is appropriate for the current background noise environment.

While the calculated gain provides the gain necessary to reach the desired SNR, without considering power limits there is a risk of providing total sound levels underneath the ear cup that exceed the threshold for hearing damage. The power, P, of the noise plus speech signal generated with a given communication channel gain is evaluated by

$$P = G^2 \cdot C(k) + E(k). \tag{3}$$

Hence, the instantaneous gain power limit of the communication channel for each subband can be found for a desired maximum SPL. To maintain speech intelligibility, the communication channel gain in each subband is set to the gain necessary to maintain the desired SNR as long as it is less than the gain power limit; otherwise the gain is set to the maximum permissible gain. The gain of the higher frequency components of the communication channel is set to the average gain calculated for the lower frequency subbands subject to ANR. Once the gain has been applied to each communication subband, the fullband communication signal is reconstructed and presented to the user.

2.6. Performance evaluation

The Speech Transmission Index (STI) is an objective method for evaluating the speech intelligibility of a communication system (Houtgast and Steeneken, 2002). It uses recordings of the original speech and the speech as detected through the communication channel to provide a value, from zero to one, which represents the expected speech intelligibility. This value can be related to the word intelligibility scores of human subjects as measured, for example, by the Modified Rhyme Test (MRT) (Yu et al., 2010). Using the STI, the effects on speech intelligibility can explored during every stage of algorithm development to help optimize communication processing. STI values from the simulations were used to obtain estimates of the expected speech intelligibility if each algorithm was built into the physical hearing protector from which the primary and secondary path models were obtained.

While the STI was developed for persons with normal hearing it has been applied to persons with common forms of hearing loss (George et al., 2010). The primary consequence of cochlear hearing loss is reduced audition of faint sounds, which hence require increased intensity to be audible in a noisy environment and thus an increased speech signal-to-noise ratio (see, for example, Moore, 2007). The STI, being ultimately based on a metric of speech signal-to-noise ratio, will in consequence tend to overestimate the intelligibility of speech for persons with hearing loss. For the purposes of this simulation based analysis, the STI results presented in this study represent speech intelligibility for normal hearing individuals.

3. Results and discussion

3.1. Performance of model hearing protector

The Optime 98 (Peltor, St. Paul, MN) hearing protector was selected for conversion into an ANR communication headset due to its superior predicted total attenuation performance over the entire frequency range. This device provided the best balance relating the larger volume underneath the ear cup, leading to higher passive

attenuation at mid-to-high frequencies, to the shorter distance between the reference and error microphones, leading to better coherence and higher predicted levels of active control at low frequencies. Fig. 5 plots the passive (PNR), predicted active (ANR), and total (TNR) performance prior to adding the active noise control electronics to the ear cup. The PNR indicates a steady attenuation of approximately 35 dB above 600 Hz, but decreasing attenuation for lower frequencies. Conversely, the ANR is predicted to yield between 10 and 20 dB of attenuation for frequencies below 600 Hz. As a result, the combined attenuation performance (TNR) indicates that the modified active hearing protector is expected to achieve over 30 dB of attenuation for frequencies within the 100–1600 Hz frequency range investigated.

Once the earmuff was selected, it was outfitted with reference and error microphones and a secondary source loudspeaker (see Section 2.1). While the probe microphone system described in Section 2.1.1 could be used as an error microphone, the added customization of the ANR system to each user was deemed undesirable. The reference microphone was located on the outside surface of the ear cup roughly along the inter-aural axis. This location was identified as best for an environment where the direction of the incoming sound is unknown. If the direction of the primary disturbance is known, for example towards the direction the subject is facing, then the reference microphone would be positioned on the ear cup to optimize the response in the desired direction.

3.2. Primary and secondary path modeling

The transfer functions of the primary and secondary path models used for all simulations are shown in Fig. 6. The primary path model closely follows the performance of the passive hearing protector, with reduced attenuation at low frequencies and increasing attenuation at higher frequencies. The secondary path transfer function indicates that the loudspeaker provided a sufficiently flat frequency response for the (lower) frequencies associated with active noise control. However, the relatively large magnitude variations at the higher frequencies involved in speech communication were attributed to operation of the speaker in a different ear cup volume from its original headphone. These gain variations in the speech communication channel were audible to listeners and so the communication signal was pre-filtered by an inverse of the secondary path estimate in an attempt to equalize the communication SPLs at the error microphone.

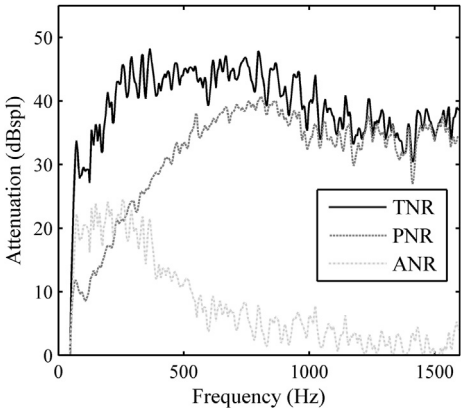

Fig. 5. Passive noise reduction (PNR), predicted active noise reduction (ANR), and total noise reduction (TNR) of the model active noise reduction communication headset.

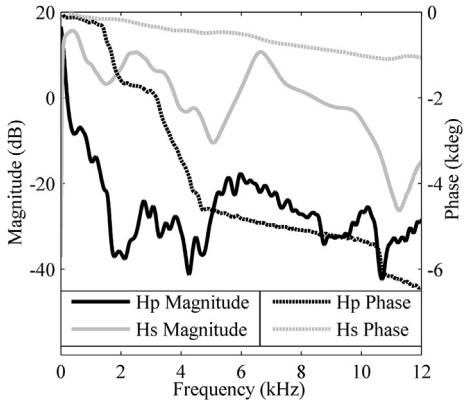

Fig. 6. Primary and secondary path models transfer function magnitude and phase responses of model ANR communication headset.

3.3. Simulation of communication headsets

3.3.1. Speech SNR performance

All simulations were developed in Matlab to compare the performance of a passive attenuation headset with fixed communication channel gain, a fullband FXLMS ANR headset with fixed communication channel gain, and the subband ANR headset with adaptive communication channel gain. Fig. 7 shows the time waveforms for each during a 180 s simulation where the first 90 s contains background noise only, to evaluate the ANR headset performance of the active control algorithms, and the second 90 s contains both noise and a speech communication signal. The background noise used in the simulations is from a Leopard tank (Steeneken and Geurtsen, 1988) and the speech signal is an MRT

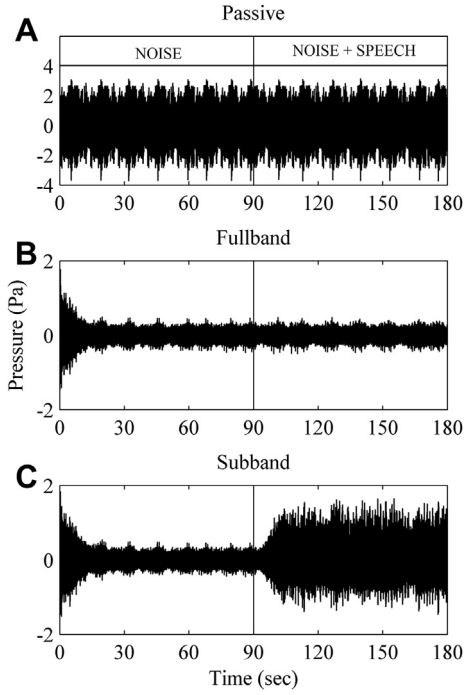

Fig. 7. Simulated time waveforms of the passive attenuation, fullband FXLMS, and subband ANR communication headsets. The fullband and subband designs both provide 10 dBA attenuation beyond a typical passive system during the noise only period, but only the subband system demonstrates SNR improvement once the speech signal is introduced at $t = 90$ s.

word set embodied in a carrier phrase (House et al., 1965). The adaptation times of both the fullband and subband systems were tuned to reach a steady state performance within approximately 15 s.

The SPL of both the environmental noise and the speech signal for this simulation were set according to the performance of the passive attenuation headset with fixed communication channel gain. The SPL of the combined noise plus speech signal was set to 80 dBA, a maximum level considered desirable for many occupational settings to prevent hearing damage, with an STI value of 0.2, which corresponds to a word intelligibility of approximately 50% for an MRT test. Since the passive system does not provide any additional attenuation beyond its starting levels, the sound pressures in Fig. 7A remain relatively constant during the noise only period. The speech signal is introduced with an SNR of less than 1 dB, which is supported by the nearly unrecognizable increase in sound pressure in Fig. 7A when comparing the 'Noise + Speech' period with the 'Noise' period. This speech SNR is insufficient for critical communications, but increasing the gain of the speech communication signal will exceed the hearing damage exposure threshold of 80 dBA.

Using the same gain levels, the fullband algorithm demonstrates a clear improvement in attenuation performance over the passive communication headset. Fig. 7B shows that after a brief convergence period, the active control system is able to provide an additional 10 dB of A-weighted attenuation of the environmental noise. However, the resulting STI value of 0.2 indicates that the speech SNR is still insufficient because the fullband algorithm is still using gain settings originally established for the passive system. The decrease in environmental noise SPL allows for an increase in the communication channel gain without the risk of exceeding hearing damage thresholds, but requires user intervention in order to restore appropriate speech intelligibility. Manual selection of communication channel gain in high noise environments commonly leads to excess sound pressures at the ear (Crabtree, 2002). User intervention could also divert attention away from other important cognitive tasks.

The subband system, seen in Fig. 7C, replicates the 10 dB of A-weighted attenuation performance of the fullband algorithm during the noise only period, while providing a significant improvement in speech intelligibility during the noise plus speech period. This system compares the speech signal power with the environmental noise power detected at the error microphone and hence recognizes the inadequate gain levels in the communication channel. Over a period of approximately 10 s, the gain levels are slowly increased to a steady state level corresponding to the specified unweighted SNR level of +12 dB. The SNR target level was established based on the +12–15 dB typical listening levels of users of communication headsets (Giguère and Dajani, 2009). The resulting STI value of 0.6 corresponds to an MRT word intelligibility score of approximately 95%, which is suitable for most critical communication environments where instructions are generally repeated to ensure understanding. By adjusting the communication channel gain in each subband, gain levels are selected to maintain speech intelligibility appropriate for the current environmental noise. This is ideal for noise environments with changing background noise levels where the user cannot focus on adjusting communication gain levels, such as a pilot preparing a plane for landing where varying air speeds can change background noise levels and disrupt critical communication with air traffic control personnel.

3.3.2. Maximum speech power

Maintaining speech communication SPLs below hearing damage thresholds is important for preserving the health of users. While the +12 dB SNR is appropriate for maintaining speech intelligibility, reduced communication channel gain settings may lead to adequate intelligibility with more reasonable exposure levels. If the A-weighted noise levels after adaptation of the fullband FXLMS algorithm are set to approximately 74 dBA, then an unweighted speech SNR of +12 dB would lead to a total SPL underneath the ear cup of approximately 84 dBA, which is considered excessive for the purposes of the simulation. In this situation, the subband system identifies that a high gain is necessary, but limits the total SPL to approximately 80 dBA. The result is a decrease in the STI from 0.9 to 0.6, which provides a minimal decrease in MRT word scores from 99% to 95%. This reduction in intelligibility would be acceptable in most critical communication environments where instructions are generally repeated for clarification and the reduced total SPL means increased acceptable exposure times for the user.

Fig. 8 shows how the gain is calculated within each of the five communication subbands. At any given time, the system calculates both the gain needed to maintain the desired +12 dB SNR (Desired Gain) and the maximum gain that can be applied to maintain a total SPL below the maximum 80 dB limit (Maximum Gain). For each input signal sample, the system selects the gain to apply to the communication channel (Selected Gain) according to the lesser of the Desired and Maximum Gain.

Most of the power of the tank noise used as the primary noise source in this simulation is concentrated in the lower frequencies. As a result of the higher power at these frequencies, the power that can be added as part of a communication channel is decreased. Subband 1 demonstrates the compromise between the desired SNR and the maximum power limit. In the first 10 s after the communication channel starts (at $t = 90$ s), the Desired and the Maximum Gain profiles nearly coincide during initialization, thus the Selected Gain follows both waveforms. During the remaining period shown, the Maximum Gain is actually less than the Desired Gain and thus the Selected Gain levels are held at the maximum limit.

The gain selected for subband 2 closely follows the desired SNR target value. In the case of subband 2, the Maximum Gain approximately follows the Maximum Gain waveform for subband 1, but based on the error and communication signal powers a lower gain is necessary to reach the desired speech SNR level, and thus the Selected Gain follows the Desired Gain.

When examining the power spectrum of the primary noise file in the frequency range of subband 3, it appears that while there is high power in the lower half of the frequency band, there is much less power in the upper half of the subband. This leads to the system calling for a Desired and Maximum Gain of over 20× for subband 3, which is higher than that of any of the other subbands that usually demonstrate a Desired Gain level between 5× and 15×. The exact reason for the higher gain level is unclear, but the Selected Gain remains below the maximum permissible SPL limit and does not create any audible speech signal distortions when reproduced by a high fidelity sound system.

The SPLs detected underneath the ear cup for subbands 4 and 5 are significantly reduced when compared with the lower frequency bands. This is due to both lower levels in the primary noise source and higher attenuation through the primary path. As a result, the Maximum Gain waveform takes values much higher than the other bands because much more communication signal power can be added before reaching hearing damage threshold limits. However, since a low gain is sufficient to maintain speech intelligibility, the lower Desired Gain is selected. The resulting signal provides adequate levels of speech intelligibility at exposure levels unlikely to cause hearing damage.

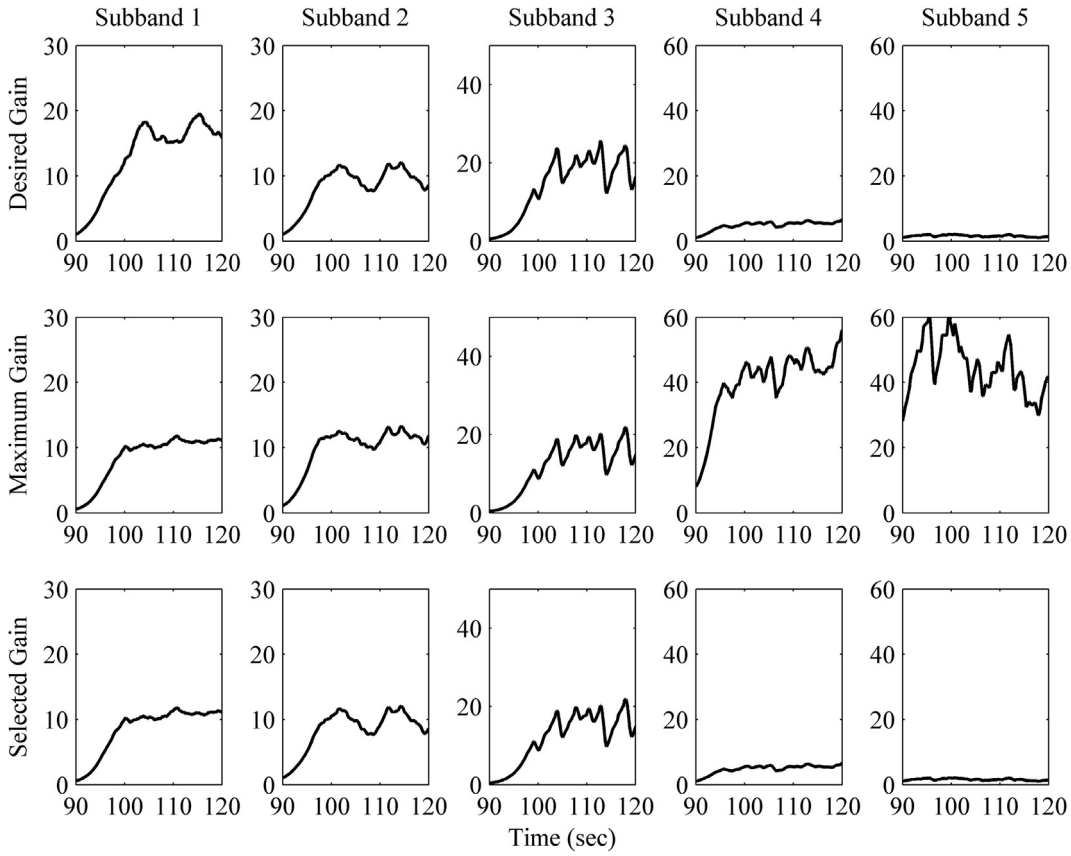

Fig. 8. The desired gain to maintain a +12 dB speech SNR (Desired Gain) and the 80 dBA maximum gain (Maximum Gain) are calculated for each of the five subbands. The Selected Gain follows the minimum values indicated by either the Desired or Maximum Gain. Note the differences in *Y*-axis scales between subbands.

3.3.3. Simulation sound pressure levels

Table 1 summarizes the performance of the three communication headset designs for the two different simulation conditions; the 12 dB target SNR and 80 dBA maximum SPL tests. The table shows the A-weighted SPL underneath the ear cup during the noise (N) and noise plus speech (N + S) periods, as well as the calculated STI values and the corresponding MRT word scores indicating the expected percentages of words correct if human psychoacoustical performance tests were conducted (Yu et al., 2010). In both simulations, the subband headset attenuation performance is comparable with the performance of a traditional feedforward fullband FXLMS headset design. The primary benefit of the subband ANR communication system is demonstrated when the speech signal is introduced. The subband system monitors the environmental noise and communication signal levels and adjusts the communication channel gain to the desired speech SNR of +12 dB in the first simulation thereby improving speech intelligibility by an expected 45% over a typical fullband system. Note that the spectral content of the noise and speech signals results in an increase in A-weighted SPL of 8 dB rather than the 12 dB unweighted increase. In the 80 dBA maximum SPL test, the subband system demonstrates the ability to maintain a maximum SPL close to the limit of 80 dBA which is 5 dB lower than the gain levels needed to reach the speech SNR target. The resulting lower SPLs reduce the likelihood of hearing damage following prolonged exposure, yet still manage to maintain an MRT score of 95%, which is adequate for most critical speech communication. While a conventional user-adjustable gain could mitigate the noise exposure issues, the ability of the subband system to adapt to changing noise conditions ensures that speech intelligibility and noise exposure limits will be maintained without user intervention. Therefore, the user can focus attention on more important tasks while maintaining comfortable communication levels with coworkers and healthy noise exposure levels.

Table 1
Comparison of the SPLs during the noise only (N) and noise plus speech (N + S) periods for the +12 dB SNR target and 80 dBA maximum SPL simulations. The subband system provides a 45% increase in word intelligibility for the +12 dB target SNR simulation over the other designs. In the 80 dBA maximum SPL simulation, the subband system demonstrates an ability to maintain a total SPL close to the desired maximum limit.

	12 dB target SNR				80 dB max SPL			
	N	N + S	STI	MRT	N	N + S	STI	MRT
Passive	80	80	0.2	50%	–	–	–	–
Fullband	70	70	0.2	50%	74	84	0.9	99%
Subband	70	78	0.6	95%	74	81	0.6	95%

4. Conclusion

Speech communication is necessary in many high-noise industrial environments to maintain worker interaction and facilitate safe working conditions. However, the hearing protection devices intended to protect worker hearing health can impede speech communication or expose workers to excessive noise exposure levels if a communication signal is presented improperly. The subband communication enhancement algorithm developed exploits information needed for active control to estimate environmental noise levels underneath the ear cup and present

a speech signal designed to optimize speech intelligibility. By maintaining a specified speech SNR in each subband, communication signal power is added where it is needed to maintain speech intelligibility without adding excess power characteristic of a communication channel with user-controlled gain. More importantly, by limiting the communication channel gains in each subband, noise exposure limits can be established to prevent exposure levels that could lead to hearing damage over extended periods of time.

Acknowledgements

This work was supported by the National Institute for Occupational Health and Safety under research grant R01 OH008669-05.

References

Bai, M., Yuanpei, L., Jienwen, L., 2001. Reduction of electronic delay in active noise control systems – a multirate signal processing approach. J. Acoust. Soc. Am. 111 (2), 916–924.

Bernstein, E., Brammer, A., Yu, G., Cherniack, M., Peterson, D., 2010. Ear cup selection for feedforward active noise reduction hearing protectors. Can. Acoust. 38 (3), 82–83.

Bockstael, A., De Coensel, B., Botteldooren, D., D'Haenens, W., Keppler, H., Maes, L., Philips, B., Swinnen, F., Bart, V., 2011. Speech recognition in noise with active and passive hearing protectors: a comparative study. J. Acoust. Soc. Am. 129 (6), 3702–3715.

Brammer, A., Bernstein, E., 2011. Method and Device for Improving the Audibility, Localization and Intelligibility of Sounds, and Comfort of Communication Devices Worn on or in the Ear. U.S. Provisional Patent Serial No. 61/546,555.

Brammer, A., Pan, G., Zera, J., Goubran, R., 1995. Application of adaptive feedforward active noise control to a circumaural hearing protector. In: Active 1995, Newport Beach, CA, July 6–8, 1995, pp. 1319–1326.

Brammer, A., Pan, G., Crabtree, R., 1997. Adaptive feedforward active noise reduction headset for low-frequency noise. In: Active 1997, New York, NY, pp. 365–372.

Brammer, A., Yu, G., Bernstein, E., Peterson, D., Cherniack, M., Tufts, J., 2009. Field measurement of sound pressure at the eardrum. In: Inter-noise 2009, Ottawa, CA, August 23–26, 2009.

Cardosi, K., Falzarano, P., Han, P., 1998. Pilot-Controller Communication Errors: an Analysis of Aviation Safety Reporting (ASRS) System Reports. U.S. Department of Transportation, Office of Aviation Research, Washington. DOT/FAA/AR-98/17.

Crabtree, R., 2002. Hercules Audio Enhancement: Project Report. DRDC Toronto, Letter Report, September 20, 2002.

George, E., Goverts, S., Festen, J., Houtgast, T., 2010. Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners. J. Speech Lang. Hear. Res. 53, 1429–1439.

Giguère, C., Dajani, H., 2009. Noise exposure from communication headsets: the effect of external noise, device attenuation and effective listening signal-to-noise ratio. In: Inter-noise 2009, Ottawa, CA, August 23–26, 2009.

Haykin, S., 2002. Adaptive Filter Theory, fourth ed.. Upper Saddle River, NJ.

House, A., Williams, C., Hecker, M., Kryter, K., 1965. Articulation testing methods: consonantal differences with a closed-response set. J. Acoust. Soc. Am. 37, 158–166.

Houtgast, T., Steeneken, H., 2002. Past, Present, and Future of the Speech Transmission Index. Soesterberg, The Netherlands.

Kuo, S., Morgan, D., 1996. Active Noise Control Systems: Algorithms and DSP Implementations. New York, NY.

LaTourette, T., Peterson, D.J., Bartis, J.T., Jackson, B.A., Houser, A., 2003. Protecting Emergency Responders: Community Views of Safety and Health Risks and Personal Protection Needs. Rand Science and Technology Institute.

Lee, K., Gan, W., Kuo, S., 2009. Subband Adaptive Filtering: Theory and Implementation. Chichester, UK.

Mitra, S., 2006. Digital Signal Processing: a Computer Based Approach. New York, NY.

Moore, B., 2007. Cochlear Hearing Loss: Physiological, Psychological, and Technical Issues. Chichester, England.

Morgan, D., Thi, J., 1995. A delayless subband adaptive filter architecture. IEEE Trans. Signal. Process. 43 (8), 1819–1830.

Nelson, P., Elliott, S., 1992. Active Control of Sound. San Diego, CA.

NIOSH, 1998. Criteria for a Recommended Standard: Occupational Exposure to Noise. National Institute for Occupational Safety and Health DHHS Publication No. 98–126.

Park, S., Yun, J., Park, Y., Youn, D., 2001. A delayless subband active noise control system for wideband noise control. IEEE Trans. Speech Audio 9 (8), 892–899.

Qiu, X., Li, N., Hansen, C., Chen, G., 2006. The implementation of delayless subband active noise control algorithms. In: Active 2006, Adelaide, Australia, September 18–20.

Ray, L., Solbeck, S., Streeter, A., Collier, R., 2006. Hybrid feedforward-feedback active noise reduction for hearing protection and communication. J. Acoust. Soc. Am. 120 (4), 2026–2036.

Shaw, E., Thiessen, G., 1962. Acoustics of circumaural earphones. J. Acoust. Soc. Am. 34, 1233–1246.

Song, Y., Gong, Y., Kuo, S., 2005. A robust hybrid feedback active noise cancellation headset. IEEE Trans. Speech Audio 13 (4), 607–617.

Steeneken, H., Geurtsen, F., 1988. Description of the RSG-10 Noise Database. TNO Institute for perception. Report No. 1988-3.

Wu, M., Qui, X., Chen, G., 2008. An overlap-save frequency-domain implementation of the delayless subband ANC algorithm. IEEE Trans. Audio Speech Lang. Proc. 16 (8), 1706–1710.

Yu, G., Brammer, A., Swan, K., Tufts, J., Cherniack, M., Peterson, D., 2010. Relationships between the modified rhyme test and objective metrics of speech intelligibility. J. Acoust. Soc. Am. 127 (3), 1903.