ALPHA FOUNDATION FOR THE IMPROVEMENT OF MINE SAFETY AND HEALTH

Revised Final Technical Report

1.0 Cover Page

Grant Number and Title: AFC518-10, "Improving Communication in Noise for Miners Wearing Hearing Protection"

Organization Name: University of Connecticut Health

Principal Investigator: Anthony J. Brammer

Contact Information: brammer@uchc.edu; Tel. (860) 679-2760; FAX. (860) 679-1989

Period of Performance: August 1, 2017 - January 31, 2019

This study was sponsored by the Alpha Foundation for the Improvement of Mine Safety and Health, Inc. (ALPHA FOUNDATION). The views, opinions, and recommendations expressed herein are solely those of the authors and do not imply any endorsement by the ALPHA FOUNDATION, its Directors and staff.

2.0 Executive Summary

According to the National Institute of Occupational Safety and Health, one in four miners has a hearing problem and, by retirement age, four out of five mine workers have impaired hearing. Studies have documented that many miners exposed to noise are unwilling to wear hearing protection for fear of failing to understand speech or hear warning sounds. In recognition of these issues, specialized hearing protectors (HPDs) have been developed for situations in which noise levels change in space or with time, such as when walking towards or around a machine or when a nearby vehicle moves or stops operating. In these circumstances, the device automatically adjusts electronically the amount of hearing protection to enable the user to hear more sounds in the environment when the environmental noise is low. Unfortunately, at the present state of development, the devices fail to improve the intelligibility of speech in the presence of noises commonly leading to overexposure in a mine (i.e., continuous mining machines, roof bolting machines) compared to when conventional hearing protectors are worn. Our mission, therefore, was to develop a method for improving communication in noise suitable for miners wearing an HPD, in order to increase the audibility of warning sounds and reduce the confusion understanding words spoken in noise. These improvements should reduce the risk of miners being struck by moving equipment and errors in speech communication with co-workers.

Our method involves first dividing the frequency spectrum of sounds (i.e., speech or warning sound + environmental noise) into narrow bands of frequencies (called "subbands"). The magnitude of the envelope of the signal in each subband is then used to adjust the instantaneous gain applied to the subband signal (i.e., modulate the signal), depending on whether the subband is judged to contain mostly desired sounds or undesired noise. The modified subband signals are then recombined to form the processed output. The complete algorithm is then repeated, leading to a time-varying gain applied to each subband and time-varying reconstruction of the original signal that amplifies the desired sounds identified.

The primary challenge for the research is to distinguish desired sounds buried in environmental noise from the noise, and hence whether the algorithm should amplify or attenuate the sounds in a given subband. The work focused on the following factors: influence of subband bandwidths, parameters influencing envelope-derived modulations, signal and envelope fidelity, the relationships between signals in different subbands, and minimizing temporal fluctuations in sound pressure introduced by the digital signal processing (so-called "musical noise"). The signal processing employed MATLAB, and the sounds were processed off-line. Time histories were constructed for use in the Modified Rhyme Test (MRT), which is a psychophysical test of speech intelligibility evaluated by the number of word correctly identified when one of six alternatives words, chosen from a prescribed set, is replayed to a listener in a series of trials. The time histories containing the MRT words were stored electronically files for subsequent evaluation by predictive methods or for replaying to subjects in listening tests.

Simulations were first employed to predict the influence of the factors described above on intelligibility in different noise environments. It was established that improving speech intelligibility in speech-spectrum shaped noise presented the most difficult challenge, and so it was employed in all subsequent work. Three algorithms designed to improve speech intelligibility were selected for further evaluation, and formal listening tests, as a result of the simulations and informal listening.

The performance of the three algorithms was first determined when an intermittent, 1000 Hz, vehicle back-up alarm was broadcasting its familiar "beep - beep - beep" sounds in an environment containing speech-spectrum shaped noise. All three algorithms improved the signal-to-noise ratio (SNR), and hence demonstrated that the audibility of the alarm will be increased. A similar performance was obtained when alarm sounds were replaced by speech.

Seventeen subjects undertook the MRT in speech-spectrum noise at an SNR of -8 dB. Two of the three algorithms resulted in increased mean word scores of 5.1% and 18.7% when compared to the original (unprocessed) speech in noise, both of which were statistically significant. To the best of our knowledge, no alternate method for improving speech intelligibility has reported as great as improvement as that obtained here. The performance of these algorithms, together with: i) the lack of a statistically significant effect on the intelligibility of speech without noise, and: ii) an improvement in warning sound SNR, which increases the audibility of a tonal alarm, summarizes the outcomes that demonstrate proof-of-concept. A disclosure of the invention is being made to the university.

We judge the Technology Readiness Level of the proof-of-concept to be level TRL 5 ("Technology demonstration while under development") of the nine-unit version of the NASA scale (see https://en.wikipedia.org/wiki/Technology_readiness_level). The performance of our algorithms, including variations of those discussed, while promising, remains to be established for different talkers (both male and female), different environmental noises, different sound levels and different SNRs. Clearly, more research is needed to resolve these issues.

Even after an algorithm is developed that can increase the intelligibility of speech spoken in a noisy environment without interfering with the audibility of warning sounds, it must be transferred to electronics capable of microminiaturization. The computational complexity of the algorithm in its present form will require careful attention to circuit design to be implemented within a small, lightweight package that could function effectively throughout a work shift and be worn as part of a miner's equipment, or attached to, or integrated into, a miner's helmet. In view of the expectation that the electronic HPD may be used in an underground mine, all aspects of the design must involve intrinsically safe circuitry and components. We note that several commercial electronic HPDs have already been approved by the Mines Safety and Health Administration for use underground.

While our research has focused on improving speech intelligibility and included evaluating a tonal alarm, it has not generally considered warning sounds. It is possible that the audibility of non-tonal warning sounds, such as "roof talk", will not be increased by the speech-focused algorithms. Should this prove to be the case, we could introduce a second algorithm operating in parallel with the speech-focused algorithm. The alarm-focused algorithm needs only to respond to specific non-tonal sounds that are known to present hazards in a given mine. We have previously developed, implemented and evaluated by listening tests such an algorithm for this application (Bernstein et al., 2014).

A detailed Appendix entitled *Response to Review* has subsequently been added to the original Final Report as a result of questions raised by Alpha Foundation reviewers.

The first topic concerns the observation that the algorithm providing best noise reduction does not provide most improvement in speech intelligibility. The explanation for this apparent inconsistency stems from the importance of the small amplitude modulations occurring in a sound. These turn out to be extremely important for understanding speech but of little consequence to the audibility of a warning sound. Consequently, the algorithm providing most attenuation of small modulations produces most noise reduction while the algorithm providing least attenuation of small modulations produces best speech intelligibility.

The second topic concerns the noises used in the study and how they compare to noises that might be experienced in a mine. Frequency spectra and sound levels reported in studies conducted in mines are provided for common mining machines. These are compared with the noises used in this research, which focused on so-called speech-spectrum shaped random noise as it is widely used in studies such as this to evaluate the intelligibility of speech spoken in a noisy environment. The reason for its almost universal use in speech and hearing in noise research is the expectation that speech-spectrum shaped noise will provide the "worst case" for evaluating the performance our algorithms. Thus, our algorithms can be expected to perform the same or better in noise with a different spectrum, including those of the machines in a mine. The intense noise produced at operator locations of mining machines will reduce the ability of an operator to understand speech, even when a talker shouts. This reduction occurs in the auditory system of the listener and not in the sounds external to the ear, and so intense noise will not affect the performance of our algorithms. In fact, to the extent that the algorithms reduce the overall noise level, they should disproportionally improve both alarm detection and speech understanding. Finally, a method is described that could be considered for improving the audibility of specific sounds in a given mine, such as "roof talk", provided recordings of the sounds to be identified can be obtained prior to using the algorithm.

The third topic concerns the technology readiness of our technology. The comparison of noises used in our study with those likely to be experienced in a mine (topic 2) provides context for assessing the applicability of the technology to mining operations. The limitations in scope of the present work and the discussion of topics in the Appendix lead to an outline of activities that will be necessary should the development be continued to reach a working prototype.

3.0 Concept Formulation and Mission Statement

Mission Statement

Our mission was to develop a method for improving communication in noise suitable for miners wearing a hearing protection device (HPD), in order to reduce confusion identifying spoken words and increase the audibility of warning sounds. These improvements should reduce the risk of miners being struck by moving equipment and errors in speech communication with co-workers.

We interpreted this mission to require demonstration of improved speech understanding in noise, while increasing the audibility of warning sounds, to establish "proof-of-concept". Thus, the first part of our mission was to develop a method for improving the intelligibility of speech in a noisy environment, as confirmed by listeners with normal hearing in a controlled psychoacoustic speech-in-noise listening test. It was originally intended that the tests be conducted using a custom-built HPD containing electronic components for digital signal processing as well as a microphone and earphone. The device had previously been designed and constructed in our laboratory (Bernstein, 2013), and was available for the present study. However, it was determined during the study that the custom-built electronics could not implement the complex algorithms being developed and could introduce distortion, as discussed in our Interim Progress Report (August, 1 2017 - January 31, 2018). This led to the custom-built electronic HPD being replaced by commercial headphones, or earphones located within the ear canal ("insert" earphones), and the signal processing being conducted off-line.

The second part of our mission statement was to ensure the algorithms developed for enhancing speech improved the audibility of warning sounds buried in noise. The long time required to develop the algorithms resulted in insufficient time to conduct listening tests with warning sounds. This led to reliance on the effectiveness of the algorithms being judged by changes in the signal-to-noise ratio (SNR), which is directly related to the audibility of tonal warning sounds (e.g., a vehicle back-up alarm) ((Zheng et al., 2007).

Health and Safety Mining Need

The US mining industry has the highest prevalence of hazardous workplace noise exposures of all industrial sectors (Tak et al., 2009). According to the National Institute of Occupational Safety and Health (NIOSH), one in four miners has a hearing problem and, by retirement age, four out of five mine workers have impaired hearing. The unwillingness of workers to wear commercially available HPDs because of fear they will not be able to understand speech or hear warning sounds has been repeatedly documented in the literature (for reviews, see Suter, 1992; Suter 2001). This contributes to the avoidance of hearing protector use by up to 50 % of some noise-exposed worker groups (McKinley et al., 2005; Morata et al., 2001). In miners' focus group sessions, the priority of underground survival was ranked well above the "nuisance" of hearing loss (Murray-Johnson et al., 2004; Patel et al., 2001). A common opinion was that hearing warning noises (e.g., from machinery and tunnel roofs - "roof talk"), and communicating with co-workers were major reasons for not wearing hearing protection. As succinctly stated by Azman and Hudak (2011), "*miners often complain of reduced audibility or confusion identifying spoken words when wearing conventional hearing protectors. This leads to an increased risk of miners being struck by moving equipment or errors in communication with co-workers*".

Failure to hear environmental and warning sounds is an additional concern for job safety for miners with subclinical hearing loss (Morata et al., 2005), which compromises audibility and has long been associated with increased risk of injury in a noisy workplace (for review, see Wilkins et al., 1987). In a study focusing on hearing acuity, noise and hearing loss accounted for more than 40% of the injuries occurring in a shipyard (van Charante et al., 1990). The elevated risk of injury when wearing existing commercial HPDs even for persons with normal hearing has also been documented (Choi et al., 2005). In their study of agricultural injuries, the relative risk of injuries to workers wearing HPDs was 2.2, and was independent of their hearing acuity. NIOSH's National Traumatic Occupational Facility Surveillance system records 204 accidental deaths of pedestrians in industry struck by forklifts from 1980 to 1994 (Collins et al., 1999). While the causes of these accidents cannot be deduced, a Fatality Assessment and Control Evaluation (FACE) report of a worker wearing an HPD, who died after being run over by a log loader reversing with its back-up alarm sounding, would appear to be an example of the failure to identify the warning sound (Anon, 1995).

NIOSH in its *Criteria for a Recommended Standard: Occupational Exposure to Noise* identified the "persistent problems" of HPDs, concluding that "*Research should also lead to the development of hearing protectors that eliminate troublesome barriers by . . . improved speech intelligibility and audibility of warning signals"* (Anon, 1998, p. 71). In this study, we have focused on reducing communication problems when the sound source (e.g., talker or warning sound) is in the same environment as a listener, and consequently on methods for improving the intelligibility of face-to-face speech communication and the audibility of warning sounds suitable for users of hearing protection. A successful method would enable the development of improved HPDs that incorporate the appropriate electronic processing of sounds. Such devices could ultimately lead to greater acceptability, and consequently wider use, of HPDs in the workplace, hence reducing the risk of noise-induced hearing loss and accidents involving an inability to talk face-to-face or hear warning sounds.

Shortcomings of Previous Technology Approaches

While there have been numerous attempts to reduce the noise of machinery used in mechanized mining, it is generally recognized that many miners remain potentially overexposed to noise (Babich and Bauer, 2006; Joy and Middendorf, 2007). During the last twenty years, specialized hearing protectors have been developed for situations in which noise levels change in space or with time, such as when walking towards or around a machine or when a nearby vehicle moves or stops operating. In these circumstances, the device automatically adjusts the amount of hearing protection to enable the user to hear more sounds in the environment when there is less environmental noise. These so-called sound level dependent HPDs (sometimes called level dependent HPDs, or sound restoration HPDs) are gaining popularity, and their applicability to mining environments has been studied (Azman & Hudak, 2011). Several have been approved by the Mines Safety and Health Administration for use underground. They employ electronically-controlled sound transmission from the environment surrounding the user to the ear, and for this purpose include a microphone outside the HPD, processing electronics, and a miniature earphone or loudspeaker located underneath the HPD, between it and the ear canal. Unfortunately, at the present state of development, the devices fail to improve the intelligibility of speech in the presence of noises commonly leading to overexposure in a mine (e.g., continuous mining machines, roof bolting machines) compared to when conventional hearing protectors are worn (Azman & Hudak, 2011). The same conclusion has been drawn in other studies using different background noises (Dolan & O'Loughlin, 2005; Plyler & Klumpp, 2003), though not in a survey of HPD preferences in an industrial setting (Tufts et al., 2011). When the background noise is sufficiently loud, a level dependent HPD is designed to cut-off electronically all environmental sounds from the user, so speech and warning sounds as well as the background noise will not be heard. The method developed here is intended to improve communication in all situations.

There have been several attempts reported in the literature to improve face-to-face communication in a noisy environment, mostly intended for application to hearing aids. A recent study described a method for reducing noise (as opposed to *improving intelligibility*) when attempting to communicate face-to-face in a noisy environment for persons wearing HPDs (Lezzoum et al., 2016). The method involved first dividing the frequency spectrum of the sounds (i.e., the combined speech plus environmental noise) into narrow bands of frequencies, commonly termed "subbands". The instantaneous magnitude of the envelope of the signal in each subband at frequencies at which speech sounds are expected to occur were then used to adjust the instantaneous gain applied to the signal in that subband. The subbands signals were then recombined and the processed sounds presented to a listener. The complete process was ongoing, leading to a time-varying gain in each subband based on the envelope of the sound pressure in that subband. Lessoum et al. (2016) observed that listeners reported hearing reduced noise and improved sound *quality* when the speech was initially mostly intelligible. The method was proposed for application to persons wearing hearing protection.

The application of time-varying subband gains to speech processing in noise is a well-established technique that has been described extensively in the scientific literature, though only infrequently for application to hearing protection (Chung et al., 2007, 2009; Lezzoum et al., 2016). Both so-called *modulation*-based methods, as described here, and amplitude-based methods have been proposed (Apoux et al., 2004; Brons et al., 2013; Chung et al., 2009; Clarkson and Bahgat, 1991; Langhans and Strube, 1982; Lorenzi et al., 1999; van Buuren et al., 1999; Wiinberg et al., 2018). Overall, the results have been inconsistent, with some studies finding a small improvement in intelligibility under some conditions of speech SNR (Clarkson and Bahgat, 1991; Lorenzi et al., 1999; Wiinberg et al., 2018), up to ~10% improvement for some SNRs and noises (Apoux et al., 2004; Chung et al., 2009), and other studies finding no improvement (Clarkson and Bahgat, 1991; Langhans and Strube, 1982, van Buuren et al., 1999).

Rationale for the Proposed Approach

We have extensive experience developing modulation-based techniques for *predicting* speech intelligibility in noise, which also require recovering speech sounds from speech buried in noise (Brammer et al., 2010, 2011; Yu et al., 2010, 2014), and for improving speech intelligibility in the communication channel of an HPD or headset worn in environmental noise (Bernstein et al., 2013; Brammer et al., 2013, 2014, 2015). Based on the literature and our experience, we believe that the performance of modulation-based noise reducing methods for improving face-to-face conversation are limited by the routines within the algorithms making the crucial decision whether the signal in a subband contains mostly noise or mostly speech (and hence whether to attenuate or amplify the sound in that subband). This belief has been confirmed experimentally by Wójcicki & Loizou (2012), who pre-mixed initially separate recordings of speech and noise, and then used prior knowledge of the contents of each subband to determine the subband gains. When the processed sounds were recombined and subsequently evaluated by listeners, substantial improvements in speech intelligibility were reported.

Wójcicki & Loizou (2012) considered the results of their experiment to demonstrate the maximum improvement in speech intelligibility obtainable if the ideal gain could be applied to each subband. It should be emphasized that the improvement in speech intelligibility was obtained using *prior knowledge* of the contents of each subband, that is, whether the signal in a subband contained mostly noise or mostly speech (and hence whether to attenuate or amplify the sound in that subband). To follow on from this research, we undertook extensive preliminary work in preparation for this study on how to identify speech when buried in noise *without* prior knowledge of the contents of each subband, as would occur in real-life situations. Under these conditions, we succeeded in simulating a method for improving the intelligibility of face-to-face speech communication and the audibility of warning sounds in environmental noise suitable for persons wearing HPDs.

The results of our initial modulation-based method for separating desired sounds from noise are shown in Figure 1. The top panels of Figure 1 show the time history of the sound pressures produced by a short segment of speech (left) and by a vehicle back-up alarm producing its characteristic repeated tonal burst of sound (right), both in the absence of background noise. The middle panels show the same sounds buried in noise. While the details of the sounds desired to be heard have been lost, close inspection reveals small changes in the envelopes of the combined sounds that may reflect the envelopes of the desired sounds (i.e., compare top and middle panels at the same times). The bottom panels show the desired signals recovered by our method from those in the middle panel, which clearly has markedly improved the SNR both for speech and the warning sound (i.e., compare waveforms in top and bottom panels).

We believe the predicted success of our initial method for separating desired sounds from noise warranted further study. Hence we proposed to study it and related ways for detecting whether a mixture of desired sounds and noise contains more desired sounds than noise, and so should be amplified and transmitted to a listener, or vice versa. Of equal import is for human subjects to confirm the predicted improvement in speech intelligibility in controlled listening tests. While it would be desirable for the performance of warning sounds to be evaluated by listeners, the effectiveness of the algorithms will be judged by changes in the SNR, which is directly related to the audibility of common tonal warning sounds (e.g., vehicle back-up alarm), for the reason given in the *Mission Statement*.

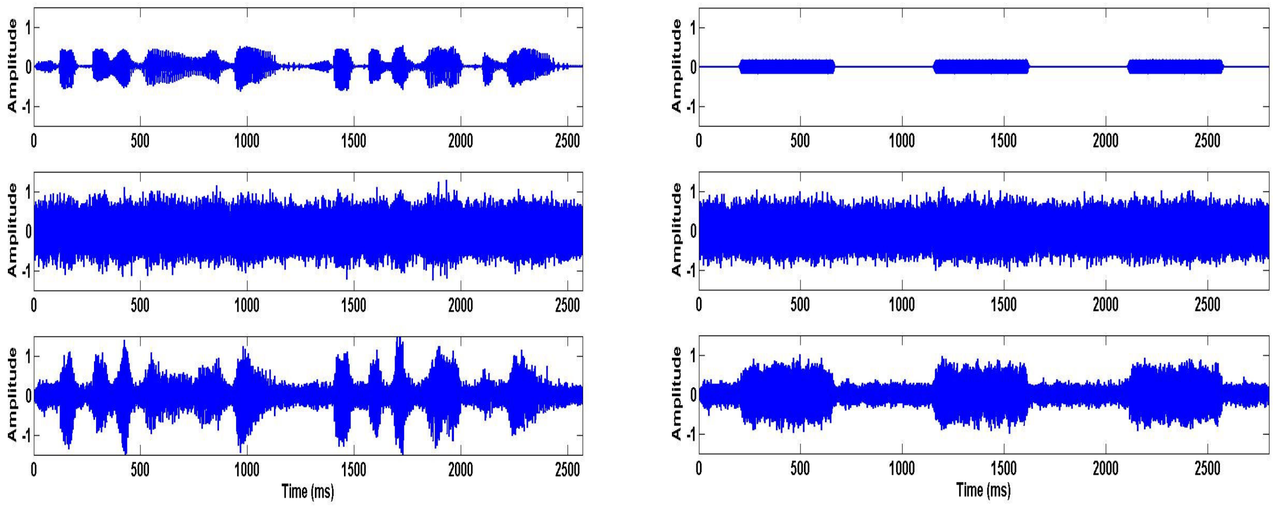


Figure 1: Sound pressure time histories of speech (left) and a vehicle back-up alarm (right) showing waveforms without interfering noise (top panel), with white environmental noise (middle panel), and speech and alarm signals recovered by signal processing from the signals in the middle panels (bottom panel).

4.0 Proof-of-Concept Technology Components

Design Strategy

Our preliminary studies suggested that the performance of modulation-based methods for distinguishing desired sounds from noise can be improved by attention to the part of the algorithm making the crucial decision whether the signal in a subband contains mostly noise or mostly desired sounds (and hence whether to attenuate or amplify the sound in that subband). Experiments conducted elsewhere, in which speech and noise were available separately and could be combined with prior knowledge of which subbands to amplify and which to attenuate, supported this conclusion (Wójcicki & Loizou (2012). We therefore proposed to continue this approach to optimize, and confirm, a modulation-based method for improving the intelligibility of face-to-face speech communication in noise and the audibility of tonal warning sounds applicable to wearers of electronic HPDs.

The primary goal of the research was to distinguish desired sounds buried in environmental noise from the noise using a filter bank consisting of a set of band-pass filters arranged in parallel with overall bandwidth compatible with that of speech sounds. The work will focus on the influence of subband bandwidths, parameters influencing envelope-derived modulations, signal and envelope fidelity, the relationships between signals in different subbands, and minimizing temporal fluctuations in sound pressure introduced by the digital signal processing, i.e., so-called "musical noise" (Apoux and Bacon, 2004; Arehart et al., 2015; Clarkson and Bahgat, 1991; Drullman et al., 1994; Esch & Vary, 2009; Greenberg et al., 2003; Noordhoek and Drullman 1997; van Buuren et al., 1999; van Schijndel et al., 2001). The signal processing will be undertaken using MATLAB, and the sounds processed off-line. Time histories suitable for use in the Modified Rhyme Test (MRT) (House et al., 1965), a psychophysical test of speech intelligibility, will be constructed and stored in wave files, for subsequent evaluation by predictive methods and for replaying to subjects wearing headphones or insert earphones.

For single-tone warning sounds, such as a vehicle back-up alarm, the audibility of the alarm is dependent on the SNR (Zheng et al., 2007), as already noted. This will be evaluated for a given noise environment, and the improvement in SNR obtained by signal processing taken as the metric of increased audibility.

Overview of Technology System

We initially proposed to confirm the concept in the laboratory using an active electronic circumaural HPD constructed for a previous project as the "test bed" (Bernstein, 2013), as already noted. However, as reported in our Interim Progress Report (August, 1 2017 - January 31, 2018), the capacity of the digital signal processor was only sufficient to construct the sixteen subbands, which, while an essential prerequisite to modifying the subband signals, did not allow for any signal manipulation. Also as noted in the Interim Progress Report, the fidelity of sound reproduction by the custom-built HPD led to questions concerning its suitability for evaluating algorithms designed to improve speech intelligibility. It thus became apparent the complexity of the algorithms would be such that the custom-built electronic HPD could not support the necessary signal processing in real time, and the fidelity of sound reproduction by the device could prove a limiting factor in listening tests. Accordingly, for these reasons, the proof-of-concept evaluation of speech buried in noise consisted of listening tests conducted using wave files constructed off-line by MATLAB when the subject wore commercial high-fidelity headphones (Sennheiser HD 580) or insert earphones (E·A·RTone type 3A or 5A).

In consequence, the components employed for the proof-of-concept evaluation were ad hoc and will not form part of a future working prototype. The primary tasks of this study to complete the proof-of-concept were thus to focus on developing the algorithms that are needed to progress from simulation to evaluation of the concept: 1) for speech, by listeners with normal hearing, and 2) for a tonal warning sound, by observing the SNR.

5.0 Proof-of-Concept Evaluation

Overview of speech intelligibility evaluation

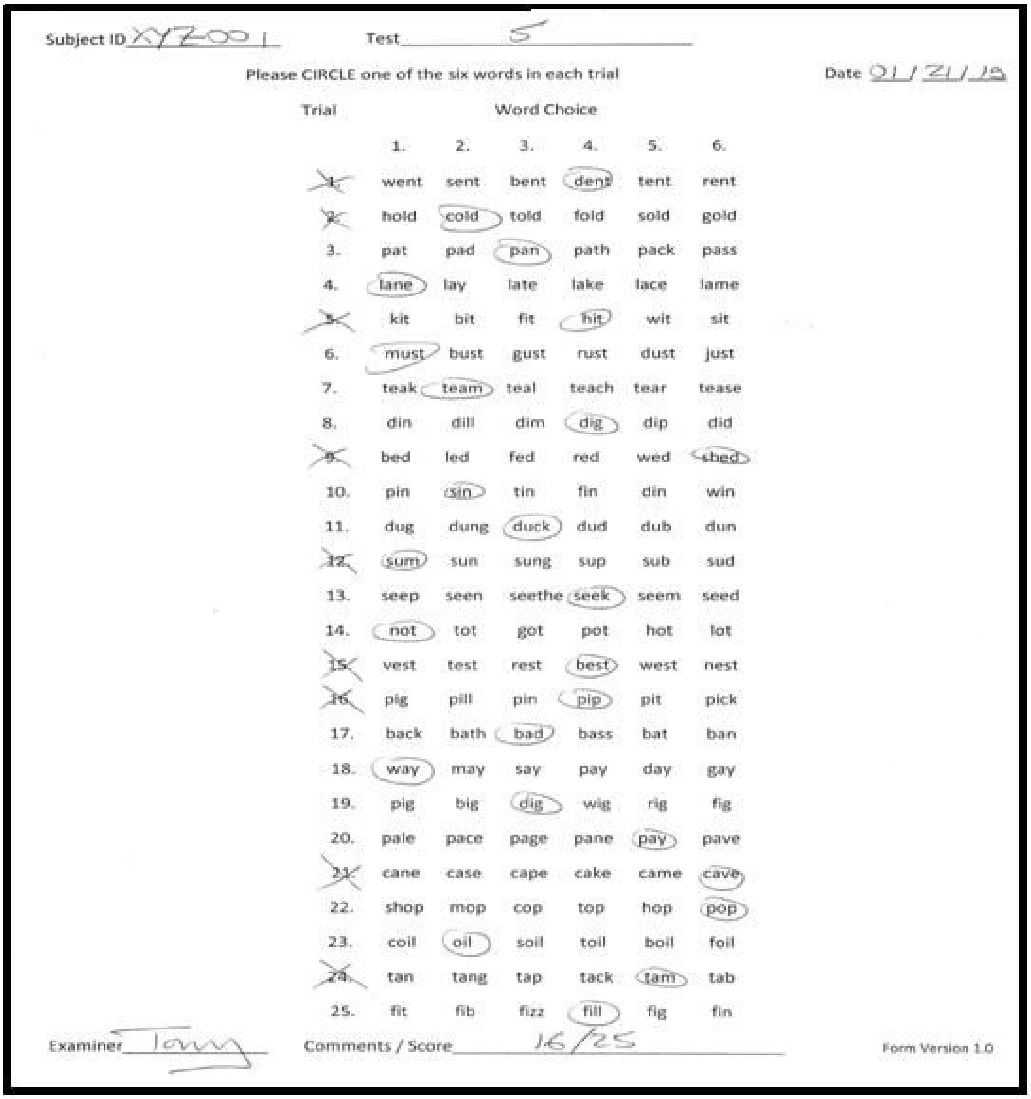
During development of the method for improving speech understanding in noise, the performance of algorithms were first obtained by prediction, as it avoids the need for human subjects and so provides results more rapidly. We have previously developed a method for this purpose, namely the speech-stimulus Speech Transmission Index (STI) (Brammer et al., 2010, 2011; Payton and Braida, 1999; Yu et al., 2010, 2014), and have extensive experience qualifying and using this metric as a measure of speech intelligibility in a range of noisy environments and speech distortions. It also enables the performance to be predicted in intense noise environments in which it is ethically unacceptable to expose human subjects (Brammer et al., 2014). The speech-stimulus STI yields similar predictions of word scores obtained in the MRT as the traditional STI in circumstances in which the latter does not fail (Brammer et al., 2011; Goldsworthy and Greenberg, 2004; IEC 60628-16, 2003).

Promising algorithms were then evaluated by formal listening tests. After consenting to participate in the study, volunteers were asked to undergo an induction procedure. This included completing a brief questionnaire about their hearing function (see Appendix *Hearing Function Questionnaire*), a medical history related to hearing, an examination of their ears and the area surrounding the ears, and asking them whether they are allergic to the plastic materials used in earphones, headphones or HPDs. The examination was followed by the initial listening tests, which were used to confirm the suitability of the volunteer for undertaking formal tests. They consisted of identifying out-of-context words presented in quiet or in background noise and, in some cases, audiometric determination of hearing thresholds using established clinical procedures (see, for example, Schlauch and Nelson, 2009). The sound and noise intensities used were chosen not to produce any effect on hearing acuity.

The listening tests were conducted in our clinic/laboratory at UConn Health by a member of the research team, and are described in more detail below. All measurements were conducted in an audiometric booth, which fulfills the ambient noise requirements for audiometric threshold determinations in ANSI S3.1, 1999. For all listening tests, volunteers were seated comfortably.

Word intelligibility in noise - Modified Rhyme test (MRT)

The MRT was used to characterize the intelligibility of individual words in a six-alternative forced choice paradigm (ANSI, 2009; House et al., 1965). This test of consonant confusion has been used extensively for its relevance to critical communications in which a single word error could have serious consequences (e.g., air traffic control, military and first responder operations) (Cardosi, 1998; Anderson et al., 1997; LaTourette et al., 2003). The word lists were those standardized for American English as spoken by a male talker (Anon, 2016). A trial consists of one of six words written on paper being randomly replayed within a carrier phrase, e.g., "Please select the word … (insert test word)". Subjects were instructed to circle the word



Subject ID XYZ001 Test 5

Please CIRCLE one of the six words in each trial Date 01/21/19

Trial	1.	2.	3.	4.	5.	6.
1.	went	sent	bent	dent	tent	rent
2.	hold	cold	told	fold	sold	gold
3.	pat	pad	pan	path	pack	pass
4.	lane	lay	late	lake	lace	lame
5.	kit	bit	fit	hit	wit	sit
6.	must	bust	gust	rust	dust	just
7.	teak	team	teal	teach	tear	tease
8.	din	dill	dim	dig	dip	did
9.	bed	led	fed	red	wed	shed
10.	pin	sin	tin	fin	din	win
11.	dug	dung	duck	dud	dub	dun
12.	sum	sun	sung	sup	sub	sud
13.	seep	seen	seethe	seek	seem	seed
14.	not	tot	got	pot	hot	lot
15.	vest	test	rest	best	west	nest
16.	pig	pill	pin	pip	pit	pick
17.	back	bath	bad	bass	bat	ban
18.	way	may	say	pay	day	gay
19.	pig	big	dig	wig	rig	fig
20.	pale	pace	page	pane	pay	pave
21.	cane	case	cape	cake	came	cave
22.	shop	mop	cop	top	hop	pop
23.	coil	oil	soil	toil	boil	foil
24.	tan	tang	tap	tack	tam	tab
25.	fit	fib	fizz	fill	fig	fin

Examiner Tony Comments / Score 16/25 Form Version 1.0

Figure 2: Example of a completed 25-trial Modified Rhyme test.

chosen on the prepared word list. There were 25 trials in each test from which a word score (i.e., number of words correct) was derived for the preset speech SNR, which was the mean SNR of the target words. An example of a completed test is shown in Figure 2.

The SNR was selected to produce word scores in the range 30 to 70% words correct in the reference condition without signal processing. Tests were also conducted without noise. A successful outcome for the proof-of-concept would be obtained by an increase in word score when an algorithm is employed compared to when it is not employed, thus demonstrating improved speech intelligibility. The statistical test of significant difference between the word scores for the two conditions was a two-sided t-test (see, for example, Bland, 2000).

Human Subjects

Seventeen healthy volunteers were invited to our clinic/laboratory, to undergo the induction procedure. Criteria for inclusion in the study were: 1) age between 18 and 50 years; 2) absence of factors that could influence the performance or acceptability (e.g., comfort) of insert earphones, or circumaural headphones; 3) no infections of the skin or external ear; 4) no impacted cerumen; 5) no middle ear infections; 6) tolerance to plastics on the skin; 7) hearing

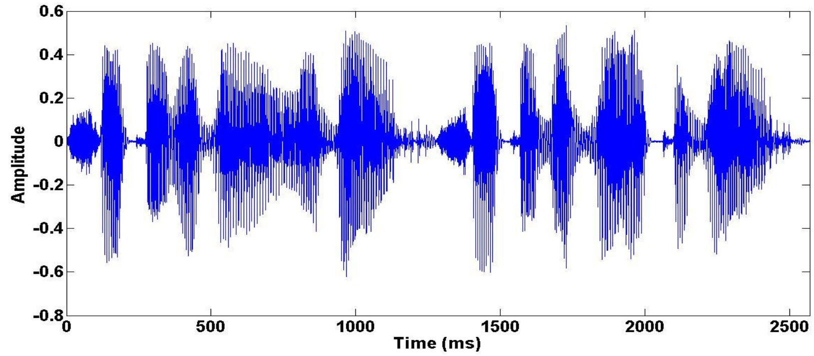


Figure 3: Sound pressure waveform of a segment of repeated speech from a speech test. (The words are: "Circle the - [test word] - now, circle the - [test word] - now").

function questionnaire score less than 6 (see Appendix *Hearing Function Questionnaire*); 8) when measured, pure-tone air conduction Hearing Threshold Levels (HLs) re ANSI S3.6-1996 of less than (i.e., more sensitive than) 20 dB HL at frequencies of 0.25, 0.5, 1, 2, 4, 6 and 8 kHz, with thresholds of individual ears that differ by less than 10 dB (Schlauch and Nelson, 2009); 9) absence of persistent tinnitus, and; 10) ability to identify similar sounding words spoken out of context in American English, read the words, and signal the word selected by completing a written form.

Audiometric evaluation of hearing thresholds was conducted on one-third of the subjects to confirm that the MRT word scores were representative of those to be expected from persons with normal thresholds. A statistical test of differences between mean word scores obtained by subjects whose audiometric thresholds were less than 20 dB HL and who reported normal hearing function compared to subjects who also reported normal hearing function but did not undergo audiometric tests found no difference between the results in 90% of MRTs conducted (two-sided t-test). When different, the results for subjects who did not undergo audiometric tests tended to reduce the probability of an algorithm producing a statistically significant improvement in intelligibility (i.e., bias to a null effect), and so they were included in the data analysis.

Algorithm Development

The sound pressure waveform (time history) of a short segment of repeated speech is shown in Figure 3. The similarity between the signal in the first 500 ms and from 1250 - 1750 ms, as well as around 1000ms and 2300 ms results from the same words being spoken. Equally visible are the differences when a different test word is inserted in the middle of the phrase (i.e., from 500 to 700 ms versus from 1800 to 2000 ms). The purpose of this work is to develop an algorithm for identifying characteristics such as these when the sounds are buried in environmental noise. The strategy adopted here is to focus on the *changes* in amplitude with time as is done visually by an observer looking at Figure 3. The signal processing establishes the envelope of the signal, which when subjected to low-pass frequency filtering forms the green

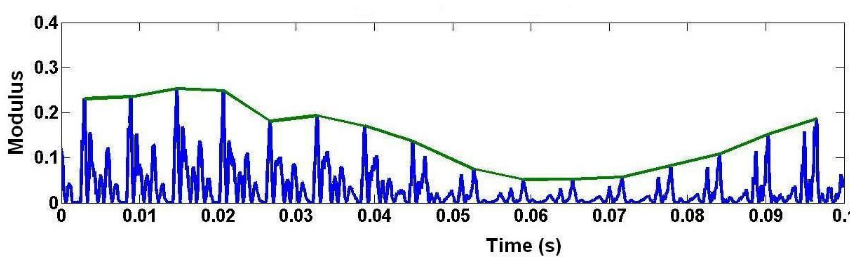


Figure 4: Example of the low-frequency envelope (green curve) of the modulus of a speech signal shown as a function of time (blue curve).

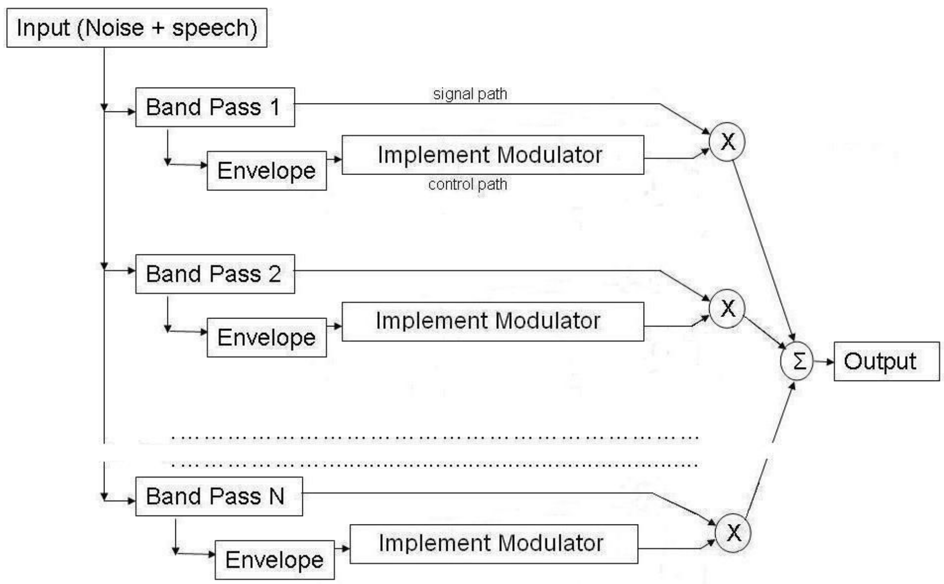


Figure 5: Concept block diagram of algorithms.

time history shown in Figure 4. It is evident from Figure 4 that increasing the low-pass filter cutoff frequency would allow the small amplitude peaks to be registered by the envelope of the signal. Our methods for distinguishing speech from noise employ the detailed characteristics of the envelope fluctuations. We have observed that the frequency spectrum of the envelope fluctuations is distinctly different for speech, tonal warning sounds and environmental noises, with important speech modulations at frequencies between about 2 and 20 Hz (Greenberg et al., 2003). These differences are detected and exploited by our algorithms.

The basic concept of our algorithms is shown in Figure 5. The input to the algorithm, identified as speech buried in environmental noise but could equally be a warning sound buried in noise, is shown to the left of the diagram. The first step involves a set of N band-pass filters arranged in parallel. These create N subbands that together span the frequency range in which speech and warning sounds occur. The frequency range chosen is from 200 Hz to 6 kHz.

Within each subband there is a signal path and a control path, as identified in the diagram for subband #1. The signal path contains the components of speech or warning sound buried in environmental noise with frequencies that fall within the range of the corresponding band-pass filter. The envelope of this signal is constructed in the manner already described (see Figure 4), and feeds the control path. Within the control path is the signal processing that manipulates the envelope to construct the modulation to be applied to the signal path. Amplitude modulation of the speech or warning sounds plus environmental noise within a subband is shown by the "X" in Figure 5. While the treatment of the envelope in the modulators is similar for each subband, the details depend on the frequency, which serves, in part, to distinguish our approach from those reported elsewhere. Finally, the outputs of all subbands are combined (shown by Σ in the diagram), and the reconstructed speech or warning sounds in environmental noise can be replayed to a listener or evaluated by predictive methods.

There is a large number of variations in signal processing that could be applied to the envelope in the modulator, so selecting those to implement requires a basis for informed decisions. Approaches that have been explored by others provide some information, but this is insufficient to define an improved modulation strategy as most previous attempts failed or produced small improvements in intelligibility. A significant contribution to solving this problem is provided by our model for predicting speech intelligibility in noise, namely the speech-stimulus STI. In interpreting its predictions, however, it is important to recognize that the model calculates speech SNRs of signal envelopes, and so is not sensitive to the phase relationships between signals. This limitation restricts predicting outcomes in some situations.

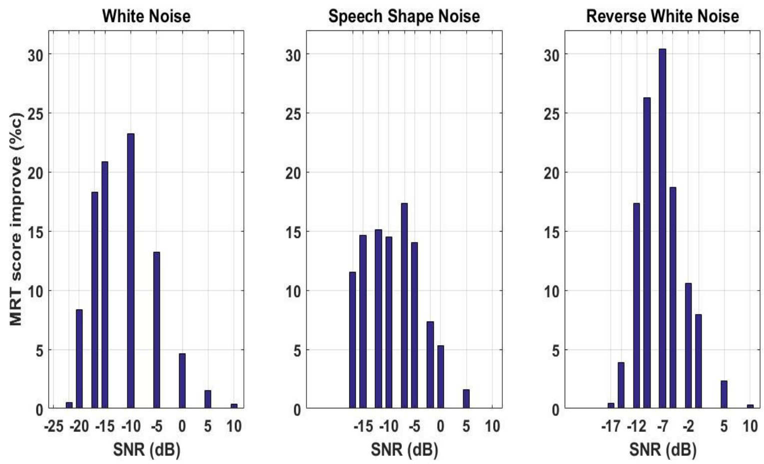


Figure 6: Predicted increase in MRT word scores (percent words correct, %c) for three environmental noises and different speech SNRs, for an algorithm employing linear envelope modulation with unity modulation gain control.

Simulations have been prepared to predict the influence on MRT word scores of subband bandwidths, parameters influencing envelope-derived modulations, signal and envelope fidelity, and the relationships between signals in different subbands. Predictions were made for speech sounds buried in three different environmental noises: white (Gaussian) noise, i.e., noise with a flat frequency spectrum; speech-spectrum shaped noise, i.e., noise with a frequency spectrum shaped to replicate the long-term average of speech, and; so-called "reverse white noise" which is white noise with a frequency spectrum that decreases by 3 dB per octave increase in frequency. The different frequency spectra were chosen to be representative of the range of those that could occur, there being no single spectrum that appears to be typical of those occurring in underground mines (Åsman and Hudak, 2011; Camargo et al., 2009, 2016). The results of the predictions were as follows.

i) Subband bandwidth. When subband bandwidths range from that corresponding to the bandwidth of an auditory filter in the cochlea to an octave, there is little influence of bandwidth on predicted word scores provided the bandwidth is no greater than one-third octave. This implies that at least sixteen subbands will be required for the total bandwidth of the filter bank to encompass the desired frequency range of 200 to 6000 Hz.

ii) Envelope-derived modulation. Linear modulation of the subband signal path derived from the envelope, as shown in Figure 4, has been employed in previous research (e.g., Lessoum et al., 2016) as well as binary modulation, that is, modulation that is either "off" on "on" (e.g., Wójcicki & Loizou, 2012). The relative merits of linear versus binary modulation were studied for white noise, speech-spectrum shaped noise and reverse white noise. The predicted word scores were consistently higher for linear rather than binary modulation for the threshold condition studied (speech energy > 1.5 x noise energy). They also depended on the spectrum of the environmental noise, as can be seen in Figure 6, which contains predictions for linear modulation. The smallest maximum increase in word score was obtained for speech-spectrum noise, doubtless because the spectrum of speech is effectively the same as that of the noise, which impedes discrimination by frequency content for both listeners and the speech-stimulus STI. Thus, attempting to improve speech understanding in speech-spectrum noise will represent the most difficult challenge of the three, and so was adopted for all future work.

iii) Signal and envelope fidelity. This subject has been implicitly treated by paragraph ii), from which the conclusion is that fidelity needs to be maintained for the modulations studied. Predictions also indicated that digital filters with infinite impulse response (IIR) would perform better than those with finite impulse response (FIR) for this application. However, this conclusion was discounted as it is known that, in contrast to FIR filters, IIR filters introduce frequency-dependent time delays, which would not be detected by the speech-stimulus STI

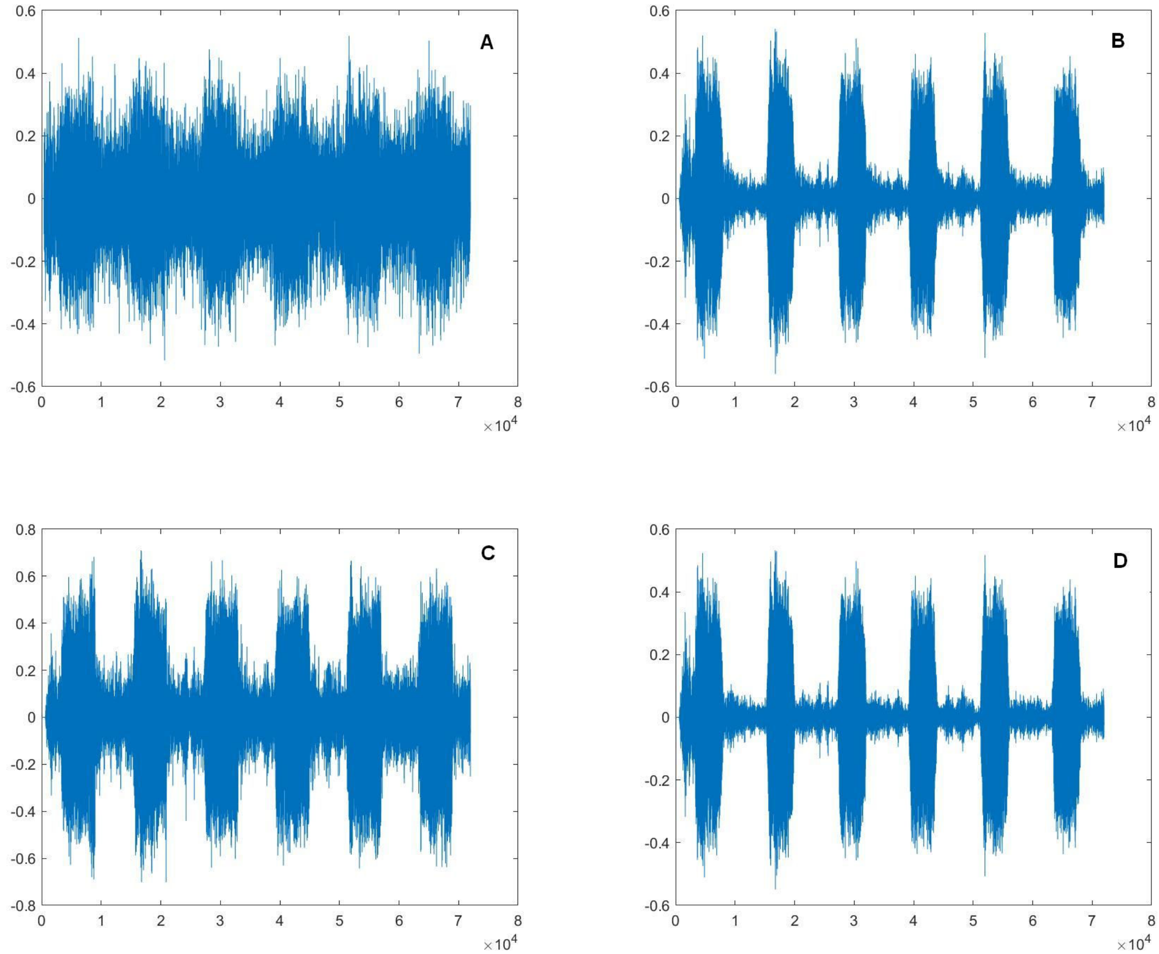


Figure 7: Sound pressure time histories (Amplitude versus Time) of an intermittent, 1000 Hz, vehicle back-up alarm in speech-spectrum shaped noise, A - unprocessed; B - processed by algorithm #1; C - processed by algorithm #2, and; D - processed by algorithm #3.

predictive model. Moreover, the relative timing of signals is an important feature of all algorithms that needs to be optimized to maintain signal and envelope fidelity.

iv) Relationships between signals in different subbands. Our model for calculating the speech-stimulus STI employs a correction for the correlation between speech sounds in adjacent subbands (Brammer et al., 2011). We were unable to confirm the need for such terms in the current predictions, which may be more a consequence of the environmental noises selected for this study rather than the absence of the interaction.

A number of algorithms were constructed as a result of the simulations. Additionally, numerous variations of each algorithm were evaluated by informal listening over approximately a six-month period as attempts were made to fine-tune each version. Three variations of what may be described as our "basic" algorithm have been evaluated by formal listening tests as well as the capacity to reduce noise. The algorithms contained sixteen subbands with frequencies extending from 200 to 6000 Hz, and are distinguished as follows.

#1. linear modulation, with linear gain control of the subband modulation derived from the subband envelope fluctuations.

#2. linear modulation, with unity gain control of the subband modulation.

#3. linear modulation, with nonlinear gain control of the subband modulation derived from the subband envelope fluctuations.

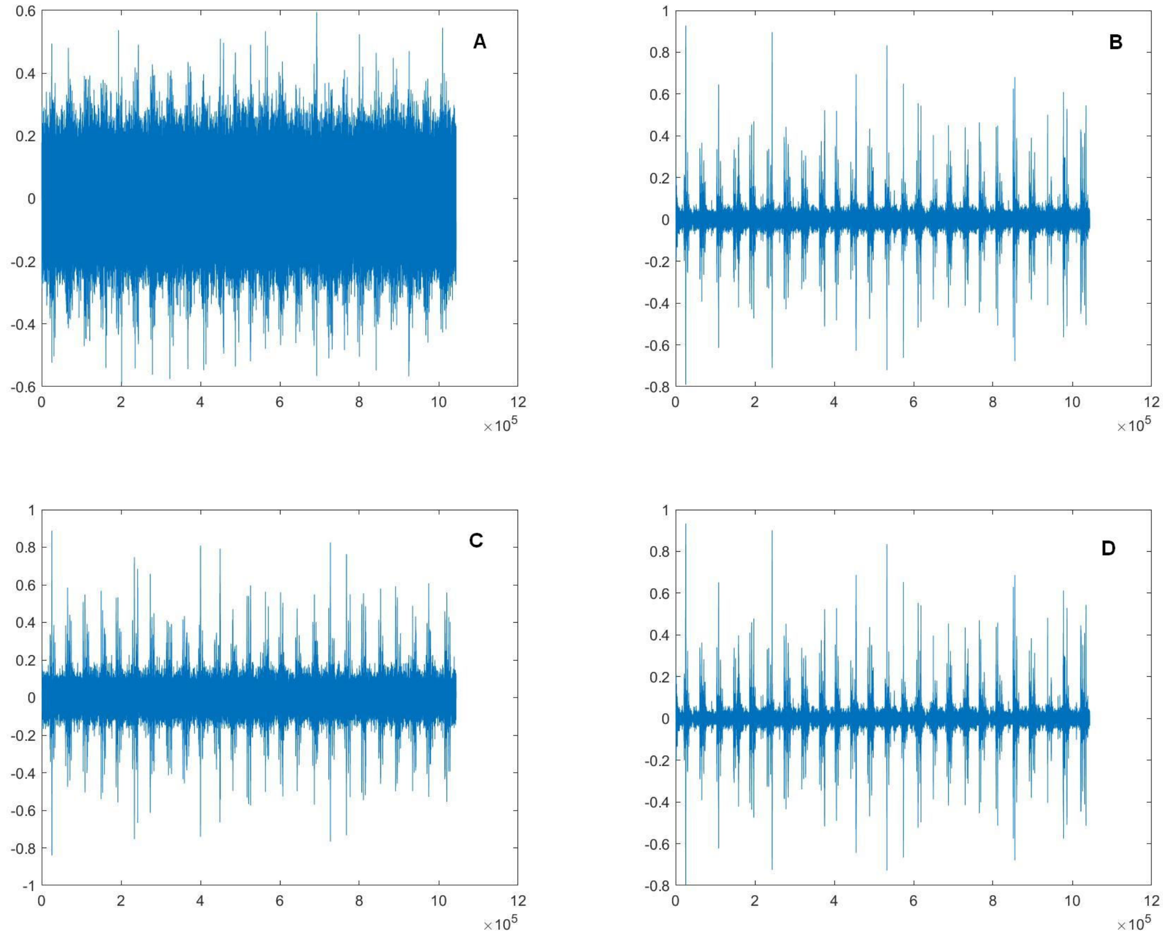


Figure 8: Sound pressure time histories (Amplitude versus Time) of speech in speech-spectrum shaped noise. The speech consists of 25 sentences, each separated by ~3 s, "Please select the word . . (insert test word)." A - unprocessed; B - processed by algorithm #1; C - processed by algorithm #2, and; D - processed by algorithm #3.

Environmental Noise Reduction

An essential feature of the performance of any method for improving the audibility of warning sounds or the intelligibility of speech in environmental noise is to reduce the intensity of the noise more than that of the sounds desired to be heard. The performance of the three algorithms when an intermittent, 1000 Hz, vehicle back-up alarm is broadcasting its familiar "beep - beep - beep" sounds of ~0.5 s duration in an environment containing speech-spectrum shaped noise is shown in Figure 7. The time history of the combined alarm sounds and noise in the environment are shown in the top left hand graph (Figure 7A), where six "beeps" are just visible. Processing these sounds by algorithm #1, shown in Figure 7B, enables the alarm sounds to be resolved and the SNR substantially improved. The same observation can be made for the other algorithms, shown in Figures 7C and 7D, though it is clear that they result in different SNRs after the signal processing. As the goal of noise reduction is to maximize the improvement in SNR after processing, the performance of the algorithms would be ranked (best to worst): first - algorithm #3, second - algorithm #1, and third - algorithm #2.

Clearly, all algorithms increase the SNR and hence demonstrate that the audibility of a tonal alarm will be improved. As all three algorithms are optimized for improving speech intelligibility in environmental noise rather than improving warning signal audibility in noise, these results accomplish the second part of our mission statement for such alarms, namely, to ensure that algorithms developed for enhancing speech improve the audibility of warning sounds in noise.

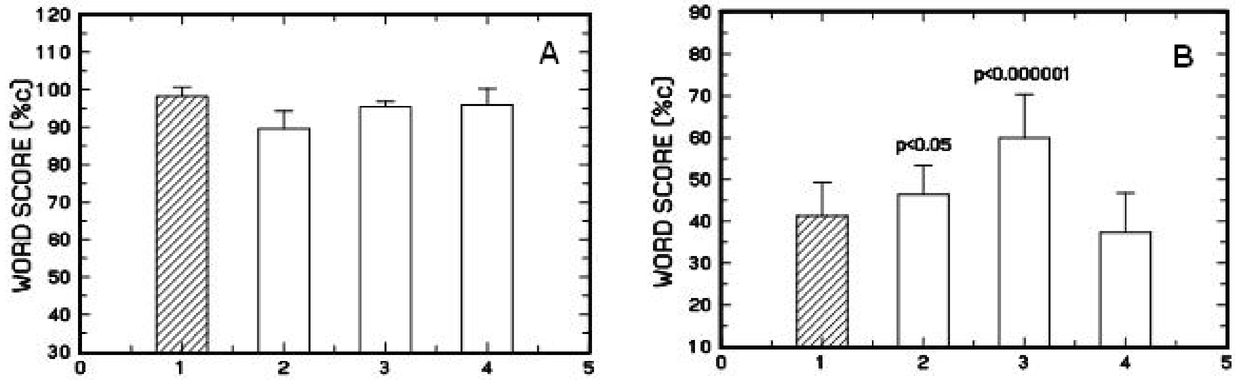


Figure 9: Mean and standard deviation word scores (percent words correct, %c): A - speech with no noise, B - speech in speech-spectrum noise at SNR of -8 dB. Column 1 - unprocessed (shaded); column 2 - processed by algorithm #1; column 3 - processed by algorithm #2, and; column 4 - processed by algorithm #

While, as just noted, the algorithms were designed to improve speech intelligibility in noise, it is instructive to examine their ability to reduce environmental noise. The performance of the three algorithms when a male talker speaks 25 sentences, each separated in time by about 3 s, in an environment containing speech-spectrum shaped noise is shown in Figure 8. The sentences consisted of a carrier phrase to which is added a test word selected from the MRT lexicon, i.e., "Please select the word . . (insert MRT word)." The sound pressure time history of the combined speech and noise in the environment before signal processing is shown in the top left hand graph (Figure 8A), where the peak speech amplitudes are visible. Processing these sounds by algorithm #1, shown by the time history in Figure 8B, enabled the speech sounds to be resolved and the SNR substantially improved. As was the case for warning sounds, the same observation can be made for the other algorithms, shown in Figures 8C and 8D, though it is clear that they result in different SNRs after the signal processing. If the goal were to maximize the improvement in speech SNR after processing, the performance of the algorithms would again be ranked (best to worst): first - algorithm #3, second - algorithm #1, and third - algorithm #2.

Speech Intelligibility

The primary objective of all our algorithms is to improve speech intelligibility in noise. This was evaluated by the Modified Rhyme test (MRT), which has been used extensively by ourselves and others to characterize the intelligibility of individual words in a six-alternative forced choice paradigm. Seventeen subjects participated in the listening tests.

While the focus is on establishing the performance of the algorithms in noise, it is also important that they do not distort speech sounds in the absence of noise. This performance requirement appears trivial to a human listener, but is almost as much a challenge to some algorithms as improving speech intelligibility in noise. The results of MRTs performed in the absence of environmental noise, expressed as the percentage number of words correct (%c), are shown in Figure 9A for each algorithm and for unprocessed speech. The mean word score for unprocessed speech, shown by the shaded bar, was 98.2% with a standard deviation of 2.5%. Algorithms #2 and #3 resulted in mean word scores that were close to the value for unprocessed speech (95.5% and 96%, respectively), while the score for algorithm #1 was somewhat lower (89.6%). When compared to the word scores for unprocessed speech, none of the word scores for the three algorithms reached statistical significance as assessed by a two-sided t-test. However, there appeared to be a tendency for the algorithms to produce slightly lower word scores than when the speech was unprocessed, suggesting that they may distort speech somewhat. This was particularly evident for algorithm #1.

The results of the MRT performed in speech-spectrum shaped noise are shown in Figure 9B. The listening tests were performed with a speech SNR of -8 dB. The mean word score for unprocessed speech, shown by the shaded bar, was 41.3% with a standard deviation of 8%. Processing the speech in noise by algorithms #1 and #2 resulted in mean word scores that were greater than the value for unprocessed speech (46.4% and 60%, respectively), while the score for algorithm #3 was somewhat lower (37.3%). The increased word scores for algorithms #1 and #2 compared to the value for unprocessed speech reached statistical significance as assessed by a two-sided t-test, with p-values indicated in Figure 9B.

The improvement in speech intelligibility produced by algorithm #2 in this test (18.7%) is greater than any reported elsewhere, to the best of our knowledge. This performance, together with: i) the lack of an statistically significant effect on the intelligibility of speech without noise (Figure 9A), ii) an improvement in warning sound SNR, which will increase the audibility of the tonal alarm (Figure 7C) and, iii) a modest reduction in environmental noise containing speech or warning sounds (Figures 7C and 8C), form the outcomes that demonstrate proof-of-concept.

As the goal of all our algorithms is to maximize speech intelligibility in noise, the performance of the algorithms would be ranked (best to worst): first - algorithm #2, second - algorithm #1, and third - algorithm #3. This ranking replaces the first ranked algorithm for noise reduction by the third. It also demonstrates that the ability of an algorithm, which has been designed for improving speech intelligibility in environmental noise, to reduce the noise is not necessarily an indicator of its ability to improve intelligibility. This observation highlights limitations of the speech-stimulus STI, which did not predict these differences, for estimating the performance of intelligibility enhancing algorithms.

It should be noted that the performance of our algorithms including variations of those discussed remains to be established for different talkers (both male and female), different environmental noises, different sound levels, and different SNRs. The evaluations will need to be by listeners with normal hearing, and with mild hearing loss. It cannot be assumed at this time that any of the algorithms will perform acceptably under all conditions without further development.

6.0 Technology Readiness Assessment

We judge the Technology Readiness Level of the proof-of-concept to be level TRL 5 ("Technology demonstration while under development") of the nine-unit version of the NASA scale (see https://en.wikipedia.org/wiki/Technology_readiness_level). As noted above, the performance of our algorithms including variations of those discussed, while promising, remains to be established for different talkers (both male and female), different environmental noises, different sound levels and different SNRs. Clearly, more research is needed to resolve these issues.

As already noted, the differences in the performance of algorithms were not predicted by the speech-stimulus STI. Consequently, all future changes to algorithms will need to be assessed by listening tests and it would be unwise to underestimate the effort that will be needed for this enterprise. This is in view of the time required to reach the present state of development, which exceeded the time and effort estimated at the commencement of the study on the basis of our prior pilot work. Nevertheless, the achievement of better speech intelligibility under the conditions of measurement described than ever reported elsewhere represents a significant milestone that should not be overlooked. Indeed, attempts to achieve the goals of our project have previously attracted some of the best minds in the field for over thirty-five years with limited success (see *Previous Technology Approaches*). A disclosure of the invention is being made to the university.

Even after an algorithm is developed that can increase the intelligibility of speech spoken in a noisy environment without interfering with the audibility of warning sounds, it must be transferred to electronics capable of microminiaturization. The computational complexity of the algorithm in its present form presents a challenge for implementation within a small, lightweight package that could function effectively throughout a work shift and be worn as part of a miner's

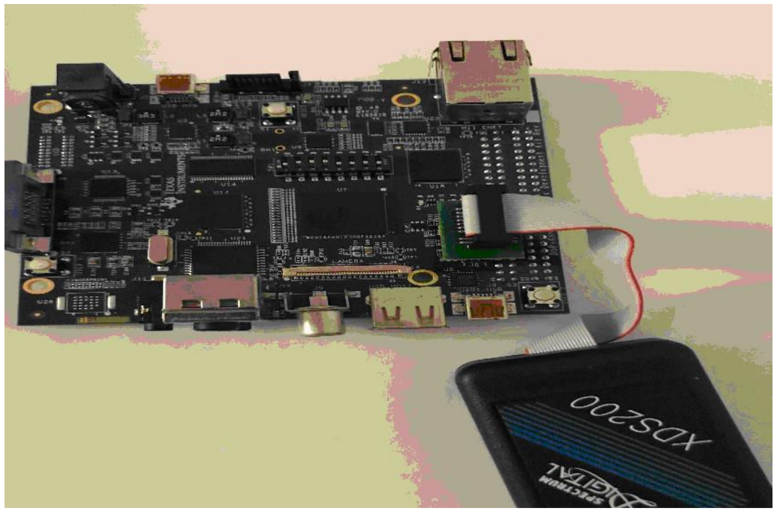


Figure 10: DSP development board (LCDK6748) with programmer (XDS200).

equipment, or attached to, or integrated into, a miner's helmet. Ultra low-powered digital signal processors (DSPs) or field-programmable gate arrays (FPGAs) required for a body-worn or helmet-mounted device will need to be identified for this application. An example of an updated, low-powered version of the DSP employed in our custom-built electronic HPD is shown as part of a development board that includes numerous peripherals, not all of which would be required, in Figure 10 (upper unit). The DSP consumes 80% less electrical power than that used in our custom-built electronic HPD, yet its computational power is increased by a third compared to the original device. The increased performance allows it to come close to implementing algorithm #2 in real time using a "brute force" software approach, which will not necessarily involve the most efficient use of computational resources. While the necessary performance can always be obtained by employing a second DSP, an FPGA may be more suited to this application in view of the amount of parallel processing (i.e., sixteen parallel channels). Coding the device(s) selected with the algorithm(s) to provide the desired performance will require expertise and time.

In view of the expectation that any electronic HPD employing our algorithms may be used in an underground mine, all aspects of the design must involve intrinsically safe circuitry and components. We note again that several commercial electronic HPDs have already been approved by the Mines Safety and Health Administration for use underground.

While our research has focused on improving speech intelligibility and included evaluating a tonal alarm, it has not generally considered warning sounds. It is possible that the audibility of non-tonal warning sounds in the environment, such as "roof talk", will not be maintained by the speech-focused algorithms. Should this prove to be the case, we could introduce a second algorithm operating in parallel with the speech-focused algorithm. The alarm-focused algorithm needs only to respond to specific non-tonal sounds that are known to present hazards in a given mine. We have previously developed, implemented using our custom-built electronic HPD, and evaluated by listening tests such an algorithm for this application (Bernstein et al., 2014). It features a memory that contains an example of the warning sound. The algorithm compares sounds in the environment to the recorded example, and selects, amplifies and transmits all similar sounds to the wearer of the HPD. This algorithm can function independently of our speech-focused algorithms, while sharing the microphone external to the HPD used to sense sounds in the environment, modest computational resources, and the loudspeaker or earphone.

7.0 Appendices

References

F. Apoux and S. P. Bacon, "Relative importance of temporal information in various frequency regions for consonant identification in noise," J. Acoust. Soc. Am. **116**, 1671-1680 (2004).

F. Apoux, N. Tribut, X. Debruille and C. Lorenzi, "Identification of envelope-expanded sentences in normal-hearing and hearing impaired listeners," Hear. Res. **189**, 13-24 (2004).

B. W. Anderson and G. R. Garinther, "Effects of active noise reduction in armor crew headsets," *Audio Effectiveness in Aviation,* AGARD-CP-596, pp. 20-1 - 20-6, (1997).

Anon., "Log loader runs over worker at stave mill," KY FACE Report 94KY161 (1995).

Anon., "Modified Rhyme Test Audio Library," National Institute of Standards and Technology (2016). www.nist.gov/ctl/pscr/modified-rhyme-test-audio-library.

Anon., "NIOSH Criteria for a Recommended Standard: Occupational Exposure to Noise," DHHS (NIOSH) Publication 98-126, National Institute for Occupational Safety and Health, Cincinnati (1998).

ANSI S3.1-1999 (R2003), "Maximum permissible ambient noise for audiometric test rooms," American National Standards Institute, New York (2003).

ANSI S3.2-1989 (R2009), "Method for measuring the intelligibility of speech over communication channels," American National Standards Institute, New York (2009).

ANSI S3.6-1996, "American National Standard specification for audiometers," American National Standards Institute, New York (1996).

K. Arehart, P. Souza, J. Kates, T. Lunner and M. Pedersen, "Relationships among signal fidelity, hearing loss, and working memory for digital noise suppression," Ear Hear. **36**, 505-516 (2015).

A. A. Azman and R. L. Hudak (2011), "An evaluation of sound restoration hearing protection devices and audibility issues in mining," Noise Control Eng. J. **59**, 622-630.

D. R. Babich and E. R. Bauer, "Summary of longwall and continuous miner section noise studies in underground coal mines," Min. Eng. **58**, 41-46 (2006).

E. R. Bernstein, "Development of a delayless subband active noise reducing hearing protector for improvement of speech intelligibiliy and detection of auditory warning sounds," Ph. D. Dissertation, University of Connecticut, Storrs CT (2013).

E. R. Bernstein, A. J. Brammer and G. Yu, "Augmented warning sound detection for hearing protectors," J Acoust. Soc. Am. **135**, EL29-EL34 (2014).

M. Bland, *An introduction to medical statistics*, Oxford University Press, Oxford U.K. (4th edition, 2015).

A. J. Brammer, G. Yu and E. R. Bernstein, "Communicating in noise when wearing hearing protection: Issues and approaches", Proc. 22nd Int. Congress Sound & Vib., July 12th – 16th, Florence, Italy (2015), pp 1-8.

A. J. Brammer, G. Yu, E. R. Bernstein, M. G. Cherniack, and D. R. Peterson, "Improving speech intelligibility in active hearing protectors and communication headsets with subband processing," Proc. Meetings Acoust. 19 (040009), 1-5 (2013).

A. J. Brammer, G. Yu, E. R. Bernstein, M. G. Cherniack, D. R. Peterson, and J. B. Tufts, "Understanding speech when wearing communication headsets and hearing protectors with subband processing," J. Acoust. Soc Am. **136**, 671-681 (2014).

A. J. Brammer, G. Yu, E. R. Bernstein, M. G. Cherniack, J. B. Tufts, and D. R. Peterson, "Intelligibility of speech corrupted by nonlinear distortion", Proc. 10th Int. Congress on Noise as a Public Health Problem, July 24th – 28th, London, UK (2011), pp 287-292.

A. J. Brammer, G. Yu, E. R. Bernstein, M. G. Cherniack, J. B. Tufts and D. R. Peterson, "Predicting the intelligibility of speech corrupted by nonlinear distortion," Canadian Acoustics **38**(3), 80-81 (2010).

I. Brons, R. Houben, W. A. Dreschler, "Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort," Ear Hear. **34**, 29-41 (2013).

H. E. Camargo, A. S. Azman and L. Alcorn, "Development of noise controls for longwall shearer cutting drums," Noise Control Eng. J. **64**, 573-585 (2016).
H. E. Camargo, P. A. Ravetta, R. A. Burdisso and A. K. Smith, "Application of phased array technology for identification of low frequency noise sources," J. Low Frequency Noise, Vib. & Active Control **28**, 237-244 (2009).
K. Cardosi, P. Falzarano and P. Han, "*Pilot-controller communication errors: An analysis of aviation safety reporting (ASRS) system reports*". DOT/FAA/AR-98/17. U.S. Department of Transportation, Office of Aviation Research, Washington (1998).
S.-W. Choi, C. Peek-Asa, N.L. Sprince, R.H. Rautiainen, K.J. Donham, G.A. Flamme, P.S. Whotten and C. Zwerling, "Hearing loss as a risk factor for agricultural injuries," Am. J. Ind. Med. **48**: 293-301 (2005).
K. Chung, "Effective compression and noise reduction configurations for application to hearing protectors," J. Acoust. Soc. Am. **121**, 1090-1101 (2007).
K. Chung, J. Tufts and L. Nelson, "Modulation-based digital noise reduction for application to hearing protectors to reduce noise and maintain intelligibility," Human Factors **51**, 78-89 (2009).
P. Clarkson and S. F. Bahgat, "Envelope expansion methods for speech enhancement," J. Acoust. Soc. Am. **89**, 1378-1382 (1991).
J. W. Collins, G. S. Smith, S. P. Baker and M. Warner, "Injuries related to forklifts and other powered industrial vehicles in automotive manufacturing," Am. J. Ind. Med. **36**, 513-521 (1999).
T. Dolan and D. O'Loughlin, "Amplified earmuffs: Impact on speech intelligibility in industrial noise for listeners with hearing loss," Am. J. Audiol. **14**, 80-85 (2005).
R. Drullman, J. M. Festen and R. Plomp, "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053-1064 (1994).
T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement systems," IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp.4409-4412 (2009).
R. Goldsworthy and J Greenberg, "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689 (2004).
S. Greenberg, H. Carvey, L. Hitchcock and S. Chang, "Temporal properties of spontaneous speech - a syllable-centric perspective," J. Phonetics **31**, 465-485 (2003).
A. S. House, C. W. Williams, M. H. L. Hecker and K. D. Kryter, "Articulation testing methods: Consonant differentiation with a closed-response set," J. Acoust. Soc. Am. **37**, 158-166 (1965).
IEC 60628-16 (3rd edition), "Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index," International Electrotechical Commission, Geneva (2003).
G. J. Joy and P. J. Middendorf, "Noise exposure and hearing conservation in U. S. coal mines - A surveillance report," J. Occup. Environ. Hygiene **4**, 26-35 (2007).
T. LaTourette, D. J. Peterson, J. T. Bartis, B. A. Jackson and A. Houser, *Protecting Emergency Responders: Community Views of Safety and Health Risks and Personal Protection Needs*, Vol. 2, Rand Science and Technology Institute (2003).
T. Langhans and H. Strube, "Speech enhancement by nonlinear multiband envelope filtering. In: *ICASSP '82, IEEE International Conference on Acoustics, Speech and Signal Processing,* edited by C. Gueguen, Institute of Electrical and Electronic Engineers, New York (1982), pp. 156-159.
N. Lezzoum, G. Gagnon and J. Voix, "Noise reduction of speech signals using time-varying and multi-band adaptive gain contol for smart digital hearing protectors," Applied Acoustics **109**, 37-43 (2016).
H. E. Lindeman, "Speech intelligibility and the use of hearing protectors," Audiology **15**, 348-356 (1976).
C. Lorenzi, F. Berthommier, F. Apoux and N. Bacri, "Effects of envelope expansion on speech recognition," Hear. Res. **136**, 131-138 (1999).
R. L. McKinley, V.S. Bjorn and J.A. Hall, "Improved hearing protection for aviation personnel," NATO Technical Report RTO-MP-HFM-123, Neuilly-Sur-Seine, France, pp. 13-1 – 13-12 (2005).

Mines Safety Health Administration web site
www.msha.gov/STATS/PART50/WQ/2004/WQ045T03.asp
T. C. Morata, A.C. Fiorini, F.M. Fischer, E.F. Krieg, L. Gozzoli and S. Colacioppo, "Factors affecting the use of hearing protectors in a population of printing workers," Noise and Health **4**, 25-32 (2001).
T. C. Morata, C. L. Themann, R. F. Randolph, B. L. Verbsky, D. C. Byrne and E. R. Reeves, "Working in noise with a hearing loss: Perceptions from workers, supervisors, and hearing conservation program managers," Ear Hear **26**, 529-545 (2005).
L. Murray-Johnson, K. Witte, D. Patel, V. Orrego, C. Zuckerman, A.M. Maxfield and E.D. Thimons "Using the extended parallel process model to prevent noise-induced hearing loss among coal miners in Appalachia," Health Education Behavior **31**, 741-755 (2004).
I. M. Noordhoek and R. Drullman, "Effect of reducing temporal intensity modulations on sentence intelligibility," J. Acoust. Soc. Am. **101**, 498-502 (1997).
D. S. Patel, K. Witte, C. Zukerman, L. Murray-Johnson, V. Orrego, A.M. Maxfield, S. Meadows-Hogan, J. Tisdale and E.D. Thimons, "Understanding barriers to preventive health actions for occupational noise-induced hearing loss," J. Health Commun. **6**, 155-168 (2001).
K. L. Payton and L. D. Braida, "A method to determine the speech transmission index from speech waveforms," J Acoust Soc Am **106**, 3637–3648 (1999).
P. Plyler and M. Klumpp, "Communication in noise with acoustic and electronic hearing protection devices," J. Am. Acad. Audiol.**14**, 260-268 (2003).
R. S. Schlauch and P. Nelson, "Pure-tone evaluation," in *Handbook of Clinical Audiology*, edited by J. Katz, L. Medwetsky, R. Burkard and L. Hood, Lippincott Williams & Wilkins, Baltimore (6th edition, 2009), pp. 30-49.
A. Suter, "The handicap resulting from noise-induced hearing loss," *Best Practices in Hearing Loss Prevention*, DHHS (NIOSH) Pub. 2001-157, Department of Health and Human Services, Washington, pp. 3-8 (2001).
A. H. Suter, Communication and Job Performance in Noise: A Review, Monograph 28, American Speech-Language-Hearing Association, Rockville (1992).
S. Tak, R. R. Davis and G. M. Calvert, "Exposure to hazardous workplace noise and use of hearing protection devices among US workers - NHANES, 1999-3004," Am. J. Industr. Med. 52, 358-371 (2009).
J. B. Tufts, M. A. Hamilton, A. J. Ucci and J. Rubas, "Evaluation by industrial workers of passive and level-dependent hearing protectors," Noise & Health, **13**, 26-36 (2011).
R. A. van Buuren, J. M. Festen and T. Houtgast, "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," J. Acoust. Soc. Am. **105**, 2903-2913 (1999).
M. van Charante, A.W. Mulder and P.G.H. Mulder, "Perceptual acuity and the risk of industrial accidents," Am. J. Epidemiol. **131,** 652-663 (1990).
N. H. van Schijndel, T. Houtgast and J. M. Festen, "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 529-542 (2001).
P. Wilkins and A. M. Martin, "Hearing protection and warning sounds in industry – A review," Applied Acoustics **21**, 267-293 (1987).
A. Wiinberg, J. Zaar and T. Dau, "Effects of expanding envelope fluctuations on consonant perception in hearing-impaired listeners," Trends in Hearing **22**, 1-12 (2018).
K. K. Wójcicki and P. C. Loizou, "Channel selection in the modulation domain for improved speech intelligibility in noise," J. Acoust. Soc. Am. **131**, 2904-2913 (2012).
G. Yu , A. J. Brammer and E. R. Bernstein, "Modeling speech intelligibility in distortion conditions using a speech based sound transmission index," J. Acoust. Soc. Am. **135**, 2164 (2014)
G. Yu , A. J. Brammer, K. Swan, J. B. Tufts, M. G. Cherniack and D. R. Peterson, “Relationships between the modified rhyme test and objective metrics of speech intelligibility,” J. Acoust. Soc. Am. **127**, 1903 (2010).
Y. Zheng, C. Giguerre, C. Laroche, C. Sabourin, A. Gagne and M. Elyea, "A psychoacoustical model for specifying the level and spectrum of acoustic warning sounds in the workplace," J. Occup. Environ. Hygiene **4**, 87-98 (2007).

Hearing Function Questionnaire

Please circle your response to the following questions.

1. Does a hearing problem cause you to feel embarrassed when meeting new people? YES SOMETIMES NO

2. Does a hearing problem cause you to feel frustrated when talking to members of your family? YES SOMETIMES NO

3. Do you have difficulty hearing when someone speaks in a whisper? YES SOMETIMES NO

4. Do you feel handicapped by a hearing problem? YES SOMETIMES NO

5. Does a hearing problem cause you difficulty when visiting friends, relatives, or neighbors? YES SOMETIMES NO

6. Does a hearing problem cause you to attend religious services less often than you would like? YES SOMETIMES NO

7. Does a hearing problem cause you to have arguments with family members? YES SOMETIMES NO

8. Does a hearing problem cause you difficulty when listening to TV or radio? YES SOMETIMES NO

9. Do you feel that any difficulty with your hearing limits or hampers your personal or social life? YES SOMETIMES NO

10. Does a hearing problem cause you difficulty when in a restaurant with relatives or friends? YES SOMETIMES NO

SCORE __________ [Examiner score: No - 0; Sometimes - 1; Yes - 2]

Response to Review

A.1 Clarification of Proof-of-Concept Evaluation Outcome

It was not obvious why the word test performed best for algorithm #2, when algorithm #2 was the worst performer for the environmental noise reduction. Please explain this apparent discrepancy.

The difference in performance of the algorithms in response to either a pure-tone warning sound or speech is an important question. The subject has been the focus of much thought since the end of the experimental work, and further explanation based on this analysis is provided here.

It is essential to appreciate the details of the signals that are being processed by the algorithms. As stated in the Final Report, the method for improving the intelligibility of speech and/or the audibility of warning sounds involves establishing and manipulating signal envelopes, that is, we focus on the rapid *changes* in amplitude of signals with time, often referred to as changes in their time histories. This is done by first establishing the modulus of the time history of the signal. This process converts the normal bipolar time history of speech (as depicted in Figure 3 of the Final Report) into a unipolar signal such as the blue curve in the example shown in Figure A1 below (forming a series of "mountain peaks" and "valleys"). An envelope of this unipolar signal is shown by the dark green time history, with instantaneous peak-to-peak magnitude indicated by the vertical red line. In this example, the dark green time history of the envelope, which is often termed the modulation, follows the changes in components of the modulus with the largest magnitudes. This is equivalent in this example to selecting the low-frequency components of the signal. Other envelopes could be drawn that include the more rapidly changing components of the modulus, that is, the high-frequency components as well as the low-frequency components. This is shown by the light green curve in Figure A1, which continues for the whole time history but has only been drawn between two adjacent large-magnitude peaks. Thus it is evident that the time history of the modulus will contain fluctuations with different frequencies and different magnitudes. We have observed that the frequency spectrum of the envelope fluctuations is distinctly different for speech, tonal warning sounds and environmental noises, and hence provides a means to identify the presence of these sounds.

All our algorithms initially contain a set of N band-pass filters arranged in parallel (see Figure 5 of the Final Report). We have employed sixteen so-called "subbands" that together span the frequency range in which speech and warning sounds occur. As described in our Report, within each subband is a signal path and a control path: the former contains the fraction of speech or warning sound buried in environmental noise with frequencies that fall within the range of the corresponding band-pass filter, while the latter contains the signal processing that manipulates the envelope to construct the amplitude modulation to be applied to the signal path (at the "X" in Figure 5).

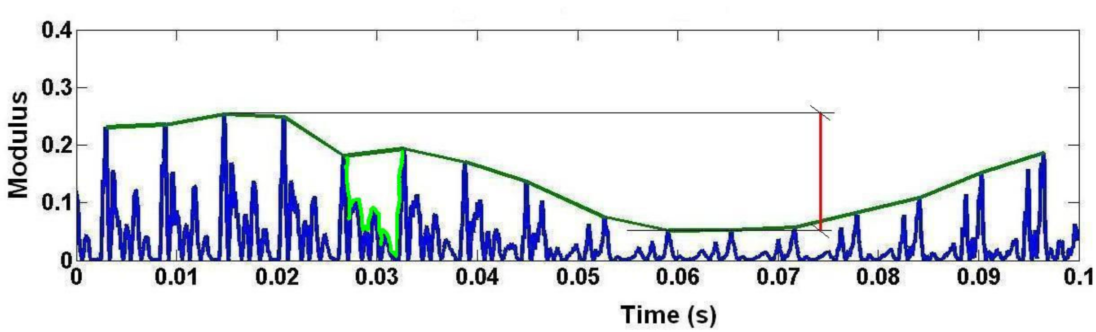


Figure A1: Example of a low-frequency envelope (dark green curve) and part of a high-frequency envelope (light green curve) of the modulus of a speech signal as a function of time (blue curve).
The instantaneous magnitude of the low-frequency envelope is shown by the red line.

To appreciate the difference in performance of the algorithms, we must focus on the signal processing within the control path. Now the speech signal responsible for the modulus shown in Figure A1 could, after band--pass filtering, represent the input to a subband that contains low- and high-frequency modulations associated with the dark and light green envelopes. Other subbands may contain mainly large magnitude envelope fluctuations and some mainly small fluctuations. For any given noise magnitude, subbands containing large magnitude envelope fluctuations (e.g., with the red line in Figure A1) will be relatively immune to noise while subbands containing small magnitude modulations will not. This is because the effect of noise is to *diminish* the magnitude of any speech (or warning sound) envelope fluctuations, which may be thought of as filling up the "troughs" in the time history of the envelopes in Figure A1. This will clearly have more effect on the overall envelope of the smaller modulations first. Hence, generally, small magnitude speech, or warning sound, modulations will be more likely to be diminished and perturbed by noise modulations when the overall signal to noise ratio (SNR) leads to the speech or warning sound being submerged in noise (as shown, for example, in Figures 7A and 8A of the Final Report). In these circumstances, it would be difficult for an algorithm to recover from the noise and/or reinforce the (comparatively small) high-frequency components of the modulations in the example of Figure 1A, but much easier for it to recover the (comparatively large) low-frequency modulations. Herein lies the challenge for algorithm design: how to identify the sounds desired to be heard when the modulations are of small magnitude.

The algorithms described in our Final Report provide three different approaches to the treatment of small envelope fluctuations. They are as follows.

#1 - linear gain control of the subband modulator in Figure 5 of the Final Report derived from the magnitude of subband envelope fluctuations. This means the gain applied by the modulator to the signal (shown by the 'X' in Figure 5 of the Report) will tend to reduce the magnitude of small envelope fluctuations in the signal more than large fluctuations. This will be in addition to any other characteristics of the signal that the modulator is enhancing, which were the same for all algorithms in our listening tests. Referring to Figure 3 of the Final Report, in which the time history of changes in magnitude of speech sounds can be seen, the effect will be that large changes in magnitude of the sounds will be made larger, and small changes in magnitude will be made smaller.

#2 - unity gain control of the subband modulator. This means the gain applied by the modulator will treat envelope fluctuations of all magnitudes the same. The output of the modulator is determined solely by the other characteristic of the signal that the modulator is enhancing.

#3 - nonlinear gain control of the subband modulator derived from the magnitude of subband envelope fluctuations. As implemented, this means the gain applied by the modulator will tend to reduce the magnitude of small signal modulations more than algorithm #1, and hence much more than algorithm #2.

The results of our first formal listening tests clearly showed that algorithm #2 performed best for improving speech intelligibility when the speech SNR was - 8 dB (see Figure 9B of our Final Report). This suggests that a gain control that treats all modulations of all magnitudes the same may be best for improving the intelligibility of speech in noise. Also, algorithms with gain control that suppress small modulations tend not to enhance speech intelligibility. However, these are not complete explanations, as the modulator is enhancing other characteristics of the signal independently of the gain and the results of our listening test involve a combination of the two control functions. For this reason, we cautioned in our Final Report that the performance of all our algorithms has not been confirmed at different speech SNRs when there could be a different combination of the two control functions, and noted that algorithm #2 did not clearly outperform the other algorithms when there was no noise (i.e., with a very different speech SNR).

Suppressing small amplitude modulations, which is done more aggressively by algorithm #3 than #1, which in turn suppresses small modulations more than algorithm #2, does, however, result in better noise suppression and improves the speech SNR, as is evident from the results in Figure 8 of our Final Report. The consequence of this signal processing is commonly described in the scientific literature as improved "clarity" of the speech. However, our results clearly demonstrate that improved speech SNR (or clarity) does not necessarily imply improved

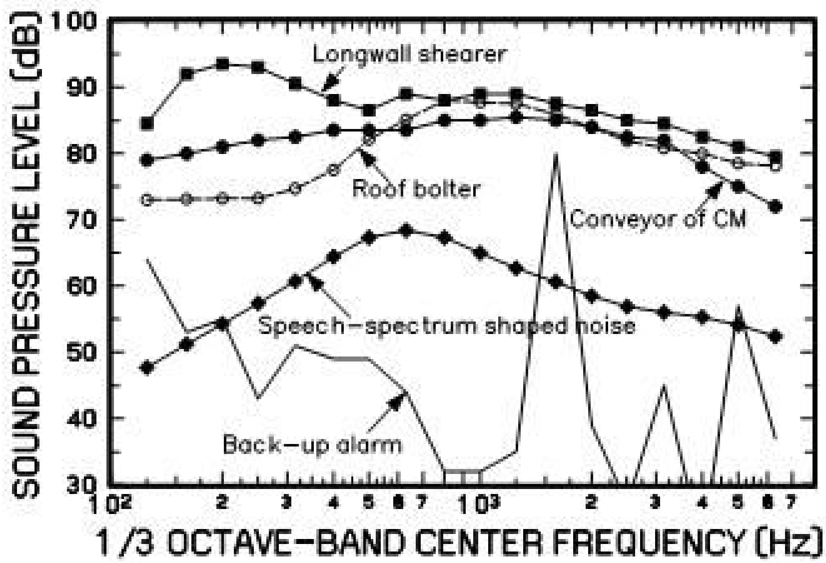


Figure A2: Frequency spectra of noise at the operator's position for the conveyor of a continuous mining machine and a roof bolter, and near a longwall shearer. Spectra are also shown for the long-term spectrum of speech and an intermittent tonal warning sound.

speech intelligibility. This reality only became apparent to us as a result of the listening tests conducted in the last month of the project.

These observations are also pertinent to appreciating how the algorithms process pure-tone warning sounds. In contrast to speech, which possesses modulations that vary in time and in frequency, an intermittent pure-tone warning sound has a consistent time-invariant modulation - it is simply the square wave envelope of the modulus. This modulation will appear primarily in the subband containing the pure-tone. When the warning sound is buried in noise, the algorithms will, as before, find it much easier to recover the comparatively large modulations rather than the small modulations. However, the small modulations associated with tonal warning sounds no longer contain important information, and so suppressing the magnitude of the subbands signals containing these modulations will disproportionally improve the SNR. Hence the algorithm that suppresses small magnitude modulations most aggressively will improve the SNR the most and result in the largest improvement in audibility. Inspection of Figure 7 in our Final Report reveals that algorithm #3, which suppresses small modulations most aggressively, does, indeed, reduce noise the most and produce the best SNR.

The distinction between improving the audibility of warning sounds and improving the intelligibility of speech is essential to an appreciation of the mechanisms at play, which involve the information contained in envelope modulations. Hence we do not believe there is a discrepancy in the results, but rather that the differences observed highlight the need for a more in-depth understanding of the mechanisms involved.

A.2 Study Noise Parameters

It would be helpful if the noise levels that were utilized in the study were compared to those in an actual mine environment. How would they be assessed relative to speech-spectrum noise or environmental noise? Would the actual noises make speech recognition more difficult? Can you describe with practical examples of speech-spectrum and environmental noise sources that would be common in a mine? How does "roof talk" fit into this description? Do you think that being able to enhance hearing such relatively minor noises as "roof talk" is feasible? Do you have noise measurement profiles taken in a mine that can be referenced in your study?

There appear to be few sources of frequency spectra for the noises encountered in mechanized mining. Detailed studies have, however, been performed of the noise at locations described as the operator's position in two underground coal mines (Camargo et al., 2016; Szary et al., 2011). Noise during normal operation is shown for three machines in Figure A2. The spectra are for the conveyor of a continuous miner (CM), a Joy QT-CM that incorporated treatments to reduce noise, for a roof bolter when drilling, and for a longwall miner in the vicinity of the shearer cutting drums. The sound levels reported for these machines were: continuous miner - 96 dBA, roof bolter - 97 dBA, and longwall miner - 98 dBA. Changes in sound levels during machine operations of up to 5 - 6 dBA have been reported (Bobick and Giardino, c1976; Szary et al., 2011). In contrast, comparatively small differences in sound level between machines from different manufacturers were found in a survey of twelve underground coal mines (Bobick and Giardino, c1976). For example, the standard deviations of the sound levels produced by 33 continuous miners when cutting and loading, and 37 roof bolters when drilling, were ±2.6 and ±3.2 dBA, respectively. The background noise in the mine when machines were not operating was not reported.

Several noises were used in our study. Three noises with different frequency spectra were employed for the initial predictions. One was selected for evaluating speech in listening tests and for estimating the audibility of a tonal warning sound. The noise chosen was so-called speech-spectrum shaped noise, that is, random noise that has a frequency spectrum shaped to approximate the long-term spectrum of speech. The spectrum of this noise is also shown in Figure A2. The shape of the spectrum is from the European Computer Manufacturers Association's Technical Report ECMA TR/105 (2012), and has been plotted with sound pressure levels for a "raised" voice as heard some one meter from the talker. The notion of evaluating speech-in-noise tests using speech-spectrum shaped noise is common in the scientific literature, and relies on the expectation that matching the spectrum of noise to that of speech will impede a listener's performance in noise the most. Hence this will provide a "worst case" for evaluating the performance of our algorithms. For any noise, the primary determinants of speech intelligibility are known to be the frequency spectrum of the noise, and the relative intensities of the speech and noise at the listener's ear. The latter are expressed by the speech signal-to-noise (SNR). Certainly all our predictions of word scores in Figure 6 of the Final Report are consistent with these expectations, and were the reason for selecting speech-spectrum shaped noise for most of our work.

The spectrum of a tonal alarm consists primarily of a narrow peak at the frequency of the pure tone, which is typically in the range from 1 to 2 kHz, as in Figure A2. The spectrum shown is for a vehicle backup alarm emitting its characteristic "beep . . beep . . beep", and is depicted at a convenient peak sound level for presentation in the graph: in practice, its peak intensity will depend on the source and the distance from the listener. It should be noted that the low-frequency components of the spectrum are more likely produced by the vehicle on which the alarm is mounted than by the alarm itself. It is evident that the spectrum of the backup alarm does not share any common features with those of the other noise sources in Figure A2. As discussed in the previous section, its modulation characteristics render it readily detectable by the algorithms developed in this study, with every expectation that its audibility will be improved by the signal processing. The performance will be primarily dependent on the relative intensities of the warning sound and the noise at the listener's ear at frequencies close to that of the pure-tone warning sound, that is, on the SNR at these frequencies. In an extremely noisy environment, the audibility of the warning sound will be reduced somewhat by high intensity noise at frequencies below that of the pure tone (so-called upward spread of masking). The process of masking occurs within the auditory system of the listener and not in the sound external to the ear. Consequently, any noise reduction by our algorithms would disproportionally improve the audibility of an alarm for operators of the mining machines.

To evaluate speech intelligibility or warning sound detection in noise by a listening test, an experimenter firstly selects a recording of the time history of the appropriate stimulus (i.e., speech or warning sound), secondly selects a recording of the time history of the noise, thirdly selects an SNR, and finally chooses an overall sound level at which to perform the test. (It is understood that the sound levels of recordings of both the stimulus and the noise can be adjusted to achieve a given SNR.) In a laboratory setting such as ours, the last mentioned consideration assumes importance as we are obligated to protect the health and safety of human subjects. For this reason, the listening tests described in our Final Report were

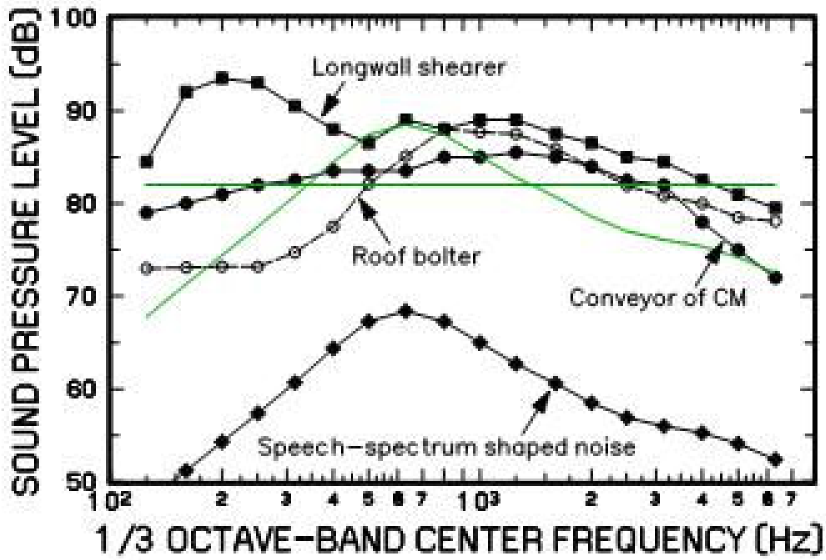


Figure A3: Frequency spectra of noise at the operator's position for the conveyor of a continuous mining machine and a roof bolter, near a longwall shearer and the long-term spectrum of speech. Also shown are spectra of test noises adjusted to the levels of the mining machines (green curves).

conducted at a "comfortable" listening level with a maximum sound level of 70 dBA, which is approximately that of the spectrum for speech-spectrum shaped noise shown in Figure A2. When listening to speech, the human auditory system functions effectively the same over a wide range of sound pressure levels from close to the threshold of hearing up to about 80 dBA (Dubno et al., 2005). This means a listening test conducted at a sound level of 80 dBA will yield the same result as a test conducted at a sound level of 40 dBA. Hence, the precise sound level selected for the listening test in the laboratory is unimportant provided it is within this range, and the results will be directly applicable to speech and noises in this range of sound levels.

Clearly, both the noise spectra and sound levels of the mining machines are different from those used in our listening tests, which begs the question as to what extent will the differences between the spectra and sound levels influence the performance of listening tests? Partial answers to this question can be inferred from the previous discussion.

Firstly, provided the sound pressure level at the listener's ear is below about 80 dBA, the performance of the algorithms would be the same as that determined in our listening tests. While the sound levels of the mining machines at the operator's location are much higher in the mines, occupational noise exposure regulations would require the use of hearing protection, which should reduce the noise of these machines at the listener's ears to 85 dBA, or less. Under these conditions, the user of a conventional (passive) hearing protector, with no electronic augmentation, is likely to experience a modest reduction in speech intelligibility from that recorded at lower sound levels for a given speech SNR. The nonlinearity is again due to masking within the auditory system of the listener, and is not due to any physical change in the sounds reaching the ear. Hence, as our algorithms operate on the sounds before they reach the ear, there is no reason to suppose the improvement in speech intelligibility they provide will be affected by this mechanism. Indeed, in as much as our algorithms reduce noise, the net improvement in intelligibility could well be greater than that obtained in the laboratory. The main difference from the listening tests conducted in our laboratory would arise from the character of the speech. It is common knowledge that two individuals attempting to communicate in a noisy environment, with sound levels of 96 - 98 dBA, would have to shout. The speech material available to us for use in listening tests is recorded at conversational levels used in quiet surroundings, and there may be differences in enunciation when shouting that could influence intelligibility. While the effects on intelligibility may be small and would equally affect listeners both when our algorithms were operating and not used, they will need to be considered in a future study.

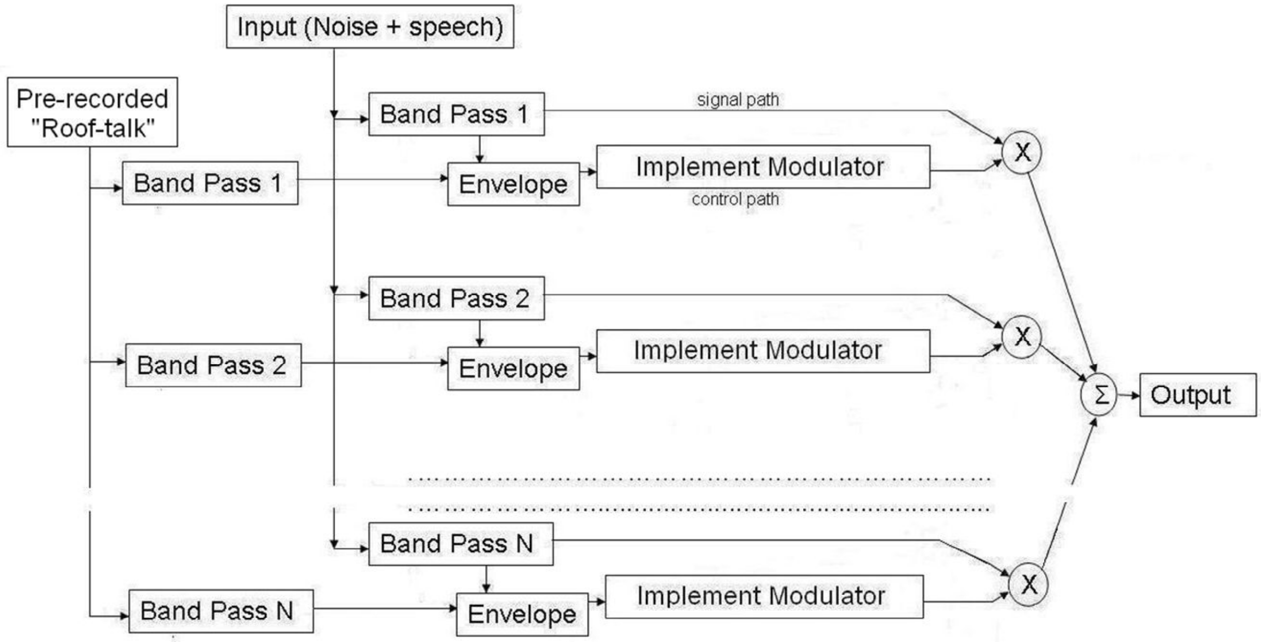


Figure A4: Concept block diagram for detecting "roof talk"

Secondly, the effect of differences in the frequency spectra between that of speech-spectrum shaped noise and those of the mining machines shown in Figure A2 may be addressed in the following way. Predictions of increases in speech intelligibility in three different noise spectra were presented in Figure 6 of the Final Report. The spectra were speech-spectrum shaped noise, white noise, and so-called reverse white noise. The frequency spectra of speech-spectrum shaped noise and reverse white noise have been drawn in Figure A3 (green curves) at sound pressure levels similar to those of the mining machines in Figure A2. While neither of the test noises is a close match to any of the mining noises, the two green curves provide a visual representation of the effect of different noise spectra on speech intelligibility. Clearly, the horizontal green spectrum is more like the continuous miner conveyor noise than the speech-spectrum shaped noise and the reverse is the case for the roof bolter. The large magnitude, low-frequency noise peak in the longwall miner spectrum is not replicated by either of the green spectra. It has already been stated that when listening to speech in noise the greatest difficulty discriminating words occurs when the frequency spectrum of the noise matches that of speech. Thus, for any given speech SNR, changing from speech-spectrum shaped noise to reverse white noise would be expected to lead to an increase in speech intelligibility. Reference to Figure 6 of the Final Report reveals that the largest increase in word scores predicted by our model when listening in speech-spectrum noise is, typically, 15 - 18 %, while the largest increase when listening in reverse white noise is, typically, 30 %. As a mean increase in word score of 18.7% was obtained in our listening tests when speech-spectrum noise was employed (see section 5.0 *Speech Intelligibility* of the Final Report), it is reasonable to expect at least this performance in listening tests, and perhaps more, when other noise spectra are used. Thus, there is an expectation that algorithm #2 will significantly improve speech understanding for the operator of the roof bolter under conditions equivalent to those tested in the laboratory (i.e., speech SNR of -8 dB), even after adjusting for the talker shouting. As the frequency spectrum of the continuous miner conveyor noise is more like that of reverse white noise, for which a larger increase in word score is predicted, there is also an expectation that algorithm #2 will significantly improve speech understanding for operators of this machine when the speech SNR is -8 dB, even after adjusting for the talker shouting. For the longwall miner, the frequency spectrum of which is unlike that of either of the green curves in Figure A3, applying the "worst case" improvement in predicted word score (15%) suggests that algorithm #2 will produce a substantial improvement in speech understanding for the operator close to the shearers of this machine, with the same caveats as for the other mining machines.

Finally, the results of listening tests conducted when there was no noise reveal that the algorithms adequately maintained the intelligibility of speech (see Figure 9A of the Final Report), thus confirming their suitability for application to noises that occur intermittently such as in mechanized mining operations.

The subject of "roof talk" was introduced in the *Technology Readiness Assessment* section of the Final Report for two reasons. Firstly, miners have explained in focus sessions their aversion to wearing hearing protection stems from fear of not hearing communications and especially "roof talk" (Patel et al., 2001). Secondly, our algorithms may not detect transient broadband noise that is believed to be "roof talk".

The solution proposed to enhance "roof talk" is to use an algorithm shown in concept in Figure A4, which is closely related to the algorithm in Figure 5 of the Final Report, and is a variant of one developed for detecting warning sounds in noise in another project (Bernstein et al., 2014). As drawn, the primary input is speech but would in this context be a warning sound buried in noise (top left hand corner of the diagram). The first step again involves a set of N band-pass filters arranged in parallel to create N subbands that together span the frequency range in which speech and warning sounds occur. Within each subband there continues to be a signal path and a control path, the former containing the components of speech or warning sounds in environmental noise with frequencies within the range of the corresponding band-pass filter, and the latter the signal processing to identify the warning sounds and construct the amplitude modulation to apply to the signal path.

In this algorithm, however, there is a second input to the control path, which consists of pre-recorded sounds that the algorithm is to detect. For the present application, the sounds will be samples of "roof talk" previously recorded in the mine in which the electronic hearing protector containing the algorithm is to be used, but could also, additionally, be other warning sounds specific to the mine. These sounds are compared with the primary input in the modulator and, if a similarity is found, the signal in the subband is amplified, combined with the outputs of other subbands that have been similarly processed and replayed to a listener.

It is envisaged prerecording of the alarm sounds to be detected will occur in comparative quiet and employ the initial stages of the algorithm, so that the alarm signatures will possess spectral features identical to those of the signals to be detected when noise is present. In practice, the recordings would be obtained before using the hearing protector as part of setting up the device for a particular workplace. The algorithm shown in Figure A4 has been successfully employed to identify a tonal warning sound in listening tests when subjects wore our electronic hearing protector (Bernstein et al., 2014).

For "roof talk" to be detected, it must consist of one or more recognizable sounds that can be recorded in a mine before commencing use of the hearing protector. The recordings would be permanently stored electronically in the memory of the device and automatically used to compare incoming sounds with the known warning sounds. For example, to detect an environmental noise such as thunder there would need to be separate recordings of nearby and distant thunder, and perhaps also at intermediate range, as it common experience that different sounds are associated with thunder depending on its proximity to the listener (e.g., a "crack" at short range, and a "rumble" at long range). It must be stressed that, at the present time, the performance of the algorithm in Figure A4 has not been evaluated on such complex sounds. Moreover, its suitability for operating in the intense noise environment experienced by operators of some machines in an underground mine is unknown. However, when used to identify a vehicle backup alarm in noise, the algorithm was able to detect the alarm under conditions that it was completely inaudible to an unaided human listener.

In summary, both sound levels and frequency spectra for common machines used in mechanized mining have been compared to the spectra of speech-spectrum noise and the sound from a vehicle backup alarm. The applicability of the results of our listening tests to other noises and in particular those highlighted in this report has been discussed in detail. Also, the possibility of enhancing specific noises such as "roof talk" has been considered. The analyses have relied on noise measurements conducted by others as reported in the scientific literature, as we have not recorded noises in a mine.

A.3 Technology Readiness

Providing a better understanding of the potential of this technology in practical terms would be helpful. Perhaps if you can more precisely put into perspective the study design compared to the noise conditions in an actual mine it might add some clarity. Do you perceive the next phase of the study would entail actual mine profiles of some sort?

The applicability of the results of our listening tests to the noises experienced by the operators of several common mining machines has been considered in detail in the previous section. As already mentioned, the motivation for choosing speech-spectrum noise as the test noise for developing the algorithms was the expectation that it would impede a listener's ability to understand speech the most, and hence provide the most difficult test of the algorithms' ability to improve intelligibility in any noise. Based on the model predictions summarized in our Final Report, it was argued that the performance of the algorithms may be better when attempting to improve the intelligibility of speech buried in other noises.

It was also pointed out in our Final Report that understanding speech when there is no noise is a trivial problem for a human listener, but constitutes almost as much a problem as improving understanding speech in noise for some algorithms. For this reason, results were obtained to document that speech intelligibility is maintained by our algorithms when there is no noise. This result also serves to demonstrate that the algorithms are suitable for noises that occur intermittently, with quiet intervals between periods of operation, such as in mechanized mining. Additionally, there may be temporal or spectral variations in the noise in the vicinity of a machine operating underground that render the speech more intelligible to a human listener than might otherwise be expected (so-called listening in the "dips" or "holes" in the noise). Such variations have not been included in our listening tests. To answer questions concerning the performance of our algorithms when processing the noise of specific mining machines more fully, it will be essential to include noises recorded during mining operations in future work. It must be appreciated, however, that a formal listening test confirming the result of any change in an algorithm or its performance in a given noise involves up to twenty listeners tested individually, and hence takes considerable time to perform. Consequently, it is not realistic to include many different noises in laboratory tests developing algorithms.

In view of the considerations highlighted here, a way forward would need to involve the following activities to reach a working prototype, some of which can be undertaken in parallel.

i) Obtaining recordings of typical noises occurring in mines. Sources might include other researchers who may possess such information (e.g., at NIOSH, Pittsburgh), or direct sound recording in mines (e.g., by a company employee, or member of our research team).

ii) Continuing the development of algorithms under "worst case" noise conditions to explore and solve the problem of small modulations. This laboratory work would initially involve one noise.

iii) Recording a corpus of phrases with the 300 test words from the Modified Rhyme Test lexicon when the talker is shouting.

iv) Evaluating and fine-tuning algorithms when operating at different SNRs and with different noises. This laboratory work should include noises recorded in a mine.

v) Evaluating algorithms in the intense noise typical of that experienced by operators of selected mining machines.

vi) Developing a hardware implementation of selected algorithms. This laboratory work would involve development of the software and fabrication of the electronic circuits necessary to enable an algorithm to operate in real time, that is, the output must occur within milliseconds of the input to enable persons using the algorithms to converse.

vii) Constructing a packaged, portable unit with integrated electronics to enable the algorithms to be connected to a hearing protector when "worn" initially by a dummy head and evaluated in a noise environment not unlike that in a mine. The use of a dummy head avoids issues of apparatus malfunction that could affect a person's hearing. A simulated mine might be adapted for this purpose to possess a suitable noise environment.

viii) Constructing a wearable unit suitable for evaluation by a miner or trainee in a noise environment not unlike that in a mine. While it is expected that all aspects of the design will involve intrinsically safe circuitry and components, it is not expected to be certified for use underground.

ix) Developing a working prototype based on the results and experience gained with the portable and wearable units.

We are fortunate to have collaborated with members of the Benjamin M. Statler College of Engineering and Mineral Resources at West Virginia University on a previous Alpha Foundation project, which provided us direct experience with extraction techniques, and health and operational issues in coal mining. The work involved some research staff touring an underground coal mine where the noise environment was observed. The difficulty communicating in the vicinity of machines as well as the potential for accidental injury from failure to hear warning sounds was confirmed. It is intended that the collaboration with members of the Benjamin M. Statler College of Engineering and Mineral Resources at West Virginia University continue in the event that the project moves forward to a working prototype, and may involve their simulated mine (see *Letter of Support* appended to our original proposal).

As for the potential of the technology beyond the mining environment, a future device is foreseen to be an advanced electronic hearing protector that could replace conventional hearing protectors for work environments in which a person's need to communicate or hear warning sounds is currently compromised by noise in the work place. With approximately 4.5 million workers in the U.S. apparently unwilling to use HPDs at present, many for fear of not hearing warning sounds (Murray-Johnson et al., 2004; Patel et al., 2001), there is reason to expect substantial benefits to occupational health from such devices in the industrial sectors NIOSH identifies noise-induced hearing loss may occur (agriculture, forestry, construction, manufacturing, mining, oil and gas extraction, and transportation). Consumers using portable electronic devices for communication (e.g., mobile phones, MP3 players, headphones), and users of headsets (e.g., pilots, call-center operators, police, emergency services), hearing aids and two-way radios may also benefit from the technology being incorporated into their devices.

A.4 References

E. R. Bernstein, A. J. Brammer and G. Yu, "Augmented warning sound detection for hearing protectors," J Acoust. Soc. Am. **135**, EL29-EL34 (2014).

T. G. Bobick and D. A. Giardino, "The noise environment of the underground coal mine," Informational Report 1034, Mining Enforcement and Safety Administration, Washington (c1976), pp. 1-26.

H. E. Camargo, A. S. Azman and L. Alcorn, "Development of noise controls for longwall shearer cutting drums," Noise Control Eng. J. **64**, 573-585 (2016).

H. E. Camargo, P. A. Ravetta, R. A. Burdisso and A. K. Smith, "Application of phased array technology for identification of low frequency noise sources," J. Low Frequency Noise, Vib. & Active Control **28**, 237-244 (2009).

J. R. Dubno, A. R. Horowitz and J. B. Ahlstrom, "Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking," J. Acoust. Soc. Am. **118**, 914-922 (2005).

ECMA TR/105. "A shaped noise file representative of speech," European Computer Manufacturers Association, Geneva (2012), pp. 1-13.

L. Murray-Johnson, K. Witte, D. Patel, V. Orrego, C. Zuckerman, A.M. Maxfield and E.D. Thimons "Using the extended parallel process model to prevent noise-induced hearing loss among coal miners in Appalachia," Health Education Behavior **31**, 741-755 (2004).

D. S. Patel, K. Witte, C. Zukerman, L. Murray-Johnson, V. Orrego, A.M. Maxfield, S. Meadows-Hogan, J. Tisdale and E.D. Thimons, "Understanding barriers to preventive health actions for occupational noise-induced hearing loss," J. Health Commun. **6**, 155-168 (2001).

M. L. Szary, Y. P. Chugh and J. Hirschi, "Noise variability in underground room and pillar coal mines," Int. J. Occup. Safety & Ergonom. **17**, 301-308 (2011).