THE DISTORTING EFFECTS OF SCBA EQUIPMENT ON SPEECH AND ALGORITHMS FOR MITIGATION

William M. Kushner, S. Michelle Harton, and Robert J. Novorita

Motorola
1301 E. Algonquin Rd., Room 2712, Schaumburg, IL 60196
phone: + (1) 847-576-0933, fax: + (1) 847-576-8378, email: kushner@labs.mot.com

ABSTRACT

High quality, reliable communications between public safety personnel is essential for accomplishing missions while maintaining their own health and safety. The SCBA mask air delivery system is essential equipment in some public service activities, and its effects on speech must be considered in public service communications system designs. An SCBA mask encloses the face and mouth and alters articulation and the speech spectrum. SCBA system noises such as alarms and air regulator noise, in addition to environmental noises, add to the deterioration of speech quality in an SCBA-fronted communication system. In digital communication systems, corrupted speech affects codec performance increasing the potential for diminished intelligibility. This paper describes a study of some of the distortion effects of SCBA equipment on speech quality. It also describes two signal processing algorithms designed to mitigate some of the effects of the distortion.

1. INTRODUCTION

Good, reliable communications between public safety personnel is essential for accomplishing their missions while maintaining their own health and safety. This is especially true for firefighters who must operate in very physically demanding environments where communications can determine life or death outcomes. Self-Contained Breathing Apparatus (SCBA) equipment is often used in firefighting activities. The SCBA mask covers the face and lips forming an acoustic chamber coupled to the vocal tract of the user that severely alters the speech spectrum [1] [2]. The tight fit of the mask restricts jaw movements making normal articulation more difficult. In addition, inhalation noises from the SCBA air regulator contaminate the speech signal. The standard bandwidth (300 – 3400 Hz) of typical communications systems contributes to the diminishment of higher frequency fricative and consonant sounds. Also, digital codecs being employed in current generation public service radio systems affect speech quality and intelligibility, especially when input speech is noisy and distorted. Because reliable communications are essential, it is important to assess the impact of SCBA equipment on communications and compensate for any adverse effects.

2. ACOUSTIC PROPERTIES AND EFFECTS OF SCBA EQUIPMENT

Typical SCBA equipment consists of an air-tight mask covering the face and mouth, and an air pressure regulator attached to the front of the mask connected to a high pressure air tank. Pictures of the inside and outside of one type of commonly used SCBA mask are shown in Figure 1.

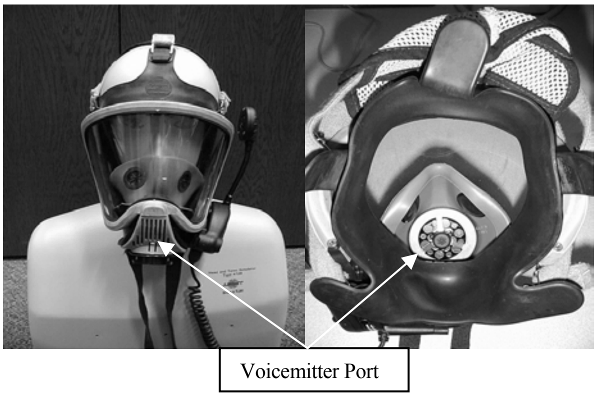

Figure 1. External and internal views of a commonly used SCBA mask showing the voicemitter port.

The mask is a rigid structure with a clear plastic faceplate and a flexible rubber seal that contacts the forehead, temples, cheeks, and chin of the wearer. Also built into the surface of the mask are one or two so-called "voicemitter" ports. These are thin metal membranes, 3 to 5 cm in diameter, designed to facilitate the transmission of speech through the mask wall while maintaining the integrity of the mask gas seal. The mask may contain an internal or external microphone and pre-amplifier for connecting directly to a radio. More commonly, a hand-held PTT microphone is held near one of the voicemitter ports for audio pickup. There are a number of different SCBA mask styles and types from various manufacturers with different acoustical properties.

The SCBA mask alters normal speech in a number of ways. The primary causes of speech distortion are the acoustic reflective and resonant properties of the mask. The enclosed spaces formed when the mask is worn become cavities with resonances in the speech pass-band. The cavities, hard internal surfaces, and short reflection distances lead to intense acoustic reflections within the mask resulting in peaks and nulls throughout the spectrum (spectral combing) [3] that can alter the perception of speech formants. Because there are many different reflective paths within the masks, the spacing and appearance of the summed peaks and nulls is quite complex. An example of the spectral distortions caused by a typical SCBA mask is illustrated in Figure 2. These spectra were derived from an SCBA mask mounted on a B&K 4128 Head and Torso Simulator (HATS) as shown in Figure 1. The HATS artificial mouth (B&K 4227) was driven by a 0.2 - 10 KHz swept sine wave source. Measurements were made using a probe mike at the mouth opening within the mask, and a B&K 4133 microphone mounted 2.5 cm outside in front of the voicemitter port.

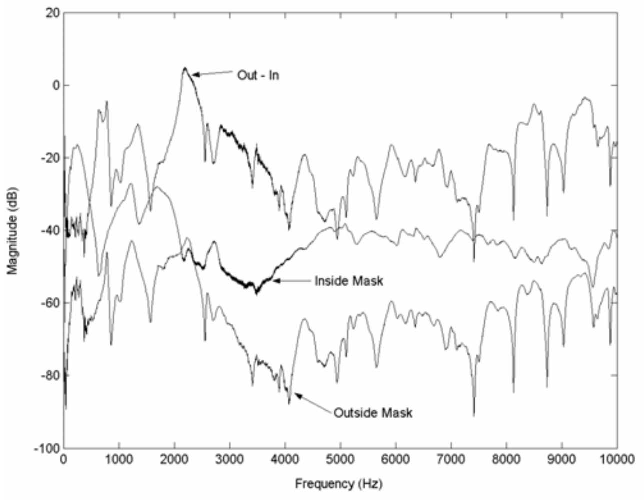

Figure 2. Spectral magnitude distortion caused by an SCBA mask.

The acoustic properties of the SCBA mask are functions of the specific design, size, and fit of each mask and thus can vary considerably across manufacturers and for individuals wearing the masks.

In addition to the SCBA mask acoustic distortions, there are a number of noises generated by the SCBA air system that contaminate the speech signal, one of which is the regulator air inhalation noise. An SCBA system is a pressure-demand air delivery system. When a user inhales, negative pressure within the mask causes the regulator valve to open allowing a rush of pressurized air to enter the mask producing a loud, broadband hissing noise that is comparable in amplitude to the speech signal. Inhalation noise is broadband and incoherent. It is also very stationary for a given mask and wearer. An example of air regulator inhalation noise and speech recorded externally from an SCBA mask is shown in Figure 3.

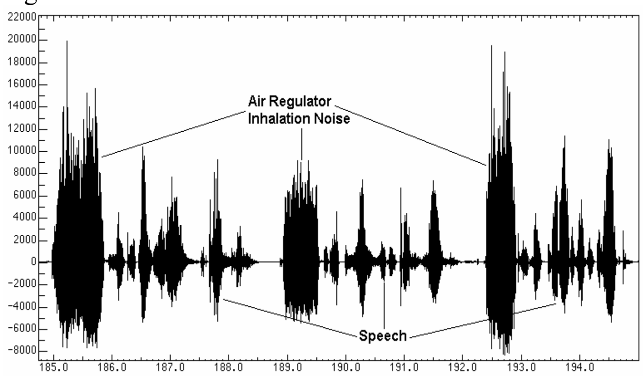

Figure 3. An illustration on SCBA air regulator inhalation noise and speech.

The air regulator inhalation noise does not directly corrupt the speech since people do not normally speak when inhaling. The noise can however, interfere with radio VOX circuits and the VADs of some digital codecs. Primarily, this noise is annoying to the listener. The effect of the SCBA mask distortion on speech is more serious as it affects both quality and intelligibility and the performance of digital codecs.

3. MITIGATING THE EFFECTS OF SCBA DISTORTION AND NOISE

Two algorithms were developed to help overcome the effects of mask distortion and noise on speech. The first called ARINA (Air Regulator Inhalation Noise Attenuator) is designed to detect bursts of air regulator inhalation noise, spectrally model it, and attenuate it. The second algorithm, AMSE (Air Mask Speech Equalizer) is designed to adaptively equalize the speech spectral distortion effects caused by the mask using the spectral model produced by ARINA. Both algorithms utilize the fact that the spectral content of the air-regulator inhalation noise is broadband and reasonably stationary for a given SCBA mask and wearer.

The basic operation of the ARINA algorithm is to model the inhalation noise using an LPC filter, inverse filter the SCBA audio signal, and then look at the filter prediction gain to determine when a close spectral match to an inhalation noise occurs. A block diagram of the algorithm is shown in Figure 4.

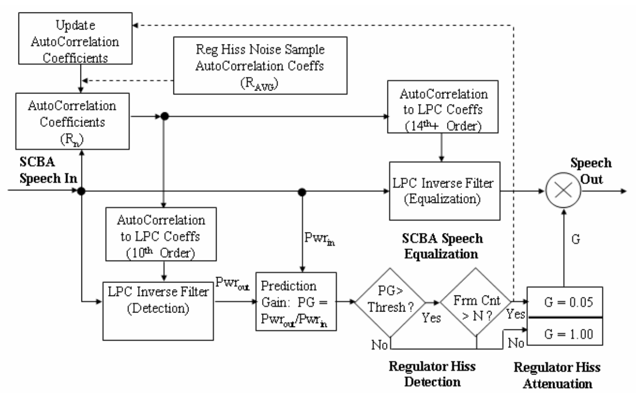

Figure 4. A block diagram of the ARINA algorithm for detecting and attenuating SCBA inhalation noise.

The detection algorithm is initialized using a set of autocorrelation coefficients derived offline from a sample of the inhalation noise. These autocorrelation coefficients are used to determine a set of 10^{th} order LPC coefficients that are then used to inverse filter the SCBA speech signal. The signal is sampled at 8 KHz and analyzed in 160 sample non-overlapping frames. The filter prediction gain is calculated every 2.5 msec sub-frame, averaged over 4 sub-frames, and then compared to an empirically derived threshold. If it exceeds the threshold, the system assumes an inhalation noise is being detected. The duration of a true inhalation noise is fairly long compared to unvoiced speech so a duration threshold test is also applied to eliminate any false detection due to speech. Thus, the prediction gain threshold must be met for N consecutive frames before detection is validated. The output of the detector is used to gate an output signal attenuator. If the inhalation noise is detected, the gain is set to a low value (e.g. 0.05); if not, the gain is set to 1.0.

The output of the detector is also used to determine when the LPC model estimate of the inhalation noise should be updated. If inhalation noise is detected the working model autocorrelation coefficients representing the air inhalation noise are updated using an averaging filter. A new set of LPC coefficients for the detection inverse filter is then recalculated from the updated autocorrelation parameters. Time delays are used to prevent model updating at the beginning and end of an inhalation noise to insure that only valid inhalation noises are used for the update.

The results of SCBA speech processed by the ARINA algorithm are shown in Figure 5. This speech was recorded from a male speaker in a quiet room. The top waveform shows the input speech. The middle waveform is the noise model filter raw prediction gain. The bottom waveform is the ARINA and AMSE processed output. The inhalation noises have been eliminated except for small artifacts at the start and end of each noise. The speech waveform is altered due to the equalization.

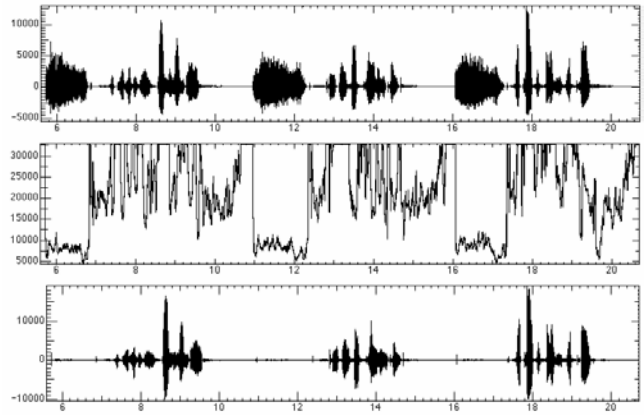

Figure 5. An SCBA speech signal with air-regulator inhalation noise (top); ARINA filter raw prediction gain (mid); and attenuated noise after ARINA and AMSE (equalization) processing (bot).

The purpose of the AMSE algorithm is to reduce the effects of SCBA mask spectral distortion using equalization methods. An estimate of the spectral magnitude acoustic transfer function for a given mask and wearer can be obtained from the inhalation noise spectrum since both signals originate in the mask, albeit not at the same source location. AMSE estimates the mask transfer function using an LPC filter and then uses the inverse filter to equalize the mask speech signal. It uses the inhalation noise detector and noise model adaptation functions of ARINA to insure that the equalizer characteristics match the current mask configuration. Thus the system can be used with any style mask and adapts to the wearer.

A block diagram of the AMSE algorithm is shown in Figure 6. As can be seen the AMSE algorithm is an extension of the ARINA algorithm with the addition of a speech equalization section shown within the dashed lines. This section is comprised of a filter estimator, a pre-filter, an equalizer, a post-filter and gain compensator. The equalization model estimator section uses the inhalation noise detection and autocorrelation parameter update portion of the ARINA algorithm. The inhalation noise reference autocorrelation coefficients are used to generate an nth order LPC model of the noise. Experiments have shown that a 14^{th} order model is suitable for some masks but a higher order can be used. The noise model coefficients are used in an inverse filter through which the speech signal is passed. The energy of the inverse filter input and output are compared and the gain of the output is adjusted to equalize the signal energy. This step is necessary to make sure any gain alterations due to the equalization process do not cause signal clipping. Additional fixed pre or post filtering to correct for any non-whiteness of the inhalation noise, or to give the speech signal a specified tonal quality to optimally match the requirements of a following specific codec or radio, is also applied in the equalization process. A key difference between the AMSE algorithm and a similar LPC approach by Vassilev [4] for equalizing diving masks is that AMSE is dynamic and continually adaptive, and requires no mask-specific offline equalization parameter estimation.

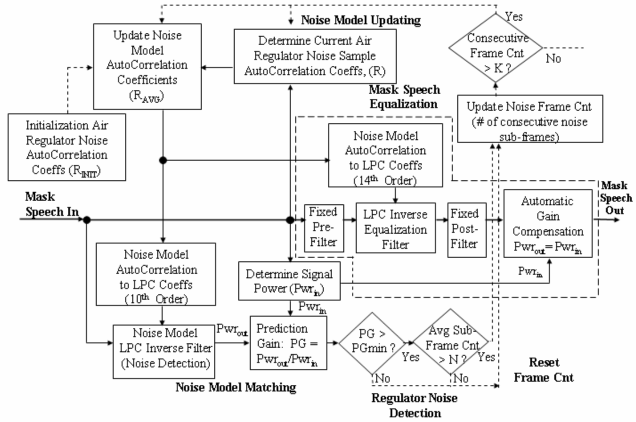

Figure 6. A block diagram of the AMSE algorithm for equalizing acoustic distortion caused by an SCBA mask.

The results of some SCBA mask speech processed by the AMSE algorithm are shown in Figures 7 and 8. Figure 7 shows the magnitude spectrum of a SCBA inhalation noise burst and the AMSE equalized output using 14^{th} and 20^{th} order inverse filters.

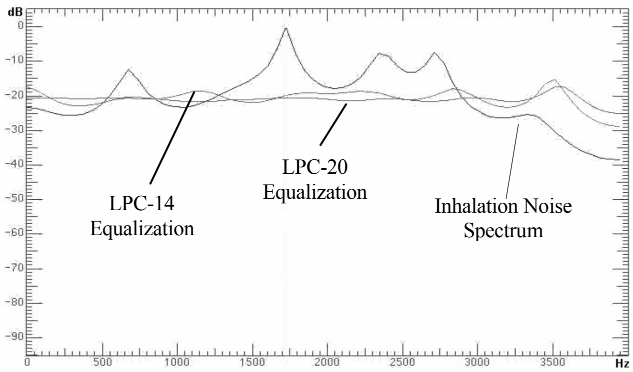

Figure 7. The magnitude spectrum of SCBA mask regulator inhalation noise before and after equalization by the AMSE algorithm using 14^{th} and 20^{th} order inverse filters.

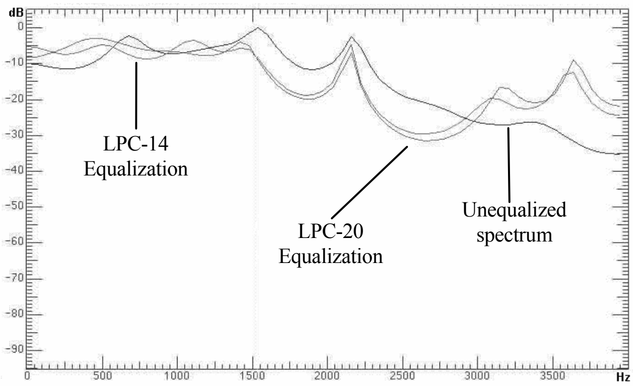

Figure 8. The magnitude spectrum of the sound /u/ in the word "two" before and after equalization by the AMSE algorithm using 14^{th} and 20^{th} order inverse filters.

Figure 8 shows the un-equalized and equalized output for the /u/ sound in the word “two” spoken in a SCBA mask. The equalization provides improved formant resolution at the high end of the spectrum and reduces the influences of the mask resonances at 650, 1725, and 2700 Hz.

4. ALGORITHM PERFORMANCE TESTING

In order to test the effectiveness of the AMSE equalization algorithm, PSQM (Perceptual Speech Quality Measure) tests [5] were performed. A test corpus consisting of a total of 192 Harvard Balanced sentences recited by 8 speakers (4 males, 4 females) was used. All 192 sentences were concatenated as a single file for processing. This audio database was played through the artificial mouth of a B&K 4128 HATS and recorded with a B&K 4133 microphone, without and with an SCBA mask mounted on the HATS (Figure 1). In addition, Gaussian white noise (GWN) and SCBA air regulator noise (ARN) samples were played and recorded in a similar manner as spectral references. The ARN sample was a free space recording of gas release from the regulator without the influence of the mask (unattached). The speech and reference noise files, originally recorded and sampled at 16 KHz, were passed trough a 3950 Hz cutoff anti-aliasing lowpass filter and downsampled to 8 KHz. Two sets of 14th order autocorrelation and LPC coefficients were derived from the spectral reference noise samples with and without the SCBA mask mounted on the HATS. These characterized the reference noise models and were used for the inverse equalization filter and post-filter respectively in the AMSE equalization algorithm. In normal operation with a live subject, the AMSE and ARINA equalization and detection model reference noise samples would be obtained from regulator inhalation noises produced during breathing. The post-filter model references would still be obtained offline since they are only influenced by the SCBA equipment and not the wearer.

PSQM test software used for testing conformed to the ITU specification as described in [5]. The PSQM test requires a reference audio file and a test file that must be time-aligned for meaningful results so all files were aligned before processing. AMSE equalization was done using both the white noise and the regulator noise recordings as reference models. The data files were also processed with and without an APCO 25 Project compliant codec operating at 4400 bps. A PSQM score of 0.0 represents a perfect match between the reference and test files and no distortion. A higher number represent greater distortion and a reduction in audio quality. Unfortunately, the PSQM score cannot be automatically related to an MOS score unless an MOS reference test is performed on the same data sets.

Table 1 shows the PSQM test results for all of the processed data. The AMSE equalization considerably improves the distortion in all cases as measured by the PSQM test. This is true for all cases and for both types of reference noise. For both the Gaussian white noise (GWN) and the air regulator noise (ARN) the distortion as measured by PSQM drops by more than half. The use of post filtering in these cases further improves the scores as might be expected since the post-filter models the actual shape of the reference noise, accounting for the fact that the recorded air regulator reference noise source is not white. Slightly better improvements are seen for the codec-processed data. Although the PSQM tool has not been characterized for testing parametric codecs, these results seem to indicate that AMSE does provide speech quality improvement with the digital speech codec. Other qualitative and quantitative tests are planned to better evaluate algorithm quality and intelligibility improvement performance.

Data Set Processing	Codec Processed	PSQM Score	% Change
SCBA Mask Speech, No Equalization	No	0.132396	0
Equalized, GWN Ref.	No	0.076759	-42
Equalized, GWN Ref., with Post Filter	No	0.037857	-71
Equalized, ARN Ref.	No	0.067550	-49
Equalized, ARN Ref., with Post Filter	No	0.054056	-59
SCBA Mask Speech, No Equalization	Yes	0.160589	0
Equalized, GWN Ref.	Yes	0.064348	-60
Equalized, GWN Ref., with Post Filter	Yes	0.052355	-67
Equalized, ARN Ref.	Yes	0.061894	-61
Equalized, ARN, with Post Filter	Yes	0.078465	-51

Table 1. PSQM score test results for SCBA corrupted speech, and with and without codec processing and AMSE algorithm equalization.

The shapes of the pre and post filters are free parameters that can have a large effect on perceived speech quality after equalization. More experiments need to be conducted to determine the sensitivity of the overall equalization to these filters. In addition, more general test methods for intelligibility and quality assessment such as the DRT, MRT, MOS, or PESQ are planned.

5. CONCLUSIONS

In this paper the acoustic distortion and noise caused by SCBA breathing equipment and two methods to mitigate the effects were discussed. The ARINA algorithm detects and removes inhalation noise from the speech signal while the AMSE algorithm reduces the spectral coloration caused by the SCBA mask. Both algorithms were shown to be effective in mitigating the effects of SCBA distortion and have been implemented in real-time hardware.

REFERENCES

[1] Bond, Z.S., Moore, T.J., and Gable, B., “Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask”, *J. Acoust. Soc. Am.*, vol. 85 (2), pp. 907-912, Feb. 1989.

[2] Vojnovic, M. and Mijic, M., “The influence of the oxygen mask on long-time spectra of continuous speech”, *J. Acoust. Soc. Am.*, vol. 102 (4), pp. 2456 – 2458, Oct. 1997.

[3] Woram, J. M., *Sound Recording Handbook*, Indianapolis, IN, Howard W. Sams & Company, 1989.

[4] Vassilev, A.V., “Improvement of the Diver Speech Intelligibility in Underwater Communications using LPC”, in *Proc. OCEANS '96,* MTS/IEEE, 'Prospects for the 21st Century', vol. 3, Sept. 1996, pp. 1504 - 1509.

[5] ITU-T Study Group 12, “Objective quality measurement of telephone-band (300-3400) speech codecs”, ITU-T Recommendation P.861, Aug. 1996.