Audio Engineering Society

Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

Automatically optimizing situation awareness and sound quality for an isolating earphone.

John Usher

Hearium Labs, San Francisco.
jusher@heariumlabs.com

ABSTRACT

Sound isolating (SI) earphones are increasingly used by the general public with portable media players in noisy urban and transport environments. The dangers of these SI earphones are becoming increasingly apparent and legislators are recommending an urgent review of their usage. The problem is that the user is removed from their local ambient scene: a reduction in their "situation awareness" that often leads to accidents involving unheard oncoming vehicles. This paper introduces a new automatic gain control system to automatically mix the ambient sound field with reproduced audio material to maintain a constant SNR and therefore enhance situation awareness. A discussion of the audio system architecture is given and an analysis of 20 different warning sounds is used to suggest suitable parameters.

1. INTRODUCTION: SOUND ISOLATING EARPHONES AND SITUATION AWARENESS

Music reproduction levels and urban noise are antagonistic: we play our personal audio devices louder to hear over the traffic and general urban noise. The same applies to voice communication, e.g. with mobile phones in loud urban environments such as bars and restaurants, where we have to either increase the level of the mobile phone to a reproduction level which is potentially damaging to our auditory system, or we have to leave the room to find a quieter place to take the phone call.

The majority of devices providing the consumer with speech or music audio do so using "in-ear earbuds", which deliver sound directly to the ear canal generating levels sufficient to perceptually mask background noise, even though the earbuds provide little ambient sound isolation. With earbuds, personal audio reproduction can reach maximum output levels in excess of 100 dB, even up to 121 dB (Fligor and Cox, 2004); enough to exceed recommended daily sound exposure levels in a number of minutes and close to the 132 dB peak level that has been shown to cause permanent acoustic trauma (Price, 1981). In fact, the Occupational Safety and Health Administration (OSHA) and National Institute of Occupational Safety and Health (NIOSH) state a damage risk criteria of 90 and 85 dBA for an 8-hour working day. Recent research (Shargorodsky et al, 2010) points to evidence that shows that the threat of noise induced hearing loss in adolescents has significantly increased over the past 10 years.[1]

Furthermore, rising population densities have continually increased sound levels in society. According to Berger (2003), 40% of the European community is continuously exposed to transportation noise of 55 dBA and 20% are exposed to greater than 65 dBA. This level of 65 dBA is considered by the World Health Organization to be intrusive or annoying, and as mentioned, can lead to users of personal audio devices increasing reproduction level to compensate for ambient noise.

The average consumer is becoming more aware of the risks of hearing damage from high-level audio reproduction levels. In the past, the user would increase playback levels due to high ambient noise levels that would otherwise create a poor signal-to-noise ratio (SNR). It is increasingly common to incorporate passive noise canceling systems by embedding the earphone loudspeaker in an acoustical isolation assembly, such as an ear plug ("insert-headphones" or "canal phones"). A typical sound-isolating "insert" earphone is described in Figure 1 that shows an insert earphone using multiple flanges to form a seal in the ear-canal, and that directs sound from a miniature loudspeaker via a tube into the ear-canal.

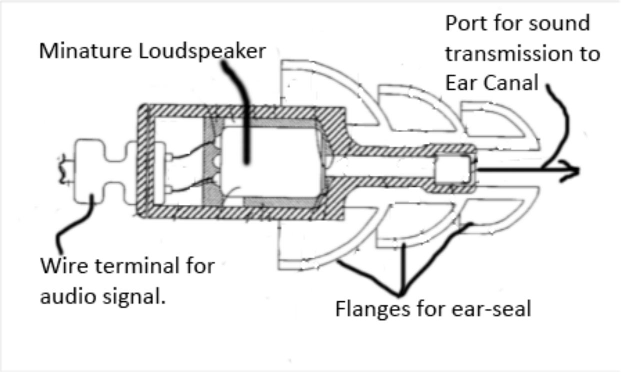

Figure 1: Cross-section of a typical sound-isolating earphone. Three flanges form an ear-seal in the ear meatus and sound is directed through a central port into the occluded ear canal from a miniature loudspeaker. Figure adapted from US patent #5,887,070.

1.1. Situation awareness

Passive sound attenuation earphones such as those shown in Fig. 1 are often used in live-sound for monitoring a music-audio mix. However, for everyday listening of reproduced audio (such as from a Portable Media Player or cell phone) these devices are potentially dangerous, as the user is acoustically isolated from ambient warning sounds such as car horns and fire alarms. In other words, the user has reduced *situation awareness (SA).*

Endsley's definition of SA is (1995*), "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"*. While some definitions are specific to the environment from which they were adapted, Endsley's definition is applicable across multiple task domains from visual to auditory modalities.

---

[1] Shargorodsky et al (2010) found that the prevalence of hearing loss among a sample of over 1500 US adolescents aged 12 to 19 years was greater in 2005-2006 compared with 1988-1994, increasing from 14.5% to 19.5%.

Acoustic isolation levels for sound isolating earphones can be as high as 30 dB at 125 Hz, to over 40 dB at frequencies above 4 kHz (Lancaster and Casali, 2004). The reduction in acoustic situation awareness with SI earphone devices is therefore obvious.

There is an unfortunately ironic relationship between the use of Hearing Protection Devices (HPDs), such as ear-muffs, and dangerous working environments. The use of HPDs by the military is a clear case of this: for example on flight-decks or by soldiers in the battlefield. Here, the very devices that are designed to protect and preserve a soldiers' hearing, i.e. HPDs, may also manifest the same sensory-perceptual impairments in their performance, even if they possess normal hearing acuity (Casali, Ahroon, & Lancaster, 2009).

Alternatively or complimentary, active noise control (ANC) systems use acoustic phase-inversion to reduce the low-frequency ambient noise transmitted to the ear-canal. ANC systems typically give up to 15 dB of attenuation for frequencies up to approximately 300 Hz. As an example of a hybrid passive and active noise control system, US patent #7039195[2] combines both a passive ear-sealing canal-phone with an active noise control system. ANC is particularly complimentary to passive SI systems when the ear seal material is foam, as foam has a low-pass attenuation characteristic (i.e. less sound attenuation at low frequencies).

US Pat. No. 2006/0182287 describes a headset monitoring system, with microphones mounted at the entrance to an occluded ear canal (i.e. Ambient Sound Microphones, ASMs), and ear canal loudspeakers mounted in the same earphone. In this patent description, the ambient microphone signals are reproduced with ear-canal loudspeakers with a user-defined gain, allowing the user to monitor both their local environment and Audio Content signals, and to *manually* control the relative level of these two signals. The present paper builds on this idea by *automatically* mixing an ambient sound signal with a reproduced speech or music signal: allowing the user to hear both warning signals in their immediate environment and reproduced sound.

To allow a listener to maintain good intelligibility of a voice or music signal reproduced with earphones, some extant work describes how to automatically adjust the level of reproduced audio in response to a changing ambient sound level. Such an "Automatic Gain Control" (AGC) system is described in Application US #2006/0262938[3] for increasing the level of reproduced audio when the background noise level is high. This earphone AGC system is similar to the AGC systems in cars: where the level of reproduced sound from the vehicle audio system is increased as the vehicle speed increases, to compensate for the masking-effects produced by increased engine and tire noise. The compressor-curve shown in Figure 2 below describes how the earphone Automatic Gain control (AGC) system operates. The AGC system is used with a headphone with an external microphone to measure the ambient sound level. An AGC circuit affects the gain of the input audio signal: e.g. an input speech audio or music audio signal. When the input audio signal level is low compared with the ambient sound level, then the audio signal is compressed: i.e. amplified to narrow the dynamic range, and thus allowing the earphone user to hear the reproduced audio even during quiet music passages.

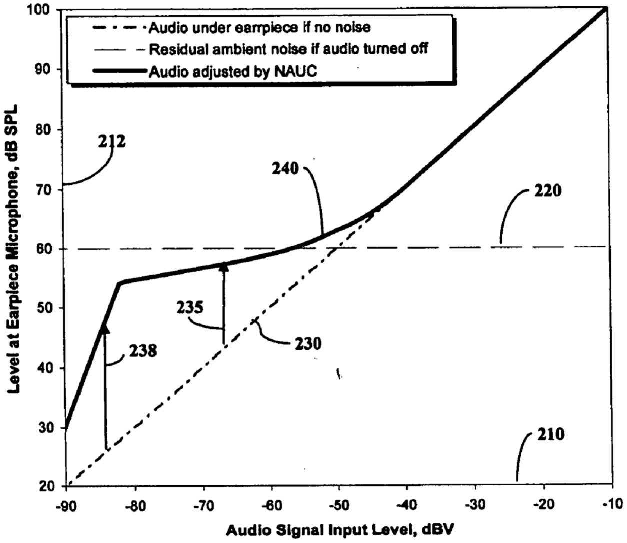

FIGURE 2: Automatic Gain Control compression curve for an earphone device described in US patent application #2006/0262938.

---

[2] "Ear Terminal", J. Svean, 2006. US patent #7039195.

[3] "Adapted audio response", D, Gauger et al, 2005, US patent application #2006/0262938.

2. HEARIUM EARPHONE

A new earphone has recently been developed to address the personal safety issues discussed in this paper: namely, passive sound attenuation so that reproduced audio can be auditioned at safe (low) listening levels in high noise environments, and to simultaneously enable sonic situation awareness of the earphone user. The earphone design is described in Figure 3. The ear-sealing mechanism is a unique air-filled bladder, much like a vascular stent used in angio-plasty. It provides approximately 30 dB of attenuation, with maximal attenuation at low-frequency (unlike conventional foam earplugs: hence active sound attenuation is not necessary). The balloon ear seal adapts to most ear canal shapes, and avoids the irritation and wax-compaction problems associated with foam or flange type ear-seals. The new earphone device has two microphones: an Ambient Sound Microphone (ASM) and an Ear Canal Microphone (ECM). The ECM detects sound in the occluded ear-canal via a port through the sealing balloon, so the sound level presented to the ear drum can be accurately determined. (The ECM can also be used for voice communication in high noise environments via the “bone conduction” acoustic path.) A miniature loudspeaker directs sound to the ear canal, via a port next to the ECM channel. An input audio signal (e.g. from a speech or music source) is automatically mixed with the ambient ASM microphone signal, so as to maintain a roughly constant signal-to-noise ratio: ensuring acoustic situation awareness regardless of the reproduction level chosen by the earphone user. The “constant SNR” system is now described in figure 4.

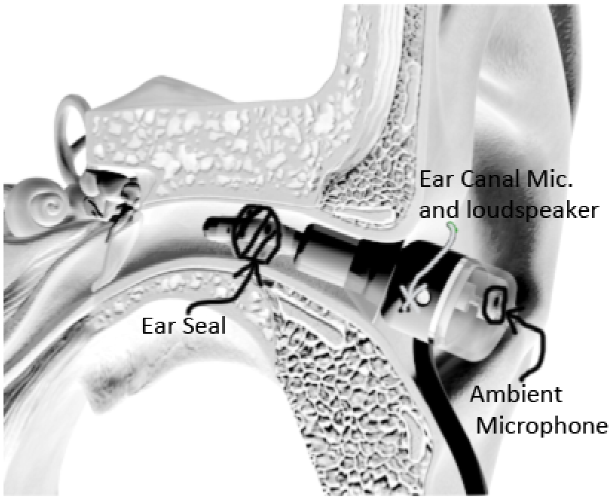

FIGURE 3: Schematic overview of a development earphone to enable passive sound isolation and situation awareness of the earphone user.

3. CONSTANT SIGNAL-TO-NOISE RATIO SYSTEM

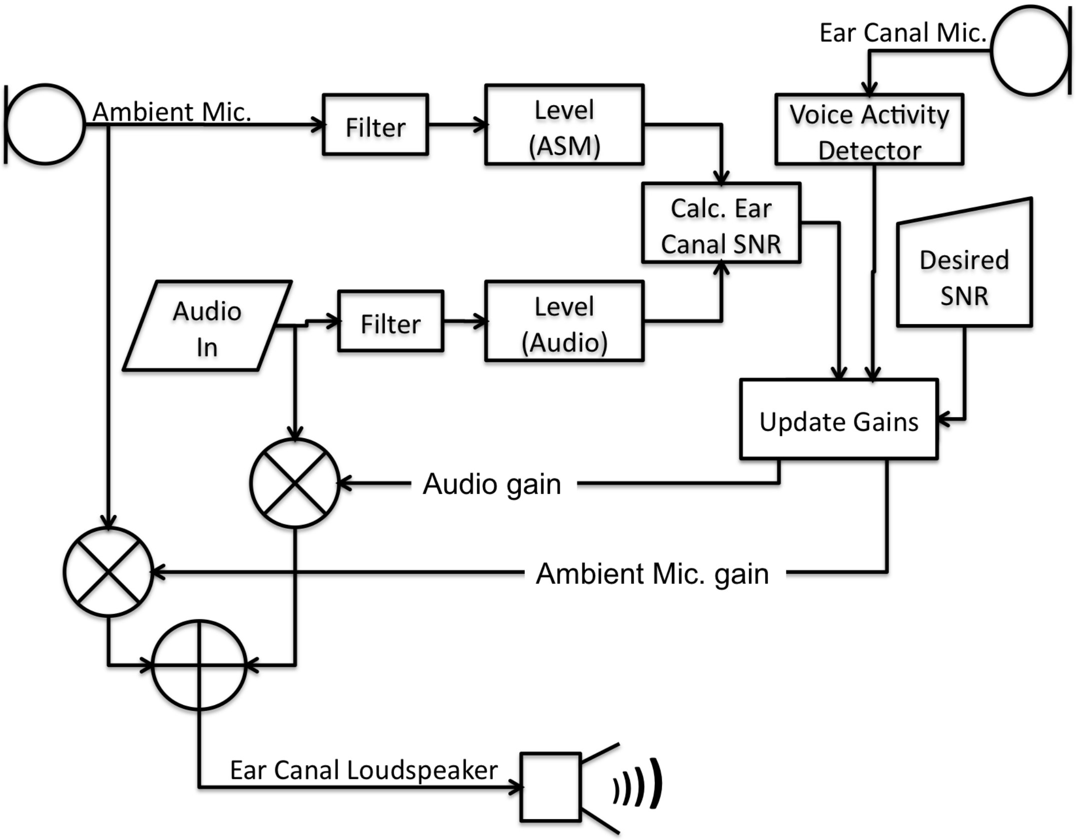

FIGURE 4: Functional overview of the Constant SNR system.

Figure 4 describes the “Constant SNR system” that maintains situation awareness of the earphone user, whilst the earphone is simultaneously listening to reproduced audio, e.g. reproduced speech or music. Let us first establish what is meant by the “signal” and the “noise” here, referring to sound reproduction with the earphone device described in figure 3. The sound pressure level generated by the reproduced audio, measured in the ear canal, is taken to be the “signal” level. The ear-canal level of sound from the users ambient environment is taken to be the “noise” level (though this is not an ideal term, as the “noise signal” may sometimes contain important sonic information such as warning alerts). The ear-canal “noise” level is affected ambient sound in the sealed ear canal that arrives via leakage through the earphone ear-seal and also by bone-conduction paths.

As already discussed, the acoustic leakage component through the ear seal is generally less than 30 dB relative to the level in the ambient sound field (hence can almost be ignored). However, the new earphone can also *actively* pass ambient sound into the ear-canal: much like a hearing-aid using the ambient sound microphone and the ear canal loudspeaker, via a signal amplifier. The total “noise” level in the ear canal is therefore a combination of the *passive* ambient leakage component and the *active* ambient sound pass-through component.

The "Constant-SNR" system is a slight misnomer, because the system does not maintain an *exactly* constant SNR. On the contrary, the system deliberately only approximates a "desired" SNR. Particularly, it is important to allow the actual ear-canal SNR to be less than the desired SNR so that sudden local sound onsets are not immediately attenuated, therefore allowing the user to hear and localize this potentially dangerous local transient sound, albeit briefly. The automatic detection of transient sounds is configured by special selection of gain time constants that affect the ambient sound signal. For instance, a slow ambient mic. gain decay would cause the ear-canal "noise" level to slowly decrease following a sudden ambient sound event.

Looking at Figure 4, we can see that the audio input and ambient microphone gains are calculated using a running estimate of the ear-canal SNR: by the noise level is calculated as discussed, by considering both the passive sound leakage and the active sound pass-through. The noise level, L_n (in Pascals) can therefore be calculated as:

$$\begin{aligned} L_n &= L_A * NRR + L_A * G_{AS} \\ &= L_A(NRR + G_{AS}), \end{aligned}$$

where L_A is the ambient sound level (in Pascals), NRR is the Noise Reduction Rating (i.e. earphone attenuation), as a no-unit scaler (i.e. linear, *not* dB, which can generally be ignored), and G_{AS} is the gain applied to the ambient sound microphone signal before it is reproduced with the loudspeaker (for the sake of simplicity, we ignore any sensitivity mismatch between the microphones and loudspeakers, i.e. we assume that if G_{AS} is unity, the ear-canal SPL generate by the ambient sound microphone signal is the same as the SPL at the microphone).

Likewise, the input audio signal level, L_s, can be calculated in a similar manner:

$$L_s = L_{s\_in} * G_S,$$

where L_{s_in} is the SPL level that would be generated in the ear-canal if the audio was directly reproduced with the ear canal loudspeaker, and G_s is the signal gain applied to the audio signal.

As shown in Fig. 4, the ambient sound and audio signal levels are calculated via frequency weighting and temporal smoothing, e.g. using A-weighting and a leaky-integrator with a time constant of approx. 100 ms.

The ear-canal SNR can therefore be calculated as a log-ratio between the signal and noise levels:

$$SNR = \log \frac{L_s}{L_n}.$$

The "desired" SNR may be obtained manually (e.g. the user can just "dial it in"), or it may be obtained automatically and dynamically adjusted, e.g. depending on whether the user is talking, on the background noise environment type, or on the reproduced audio type (e.g. speech vs music). The use of the ear canal microphone in the earphone is particularly conducive for accurately detecting user voice activity (e.g. by correlation with the other earphone), and can be configured to be used as an automatic "push to talk" switch when listening to music: e.g. muting the music playback and activating the ambient sound pass-through to talk to a nearby individual. Furthermore, we may which to disable the ambient sound level calculation whilst the user is talking (i.e. so the user voice level is not factored into the level estimate). The mismatch between the desired and actual ear-canal SNR is used to adjust the ambient microphone and audio signal gains so as to iteratively force the SNR mismatch to zero.

Various operating modes can be used to control the rate of change of the ambient microphone and audio signal gains, depending on the degree of signal distortion tolerated or the user circumstances. For instances, in a particular "high quality" mode of operation, we may wish to only adjust the ambient sound signal gain, so as to eliminate distortion artifacts from modulating the audio signal level (i.e. to minimize compressive "pumping artifacts). Alternately, for a "critical mission" scenario, it may be imperative to maintain a high SNR, so that incoming audio messages can be continuously heard (e.g. on a flight deck or space-ship).

We shall now look at how different configurations for the gain time constants affect detection of a sudden ambient sound event, and discuss suitable time constants for detection of different warning sounds.

4. SIMULATION OF THE C-SNR SYSTEM USING MUSIC AUDIO AND AN AMBIENT SOUND RECORDING

Offline simulations of the Constant SNR system were conducted to investigate the change in ear-canal SNR when different rise and fall gain time constants were used for the ambient sound microphone signal path. In the simulations, the audio signal level was not adjusted, constant with the “high quality music listening mode”, as previously discussed.

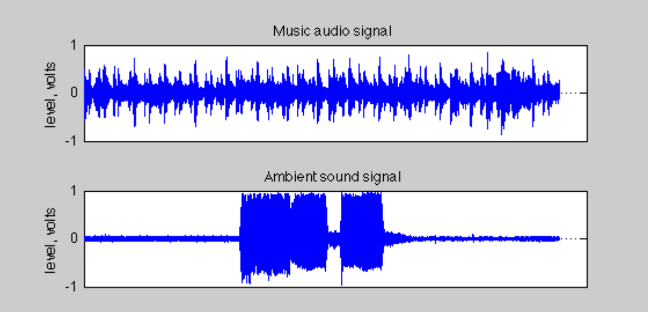

FIGURE 5 (above): Test input signals for simulations of C-SNR AGC system. The music audio signal was 15 seconds from the song “Higher Ground” by Stevie Wonder. The ambient sound signal was a recording from a street with a sudden car-tire screech with two car horn honks.

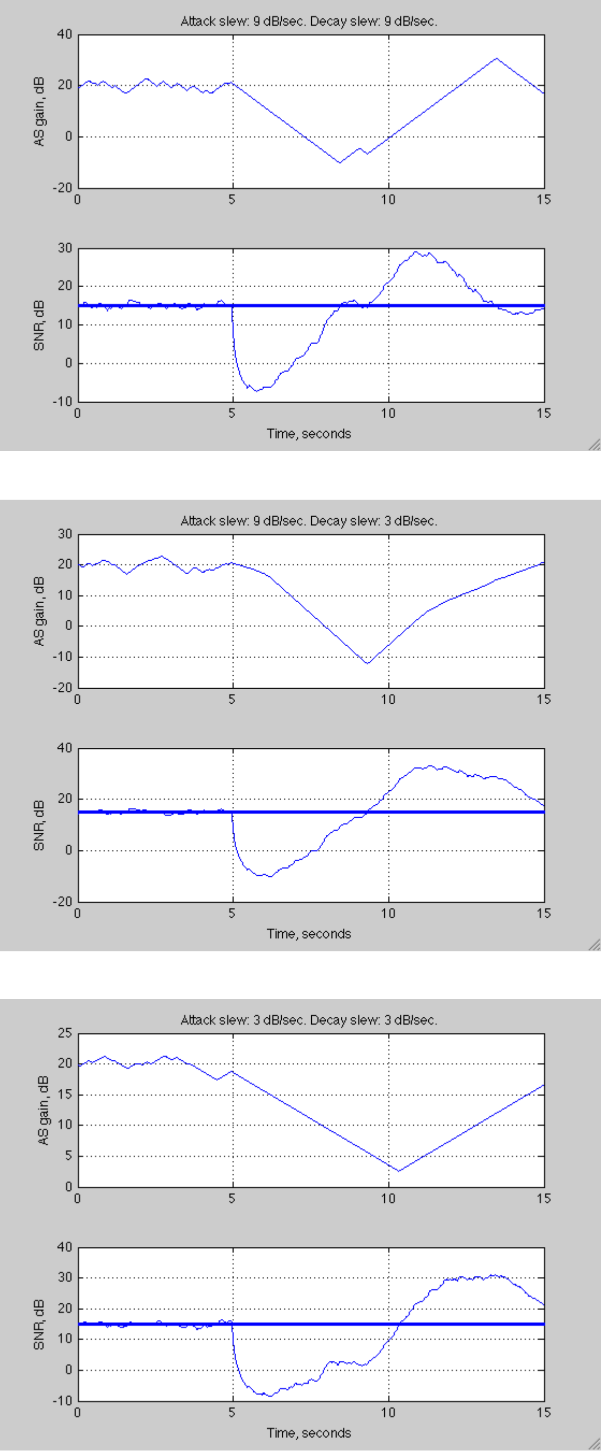

FIGURE 6 (right): Simulations showing Ambient Sound microphone gain and ear-canal SNR, with target SNR = 15 dB. Three simulations are shown for different ambient sound mic. gain attack and decay times.

5. DISCUSSION

Figure 6 clearly shows the general affect of the attack/decay rate for the ambient sound gain on the ear canal SNR. For low gain-change-rates (i.e. low *slew* values), we see that the SNR is tracked less robustly. However, the advantages of this slower tracking are also revealed: as we can see that the SNR remains lows for the duration of the ambient sound event, and the earphone user would thus be able to hear both events.

The rate of change of the ambient gain coefficient (i.e. G_{AS}) is a matter for careful selection. A brief analysis of typical ambient warning sounds show that ambient "Sounds of interest", e.g. horns, typically have extremely fast onset times.

20 recordings of horns were obtained from the freesound.org project[4] and their rise-times were analyzed. The initial 300 ms of the time-smoothed temporal envelopes of these recordings are shown in Figure 7. The onset times were then calculated by analysis of the gradient from the envelope maxima to the -6 dB point of the envelope, as shown in the box and whisker plot in Fig. 8.

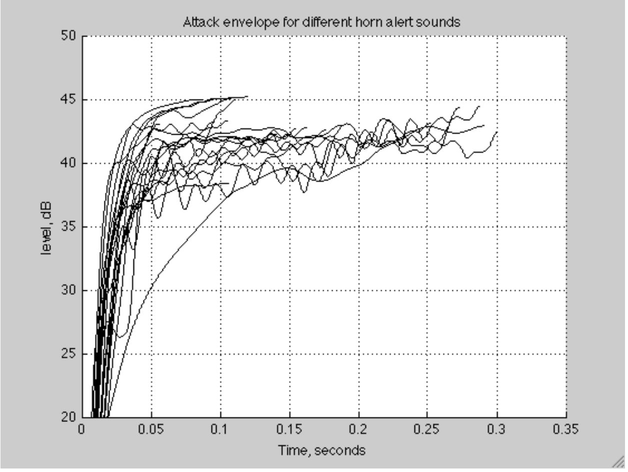

FIGURE 7: Initial 300 ms of 20 different warning horn sound recordings. Audio was normalized to equal RMS level. Envelope smoothed using 100 ms Hanning window. Envelope data is not shown past the maxima.

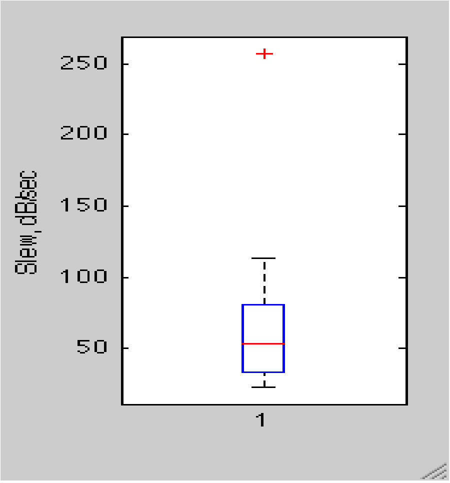

FIGURE 8: Box and whisker plot showing distribution of 20 horn rise times or slew, in dB/sec (same data as Fig. 7). Outlier is due to the oscillating envelope of one horn.

The onset times for the horn are mostly within the range of 40-75 dB/sec: indicating the decay slew rate for the ambient sound microphone should be kept much lower than this range in order for the earphone user to maintain situation awareness for similar transient sound events.

6. CONCLUSION

The purpose of this paper is to draw attention to a unique way of automatically detecting transient onsets in an earphone users' environs, and ensuring that the individual remains a safe degree of sonic situation awareness. Rather than defining a precise recipe for the circuitry of the "constant signal to noise ratio" system, the present paper has described a general audio system architecture for automatically mixing the ambient sound field with reproduced audio. The C-SNR system is ideally suited for use with a sound isolating earphone containing an ambient and (optionally) ear canal microphone, and a miniature loudspeaker directed at the occluded ear canal. The system does not maintain an *exactly* constant SNR: On the contrary, it deliberately only approximates a "desired" SNR. Particularly, it is important to allow the actual ear-canal SNR to be *less* than the desired SNR so that sudden local sound onsets are not immediately attenuated, therefore allowing the user to identify and localize this potentially dangerous local transient sound. The

---

[4] Sounds were selected from www.freesound.org.

automatic detection of transient sounds is configured by special selection of gain time constants that affect the ambient sound signal. Future work will discuss how the hardware architecture can further protect hearing and maintain personal safety of the earphone user.

7. REFERENCES

Berger, E.H., Royster, L.H., Royster, J.D., Driscoll, D.P., Layne, M. (2003) *The Noise Manual*. Revised 5th ed. Fairfax, VA: American Industrial Hygiene Association (AIHA) Press.

Casali, J. G., Ahroon, W. A., and Lancaster, J. (2009). *A field investigation of hearing protection and hearing enhancement in one device for soldiers whose ears and lives depend upon it*. Noise and Health Journal, 11(42) 69-90.

Endsley, M.R. (1995). *Toward a theory of situation awareness in dynamic systems*. Human Factors 37(1), 32–64.

Fligor, B.J. and Cox, L.C. (2004). *Output levels of commercially available portable compact disc players and the potential risk to hearing*. Ear and Hearing; 25:513–527.

Lancaster, J. A. and Casali, J. G. (2004). *REAT and MIRE Insertion Loss Comparison for Eight Headphones of Various Passive and ANR Designs, Inclusive of NRR and Spectral Attenuation*, Final Report (ISE Department Technical Report 200402, Audio Lab Number 9/07/04-2-HP). Blacksburg, VA: Virginia Tech, Grado Department of Industrial and Systems Engineering, Auditory Systems Laboratory.

Price, G. (1981). *Implications of a critical level in the ear for assessment of noise hazard at high intensities.* Journal of the Acoustical Society of America; 69: 171–177.

Shargorodsky, J., Curhan, S.G., Curhan, G. C. and Eavey, R. (2010). *Change in Prevalence of Hearing Loss in US Adolescents*. Journal of the American Medical Association; 304: 772-778.