Applying Active Noise Control Technique for Augmented Reality Headphones

Rishabh Ranjan[1]; Gan Woon-Seng[2]; Chong Yong-Kim[3]

Digital Signal Processing Laboratory

School of Electrical and Electronic Engineering,

Nanyang Technological University, Singapore

ABSTRACT

With the increased popularity of wearable devices, we are seeing head gears (such as Google glass) that provide augmented visual information. We can also extend augmented information to audible signals to guide users in finding their way, alert user to prominent danger in noisy environment, or simply creating virtual-and-augmented reality gaming platform. In this paper, we outline new approaches in combining real sonic environment and augmented virtual sound source through an open-ear headphones. Microphones are positioned in the headphones to capture the sonic information as well as the signal playback to the headphones. Active noise control (ANC) techniques have conventionally been used for noise cancelling headphones, and this paper shows how we can apply ANC techniques in augmented reality headphones to compensate for the sonic difference between augmented and real sound objects and provide a seamless combination of the two. Some interesting new applications using the augmented reality headphones can be realized with this augmented reality headphones, and open up new possibilities for other interactive applications.

Keywords: Augmented reality, Active noise control

1. INTRODUCTION

Augmented reality (AR) is enhancing our daily life experiences by combining virtual reality with the real world environments, especially with the advent of wearable devices, such as Google glass. These AR devices mainly provide visual aids to the listeners in an augmented reality environment (ARE). In recent years, we are seeing spatial audio information being included in the ARE via headphones to provide listeners an immersive natural sound experience combining virtual and real sounds, alert them to prominent danger in a noisy environment, and adding realism in the augmented reality gaming. The main purpose of augmented reality audio via headphones is to reproduce the virtual auditory scenes as close as possible to real sounds, as well as a seamless combination with real auditory environments.

This paper outlines an active noise control (ANC) based approach to combine real sonic environment with augmented virtual sounds with the help of open-air headphones. As contrast to close-back headphones, which require compensation for headphones attenuation, open-air headphones do not isolate most of the direct sound and allow a more natural listening. In this paper, we use two pairs of microphones (one external and one internal) that are positioned in the external and internal of the headphones' shell. The external microphone-pair captures the real sound while the internal microphone-pair captures both the real sound and the virtual sound that are played from the emitters of the open-air headphones.

ANC techniques are generally used in noise cancelling headsets to cancel the primary ambient noise coming from fan, generators, engines, transformers, etc. In this paper, we show how ANC

[1] rishabh001@ntu.edu.sg
[2] ewsgan@ntu.edu.sg
[3] ekychong@ntu.edu.sg

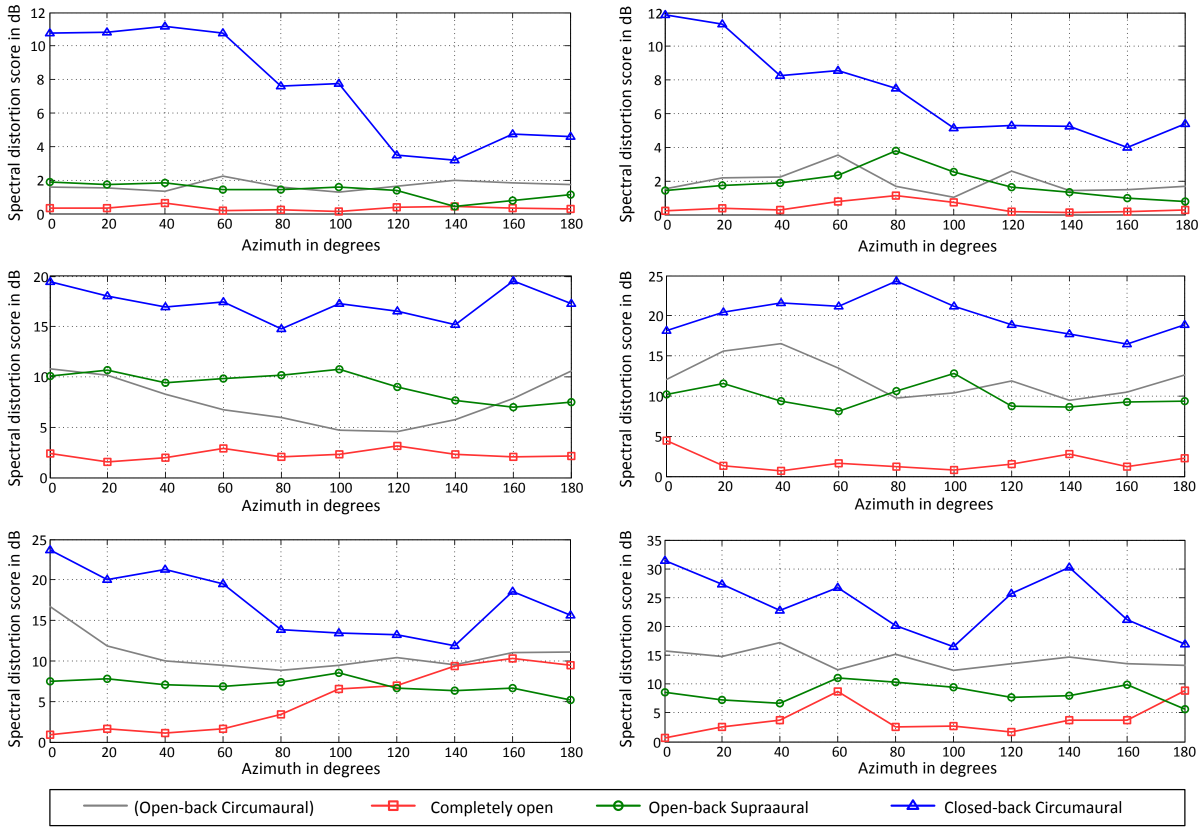


Figure 1 Spectral distortion (SD) score for the different headphones in low frequency (0.1-1.5 KHz) – Top; mid frequency (1.5-7 KHz) – Middle; and high frequency (7-16 KHz) – Bottom; for contralateral (left) and ipsilateral (right) ears.

techniques can be applied in AR headsets, enabling natural listening with headphone playback in the presence of real auditory environments. A modified version of the filtered-x normalized least mean square (FxNLMS) adaptive algorithm is proposed along with online adaptive estimation of external sounds to ensure augmented headphone playback signals perceived as real as possible even in the presence of real sonic environments. Simulation results showed improvements in convergence rate and steady state performance.

The remainder of this paper is organized as follows. Section 2 discusses the different headphones isolation characteristics focusing on the suitability of open headphones in AR applications. Section 3 introduces an ANC technique to be employed in the proposed AR headset, followed by simulation results in Section 4. Section 5 concludes the paper with brief discussion on the use of proposed AR headset in potential application scenarios.

2. HEADPHONES ISOLATION CHARACTERISTICS

When augmented reality sound objects are presented to a listener using AR headsets, the first and foremost requirement is that real external sounds should reach listeners' ears as naturally as possible. In other words, AR headsets should not hinder the sound path of any real sound sources. But, in practice, headsets do alter the direct sound spectrum due to the reflections with the headphones outer shell that also attenuates the sound signal entering the ear. We compare the spectral variations in the direct sound spectrum for different types of headphones using a widely used objective metric [1,2] known as the spectral distortion (SD) score:

$$SD = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(20log\frac{|H(f_k)|}{|\hat{H}(f_k)|}\right)^2} \qquad [\text{dB}], \tag{1}$$

where, $H(f)$ is the magnitude response of direction sound spectrum without headphones, $\hat{H}(f)$ is the magnitude response for the head-related transfer function (HRTF) measured with headphones present on the dummy head, and K is the number of frequency bins in the observed range (100 Hz – 16 kHz).

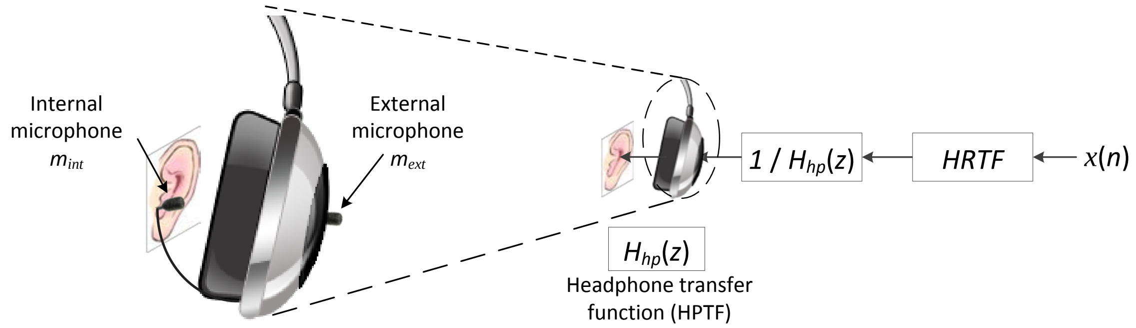


Figure 2 Binaural synthesis via headphone for virtual sound reproduction and AR headset structure

Thus, the smaller the spectral distortion is, the closer is the measured HRTFs to the reference direct sound spectrum. Figure 1 shows the spectral distortion scores in different frequency bandwidths for four different types of headphones (open-air circumaural, open-air supraaural, completely open, and closed-back circumaural) and 0 to 180 degree azimuths. Closed-back circumaural headphones isolate direct sound the most over the entire frequency range with the highest SD scores, as shown in Figure 1. As expected, completely open headphones with its drivers positioned externally, is closest to the reference direct sound spectrum with the least spectral distortion for all the three frequency regions. In low frequency, mean SD score of around 8 dB is observed for closed-back headphones, which is significantly more than the mean SD scores (less than 2 dB) for other open-air headphones. In mid-frequency region, spectral distortions for closed-back circumaural headphones are roughly 10 dB more than the open-air circumaural and open-air supraaural headphones, while around 15 dB more than the completely open headphones. In the high frequency regions, open-air supraaural headphones is found to be closer to the direct sound spectrum than the open-air circumaural headphones with around 5 dB lesser SD scores especially for ipsilateral ear. Additionally, it is also observed that the frontal pinna notches in high frequencies [3], which provide important cue for frontal perception, are suppressed in the case of open-air supraaural headphones and closed-back circumaural headphones. The reason why frontal pinna notches are suppressed in the supraaural headphones is due mainly to the direct projection of the sound signal to the ear opening without any filtering by the listener’s pinnae reflection. For the closed-back headphones, strong isolation by the headphones shell possibly results in the missing pinnae notches. To summarize, open-air headphones are more suitable for the choice of AR headsets than closed-back headphones as they allow more of the direct sound to pass through and reach listeners’ ear without much attenuation. Closed-back headphones isolate the direct sound and there is a need to compensate for the passive attenuation in order to perceive the real sound sources with higher realism. As shown by authors in [4], generic equalization methods for in-ear headphones can be used to compensate for the headphone isolation. In this paper, we will focus on the use of an open-air AR headset to circumvent additional processing for real sound source perception and facilitates natural listening. In the next section, we shall introduce the ANC based approach to reproduce the augmented reality auditory scenes using headphone playback.

3. ANC TECHNIQUES IN AN AR HEADSET

The main challenges for the AR headset are to reproduce the virtual signals as real as possible, and at the same time, the internally generated virtual sources should be seamlessly blended in the listeners’ local environment. Information about the sound propagation from source position to listeners’ ears in a given environment is needed to emulate the virtual sound reproduction as close to the physical sound objects. HRTF measured in a room environment is used in binaural synthesis via headphones to create an illusion of sound coming from surroundings. Therefore, head-related impulse response (HRIR) filters the source signal to account for the natural propagation from source to the listener, as shown in Figure 2. In addition, these transfer functions has to be measured for every individual before playing back to the headphones to create a realistic sound experience to the listener. However, the presence of headphones also colors the intended HRTF filtered signal, resulting in an unnatural sound perception. Consequently, the desired HRTF must be equalized to compensate for the headphone transfer function (HPTF) using an inverse filter, as shown in Figure 2. It should be noted

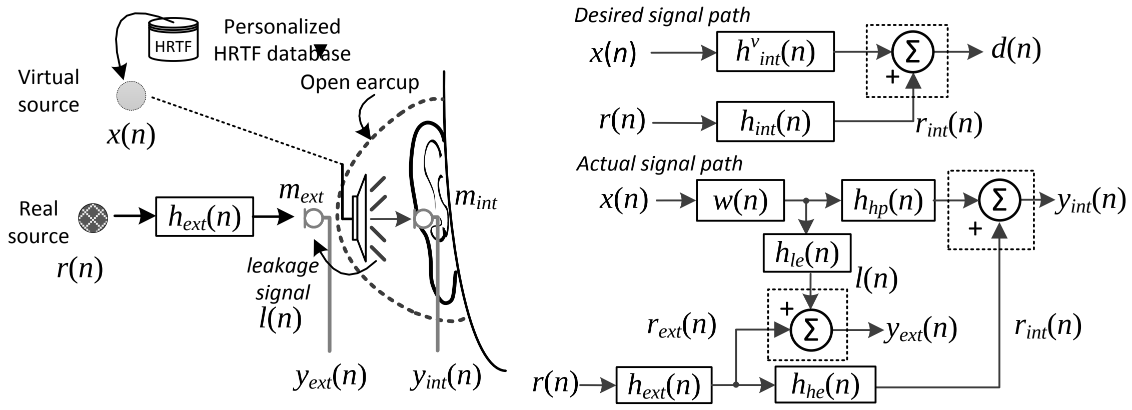


Figure 3 Augmented reality scenario with both virtual and real source present and corresponding signal flow block diagram (Extracted from (6))

that HPTF also changes with every reposition of headset and is unique due to the pinnae, and thus, individualized headphone equalization is required for acoustically transparent headphone listening [5]. Both individualized HRTFs and HPTFs are generally measured using the binaural microphones fixed near ear canal and subsequently used in binaural synthesis via headphones. Unfortunately, binaural synthesis using most of today's commercial headphones suffers from large localization errors and reproduction inaccuracies due to the use of non-individualized transfer functions.

In this paper, we consider an AR headset structure that was first proposed in [6] with two pairs of microphones mounted in open-ear headphones, as shown in Figure 2. On each side of the headphones ear cup, internal (m_{int}) and external microphones (m_{ext}) are positioned near the ear opening and just outside the ear cup, respectively, as shown in Figure 2. The AR headset with open-air ear cup will ensure natural perception of external sounds and with the help of m_{ext}, we can capture external sounds coming from surrounding as well as adapting the augmented headphone playback signals to the real sources. In addition, individualized transfer functions can be readily measured using the internal microphones of the headset. This work focuses on the use of adaptive equalization method to compensate for the individual HPTF, as well as adapting to any change in HPTF due to headset repositioning in the presence of physical sound sources. Figure 3 shows the most general scenario for a user wearing AR headset when virtual and real sound sources are present in the augmented reality environment along with corresponding signal flow block diagram. Impulse responses, $h_{int}(n)$ and $h_{ext}(n)$ are the head related impulse responses (HRIR) measured in a listener environment at the two microphone positions m_{int} and m_{ext}, respectively, for an external sound source. Similarly, $h_{int}^{v}(n)$ and $h_{ext}^{v}(n)$ represent the desired HRIRs corresponding to virtual sound reproduction selected from the personalized HRTF database for the intended virtual sound source position. As shown in Figure 3, real external signal $r(n)$ propagates through the air and reaches the listeners' air after passing through the passive headset structure. $h_{he}(n)$ represents the headphone effect impulse response accounting for the headphones effect between the two microphone positions. Headphone-effect(passive) transfer function is thus, expressed as:

$$H_{he}(z) = \frac{H_{int}(z)}{H_{ext}(z)} \quad . \tag{2}$$

Ideally, we would prefer virtual sources to be reproduced exactly like real sound sources by modeling the transfer function, $H_{int}(z)$, as shown by the desired signal path in Figure 4. But due to the HPTF, represented by $h_{hp}(n)$, an equalized filter $w(n)$ is used to convolve the virtual monaural signal such that

$$W(z) = \frac{H_{int}^{v}(z)}{H_{hp}(z)} \quad . \tag{3}$$

In addition, real sound sources from surrounding are also acoustically added along with the virtual signals at listeners' ear and are acquired simultaneously by m_{int}, as shown in the Figure 4. However, real sound source, $r_{int}(n)$ at m_{int} behaves as an undesired noise to the adaptive equalization of $w(n)$ and need to be removed using an estimate of $r_{int}(n)$. An accurate estimate of $r_{int}(n)$ can be easily

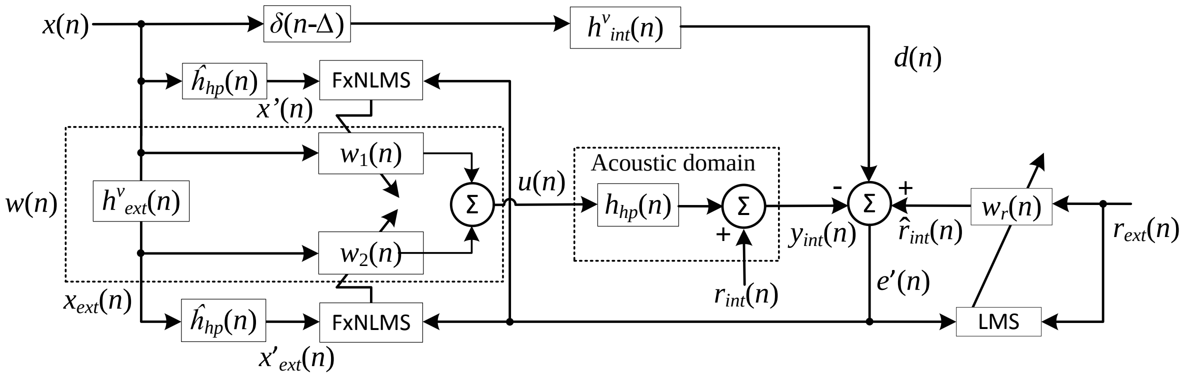


Figure 4 Block diagram of proposed hybrid adaptive equalizer with adaptive estimation of real signals (From [6])

obtained using a reference signal having some prior information. Therefore, real signal $r_{ext}(n)$ captured just outside the headset at m_{ext}, is used to estimate the $r_{int}(n)$. However, a leakage signal is also acquired from inside of headset due to the headphone playback signal and might affect the estimation process. Headphone leakage transfer function is represented by $h_{le}(n)$, as shown in the Figure 3. In this paper, we assume the leakage signal to be negligible to simplify the estimation process. Next, we present the adaptive equalization approach applying ANC techniques for the AR headset.

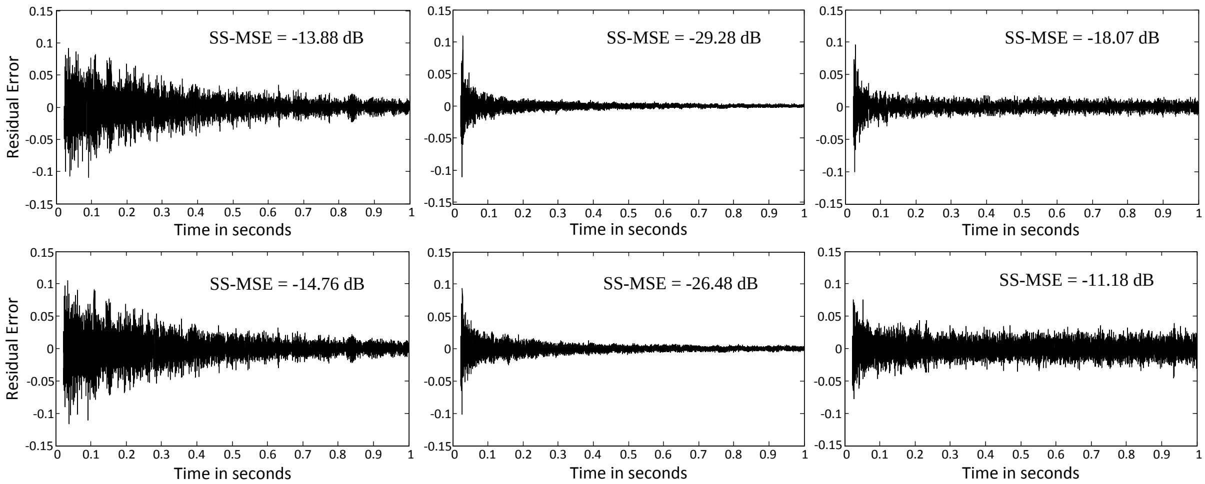
Figure 4 shows the proposed hybrid adaptive equalizer with adaptive estimation of real signals based on the filtered-x normalized least mean square algorithm (FxNLMS). Normalized version of the filtered-x least mean square algorithm (FxLMS) is used since FxLMS suffers from slower convergence because of the presence of a filter in front of the adaptive filter in the secondary path. The aim of the adaptive equalizer is to achieve faster convergence rate and least steady state error performance for AR headset to be used in practical scenarios, such that virtual source is perceived as realistic as real source. The presented adaptive equalizer is a simple combination of conventional FxNLMS and a modified FxNLMS presented by the authors in [6]. The modified FxNLMS further improves the convergence rate of the conventional FxNLMS by using an additional filter in the form of $h_{ext}^{v}(n)$ in the secondary path. It should be noted that since HRIR $h_{ext}^{v}(n)$ is measured just outside the ear cup, it contains all the spatial information, as well as individual characteristics due to head, torso but minus the pinnae reflections. Therefore, the prior information in the secondary path about the desired signal simplifies the complexity of the adaptive filter and speed up its convergence. The primary path delay must be at least equal to that of the secondary path for adaptive filter to converge [7]. Hence, a forward delay (Δ) is also introduced in the primary path due to the additional filter in secondary path and is usually chosen similarly to the secondary path delay. However, a comparison study of the frequency responses of the estimated secondary path transfer function for the conventional and modified FxNLMS with that of the primary path reveals spectral distortions in high frequencies for the modified FxNLMS, while conventional approach deviates from desired response in low frequencies. As a result, we present a hybrid adaptive equalizer takes benefit of both the FxNLMS approaches to optimize the steady state performance with fast convergence rate. As shown in the Figure 4, $w(n)$ represents the equivalent hybrid adaptive filter, which comprises of two adaptive filters, $w_1(n)$ and $w_2(n)$ corresponding to the conventional and the modified FxNLMS. Further, $w_r(n)$ is used to adaptively estimate the external signal $r_{int}(n)$ using the LMS algorithm based on the reference signal acquired at $r_{ext}(n)$. The relevant equations including weight update equations and optimal solution of the control filters for the hybrid adaptive equalizer is shown in Table 1. As shown, secondary source signal $u(n)$ is computed by summing the two outputs of two adaptive filters. It should be noted that the two adaptive filters are of different length with that of modified FxNLMS is shorter taps due to the fact most of the spatial information about desired signal is there in the secondary path. The augmented superimposed signal captured by the internal microphone is expressed by $y_{int}(n)$ as shown in third row of Table 1. Residual error signal $e'(n)$ is due to both the real and virtual sources presented in the ARE. For virtual source, equalized virtual signal is subtracted from the desired virtual signal, while for real source, real signal received at m_{int} is subtracted from the estimated real signal. For convergence of the hybrid adaptive process, both the error components of the residual signal must be minimized. The optimal solution for equivalent representation of hybrid adaptive filter $w(n)$ is expressed in the z-transform domain. It is found that optimal solution can be expressed as linear combination of the optimal solutions for the conventional and modified FxNLMS adaptive filters such that overall residual error is minimized. Optimal solution for the adaptive estimation filter $w_r(n)$ is found to be equal to the headphone-effect transfer function for corresponding real source direction. This gives us an insight that, if the real source's positions are precisely known then, $\hat{r}_{int}(n)$ can be exactly estimated as $r_{int}(n)$. Since, in practice the exact source position may not be known, alternatively, an average headphone-effect filter can also be used as off-line estimation instead of adaptive estimation. The performance of hybrid adaptive equalizer with and without adaptive estimation is compared in the following section.

Table 1 List of relevant equations for hybrid adaptive equalizer

Description	Equation
Desired signal $d(n)$	$d(n) = h_{int}^{v}(n) * x(n-\Delta)$
Secondary source signal $u(n)$	$u(n) = \mathbf{w_1^T}(n)\mathbf{x}(n) + \mathbf{w_2^T}(n)\mathbf{x_{ext}}(n)$, where, $\mathbf{x}(n) = [x(n) \;\; x(n-1) \ldots\ldots\ldots.. x(n-L_1+1)]^T$. and $\mathbf{x_{ext}}(n) = [x_{ext}(n) \;\; x_{ext}(n-1) \ldots x_{ext}(n-L_2+1)]^T$.
Augmented superimposed signal $y_{int}(n)$	$y_{int}(n) = x_{int}(n) + r_{int}(n)$, where, $x_{int}(n) = h_{hp}(n) * u(n)$
Error signal $e'(n)$	$e'(n) = \{d(n) + \hat{r}_{int}(n)\} - y_{int}(n)$ $= \{d(n) - x_{int}(n)\} + \{-r_{int}(n) + \hat{r}_{int}(n)\}$
Weight update equations	$\mathbf{w_1}(n+1) = \mathbf{w_1}(n) + \mu \frac{\mathbf{x}'(n)}{\lVert \mathbf{x}'(n) \rVert^2} e'(n)$, $\mathbf{w_2}(n+1) = \mathbf{w_2}(n) + \mu \frac{\mathbf{x}'_{\boldsymbol{ext}}(n)}{\lVert \mathbf{x}'_{\boldsymbol{ext}}(n) \rVert^2} e'(n)$. $\mathbf{w_r}(n+1) = \mathbf{w_r}(n) - \mu\, \mathbf{r_{ext}}(n)\, e'(n)$.
Optimal solution for hybrid adaptive filter $w(n)$	$\left. \begin{array}{l} W^o(z) = \alpha\, W_1^o(z) + \beta\, H_{ext}^{v}(z) W_2^o(z), \\ \text{where,} \\ W^o(z) = W_1^o(z) = \dfrac{H_{int}^{v}(z) z^{-\Delta}}{H_{ext}^{v}(z) H_{hp}(z)} \\ \text{and,} \\ W_2^o(z) = \dfrac{H_{int}^{v}(z) z^{-\Delta}}{H_{ext}^{v}(z) H_{hp}(z)} \end{array} \right\} \forall\; \alpha + \beta = 1 ; \quad 0 \le \alpha, \beta \le 1$
Optimal solution for adaptive estimation filter $w_r(n)$	$W_r^o(z) = \frac{R_{int}(z)}{R_{ext}(z)} = \frac{H_{int}(z)}{H_{ext}(z)} = H_{he}(z)$

4. RESULTS

In this section, we present simulation results for the hybrid adaptive equalizer based on HRTF measurements carried out on a dummy head in a semi-anechoic chamber. HRTF were measured for 0 to 180 degree azimuths and subsequently used as target transfer functions in the adaptive equalization of AR headset. A white noise signal is used for the hybrid adaptive equalizer to find the optimum weights of the two adaptive filters for virtual source reproduction. For adaptive estimation of real signal, an



(a) Conventional FxNLMS with no real source present

(b) Hybrid adaptive equalizer with no real source present

(c) Hybrid adaptive equalizer with real source present

Figure 5 Performance of hybrid adaptive equalizer with and without real source present (Top: Contralateral ear; Bottom: Ipsilateral ear)

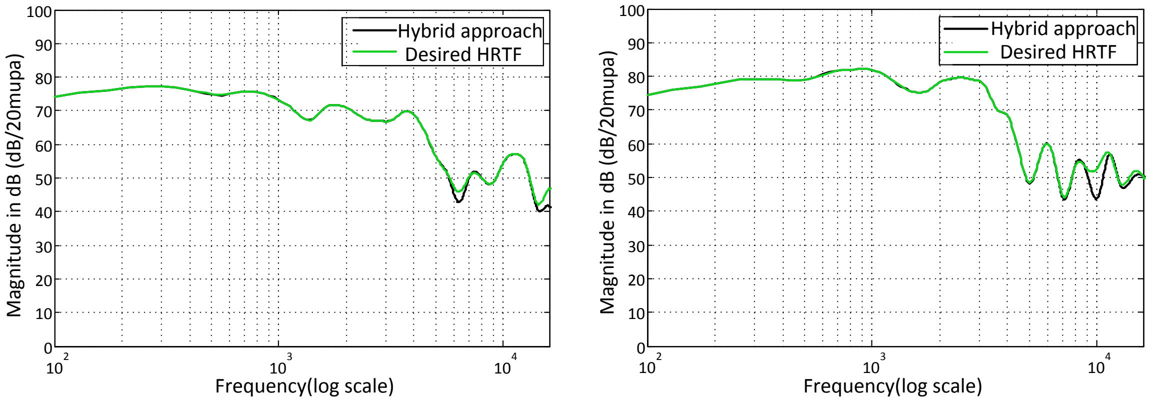
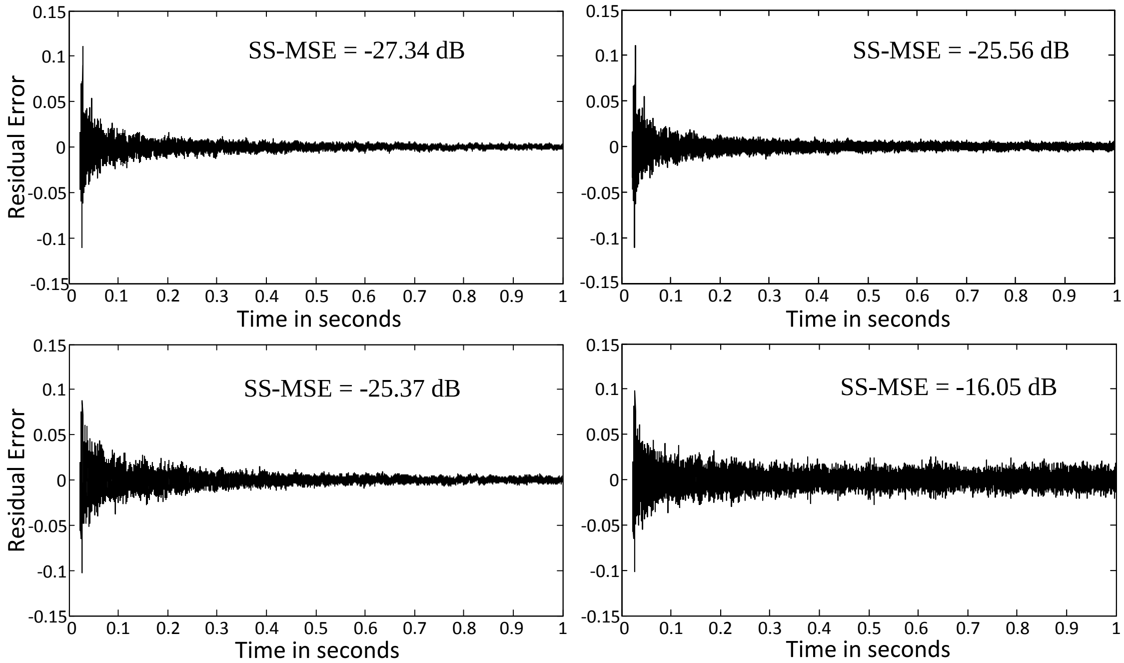


Figure 6 Estimated secondary path response versus desired response of hybrid adaptive equalizer with no real source present (Left: Contralateral ear; Right: Ipsilateral ear)

uncorrelated random noise signal is used in the simulation. A simulation scenario was created with virtual sound coming from 40^o azimuth at 1.4 m from the dummy head and real sound source is emulated by a loudspeaker playback from 0^o azimuth. The two adaptive filters were chosen of length 1024 and 256 taps for the conventional FxNLMS and modified FxNLMS, respectively. Length of the target impulse responses, $h^v_{int}(n)$ and $h^v_{ext}(n)$, which are used in the primary and secondary path respectively, were chosen as 1024 taps. HPTF $h_{hp}(n)$ is modeled as 256 taps filter. Open-back AKG K702 headphone is used to construct the AR headset prototype with four microphones installed (two microphones on each side of ear cup).

Figure 5 compares the performance of adaptive equalizer in the absence of any real source and in the presence of real source without any estimation of real signals. The proposed hybrid approach performs much better than the conventional FxNLMS with faster convergence rate as well as optimum steady state error reduction as shown in Figure 5 (a) and Figure 5 (b). It is also observed that due to the presence of real source with no estimation applied, the hybrid approach results in large steady state errors as compared to when there is no real source present. Hybrid adaptive equalizer in the absence of real signals results in 29 dB and 26 dB attenuation of residual signal respectively, for contralateral and ipsilateral ears as against 18 dB and 11 dB in the presence of real signals as shown in Figure 5 (b) and Figure (c). This implies that the effect of real signals must be removed from the convergence process to avoid large steady state errors. Figure 6 shows the frequency response of estimated secondary path transfer functions compared with that of the desired primary path response. It is evident that the hybrid approach closely matched with the desired response, which is also corroborated by the high steady-state error reductions in Figure 5. Figure 7 shows the performance of the hybrid adaptive



(a) Hybrid adaptive equalizer with adaptive estimation

(b) Hybrid adaptive equalizer with off-line estimation

Figure 7 Performance of hybrid adaptive equalizer with adaptive estimation and off-line estimation (Top: Contralateral ear; Bottom: Ipsilateral ear)

equalizer using online adaptive estimation, as well as off-line estimation using an average $h_{he}(n)$ filter as described in the previous section. Both approaches in Figure 7 use the estimation of real signals to remove their effects results in better steady state performance when compared to Figure 5 (c), where no estimation is performed. Furthermore, the proposed approach using online adaptive estimation performs significantly better than the one when using average filter for offline estimation, especially for the ipsilateral ear, as shown in Figure 7 (a) and (b). This might be due to the fact that there is large variation in headphone-effect transfer function with head movements and the average filter is not able to compute exact estimate of the real signals. To summarize, hybrid adaptive equalizer works equally well even in the presence of real sound sources and the reproduced virtual sources are perceived as close to the real sound source[6].

5. CONCLUSIONS

In this paper, we presented ANC techniques based on the FxNLMS algorithm to enable natural listening in an augmented reality environment via AR headset. The AR headset is constructed from open-air headphones with two microphones (one internal and one external) attached on each side. Open-air headphones were chosen over closed-back headphones as they allow external sound from physical sources reach listeners' ears without much attenuation. For virtual source reproduction, hybrid adaptive equalizer is presented, such that the AR headset is equalized to an individual with the help of personalized HRTFs measured in listener environment. The proposed hybrid approach is able to attenuate the residual error signal by more than 25 dB, thus implying that virtual source is reproduced alike real source. Even in the presence of real signals, proposed approach works equally well using an online adaptive estimation of real signals, such that they do not affect the convergence process of hybrid adaptive filter. Informal listening tests validated that virtual sounds could not be differentiated from the real sounds when presented over AR headset.

There are various application scenarios for augmented reality via headset. For example, (i) in binaural telecommunication, where any person located at far end can be made virtually immersed around a user wearing AR headset, giving a feeling of being there at any time and at any place; (ii) augmented reality gaming in a real environment with spatial audio cues; (iii) assistive listening devices that are worn to communicate with others in a physical setting, with augmented cues alerting any impending danger; (iv) audio-visual aids for visually impaired, where spatial audio cues are guiding the users to navigate around obstacles; (v) audio-sticker application to leave audio message attached to a particular space or objects; and (vi) audio guides in museum or tourist areas to provide recorded message tagging to art pieces or sign-posts.

REFERENCES

[1] Nishino T, Inoue N, Takeda K, Itakura F. Estimation of HRTFs on the horizontal plane using physical features. Applied Acoustics. 2007;68(8):897-908.

[2] Qu T, Xiao Z, Gong M, Huang Y, Li X, Wu X, editors. Distance dependent head-related transfer function database of KEMAR. Audio, Language and Image Processing, 2008 ICALIP 2008 International Conference on; 2008: IEEE.

[3] Hebrank J, Wright D. Spectral cues used in the localization of sound sources on the median plane. The Journal of the Acoustical Society of America. 2005;56(6):1829-34.

[4] Härmä A, Jakka J, Tikander M, Karjalainen M, Lokki T, Hiipakka J, et al. Augmented reality audio for mobile and wearable appliances. Journal of the Audio Engineering Society. 2004;52(6):618-39.

[5] Brinkmann F, Weinzierl S. Individual headphone compensation for binaural synthesis. 2011.

[6] Ranjan R, Gan W-S. Natural Listening in Augmented Reality using Active Noise Control Techniques. IEEE Journal of Seletected Topics in Signal Processing.[Submitted].

[7] Kuo SM, Morgan D. Active noise control systems: algorithms and DSP implementations: John Wiley & Sons, Inc.; 1995.