ASSISTIVE LISTENING HEADSETS FOR HIGH NOISE ENVIRONMENTS: PROTECTION AND COMMUNICATION

Sven Nordholm*, Alan Davis$^\diamond$, Pei Chee Yong*, Hai Huyen Dam†

*Curtin University, Department of Electrical and Computer Engineering, Bentley WA 6102 Australia
$^\diamond$Sensear Pty Ltd, 197-199 Great Eastern Highway, Belmont WA 6104 Australia
†Curtin University, Department of Mathematics and Statistics, Bentley WA 6102 Australia
{S.Nordholm, P.Yong, H.Dam}@curtin.edu.au, $^\diamond$alan.davis@sensear.com

ABSTRACT

In industrial noise environments, the use of assistive listening headsets is a means to provide adequate access to voice communication while wearing hearing protection. This paper presents a performance evaluation and comparison of two different methods to provide the binaural speech enhancement in real industrial noise scenarios. The investigated binaural methods based on differential beamforming and multichannel Wiener filter show different strengths and weaknesses. A transient noise suppression algorithm is also proposed and evaluated. Performance evaluation shows that this algorithm, together with the binaural multi-channel Wiener filter approach, can successfully reduce the hammering noise. This can be observed from the PESQ scores and the signal characteristics.

Index Terms— Assistive listening, speech enhancement, differential microphone array, binaural multi-channel Wiener filter, impulsive noise reduction

1. INTRODUCTION

In industrial environments such as mine sites and oil-and-gas industry, workers are exposed to noise for approximately 8 hours daily with sound pressure levels exceeding 100 dBA [1]. In such environments, speech communication would cease, or could only be performed by shouting. Apart from that, long exposure to high noise levels on a daily basis would not only cause permanent damage to the hearing, but could also cause serious injury or even death as safety could be compromised. Therefore, it is crucial to develop an assistive listening headset that can simultaneously protect the hearing of the users while providing spatial awareness, and enable intelligible speech communication capability in high noise environments.

The main considerations when designing an assistive listening headset are; the ability to suppress noise without generating too much speech distortion, and to preserve the binaural cues of all sound sources from the surrounding. In order to achieve those criteria, two main classes of binaural noise reduction techniques have been investigated. In the first class, the multi-microphone signals are used to calculate a real-valued spectral gain, namely the binaural post-filter (BPF), where such identical gain is applied to the reference microphones corresponded to both sides of the device [2, 3, 4, 5]. These concepts are similar to single-channel speech enhancement [6] as it allows the binaural cues of both speech and noise components to be perfectly preserved. The output signals

will potentially generate higher speech distortion and noise artefacts known as the musical noise, leading to limited speech intelligibility improvements. The second class of binaural techniques perform true array processing, which combines spatial and spectral filtering, by applying a complex-valued filter to all microphone signals on both sides of the device [7, 8, 9, 10, 11, 12]. These works are employing extended versions of single-side multi-channel speech enhancement algorithms [13, 14, 15]. These techniques ensures better noise reduction performance, with a drawback that the binaural cues of the noise component may not be preserved.

The scope of this paper is twofold: First, the performance evaluation of two distinct binaural speech enhancement algorithms, and second, a transient interference suppression algorithm is proposed. The first framework, namely the DMA-BPF approach, applies identical Differential Microphone Array (DMA) at both sides of the device and then a single-channel algorithm. The second technique is a modified binaural MWF approach which is able to select the sources that are present in the environment dynamically to improve binaural cues preservation. A comparison in terms of speech quality improvement and binaural cues preservation is conducted. Besides that, the proposed transient noise suppression method aims at preserving the main waveform characteristics of the impulse noise while reducing the output to a safe level, without using training.

The paper is organized as follows. Section 2 presents the algorithms of the two binaural speech enhancement frameworks. Section 3 develops the proposed transient noise suppression method. Section 4 presents the performance measurement and results, and Section 5 concludes the paper. Section 6 shows the relation between the contribution in this paper and prior works in the field.

2. BINAURAL SPEECH ENHANCEMENT FRAMEWORK

2.1. Signal model and notation

Consider a configuration of $2L$ microphones, with $L = 2$ of them mounted on each side of the device. The l-th microphone signal on the left and the right sides are defined in short-time Fourier transform (STFT) domain as

$$
\begin{aligned}
Y_{\text{left},l}(k,m) &= X_{\text{left},l}(k,m) + V_{\text{left},l}(k,m) \\
Y_{\text{right},l}(k,m) &= X_{\text{right},l}(k,m) + V_{\text{right},l}(k,m) \quad l = 1,2
\end{aligned} \tag{1}
$$

where $X_{\text{left},l}(k,m)$ and $X_{\text{right},l}(k,m)$ represent the speech components in the microphone signal, while $V_{\text{left},l}(k,m)$ and $V_{\text{right},l}(k,m)$ represent the noise components. Here, k and m denote the frequency bin index and time-frame index, respectively. They will be omitted

in the following sections for notational convenience. Moreover, instead of listing all equations that are the same for both sides of the device, only those at the left side are shown whenever possible due to space constraint.

The L-dimensional stacked microphone signal vectors $\mathbf{y}_{\text{left}}$ and $\mathbf{y}_{\text{right}}$, and the $2L$-dimensional signal vector $\mathbf{y}$ are given as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_{\text{left}} & \mathbf{y}_{\text{right}} \end{bmatrix}^T \tag{2}$$

with

$$\begin{aligned} \mathbf{y}_{\text{left}} &= \begin{bmatrix} Y_{\text{left},1} & Y_{\text{left},2} & ... & Y_{\text{left},L} \end{bmatrix}^T, \\ \mathbf{y}_{\text{right}} &= \begin{bmatrix} Y_{\text{right},1} & Y_{\text{right},2} & ... & Y_{\text{right},L} \end{bmatrix}^T \end{aligned} \tag{3}$$

where T denotes the transpose operator. The correlation matrix of speech plus noise $\mathbf{R}_y$, the clean speech correlation matrix $\mathbf{R}_x$, and the noise correlation matrix $\mathbf{R}_v$ are defined as

$$\mathbf{R}_y = E\{\mathbf{y}\mathbf{y}^H\}, \ \mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}, \ \mathbf{R}_v = E\{\mathbf{v}\mathbf{v}^H\}, \tag{4}$$

where $E\{\cdot\}$ denotes the expectation operator, and H denotes the conjugate transpose operator. The $2L$-dimensional signal vectors $\mathbf{x}$ and $\mathbf{v}$ are defined similarly as $\mathbf{y}$. The speech and the noise components are assumed uncorrelated, such that $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$. The front microphones of both sides will be used as the so-called reference signals. For conciseness, the reference microphone signal at the left side is defined as

$$Y_{\text{left}} = \mathbf{e}_{\text{left}}^H \mathbf{y} = X_{\text{left}} + V_{\text{left}} \tag{5}$$

where X_{left} and V_{left} are speech and noise components, respectively. Here, $\mathbf{e}_{\text{left}}$ is a $2L$-dimensional vectors with only one element equal to 1 and the other elements equal to 0.

2.2. Differential Microphone Array Binaural Post Filter Framework

The first presented framework is an extension of the BPF technique by adding in the DMAs. The DMAs serve as null-steering beamformers to suppress signals originating from a certain direction where the undesired source is located [16]. In this paper, this is done by utilising the first-order DMAs which incorporate Head-Related Transfer Functions (HRTFs). For simplicity, only the case of a single sound source located at ϕ_s is considered. Since the distance between the microphones of a DMA is inherently small, it can be assumed that the corresponding HRTFs have an identical magnitude and only differ in phase. Hence, they can be related to each other at the left side as

$$H_{\text{left},2}(e^{j\Omega}, \phi_s) = H_{\text{left},1}(e^{j\Omega}, \phi_s)e^{-j\Omega f_s \frac{d}{c} \cos \phi_s} \tag{6}$$

where Ω is the angular frequency, f_s is the sampling frequency, d is the distance between microphones, ϕ_s is the angle of the sound source and c is the speed of sound. The output signal of the DMA at the left side is then given by

$$\begin{aligned} &Y_{\text{DMA,left}}(e^{j\Omega}, \phi_s) \\ &= S(e^{j\Omega})H_{\text{left},1}(e^{j\Omega}, \phi_s) - S(e^{j\Omega})H_{\text{left},2}(e^{j\Omega}, \phi_s)e^{-j\Omega f_s \tau} \\ &= S(e^{j\Omega})H_{\text{left},1}(e^{j\Omega}, \phi_s)\left(1 - e^{-j\Omega f_s\left(\tau + \frac{d}{c}\cos \phi_s\right)}\right) \end{aligned} \tag{7}$$

where τ is a delay for processing. If $\tau = d/c$, the DMA will have a null at $\phi_s = \pi$, while for $\phi_s = 0$, it is a high pass filter. Now, for

each DMA output signal (left and right), the respective noise components are estimated and used to design two independent single-channel noise suppression filters. Both of the filters are merged to a single binaural gain function applied to both channels, as given by

$$\mathcal{G} = \frac{1}{2}\left(G_{\text{left}} + G_{\text{right}}\right) \tag{8}$$

where G_{left} and G_{right} are gain functions obtained from the corresponding sides. The role of $\mathcal{G}$ is to pick up all acoustic source signals from the environment and maintain those with speech while suppressing the background noise. Finally, the output signal at the left side is obtained by

$$Z_{\text{DMA}-\text{BPF,left}} = \mathcal{G}Y_{\text{DMA,left}}. \tag{9}$$

2.3. Binaural MWF approach

The MWF method has been widely used for binaural speech enhancement given that it produces a minimum mean square error (MMSE) estimate of the speech component in the reference microphone at respective sides, simultaneously reducing noise and limiting speech distortion [9, 12]. In [17], a modified binaural MWF approach has been presented, which has a weighted average cost function where the first term is weighted by the probability that speech is present, while the second term is weighted by the probability that speech is absent. It is given as

$$\begin{aligned} \mathcal{J}_{\text{MWF}_\lambda-\text{SPP}}(\mathbf{w}) = \ &p \times E\left\{ \left\| \begin{bmatrix} X_{\text{left}} - \mathbf{w}_{\text{left}}^H\mathbf{y} \\ X_{\text{right}} - \mathbf{w}_{\text{right}}^H\mathbf{y} \end{bmatrix} \right\|^2 |\mathcal{H}_1 \right\} \\ &+ (1-p) \times E\left\{ \left\| \begin{bmatrix} \mathbf{w}_{\text{left}}^H\mathbf{y} \\ \mathbf{w}_{\text{right}}^H\mathbf{y} \end{bmatrix} \right\|^2 |\mathcal{H}_0 \right\} \end{aligned} \tag{10}$$

where $\mathcal{H}_1$ and $\mathcal{H}_0$ denote the probability of speech presence and absence, respectively at every frequency bin and frame. The probability p is given by

$$p = \frac{p_{\text{left},r_{\text{left}}} + p_{\text{right},r_{\text{right}}}}{2} \tag{11}$$

with $p_{\text{left},r_{\text{left}}}$ and $p_{\text{right},r_{\text{right}}}$ denote the conditional probability that speech is present at the corresponding reference channels respectively at the left and the right, while $1-p$ is the combined conditional probability that speech is absent. The solution of Eq. (10) for the left side of the device is then given by

$$\mathbf{w}_{\text{MWF}_\lambda-\text{SPP,left}} = (p\mathbf{R}_y + (1-p)\mathbf{R}_v)^{-1} p\mathbf{r}_{yx,\text{left}}. \tag{12}$$

Both speech plus noise correlation matrix and noise correlation matrix can be updated continuously by employing the conditional speech presence probability

$$\begin{aligned} \hat{\mathbf{R}}_v[m] &= (1 - \alpha_{vv}(1 - p[m]))\,\hat{\mathbf{R}}_v[m-1] \\ &\quad + \alpha_{vv}(1-p[m])\,\mathbf{y}[m]\mathbf{y}^H[m] \end{aligned} \tag{13}$$

$$\hat{\mathbf{R}}_y[m] = (1 - \alpha_{yy}p[m])\,\hat{\mathbf{R}}_y[m-1] + \alpha_{yy}p[m]\mathbf{y}[m]\mathbf{y}^H[m] \tag{14}$$

Here, the values of both smoothing factors α_{vv} and α_{yy} are chosen carefully to reflect the degree of stationarity of speech and noise.

The estimate of the cross-correlation vector at the left side is defined as

$$\hat{\mathbf{r}}_{yx,\text{left}}[m] = (1 - \alpha_x)\hat{\mathbf{r}}_{yx}[m-1] + \alpha_x\mathbf{y}[m]\mathcal{G}[m]Y_{\text{left}}^*[m] \tag{15}$$

where $\mathcal{G}$ is the single-channel gain function defined similarly as Eq. (8), and α_x is the smoothing factor. The output signals are obtained by filtering and summing all microphone signals

$$Z_{\mathrm{MWF}_\lambda,\mathrm{left}} = \mathbf{w}_{\mathrm{left}}^H \mathbf{y} \qquad (16)$$

where $\mathbf{w}_{\mathrm{left}}$ and $\mathbf{w}_{\mathrm{right}}$ are $2L$-dimensional complex weight vectors.

3. TRANSIENT NOISE SUPPRESSION

Apart from the stationary noise generated from machinery, there are always transient interferences in industrial environments such as hammering noise. A transient is a high amplitude, short-duration sound featured by a wide spread over the frequency domain. Thus, traditional speech enhancement algorithms that assume pseudo-stationary noise is not applicable for transient interferences due to their fast changing nature. Here, a method to tackle the problem is proposed by first applying a long term smoothing to the magnitude of the output signal

$$\psi_{\mathrm{left}}(k,m) = \alpha_\psi \psi_{\mathrm{left}}(k,m-1) + (1-\alpha_\psi)|Z_{\mathrm{left}}(k,m)| \quad (17)$$

where α_ψ denotes the smoothing constant. The impulse signal suppression gain at the left side $F_{1,\mathrm{left}}$ is computed by finding the average over all frequency bins of the ratio between the smoothed signal and its weighting with the instantaneous output signal

$$F_{1,\mathrm{left}}(m) = \frac{1}{K} \sum_{k=1}^{K} \frac{\psi_{\mathrm{left}}(k,m)}{\beta_1 |Z_{\mathrm{left}}(k,m)| + \psi_{\mathrm{left}}(k,m)} \qquad (18)$$

where K is the number of bands, and β_1 denotes a parameter to adjust the amplitude level of the output signal. If a strong transient noise exists in a time-frame, $F_{1,\mathrm{left}}$ will take a very small value. The proposed algorithm to reduce the transient noise to a safe level is described in the following pseudo-code

For each m-th frame, k-th frequency bin, calculate $F_{1,\mathrm{left}}(m)$
    If $F_{1,\mathrm{left}}(m) < \kappa$ **then**
        $F_{2,\mathrm{left}}(m) = \frac{1}{K}\sum_{k=1}^{K} \frac{\psi_{\mathrm{left}}(k,m)}{\beta_2|Z_{\mathrm{left}}(k,m)| + \psi_{\mathrm{left}}(k,m)}$
        $\delta_c = 0$
    Otherwise
        $F_{1,\mathrm{left}}(m) = 1$
    End if
    If $\delta_c < \delta_h$ **then**
        $Z_{\mathrm{left}}(k,m) = F_2(m) \cdot Z_{\mathrm{left}}(k,m)$
        $\delta_c = \delta_c + 1$
    Otherwise
        $Z_{\mathrm{left}}(k,m) = F_1(m) \cdot Z_{\mathrm{left}}(k,m)$
    End if

where $F_1(m)$ and $F_2(m)$ are the average of $F_{1,\mathrm{left}}(m)$, $F_{1,\mathrm{right}}(m)$, and $F_{2,\mathrm{left}}(m)$, $F_{2,\mathrm{right}}(m)$, respectively. The parameter $\beta_2 > \beta_1$ is a fixed number to control the level of impulse noise suppression, κ is a fixed value, δ_c and δ_h are parameters for sample counter and hangover time to control and avoid stagnation of the updates.

4. PERFORMANCE EVALUATION

4.1. Performance measures

Performance evaluation for speech quality includes the comparison of the binaural cues and the noise reduction performance. The noise reduction performance is measured by the intelligibility frequency weighted segmental SNR (IFWSNRseg) measure [18, 19]

$$\mathrm{IFWSNRseg} = \frac{10}{M} \sum_{m=0}^{M-1} \frac{\sum_{k=0}^{K-1} B_k \log_{10} \frac{\mathcal{A}^2(k,m)}{|\mathcal{A}^2(k,m) - \hat{\mathcal{A}}^2(k,m)|}}{\sum_{k=0}^{K-1} B_k} \tag{19}$$

where B_k is the ANSI SII weight placed on the kth frequency bin [20], M is the number of frames, $\mathcal{A}(k,m)$ and $\hat{\mathcal{A}}(k,m)$ are spectrum amplitudes of the clean speech signal and enhanced speech signal, respectively. Each frame has a threshold of $-10\,\mathrm{dB}$ lower bound and a 35 dB upper bound to discard non-speech frames. The segmental noise attenuation (NATTseg) and the segmental speech preservation (SPREseg) measures are utilized to study if a difference in IFWSNRseg is due to more noise reduction or less speech distortion. Both are given, respectively, by [21]

$$\mathrm{NATTseg} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log_{10} \frac{||\mathbf{v}_t(m)||^2}{||\tilde{\mathbf{v}}_t(m)||^2}, \qquad (20)$$

$$\mathrm{SPREseg} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log_{10} \frac{||\mathbf{x}_t(m)||^2}{||\mathbf{x}_t(m) - \tilde{\mathbf{x}}_t(m)||^2}. \qquad (21)$$

Here, $\mathbf{v}_t(m)$ and $\mathbf{x}_t(m)$ are m-th frame time-domain vectors for the noise and the clean speech signal, respectively. The signals $\tilde{\mathbf{v}}_t(m)$ and $\tilde{\mathbf{x}}_t(m)$ indicate both noise and clean signals processed with the same corresponding filters as used to enhance the noisy signal. The widely-used and the perceptual evaluation of speech quality (PESQ) measure has also been included for performance comparison [19]. For all measurements, results of the reference channels from the left and right were averaged to obtain a single value.

The binaural cues were evaluated using the ITD and the ILD measures. The ITDs here are computed by interpolating the envelope of the cross-correlation, and the absolute ITD errors are then given by [17, 22]

$$\begin{aligned} \Delta\mathrm{ITD}_x &= |\mathrm{ITD}_x^{\mathrm{in}} - \mathrm{ITD}_x^{\mathrm{out}}| \\ \Delta\mathrm{ITD}_v &= |\mathrm{ITD}_v^{\mathrm{in}} - \mathrm{ITD}_v^{\mathrm{out}}|. \end{aligned} \qquad (22)$$

The ILDs are obtained by evaluating the logarithm of the power ratio between the respective signals at the left and right sides. The ILDs of the input speech and noise are given as

$$\begin{aligned} \Delta\mathrm{ILD}_x^{\mathrm{in}} &= \frac{1}{K} \sum_{k=1}^{K} 10 \log_{10} \frac{\hat{P}_{x_{\mathrm{left}}}(k)}{\hat{P}_{x_{\mathrm{right}}}(k)} \\ \Delta\mathrm{ILD}_v^{\mathrm{in}} &= \frac{1}{K} \sum_{k=1}^{K} 10 \log_{10} \frac{\hat{P}_{v_{\mathrm{left}}}(k)}{\hat{P}_{v_{\mathrm{right}}}(k)} \end{aligned} \qquad (23)$$

where $\hat{P}_{(\cdot)}$ is the power spectrum estimate of the corresponding signals obtained using Welch method. The ILDs of the processed speech and noise are obtained in a similar way as Eq. (23).

4.2. Experimental setup and results

Measurements are performed with $L = 2$ microphones (with inter-element space of 1 cm) embedded at each side of a pair of earmuffs on a KEMAR manikin so that the head-shadowing effect is included. The head and torso simulator is placed close to the center of a room with dimensions $3.05\,\mathrm{m} \times 3.05\,\mathrm{m}$, with a reverberation time T_{60} of approximately $0.2\,\mathrm{s}$. The loudspeakers are positioned at 1 m from the center of the head, with the speech source located at $0°$ and the noise rendered at $45°$, $90°$, $135°$, $180°$, $225°$, $270°$ and $315°$ to the left of the head. The speech signals consists of 5 (2 male and 3 female) sentences with length ranging from 11 s to 22 s. The signals
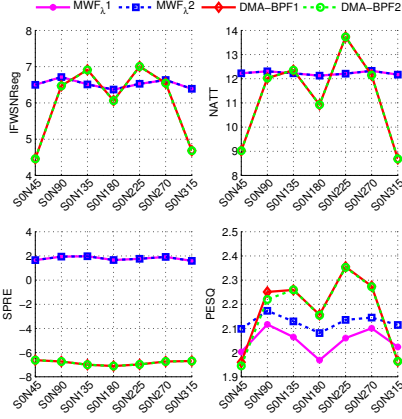
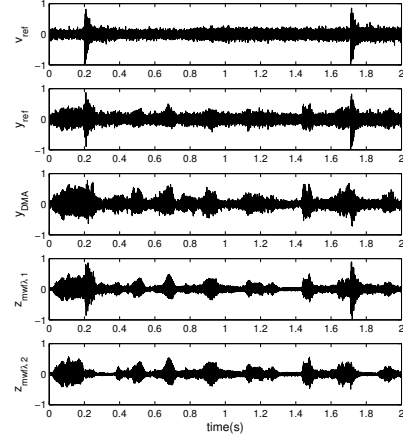**Fig. 1**. Noise reduction performance for input SNR 0 dB

Fig. 2. ITD and ILD results for input SNR 0 dB

Fig. 3. Signals Comparison at input SNR 0 dB
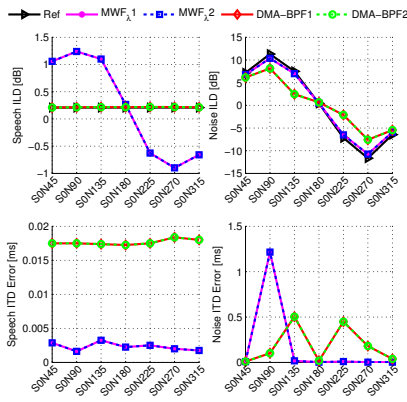
are sampled at $f_s = 16\,\text{kHz}$. An STFT length of $K = 512$ is used with a frame rate $R = 256$ and square-root Hann windowing. The noise is a cement factory noise mixed with hammering noise.

Performance comparison are conducted between the DMA-BPF approach and the binaural MWF$_\lambda$ method. For those of the presented methods where the single-channel speech enhancement technique is required, for G in Eqs. (8) and (15) the modified sigmoid (MSIG) gain function from [23] has been used. The parameters of the gain function are the same as shown in [23], except the parameters for the *a priori* SNR estimate, which are $\alpha_y = 0.17$, $\beta = 0.98$ for DMA-BPF, and $\alpha_y = 0$, $\beta = 0.9$ for MWF$_\lambda$. For the noise and the conditional SPP estimates, the algorithm in [24] is used. The values of smoothing factors for MWF$_\lambda$ are $\alpha_{yy} = 0.17$, $\alpha_{vv} = 0.98$ for Eqs. (14) and (13), respectively, and $\alpha_x = 0.17$ for Eq. (15). For the transient noise suppression algorithms in Sec. 3, the assigned values for the parameters are $\beta_1 = 0.2$, $\beta_2 = 1$, $\kappa = 0.6$, $\delta_c = \delta_h = 10$. Note that for the results, MWF$_\lambda$1 and DMA-BPF1 represent the method without applying the impulse noise suppressor in Section 3, while MWF$_\lambda$2 and DMA-BPF2 are with the ones with the impulse noise suppression method.

Figs. 1 and 2 depict the results of noise reduction performance and binaural cues preservation, respectively at 0 dB input SNR. As observed in Fig. 1, MWF$_\lambda$ performs similar or better than DMA-BPF in terms of the amount of noise reduction, except for the configurations where the noise is coming from $\pm135°$. The SPREseg

results show that MWF$_\lambda$ can preserve more speech components, in other words less speech distortion compared to DMA-BPF. The DMA-BPF approach has however better PESQ performance when compared with MWF$_\lambda$. In terms of binaural cues preservation, the speech ILDs of MWF$_\lambda$ differ slightly from the reference channels', while the noise ILDs of DMA-BPF have larger distortion compared to MWF$_\lambda$ due to the nature of DMAs. As for ITD measurement, both methods have recorded small delays in noise ITDs at different noise configurations.

The transient noise suppression method does not change the results of the performance measures, except for PESQ scores. The PESQ results show improvement in speech quality for $Z_{\text{MWF}\lambda2}$ when compared to $Z_{\text{MWF}\lambda1}$. As for DMA-BPF1 and DMA-BPF2, not much difference can be seen as they actually suppress the unwanted transient noise already. These results can be confirmed from the signals as shown in Fig. 3. The hammering noise that exists in the output signal $Z_{\text{MWF}\lambda1}$ has been reduced in $Z_{\text{MWF}\lambda2}$.

5. CONCLUSIONS

A performance evaluation has been made with respect to binaural cues preservation and noise reduction performance for the DMA-BPF and the modified binaural MWF algorithms, both use an improved single-channel speech enhancement algorithm and a transient noise suppression post-filter. The results show that for evaluated noise conditions, the binaural MWF method has the advantage of generating lower speech distortion with high noise reduction, while the binaural DMA-BPF method has better PESQ scores. The transient suppression method also further improves the output signal of the binaural MWF method by reducing the impulse noise to a level similar to the speech signal.

6. RELATION TO PRIOR WORK

In this paper, two multichannel binaural speech enhancement methods are introduced, evaluated and compared in terms of binaural cues preservation and noise reduction performance. The presented frameworks are based on the extensions, respectively, of (i) the DMA-BPF in [25] with improved noise estimation and speech enhancement, and (ii) the modified binaural MWF algorithm [17]. In addition, this paper also includes a novel transient suppression algorithm as a post filter of both frameworks to deal with the real-world noise scenarios.

7. REFERENCES

[1] D. I. McBride, "Noise-induced hearing loss and hearing conservation in mining," *Occupational Medicine*, vol. 54, no. 5, pp. 290–296, 2004.

[2] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Trans. on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1141–1154, 2011.

[3] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Process.*, vol. 2006, pp. 175–175, 2006.

[4] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. 18th European Signal Process. Conference (EUSIPCO'10)*, Aalborg, Denmark, Feb. 2010, pp. 988–993.

[5] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Communication*, vol. 39, no. 1, pp. 111–138, 2003.

[6] P. C. Yong, S. Nordholm, H. H. Dam, Y. H. Leung, and C. C. Lai, "Incorporating multi-channel Wiener filter with single-channel speech enhancement algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'13)*, Vancouver, Canada, May 2013.

[7] R. Aichner, H. Buchner, M. Zourub, and W. Kellermann, "Multi-channel source separation preserving spatial information," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'07)*, Honolulu, USA, Apr. 2007, vol. 1, pp. I–5.

[8] R. Nishimura, Y. Suzuki, and F. Asano, "A new adaptive binaural microphone array system using a weighted least squares algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'02)*, Orlando, USA, May 2002, vol. 2, pp. II–1925.

[9] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7, pp. 636–656, 2007.

[10] A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids," *IEEE Trans. on Signal Process.*, vol. 53, no. 3, pp. 911–925, Mar. 2005.

[11] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. on Signal Process.*, vol. 55, no. 4, pp. 1579–1585, 2007.

[12] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Proc. Int. Workshop Acoust. Echo and Noise Control (IWAENC'06)*, Paris, France, Sep. 2006.

[13] S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: an analytical evaluation," *IEEE Trans. on Speech and Audio Process.*, vol. 7, no. 3, pp. 241–252, May 1999.

[14] S. Y. Low, N. Grbic, and S. Nordholm, "Speech enhancement using multiple soft constrained subband beamformers and non-coherent technique," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'03)*, Hong Kong, Apr. 2003, vol. 5, pp. V–489.

[15] H. H. Dam, H. Q. Dam, and S. Nordholm, "Noise statistics update adaptive beamformer with psd estimation for speech extraction in noisy environment," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 16, no. 8, pp. 1633–1641, Nov. 2008.

[16] M. Buck, "Aspects of first-order differential microphone arrays in the presence of sensor imperfections," *European Transactions on Telecommunications*, vol. 13, no. 2, pp. 115–122, 2002.

[17] P. C. Yong, S. Nordholm, and H. H. Dam, "Effective binaural multi-channel processing algorithm for improved environmental presence," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 22, no. 12, pp. 2012–2024, Dec. 2014.

[18] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance.," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[19] P. Loizou, *Speech Enhancement Theory and Practice*, CRC Press, Boca Raton, FL, 2007.

[20] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," Jun. 1997.

[21] R. C. Hendriks and R. Martin, "MAP estimators for speech enhancement under normal and rayleigh inverse gaussian distributions," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 918–927, Mar. 2007.

[22] T. J. Klasen, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural multi-channel wiener filtering for hearing aids: preserving interaural time and level differences," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP'06)*, Toulouse, France, May 2006, vol. 5, pp. V145–V148.

[23] P. C. Yong, S. Nordholm, and H. H. Dam, "Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement," *Speech Communication*, vol. 55, no. 2, pp. 358–376, Feb. 2013.

[24] P. C. Yong, S. Nordholm, and H. H. Dam, "Noise estimation based on soft decisions and conditional smoothing for speech enhancement," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC'12)*, Aachen, Germany, Sep. 2012, pp. 4640–4643.

[25] M. Bürger, "Binaural noise suppression techniques for assistive listening under extreme industrial noise conditions," M.S. thesis, University of Erlangen-Nuremberg, 2012.