Perceptual Active Equalization of Multi-frequency Noise

Juan Estreder, Gema Piñero[a], Miguel Ferrer[b], Maria de Diego[c] and Alberto Gonzalez[d]
Inst. Telecommunication and Multimedia Applications (iTEAM),
Universitat Politècnica de València, Spain

Keywords: Active Noise Equalization, Frequency Masking, Perceptual Equalization, Multi-channel Adaptive Filtering.

Abstract: In this paper we propose a novel multi-channel active noise equalizer (ANE) when music or speech signals are present inside the same room. Our perceptual ANE (PANE) can benefit from the masking effect of the music emitted carrying out a perceptual equalization (PEQ) of the undesired ambient noise. Our PEQ strategy automatically adapts the spectral profile of the ambient noise recorded at the error microphones to the masking threshold of the audio signal recorded at that same point. We present a real-time experiment carried out in our laboratory that simulates the position of a driver in a car to test the PANE with different audio signals. The experimental results are compared with two alternative strategies: the full cancellation (FC) profile that corresponds to an active noise cancellation strategy, and the hearing threshold (HT) profile that corresponds to an ANE system whose gains mimic the human audibility threshold. Both FC and HT profiles are independent of the music presented in the room. Results show that the noise power measured at the microphones is higher for the PEQ profile, but always below the masking threshold of the music, getting almost unnoticeable. However, the emitted anti-noise power in the case of PEQ is 15 dB lower compared to HT and FC profiles for frequencies above 300 Hz. This performance leads to a reduction of noise pollution in the room and a lower power consumption of the system loudspeakers. In addition, the PEQ profile provided by the novel PANE system is a versatile approach that can reduce the perceived noise as much as the user decides, even reaching the same performance than the HT or FC profiles if needed. Therefore, the PANE system is a versatile real-time alternative to the classic active noise cancellation systems for multi-frequency noise.

1 INTRODUCTION

The way we perceive sounds has been studied from a systematic point of view over the last few decades giving rise to the field of psychoacoustics (Fastl and Zwicker, 2007; Pickles, 2012). This approach to the human hearing system models the process that a sound undergoes from the moment it enters our outer ear until it reaches our neurological system and beyond. This knowledge has been widely used in several signal processing areas, such as coding (Brandenburg and Johnston, 1990), equalization (Välimäki and Reiss, 2016), personal sound zones (Donley et al., 2016), or noise cancellation (Wang et al., 2018; Mosquera-Sánchez et al., 2018). Generally speaking, the field of pyschoacoustics provides a more realistic approach to any sound or noise processing since it takes into account how our hearing system perceives sounds (Fletcher, 1940).

From this perceptual point of view, we revisit in this paper the active noise equalizer that was first proposed by Kuo (Kuo and Tsai, 1994). This very first idea of an active noise control system that could shape the spectrum of the residual noise was further developed for multi-frequency noise and named "adaptive noise equalizer" (Kuo and Ji, 1995). Afterwards it was extended for a multi-channel active noise equalizer (ANE) in (Gonzalez et al., 2006), where the equalization parameters that controlled each single frequency were arbitrarily selected. In the present paper we revisit the algorithm proposed in (Gonzalez et al., 2006) and go a step further introducing a perceptual strategy to control the ANE. This strategy benefits from the masking effect that a sound (music or speech) produces over the ambient noise, as for example it happens within a car where its entertainment system can emit music and can also be used to actively control the car noise.

[a] https://orcid.org/0000-0002-8719-8106
[b] https://orcid.org/0000-0002-8743-1887
[c] https://orcid.org/0000-0001-9948-3396
[d] https://orcid.org/0000-0002-6984-3212

Estreder, J., Piñero, G., Ferrer, M., de Diego, M. and Gonzalez, A.
Perceptual Active Equalization of Multi-frequency Noise.
DOI: 10.5220/0010648100390047
In *Proceedings of the 18th International Conference on Signal Processing and Multimedia Applications (SIGMAP 2021)*, pages 39-47
ISBN: 978-989-758-525-8

The idea of using the human perception model to improve some performance of an active noise cancellation (ANC) system has been previously studied. In (Wang et al., 2012), a two-stage ANC approach is proposed where a carefully designed masking signal is used to mask the residual noise, whereas (Wang et al., 2018) takes advantage of the temporal post-masking effect in order to improve the noise loudness at the ANC output. The patent by Doclo (Doclo, 2016) controls the filter shape of the ANC system according to some audio signal, which is the same idea that the one explored in this paper. However, in (Doclo, 2016) the design of the ANC filters are obtained through an iterative optimization process, which is difficult to achieve in real-time systems (increasing the active performance in a certain frequency region typically reduces the active performance in another frequency region).

In this paper, a multi-channel real-time ANE system (Gonzalez et al., 2006) that shape the residual noise according to a certain masking threshold is proposed, and its performance is compared to other two profiles for different types of musical sounds. We have selected excerpts from two different music genres, one *Latino*-type and one *Jazz*-type. The noise is a multi-frequency noise similar to that generated by a diesel engine and other rotating machines (Ho et al., 2020). The system performance in terms of noise attenuation, mean squared error of the recorded music signal and anti-noise power emitted by the loudspeakers are analysed. It can be appreciated how the perceptual equalization based on the ambient music can reduce the emitted anti-noise power without compromising the perceived quality of the music.

The rest of the paper is organized as follows: In Section 2, relevant concepts regarding active noise control and the computation of the masking threshold are reviewed. In Section 3, the strategy to implement a perceptual ANE is stated. In Section 4, the real-time ANE experiment mounted in our laboratory is described and their results are shown, together with a discussion on the system performance obtained for the different equalization profiles and types of music signals. Finally, Section 5 presents the main conclusions of the paper.

2 SYSTEM MODEL

This section states the model of the ANE system and describes the process to compute the masking threshold of a sound in the frequency domain, also known as simultaneous masking (Fastl and Zwicker, 2007). Once these previous concepts are explained, the principles of the perceptual equalization strategy will be described in Section 3.

2.1 Active Noise Equalization

As said before, the ANE system presented in this paper is based on the multi-channel model described in (Gonzalez et al., 2006). Fig. 1 shows the block diagram of the ANE system designed to equalize the multi-frequency noise. The multi-channel system is formed by J secondary sources and L error sensors (or microphones), although Fig. 1 describes the signal processing for one generic error sensor l and one generic loudspeaker j. The output of the sinusoidal generator is the reference signal $x(n)$ composed by K sine waves denoted by $x_k(n)$ with $k = 1,\ldots,K$. Therefore, the adaptive filter $W_j(z)$ is formed only by two-coefficients for each component and secondary source (Kuo and Ji, 1995): $W_j(z) = [w_{j1,1}, w_{j1,2}, \ldots, w_{jK,1}, w_{jK,2}]$.

The noise arriving at the lth microphone is denoted by $d_l(n)$. Then, the recorded signal at the lth error sensor is expressed as

$$e_l(n) = d_l(n) + \sum_{k=1}^{K}(1-\beta_k)\sum_{j=1}^{J}[s_{lj}(n) * y_{jk}(n)], \quad (1)$$

where β_k is the attenuation associated to reference signal $x_k(n)$, $s_{lj}(n)$ is the impulse response of the FIR filter that models the electroacoustic path (or room impulse response, RIR) between the jth loudspeaker and the lth microphone, and $y_{jk}(n)$ is the kth component of the adaptive filter output $y_j(n)$. Notice that the RIRs $s_{lj}(n)$ must be estimated for the design of the adaptive filters, being their estimates denoted by $\tilde{s}_{lj}(n)$ in Fig. 1. We assume that the estimated RIRs have a negligible estimation error, that is, $\tilde{s}_{lj}(n) \approx s_{lj}(n), \forall l, j$.

It was shown in (Gonzalez et al., 2006) that the system transfer function in steady state for the kth reference signal is given by

$$H_l(e^{j\omega_k}) = \frac{E_l(e^{j\omega_k})}{D_l(e^{j\omega_k})} = \beta_k,\ l = 1,\ldots,L, \quad (2)$$

which will be intuitively shown in the following[1].

Let us express (1) in the frequency domain for every ω_k such that $k = 1,\ldots,K$, and taking into account that $Y_j(e^{j\omega_k}) = W_j(e^{j\omega_k})X(e^{j\omega_k})$ according to the model shown in Fig. 1:

$$E_l(e^{j\omega_k}) = D_l(e^{j\omega_k}) + (1-\beta_k)\times \sum_{j=1}^{J} S_{lj}(e^{j\omega_k})W_j(e^{j\omega_k})X(e^{j\omega_k}). \quad (3)$$

[1]For the proof of convergence, see Section V of (Gonzalez et al., 2006)

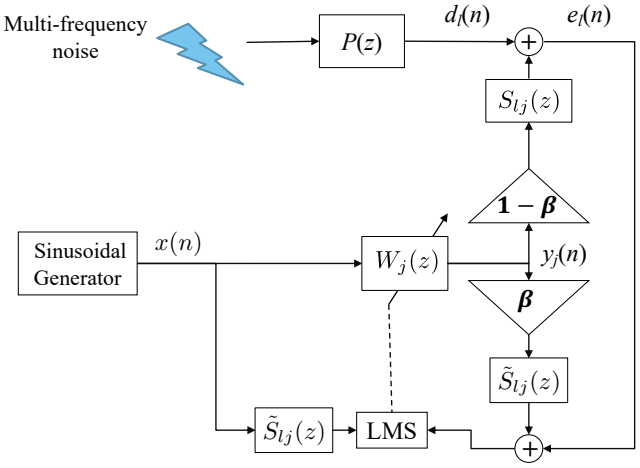

Figure 1: ANE system model where only the signal processing within the jth secondary source is shown.

Substituting $E_l(e^{j\omega_k}) = \beta_k D_l(e^{j\omega_k})$ from (2) in (3) and assuming that $D_l(e^{j\omega_k}) = P_l(e^{j\omega_k})X(e^{j\omega_k})$, where $P_l(z)$ is the primary acoustic path between the noise source and the lth microphone shown in Fig. 1, then (3) can be rewritten as

$$-(1-\beta_k)P_l(e^{j\omega_k})X(e^{j\omega_k}) = (1-\beta_k)\times \sum_{j=1}^{J} S_{lj}(e^{j\omega_k})W_j(e^{j\omega_k})X(e^{j\omega_k}), \quad (4)$$

which leads to

$$\sum_{j=1}^{J} S_{lj}(e^{j\omega_k})W_j(e^{j\omega_k}) = -P_l(e^{j\omega_k}), \quad (5)$$

that is, the primary path of the unwanted noise is compensated by the combined response of all the adaptive filters convolved with their respective secondary paths in time. Therefore, when the adaptive filters designed in the ANE system fulfill (5) in steady state, the signals captured by the error sensors are expressed as

$$E_l(e^{j\omega_k}) = \beta_k D_l(e^{j\omega_k}), \quad (6)$$

for $l = 1,\ldots,L$, $j = 1,\ldots,J$ and $k = 1,\ldots,K$.

The parameter β_k is the attenuation of the kth component of the unwanted noise with respect to its original amplitude. In Fig. 1 the parameters β_k are gathered in the vector

$$\boldsymbol{\beta} = \begin{bmatrix}\beta_1 & \beta_2 & \ldots & \beta_K\end{bmatrix}. \quad (7)$$

2.2 Masking Threshold Computation

In this section we will briefly explain how the masking threshold of a sound is computed. This concept has been previously used in (Rämö et al., 2012; Rämö et al., 2013; Estreder et al., 2018) to carry out the perceptual equalization of audio signals in the presence of an unwanted ambient noise.

The simultaneous (or frequency) masking effect occurs within the human hearing system when two sounds are presented at the same time, one of them being the masker signal and the other the maskee signal. This effect causes the maskee signal to be perceived less louder than if it were presented alone (Fastl and Zwicker, 2007; ISO Central Secretary, 2014). Moreover, depending on the masking threshold of the masker signal, the maskee can even get unnoticed by the listener (Johnston, 1988; Moore, 1985).

Fig. 2 shows the block diagram (Rämö et al., 2013; Estreder et al., 2018) to estimate the masking threshold of the mth time frame of a broadband signal $r_m(n)$ composed by L_s samples. The "Spectral Analysis" block estimates its Power Spectral Density (PSD), $P_m(k)$, over frequencies $f_k = \frac{k}{N} f_s$, where $k = 0,\cdots,N/2$, being N the size of the fast Fourier transform (FFT) and $f_s = 44100$ Hz the sampling rate. In our case $N = 2L_s$, being L_s/f_s approximately 200 ms.

Once $P_m(k)$ is obtained, the "Map to Bark Scale" block maps the frequency bins into the Bark domain using (Zwicker and Terhardt, 1980):

$$\nu = 13\arctan\left(\frac{0.76f}{1000}\right) + 3.5\arctan\left(\frac{f}{7500}\right)^2, \quad (8)$$

where ν denotes the index of the Bark band, $\nu = 1,2,\ldots,N_c$, and N_c is the number of Bark bands to process, in our case $N_c = 25$. This block obtains the energy per Bark band as (Johnston, 1988)

$$G_m(\nu) = \sum_{k=\inf(\nu)}^{\sup(\nu)} P_m(k), \quad \nu = 1,\cdots,N_c, \quad (9)$$

where $\inf(\nu)$ and $\sup(\nu)$ correspond to the frequency bin of the lower and upper boundary of the νth Bark band, respectively.

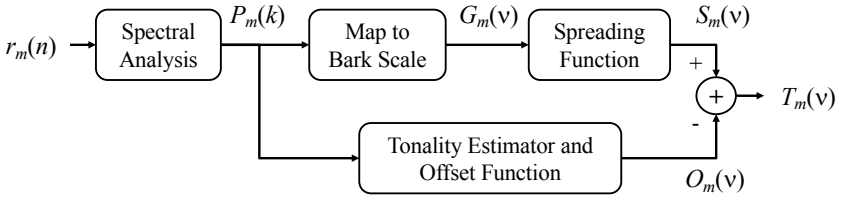

Figure 2: Procedure to estimate the masking threshold $T_m(\nu)$ of a time frame of a sound $r_m(n)$.

The "Spreading Function" block models the masking caused by the energy of a certain Bark band over the rest of the bands. From the set of previously proposed spreading functions (Bosi and Goldberg, 2012), we have used that of (Schroeder et al., 1979). In order to compute the overall spread masking function $S_m(\nu)$, we showed in (Estreder et al., 2018) that a linear addition of the curves produced by the spreading function in (Schroeder et al., 1979) obtains similar results to the exponential addition of (Rämö et al., 2012), which has also been confirmed in (Moore, 1985) for multi-frequency signals. Therefore, for each Bark band ν, their masking is computed as

$$S_m(\nu) = \sum_{\eta=1}^{N_c} b_\eta(\nu), \tag{10}$$

where $b_\eta(\nu)$ is the masking intensity produced by Bark band η over Bark band ν computed from their corresponding energies (9) as in (Estreder et al., 2018).

Regarding the lower branch of Fig. 2, it calculates the spectral offset $O_m(\nu)$, which models the influence of the type of spectral content on the final masking: if most of the energy is concentrated in a small portion of the band (tone-like masking), this value will be high, whereas if the energy is equally spread along the band (noise-like masking), this parameter will be low. In (Gray and Markel, 1974) the Spectral Flatness Measure (SFM) is used to model the tone-like or noise-like masking (for further details see also (Johnston, 1988; Rämö et al., 2013)). Finally the masking threshold $T_m(\nu)$ of time frame $r_m(n)$ at the νth Bark band is estimated as

$$T_m(\nu) = 10^{\left[\log_{10}(S_m(\nu)) - \frac{O_m(\nu)}{10}\right]}. \tag{11}$$

3 PERCEPTUAL ACTIVE NOISE EQUALIZER

As we stated in Section 1, our goal is to design a perceptual equalizer as in (Rämö et al., 2012; Estreder et al., 2018), but with an alternative strategy: instead of equalizing the audio signal to mask the noise, the ambient noise will be equalized by the ANE system of Fig. 1 in order to get masked by the audio signal.

Fig. 3 shows the block diagram of the Perceptual Active Noise Equalizer (PANE). The subscript m in all the signals of Fig. 3 denote the mth time frame of duration L_s samples. The audio signal (speech or music) emitted by all the speakers is $a(n)$, whereas, according to (1), the anti-noise signal emitted by the jth loudspeaker of the ANE is expressed as

$$y_j^{\beta}(n) = \sum_{k=1}^{K} (1-\beta_k) y_{jk}(n). \tag{12}$$

The values of vector $\boldsymbol{\beta}$ are computed in the "Attenuation Estimation" block according to the masking threshold of the music and the current level of the ambient noise at each error sensor. The measured signal at the lth error sensor (1) is now expressed as

$$\begin{aligned} e_l(n) &= d_l(n) + \sum_{j=1}^{J} [s_{lj}(n) * y_j^{\beta}(n)] + a_l(n) \\ &= d'_l(n) + a_l(n), \end{aligned} \tag{13}$$

where $d'_l(n)$ is the total ambient noise measured by microphone l and

$$a_l(n) = a(n) * \sum_{j=1}^{J} s_{lj}(n), \tag{14}$$

is the resulting audio signal propagated from the J loudspeakers to the lth microphone through the corresponding electroacustic paths $s_{lj}(n)$ of Fig. 1.

As said before, $s_{lj}(n)$ has been previously estimated in the ANE system. Since $\tilde{s}_{lj}(n) \approx s_{lj}(n)$, we can obtain the estimate of the time frames of both signals, $a_l(n)$ and $d'_l(n)$ in (13), as

$$\begin{aligned} \tilde{a}_{l,m}(n) &= a_m(n) * \sum_{j=1}^{J} \tilde{s}_{lj}(n), \\ \tilde{d'}_{l,m}(n) &= e_{l,m}(n) - \tilde{a}_{l,m}(n). \end{aligned} \tag{15}$$

Notice that the frame of the audio signal chosen to compute the masking threshold at the input of Fig. 3 is $a_m(n)$, whereas the one that will be emitted by the loudspeakers is $a_{m+1}(n)$. In a real-time experiment as the one presented in Section 4, the audio signal can

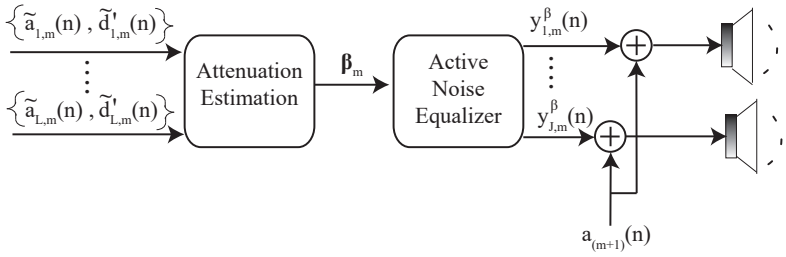

Figure 3: Perceptual Active Noise Equalizer.

be originated from a streaming web service or a radio station, thus, the masking threshold computation must be carried out on the same audio frame that is being reproduced (m), and its resulting $\boldsymbol{\beta}_m$ will affect the next time frame ($m+1$). In case the audio signal is originated from a local memory or disk, then the estimate $\tilde{a}_{l,m}(n)$ can be pre-computed prior to the reproduction, and $a_m(n)$ will be emitted by the loudspeakers instead of $a_{m+1}(n)$.

Finally, the computation of attenuation vector $\boldsymbol{\beta}_m$ for the PANE together with other two possible strategies will be explained in the following.

3.1 Perceptual Equalization (PEQ)

The perceptual equalization (PEQ) profile is the attenuation profile obtained by the "Attenuation Estimation" block of the PANE shown in Fig. 3. Considering expressions (6) and (13), the noise term at each error sensor for frequency f_k can be expressed as $D'_{l,m}(e^{j\omega_k}) = \beta_{k,m} D_{l,m}(e^{j\omega_k})$. Therefore, the values of $\beta_{m,k}$ are estimated as:

$$20\log_{10}(\beta_{m,k}) = \min_l \left(\left[T^{\mathrm{A}}_{l,m}(\nu) - G^{\mathrm{D}'}_{l,m}(\nu) \right] - g_k, 0 \right), \ \forall f_k \in B_\nu, \tag{16}$$

where $T^{\mathrm{A}}_{l,m}(\nu)$ is the masking threshold of the audio signal and $G^{\mathrm{D}'}_{l,m}(\nu)$ is the energy per Bark band of the noise signal, both referred to error sensor l and expressed in dB, and B_ν is the bandwidth (8) of the Bark band ν in Hertz. The parameter g_k is an extra gain that can be used to force the noise power to fall g_k dB below the masking threshold of the audio signal. This value is usually set to $2-3$ dB in practice to take into account the power variations of the audio signal within the time of a frame. Nevertheless, g_k can also be considered as a control parameter that can shape the noise PSD according to the user preferences.

Notice that the strategy stated in (16) computes $\beta_{m,k}$ as the minimum value considering all the error sensors. An alternative strategy could be to compute the attenuation for each sensor as $20\log_{10}(\beta_{m,k}(l)) = \left[T^{\mathrm{A}}_{l,m}(\nu) - G^{\mathrm{D}'}_{l,m}(\nu) \right] - g_k$ and then compute $\beta_{m,k}$ as their average over l. Additional strategies or constraints can be used depending on the music genre and the locations of the error sensors with respect to the noise source.

3.2 Hearing Threshold (HT)

The hearing threshold (HT) profile aims to force the noise power to remain below the standard curve of the human hearing threshold established in (ISO Central Secretary, 2014). Therefore, the values of $\beta_{m,k}$ are estimated as:

$$20\log_{10}(\beta_{m,k}) = \min_l \left(L_{TH}(e^{j\omega_k}) - G^{\mathrm{D}'}_{l,m}(e^{j\omega_k}), 0 \right), \tag{17}$$

where $L_{TH}(e^{j\omega_k})$ is the value in dB of the hearing threshold at frequency f_k and $G^{\mathrm{D}'}_{l,m}(e^{j\omega_k})$ is the same noise term that in (16), but expressed in the frequency domain.

3.3 Full Cancellation (FC)

The main objective of the full cancellation (FC) profile is to cancel the noise, or at least attenuate it as much as possible regardless the existence of the audio signal. Therefore,

$$\beta_{m,k} = 0 \ , \ \forall m,k \tag{18}$$

4 REAL-TIME EXPERIMENT

A multi-channel PANE system has been built in the listening room located at the Audio Lab[2] of the Institute of Telecommunications and Multimedia Applications (iTEAM) of the Technical University of Valencia. Fig. 4 shows a picture of the 2×2 multi-channel system formed by $L = 2$ microphones acting as error sensors, and $J = 2$ loudspeakers placed as if they were mounted at the headrest sides of a car seat. A third loudspeaker (not shown in the picture), placed oppo-

[2] https://gtac.webs.upv.es/

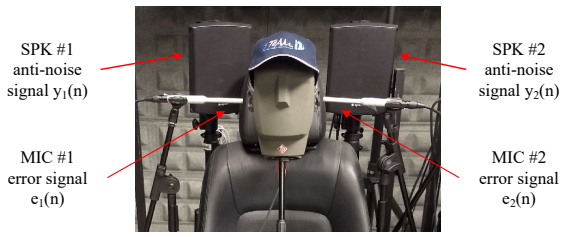

Figure 4: ANE Prototype mounted in the iTEAM Audio Lab.

site to the seat, was used to generate the unwanted noise formed by $K = 11$ frequency components such that $f_k = (k+1) \cdot 47.05$ Hz, $k = 1, \ldots, 11$, i.e., being its last frequency $f_{11} = 564.6$ Hz.

A computer equipped with an *Intel i7-8700* processor and 32 GB of RAM has been used to run the algorithms in real-time. The signal processing is entirely computed by Matlab 2018b, whereas its audio processing toolbox communication drives the record and generation of signals through a Roland Octa-Capture soundcard. The sample rate was 44100 Hz. The microphones are the Behringer ECM8000 model, the loudspeaker for the primary noise is an Event Project Studio 6 and the two loudspeakers of the ANE system are model SDQ5P from Apart.

4.1 Calibration and Estimation of Secondary Paths

The masking threshold computation needs a precise calibration of the whole audio system previous to carry out the experiments. Firstly we have calibrated both microphones with the aid of a pistonphone signal (Frederiksen, 2008). Secondly, we have estimated the secondary paths, $s_{lj}(n)$, $l, j = 1, 2$, from each loudspeaker to each microphone using a maximum length sequence (MLS) signal (Vanderkooy, 1994) of 16384 samples. The estimated secondary paths have been trimmed to a length of 4096 coefficients, keeping the 99.95% of their total energy. Finally, every estimated $\tilde{s}_{lj}(n)$ has been normalized according to the real SPL of the corresponding measured signal in microphone l. The estimated secondary paths $\tilde{s}_{lj}(n)$ include the electroacoustic characteristics of the loudspeakers, microphones and audio sound card plus the room impulse response.

4.2 Experiments

In this section a study has been carried out in order to compare the different profiles of Section 3. For the PEQ profile a value of $g_k = 3$ dB has been used for all the frequencies.

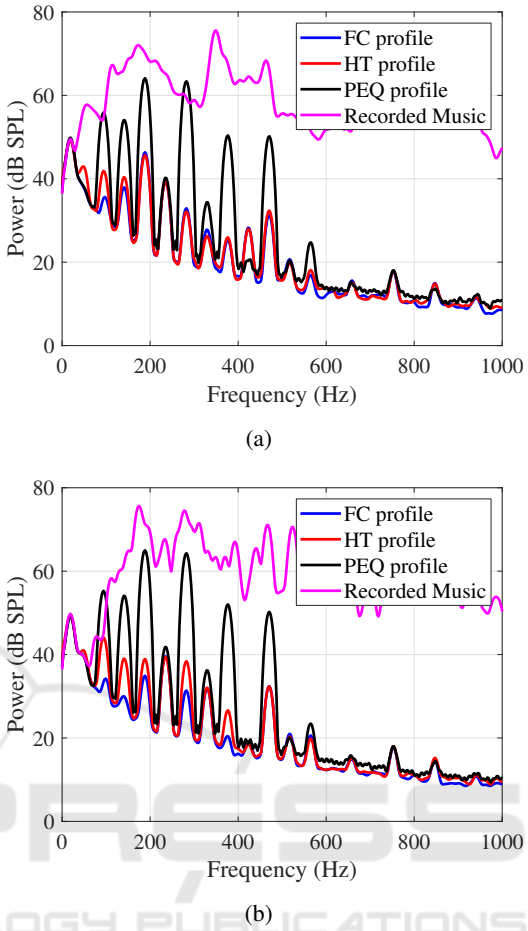

Figure 5: PSD of the recorded noise at Mic #1 for the three equalization profiles (a) when the "Latino" song is playing, and (b) when the "Jazz" song is playing. The PSD of the recorded song is also shown.

As audio signals, we have used two excerpts of different musical genres. The first audio signal is a 30 s excerpt from the song "Con altura" performed by Rosalia, J Balvin and El Guincho (2019) that we have labelled as "Latino" song. The second audio is an instrumental 36 s excerpt from the song "Alfonsina y el mar" performed by The Jazz Chamber Trio (2005) that we have labelled as "Jazz" song. The audio signals are reproduced with a level such that they are recorded at 83 dB SPL on average, while the noise signal is reproduced such that it is recorded at 76 dB SPL on average.

4.3 Results and Discussion

Fig. 5 shows the PSD of noise signal $d'_1(n)$ measured at the microphone labelled as Mic #1 in Fig. 4 for every equalization profile, and the PSD of its corresponding audio excerpt, $a_1(n)$. All the PSDs are given

in dB SPL. We have considered only the last 10 s of all the signals in order to assure that the adaptive filters $W_j(z)$ of the ANE system of Fig. 1 have converged. Although the results are only referred to Mic #1, the noise and music PSDs measured at Mic #2 were very similar. The frequency is represented only from 0 Hz to 1 kHz in order to focus on the part of the spectrum where the noise is present, being its last component located at $f_{11} = 564.6$ Hz.

As expected, the profile that provides a lower recorded noise power for both music signals is the FC, whose strategy is to cancel the noise in all the frequencies. Regarding the HT profile in Fig. 5, it presents a similar profile for both "Latino" and "Jazz" signals, with slight variations for the components around 200 Hz due to experimental variations. Compared to the FC profile, it allows for more noise content at the lowest frequencies where the human hearing system is less sensitive.

Regarding the PEQ profile in Fig. 5, the amount of noise power recorded at the error sensor is in the range of $10-20$ dB higher than the HT profile for most of the components and for both audio signals. Comparing the PEQ noise curve to the music PSD, it can be appreciated how the PEQ is far below the music in some frequencies, but not in all of them. This is due to the fact that their attenuation values, $\boldsymbol{\beta}_m$, are designed with respect to the masking threshold of the music, not with respect to its PSD. The masking threshold is computed through a non-linear process over the music PSD as Fig. 2 describes, and it is represented in the Bark domain, thus, comparing the music masking threshold with the noise PSD would have no sense. Nevertheless, from the PEQ curves in Fig. 5, we can state that the PEQ strategy achieves its goal of actively equalizing the ambient noise to get masked by the music. All the signals involved in this experiment have been saved as "wav" files and are available to be listened at http://gpinyero.webs.upv.es/Experiments-sigmap2021/.

Fig. 6, by contrast, shows the PSDs of the anti-noise signal $y_1^{\beta}(n)$ emitted by the speaker labelled as Spk #1 in Fig. 4. As in Fig. 5, the curves represent the PSD considering only the last 10 s of the corresponding signals. The power is represented in dB full scale (FS), where FS means that the power level is referred to a signal amplitude range limited to $[-1,1]$.

It can be appreciated that FC and HT profiles emit the same power for all the frequencies, although the noise power received at the error sensors shown in Fig. 5 is higher for HT in most of the frequency components. Therefore, a certain relaxation in the noise cancellation strategy apparently does not correspond to the same amount of anti-noise power saved by the

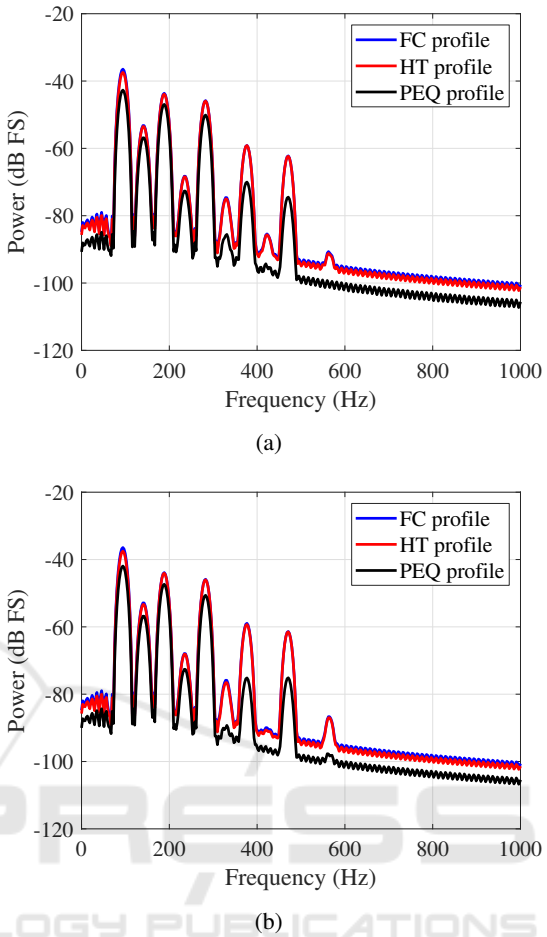

Figure 6: PSD of the anti-noise signal emitted by Spk #1 for the three equalization profiles (a) when the "Latino" song is playing, and (b) when the "Jazz" song is playing.

ANE system. Notice that Fig. 5 represents the PSD of $d_1'(n) = d_1(n) + \sum_{j=1}^{J}[s_{1j}(n) * y_j^{\beta}(n)]$ in (13), which depends on the level of the ambient noise captured at Mic #1, $d_1(n)$, and the combined response of the anti-noise signals, $y_j^{\beta}(n), j = 1,2$, propagated through their corresponding secondary paths, $s_{1j}(n)$. Therefore, it can be thought that, due to particular experimental conditions, a minimum anti-noise power is needed to actively affect the undesired ambient noise.

Regarding the curve of the PEQ profile in Fig. 6, The anti-noise PSD is only $3-5$ dB lower than HT PSD for all the components below 300 Hz. However, for the components above 300 Hz, the saved power reaches significant values around 10 dB for the "Latino" song and 15 dB for the "Jazz" song. This is a relevant result for our proposed PANE scheme, since the anti-noise emitted within a reduced room, as a car cabin, propagates all over the space and can negatively enhance the ambient noise at positions differ-

ent from the seat where the noise is being controlled. In this sense, any reduction on the amount of energy noise emitted by the secondary sources would help to actively control the noise over multiple regions within a room (Zhang et al., 2020).

Finally, Tables 1 and 2 show the Mean Square Error (MSE) and the Perceptual Evaluation of speech quality (PESQ) values of the recorded audio signal at Mic #1 respectively, being the reference signal in both cases the audio excerpt recorded at Mic #1 when the ambient noise is not present. Apart from the results obtained by the three equalization strategies (FC, HT, PEQ), the results obtained when no noise equalization is carried out are presented in the row labelled "ANE-OFF".

The MSE is an objective measure whereas the PESQ is intended to measure the subjective quality of a speech signal through an index ranging from 0.5 to 4.5 (ITU-T, 2001). Comparing the PESQ to the Mean Opinion Score (MOS) that ranges from 1 to 5, a PESQ value of 4.5 can be considered of "excellent" quality, whereas PESQ values below 3 are usually considered "fair" for speech communications, but "poor" for audio reproduction. Nevertheless, although the PESQ is not validated for other applications than speech communication (ITU-T, 2001), it is widely used in many speech and audio applications as echo cancellation, audio coding and speech enhancement.

Table 1: MSE of the recorded audio signal at Mic #1.

	Latino	Jazz
FC	-27.3 dB	-28.84 dB
HT	-27.8 dB	-28.3 dB
PEQ	-15.13 dB	-15.75 dB
ANE-OFF	-5.59 dB	-6.93 dB

Table 2: PESQ of the recorded audio signal at Mic #1.

	Latino	Jazz
FC	4.43	4.49
HT	4.41	4.48
PEQ	3.51	4.17
ANE-OFF	2.86	3.63

Table 1 shows that the FC and HT profiles achieve an MSE around -27 dB, whereas Table 2 shows that their corresponding PESQ values are close to 4.5, meaning that the audio signals are perceived as if no ambient noise were present. On the other hand, the PEQ profile achieves PESQ values of 3.5 for the "Latino" song and 4.17 for the "Jazz" song, although their respective MSE values are both $12-13$ dB higher than their corresponding HT values. Nevertheless, the PEQ profile presents a relevant reduction of $9-10$ dB of the MSE with respect to the "ANE-OFF" condition, and a "Good / Excellent" quality in the PESQ scale.

As a final remark, the results here obtained for the PEQ profile, even if they are related to the particular conditions of the experimental study, provide very interesting insights regarding the novel PANE system shown in Fig. 3:

1. The PEQ can be implemented in real-time without any additional latency on the audio reproduced by the entertainment system.
2. The PEQ obtains a "Good / Excellent" audio quality compared to the same audio signals measured in absence of the ambient noise.
3. The PEQ achieves a significant reduction in the anti-noise power emitted by the loudspeakers, although the amount of saved power depends on the particular frequency range considered.
4. The PEQ is a versatile approach that allows to exchange the amount of ambient noise power measured at the microphones with the amount of anti-noise power emitted by the loudspeakers, just by controlling the parameter g_k in the design of β_k (16): the more the value of g_k is increased, the more the PEQ profile get close to HT or FC profiles.

5 CONCLUSIONS

In this paper a novel perceptual active noise equalizer (PANE) has been presented. The PANE intends to reduce the power of a multi-frequency ambient noise in such a way that it cannot be perceived when a music signal is present thanks to the masking effect of the human hearing system. A real-time experiment has been presented and the PANE performance has been compared to other strategies of the ANE system as the so-called full cancellation (FC) and hearing threshold (HT) profiles. Although the measured noise power level is higher for the PEQ profile, the emitted anti-noise power is lower compared to HT and FC profiles. This performance leads to a reduction of noise pollution in the room and a lower power consumption of the system loudspeakers. In addition, the PEQ profile is a versatile approach that can reduce the perceived noise as much as the user decides, reaching the same performance than the HT or FC profiles. Finally, the experiment presented in this paper has validated the PANE system as a real-time versatile alternative to the classic active noise cancellation systems for multi-frequency noise. Further research is being carried out in our laboratory to extend the PEQ to the

case of broadband noises, as road and wind noises in cars.

ACKNOWLEDGEMENTS

This work has been partially supported by EU together with Spanish Government through RTI2018-098085-BC41 (MCIU/AEI/FEDER) and RED2018-102668-T, and Generalitat Valenciana through PROMETEO/2019/109.

REFERENCES

Bosi, M. and Goldberg, R. (2012). *Introduction to digital audio coding and standards*. Springer Science & Business Media.

Brandenburg, K. and Johnston, J. (1990). Second generation perceptual audio coding: the hybrid code. In *Audio Engineering Society Convention 88*.

Doclo, S. (2016). Active noise reduction method using perceptual masking, Patent US20110026724A1.

Donley, J., Ritz, C., and Kleijn, W. (2016). Improving speech privacy in personal sound zones. In *2016 IEEE Intl. Conf. on Acoustics, Speech and Signal Proc.*, pages 311–315.

Estreder, J., Piñero, G., Aguirre, F., de Diego, M., and Gonzalez, A. (2018). On perceptual audio equalization for multiple users in presence of ambient noise. In *2018 IEEE 10th Sensor Array and Multichannel Signal Proc. Workshop (SAM)*, pages 445–449.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Fletcher, H. (1940). Auditory patterns. *Reviews of modern physics*, 12(1):47.

Frederiksen, E. (2008). Microphone calibration. In *Handbook of Signal Processing in Acoustics*, pages 1293–1312. Springer New York, New York, NY.

Gonzalez, A., de Diego, M., Ferrer, M., and Piñero, G. (2006). Multichannel active noise equalization of interior noise. *IEEE Trans. on Audio, Speech and Language Proc.*, 14(1):110–122.

Gray, A. and Markel, J. (1974). A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 22(3):207–217.

Ho, C.-Y., Shyu, K.-K., Chang, C.-Y., and Kuo, S. M. (2020). Efficient narrowband noise cancellation system using adaptive line enhancer. *IEEE/ACM Trans. on Audio, Speech, and Language Proc.*, 28:1094–1103.

ISO Central Secretary (2014). Acoustics – normal equal-loudness-level contours. Standard, International Organization for Standardization.

ITU-T (2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *ITU-T Recommendation P.862*.

Johnston, J. (1988). Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, 6(2):314–323.

Kuo, S. and Ji, M. (1995). Development and analysis of an adaptive noise equalizer. *IEEE Trans. on Speech and Audio Proc.*, 3(3):217–222.

Kuo, S. and Tsai, J. (1994). Residual noise shaping technique for active noise control systems. *The Journal of the Acoustical Society of America*, 95(3):1665–1668.

Moore, B. (1985). Additivity of simultaneous masking, revisited. *The Journal of the Acoustical Society of America*, 78(2):488–494.

Mosquera-Sánchez, J., Sarrazin, M., Janssens, K., de Oliveira, L., and Desmet, W. (2018). Multiple target sound quality balance for hybrid electric powertrain noise. *Mechanical Systems and Signal Proc.*, 99:478–503.

Pickles, J. (2012). *An introduction to the physiology of hearing*. Emerald Group Pub Bingley, Bingley, UK, 4th ed. edition.

Rämö, J., Välimäki, V., Alanko, M., and Tikander, M. (2012). Perceptual frequency response simulator for music in noisy environments. In *AES Conference: 45th Intl. Conf.: Applications of Time-Frequency Proc. in Audio*.

Rämö, J., Välimäki, V., and Tikander, M. (2013). Perceptual headphone equalization for mitigation of ambient noise. In *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, pages 724–728.

Schroeder, M., Atal, B., and Hall, J. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *The Journal of the Acoustical Society of America*, 66(6):1647–1652.

Välimäki, V. and Reiss, J. (2016). All about audio equalization: Solutions and frontiers. *Applied Sciences*, 6(5).

Vanderkooy, J. (1994). Aspects of MLS measuring systems. *J. Audio Eng. Soc*, 42(4):219–231.

Wang, T., Gan, W.-S., and Chong, Y.-K. (2012). Psychoacoustic active noise control system with auditory masking. In *Proc. 2012 Asia Pacific Signal and Information Proc. Assoc.*, pages 1–6.

Wang, Y., Feng, T., Wang, X., Guo, H., and Qi, H. (2018). An improved LMS algorithm for active sound-quality control of vehicle interior noise based on auditory masking effect. *Mechanical Systems and Signal Processing*, 108:292–303.

Zhang, J., Sun, H., Samarasinghe, P. N., and Abhayapala, T. D. (2020). Active noise control over multiple regions: Performance analysis. In *2020 IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pages 8409–8413.

Zwicker, E. and Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, 68(5):1523–1525.