REAL-TIME SYSTEM FOR SOUND ENHANCEMENT IN NOISY ENVIRONMENT

Stefania Cecchi

Dep. of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
s.cecchi@staff.univpm.it

Valeria Bruschi

Dep. of Information Engineering
Università Politecnica delle Marche
Ancona, Italy
v.bruschi@staff.univpm.it

Paolo Peretti

Leaff Engineering
Ancona, Italy
peretti@leaff.it

Ferruccio Bettarelli

Leaff Engineering
Ancona, Italy
bettarelli@leaff.it

ABSTRACT

The noise can affect the listening experience in many real-life situations involving loudspeakers as a playback device. A solution to reduce the effect of the noise is to employ headphones, but they can be annoying and not allowed on some occasions. In this context, a system for improving the audio perception and the intelligibility of sounds in a domestic noisy environment is introduced and a real-time implementation is proposed. The system comprises three main blocks: a noise estimation procedure based on an adaptive algorithm, an auditory spectral masking algorithm that estimates the music threshold capable of masking the noise source, and an FFT equalizer that is used to apply the estimated level. It has been developed on an embedded DSP board considering one microphone for the ambient noise analysis and two vibrating sound transducers for sound reproduction. Several experiments on simulated and real-world scenarios have been realized to prove the effectiveness of the proposed approach.

1. INTRODUCTION

Noise perception is becoming a huge problem in everyday life. Recent studies performed during the covid 19 lockdown have demonstrated how much noise surrounds our lives and how much we no longer realize this aspect [1]. In [2], it is demonstrated how construction noise in urban areas affects human health negatively. Techniques for noise reduction can be applied to enhance the acoustic comfort of indoor environments [3] or inside vehicles [4, 5]. There are several types of noise: a general classification could be between localized noises (e.g., noise perceived near the listener and generated by rotational elements, such as engines and fans) or diffuse noises (e.g., the noise generated by a source far from the listener). Localized noise can be reduced with an active noise control (ANC) algorithm [6, 7] that can cancel the noise generating the same signal in a specific position with a phase inversion. Usually, adaptive filters [8] are employed in ANC systems to adjust the canceling process according to time variations of the acoustic noise [9, 10]. In particular, since the noise is well localized two microphones are involved, i.e., a reference microphone that captures the primary noise, and an error microphone that controls the canceling zone (feedforward ANC solution) [11, 12]. In the case of diffuse noise, it is not possible to apply this kind of algorithm since a reference microphone can not be positioned. A feedback ANC structure that has only the error microphone can be used [10, 13] but the cancellation performance is limited by frequency range. So, in those cases where a sound source is present, it is possible to modify the sound material in order to attenuate the noise perception rather than noise cancellation. Typical examples can be a person who is listening to music while traffic noise is present (external noise) or a person who is listening to music in the living room while the extractor fan is active in the near kitchen (internal noise). In these cases, it is possible to apply an auditory masking effect to modify the sound reproduction thus decreasing listener noise perception.

The auditory masking effect is deeply described in [14]. It is defined as the psychoacoustic effect capable of overcoming and thus masking a sound signal with another sound signal. The reported example of Zwicker is about two persons who are speaking with a low speech power in a quiet street. If during the conversation, a truck passes by, the generated noise overcomes the voices and the conversation is altered. If the same speech power is maintained, the conversation can not be performed. To have a conversation, it is necessary to increase the speech power intensity. Considering the same principle, it is possible to calculate a masking threshold [15] to be applied to a masker signal (i.e., the masking sound) and this threshold represents the levels under which a signal (called maskee) becomes inaudible because it is being masked completely or partially [14]. The masking effect is often investigated in the Bark bands, known as critical bands [16], evaluating the masking threshold. In [17], models for the prediction of the masking threshold for steady sounds can be found, while in [18] time-varying sound signals are involved. However, these models become ineffective when complex signals are involved.

The auditory masking effect has been mainly used in audio codec implementation [16, 19, 20]. The idea is based on the exploitation of the masking effect to reduce the information that has to be encoded. In [21, 22], the masking effect has been used for signal decorrelation in stereo acoustic echo cancellation (AEC) systems. It is based on the introduction of a masked noise in a specific frequency range in order to create a difference between the stereo signals. Over the years, the idea of auditory masking has been extended to the enhancement of headphone sound systems. Indeed, in [23, 24], an adaptive equalizer based on auditory mask-

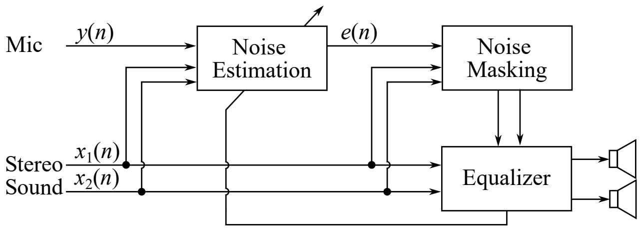

Figure 1: *Scheme of the proposed system.*

ing has been presented for headphone sound reproduction. This method allows for the definition of an equalization curve derived from the masking threshold. Recently, other perceptual equalizers based on auditory masking have been proposed in [25, 26, 27]. The technique presented in [25] is based on the system of [16] and adapted to the multi-user case. The masking threshold calculation of [25] is used in [26] to realize a perceptual active noise equalizer tested with real-time experiments that simulate a car environment. Finally, in [27], an improvement to the tonality metric calculation is introduced to enhance the masking effect for complex sounds.

In this context, a real-time system implementation for improving the audio perception and the intelligibility of sounds in a domestic noisy environment is proposed exploiting the principle of auditory spectral masking. The system which comprises a noise estimation block, a noise masking algorithm, and an equalizer, has been developed on an embedded DSP board, considering one microphone for the ambient noise analysis and two vibrating sound transducers for sound reproduction. The proposed implementation has been evaluated through simulated tests and in real-world scenarios. Experimental results have confirmed the effectiveness of the proposed approach.

The paper is organized as follows. Section 2 reports the overall scheme of the implemented algorithm based on spectral masking effect. Section 3 describes the implementation of the algorithm on a DSP board. Section 4 shows simulated and real results while Conclusions are drawn in Section 5.

2. ALGORITHMS DESCRIPTION

The scheme of the proposed system is shown in Fig. 1. The detailed description of the involved algorithms is reported in the following.

2.1. Noise Estimation

The noise estimation is performed following the frequency-domain adaptive filter of [28], where an acoustic echo canceller (AEC) is implemented. In particular, the adaption is performed between the microphone signal $y(n)$, which comprises the ambient noise and the equalized audio signal modified by the room environment, and the original audio stereo signals $x_1(n)$ and $x_2(n)$ filtered by the equalization curves, obtained, in the frequency domain, as follows:

$$X_{\mathrm{e}}(k,n) = X_1(k,n) * E_{\mathrm{q1}}(k) + X_2(k,n) * E_{\mathrm{q2}}(k), \quad (1)$$

where $X_1(k,n)$ and $X_2(k,n)$ are the FFT of the nth frame of the stereo input signals and $E_{\mathrm{q1}}(k)$ and $E_{\mathrm{q2}}(k)$ are the equalization curves derived as it will be explained in Sec. 2.3.

The estimation of the ambient noise is obtained as the error of the adaption algorithm defined as

$$e(n) = y(n) - \tilde{d}(n), \quad (2)$$

where $\tilde{d}(n)$ is the estimated signal, computed as the inverse FFT of the product between $X_{\mathrm{e}}(k,n)$ and the adaptive filter $W(k,n)$ as

$$\tilde{d}(n) = \mathrm{IFFT}\Big\{X_{\mathrm{e}}(k,n) * W(k,n)\Big\}. \quad (3)$$

The adaptive filter that estimates the room impulse response, is calculated as

$$W(k,n+1) = W(k,n) + 2\mu \frac{X_{\mathrm{e}}^*(k,n) * E(k,n)}{\epsilon + P(k,n)}, \quad (4)$$

where $X_{\mathrm{e}}^*(k,n)$ is the complex conjugate of $X_{\mathrm{e}}(k,n)$, $E(k,n)$ is the error signal defined by the Eq. (2) in the frequency domain, $\mu = 0.01$ is the step size, and $\epsilon = 10^{-15}$ is a small constant to prevent the zero-division. $P(k,n)$ is the power of the signal computed as follows:

$$P(k,n+1) = \alpha P(k,n) + (1-\alpha)\big|X_{\mathrm{e}}(k,n)\big|^2, \quad (5)$$

where $\alpha = 0.9$.

2.2. Noise Masking

The noise masking algorithm is mainly based on the approach presented in [23] with the introduction of some modifications. This technique exploits the psychoacoustic masking phenomenon to analyze the noise that must be suppressed. The masking threshold, defined as the sound pressure level of the noise that must be masked, is calculated by applying the following steps [16, 29]:

1. a windowing procedure is applied to the masker signal (i.e., the masking sound), calculating its short-time Fourier transform (STFT);
2. the power spectrum of each frequency component is calculated;
3. the frequency scale is mapped into the Bark domain and the energy of each critical band is computed;
4. a memoryless smoothing filter is applied to the noise signal;
5. a masking function is calculated depending on a spreading function (SF);
6. the masking threshold is determined by a comparison with the absolute threshold of hearing.

First, the STFT of the signal is computed by windowing each segment of the signal and calculating the DFT. In this way, the power of the nth frame can be determined as

$$Q(k,n) = \mathrm{Re}\big\{X(k,n)\big\}^2 + \mathrm{Im}\big\{X(k,n)\big\}^2 = \big|X(k,n)\big|^2, \quad (6)$$

where $X(k,n)$ is the DFT of the nth signal frame and k represents the frequency bin. Due to the nonlinear frequency resolution of the human auditory system, the power spectrum $Q(k,n)$ is converted into the Bark frequency scale by grouping several bins as follows [30]:

$$B(i,n) = \sum_{k=k_{\mathrm{low}}(i)}^{k_{\mathrm{high}}(i)} Q(k,n), \quad (7)$$

where $k_{\mathrm{low}}(i)$ and $k_{\mathrm{high}}(i)$ are the bin numbers of the lower and the higher limit of the ith Bark band, respectively, with $i = 1, ..., 25$. Then, a spectral memoryless smoothing filter is applied to the

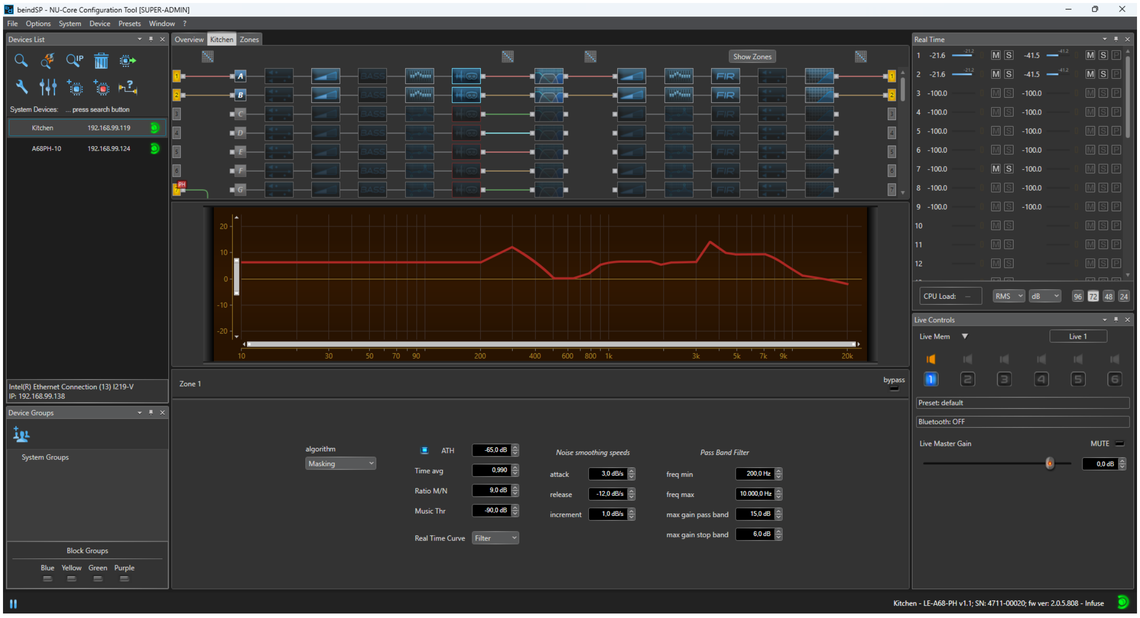

Figure 2: *Software interface with an example of equalization curve derived from auditory noise masking procedure.*

noise signal, obtaining the smoothed Bark power spectrum $B_{\mathrm{sm}}(i,n)$ as follows [30]:

$$B_{\mathrm{sm}}(i,n) = \\ = \begin{cases} \max\left\{L_{\mathrm{m}}, B_{\mathrm{sm}}(i,n-1)C_{\mathrm{inc}}\right\} & \text{if } B(i,n) > B_{\mathrm{sm}}(i,n-1) \\ \max\left\{L_{\mathrm{m}}, B_{\mathrm{sm}}(i,n-1)C_{\mathrm{dec}}\right\} & \text{else,} \end{cases} \tag{8}$$

where the minimum noise level is $L_{\mathrm{m}} = -120$ dB, the increment coefficient is $C_{\mathrm{inc}} = 6$ dB/s, and the decrement coefficient is $C_{\mathrm{dec}} = -6$ dB/s. The masking function is represented by $C(i,n)$, obtained from the convolution of the smoothed discrete Bark spectrum $B_{\mathrm{sm}}(i,n)$ with the spreading matrix $S(i,m)$, as follows [30]:

$$C(i,n) = \sum_{m=1}^{25} B_{\mathrm{sm}}(i,n) 10^{S(i,m)/10}, \tag{9}$$

where the spreading function $S(i,m)$ is defined as

$$S(i,m) = 15.81 + 7.5\big((i-m)+0.474\big) + \\ - 17.5\sqrt{1+\big((i-m)+0.474\big)^2}, \tag{10}$$

Figure 3: *DSP board used for the real-time implementation.*

with $i = 1, \dots, 25$ and $m = 1, \dots, 25$. Therefore, $S(i,m)$ is a matrix with the dimension of 25×25. Finally, the masking threshold $T(i,n)$ is calculated as follows [30]:

$$T(i,n) = \sqrt{\max\left\{C(i,n), 10^{T_{\mathrm{q}}(i)/10}\right\}}, \tag{11}$$

where $T_{\mathrm{q}}(i)$ is the absolute threshold of hearing mapped into the Bark scale, computed as

$$T_{\mathrm{q}}(f) = 3.64\left(\frac{f}{1\mathrm{kHz}}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1\mathrm{kHz}}-3.3\right)^2} + \\ + 10^{-3}\left(\frac{f}{1\mathrm{kHz}}\right)^4. \tag{12}$$

The masking threshold defined in Eq. (11) is reported in the linear frequency domain and is used to design the equalization function.

2.3. Equalizer

The equalizer is designed as an FFT-based equalizer [31], which consists of the design of a target frequency response derived from the masking threshold. The magnitude response of equalization curve $|E_{\mathrm{q}}(k)|$ is obtained by mapping the masking threshold $T(i,n)$ of each ith Bark band into the linear frequency domain as

$$|E_{\mathrm{q}}(k)| = T(i,n) \quad \text{for} \quad k = k_{\mathrm{low}}(i), \dots, k_{\mathrm{high}}(i), \tag{13}$$

with $k = 0, \dots, N/2$, where $N = 512$ is the FFT length. Then, a complex smoothing of the magnitude frequency response is performed following the approach of [32]. The phase response of the equalizer is computed as

$$\angle E_{\mathrm{q}}(k) = e^{j\pi k}. \tag{14}$$

Finally, the stereo input signal is filtered by the equalization curve and reproduced by the loudspeakers, as shown in Fig. 1.

3. REAL-TIME IMPLEMENTATION

The implementation of the system described in the previous Section has been accomplished through a DSP platform named LE-A68PH (Figure 2 and 3). In particular, Figure 2 shows the software interface dedicated to the masking noise application with an example of an equalization curve and Figure 3 shows the hardware of the DSP board used for the real-time implementation.

Going more in detail, this system is powered by the ADSP-21489 SHARC DSP which is a processor in Analog Devices single core SHARC family. It has a 256 Mbit SDRAM PC133 and 16 Mbit Serial FLASH and it is also capable of managing up to 6 balanced/unbalanced audio inputs and up to 8 balanced/unbalanced audio outputs. The LE-A68PH has been designed for maximum flexibility and performance: 450 MHz DSP core speed, audio clock frequency up to 96kHz, top class audio components, network configuration utility, easy interface to any amplifier module.

4. EXPERIMENTAL RESULTS

The evaluation of the proposed implementation has been performed by several tests both considering simulated signals and real-world scenarios.

4.1. Simulated tests

After the implementation on the DSP board, a simulation has been performed controlling the board with a PC and emulating a real scenario. The functionalities of the system have been tested considering a sound track as masker sound and three noise signals to be masked, i.e.,

- Pure tone at 500 Hz;
- White noise filtered between 1 kHz and 2 kHz;
- White noise filtered between 200 Hz and 1 kHz.

The experimental tests have been performed using a sampling rate of 48 kHz. Figure 4, Figure 5, and Figure 6 show the obtained results in terms of frequency response. Figure 4 shows the results obtained with the pure tone at 500 Hz. The pure tone noise to be masked is depicted by the green line, the music sound signal used as masker is represented by the yellow line, and the equalized sound is displayed by the blue line. It can be seen that the sound signal is equalized by the algorithm in order to cover the undesired noise. Figure 5 shows the performance of the system when the white noise from 1 kHz to 2 kHz (green line) must be masked by the sound signal (yellow line). Also in this case, the implemented approach correctly equalizes the sound (blue line). Finally, similar observations can be examined in Figure 6, where a white noise from 200 Hz to 1 kHz (green line) is involved. The yellow line shows the initial masking signal and the blue line is the equalized one, which follows the spectral behavior of the undesired noise.

4.2. Real-World Experiment

The developed system has been installed in a real room. In particular, the DSP board has been connected to a non-traditional sound reproduction system formed by vibrating sound transducers mounted on rigid panels [33], as shown in Figure 7. The transducers are Revolution Acoustics MultiducersTM SSP6 used also in [33] and they have a frequency range of 45Hz-20kHz and an impedance of 8 Ohm. In addition, the response of the panel is

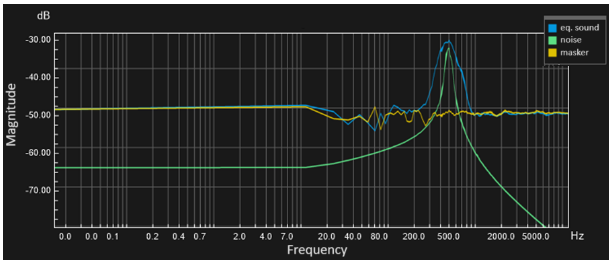

Figure 4: *First simulated test considering a pure tone (green line) to be masked, a sound signal (yellow line) used as masker and sound signal after the equalization procedure (blue line).*

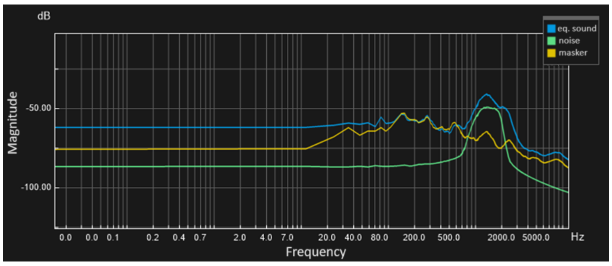

Figure 5: *Second simulated test considering white noise filtered between 1kHz and 2 kHz (green line) to be masked, a sound signal (yellow line) used as masker and sound signal after the equalization procedure (blue line).*

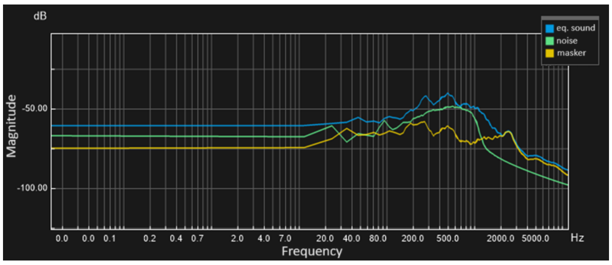

Figure 6: *Third simulated test considering white noise filtered between 200Hz and 1 kHz (green line) to be masked, a sound signal (yellow line) used as masker and sound signal after the equalization procedure (blue line).*

equalized following the procedure of [33] to enhance the sound reproduction and compensating the non-ideal effects introduced by the panel itself. A TV has been mounted in front of the panels while on the opposite side of the TV there is a kitchen. The microphone for the ambient noise analysis has been installed on the roof, as shown in Figure 8. The microphone is an Audio-Technica ES945 and it is located near the listener position, in front of the TV. The ambient noise captured by the microphone comprises the noise generated by the working extractor fan of the kitchen and propagated in the open space room. Most of the power of the captured noise is concentrated in the frequency range from 20 Hz to 3 kHz. Figure 9 shows with the red line the frequency response of the original soundtrack and the equalized sound with the green line. The equalized signal is obtained by applying the equalization curve shown in Figure 2, which includes the equalization of the

Figure 7: *Installation of the vibrating sound transducers behind the wall where the TV is installed.*

Figure 8: *Installation of the microphone in the roof near the listener position.*

panel and the masking procedure. It is evident how the original signal is equalized by the algorithm thus improving the perception of the sound generated by the system.

For the final objective evaluation, the normalized signal-to-noise ratio (NSNR) [30] has been considered. For each frequency bin k, the $\mathrm{NSNR}(k)$ is computed as

$$\mathrm{NSNR}(k) = 10\log_{10}\left\{\frac{P_{\mathrm{S}}(k)}{P_{\mathrm{N}}(k)}\right\}\ \mathrm{dB}, \tag{15}$$

where $P_{\mathrm{S}}(k)$ and $P_{\mathrm{N}}(k)$ are the power spectral density of the signal and the noise, respectively, calculated as

$$P_{\mathrm{X}}(k) = \frac{1}{L}\left|X(k)\right|^2, \tag{16}$$

where L indicates the length of the signal X that could be the masker signal, depicted by S in Eq. (15), or the noise, described by N in Eq. (15). The NSNR represents an objective measure that quantifies how much the environmental noise corrupts the audio signal. High levels of NSNR are preferred implying that the signal is less corrupted by the noise. Figure 10 shows the obtained results: the blue line represents the NSNR calculated considering the signal reproduced by the panel, including the only equalization of the panel, while the red line represents the NSNR calculated considering the enhanced signal after the application of the mask estimated by the presented algorithms. It is evident that good results are achieved in the frequency band of the undesired noise signal (i.e., below 3 kHz), achieving an enhancement of 10 dB, thus confirming the feasibility of the proposed approach. Finally, informal listening tests have been performed involving 5 expert listeners who work in the audio field and are familiar with subjective evaluations. All the listeners have noticed a great improvement in sound reproduction and noise suppression, demonstrating the capability of the system to improve sound perception and also the intelligibility of sounds in this noisy environment.

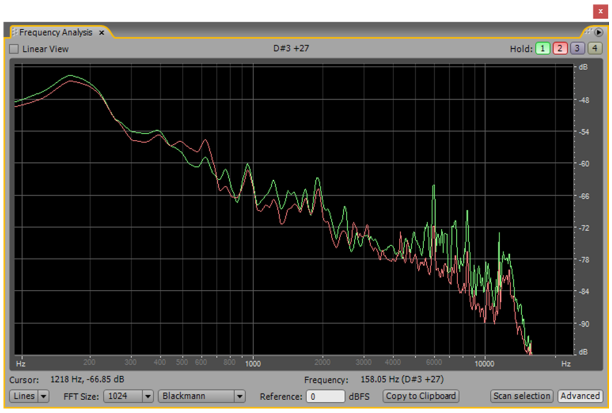

Figure 9: *Frequency response of signal used for the masking procedure (red line) and its equalized version (green line).*

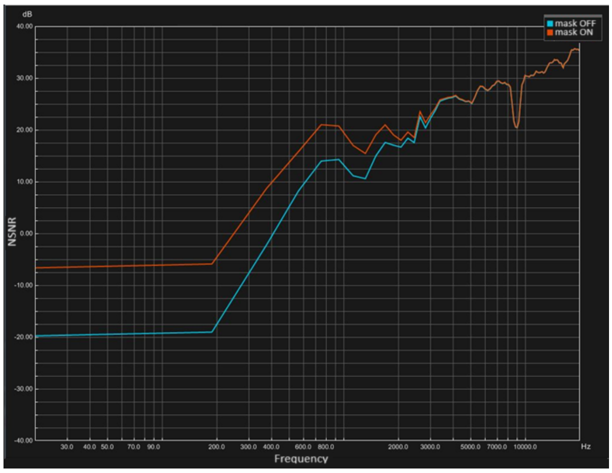

Figure 10: *NSNR behaviour calculated for the real experiment.*

5. CONCLUSIONS

In this paper, a real-time system for audio enhancement in noisy environments has been presented. The system consists of three main algorithms: an adaptive noise estimator, an auditory spectral masking technique, and an FFT-based equalizer. The system has been implemented on a DSP board and tested in a real room using vibrating sound transducers for reproduction. The performance of the proposed system has been evaluated in terms of normalized signal-to-noise ratio, showing a noise reduction of up to 10 dB and proving the method's effectiveness. Future works will be oriented to test the system in more challenging scenario like yacht environment or automotive system and also considering formal listening test with at least 10 listeners.

6. ACKNOWLEDGMENTS

This research has received funding from the project Vitality – Project Code ECS00000041, CUP I33C22001330007 - funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - 'Creation and strengthening of innovation ecosystems,' construction of 'territorial leaders in R&D' – Innovation Ecosystems - Project 'Innovation, digitalization and sustainability for the diffused economy in Central Italy – VITALITY' Call for tender No. 3277 of 30/12/2021, and Concession Decree No. 0001057.23-06-2022 of Italian Ministry of University funded by the European Union – NextGenerationEU. This work has also been supported by the financial program DM MiSE 3370 27 October 2023, project "Epignosis" F/340008/01/X59.

7. REFERENCES

[1] Ayça Şentop Dümen and Konca Şaher, "Noise annoyance during COVID-19 lockdown: A research of public opinion before and during the pandemic," *J. Acoust. Soc. Am.*, vol. 148, no. 6, pp. 3489–3496, Dec. 2020.

[2] Mostafa Mir, Farnad Nasirzadeh, Hannah Bereznicki, Peter Enticott, SangHyun Lee, and Anthony Mills, "Construction noise effects on human health: Evidence from physiological measures," *Sustainable Cities and Society*, vol. 91, pp. 104470, 2023.

[3] Bhan Lam, Kelvin Chee Quan Lim, Kenneth Ooi, Zhen-Ting Ong, Dongyuan Shi, and Woon-Seng Gan, "Anti-noise window: Subjective perception of active noise reduction and effect of informational masking," *Sustainable Cities and Society*, vol. 97, pp. 104763, 2023.

[4] Shuai Zhang, Lijun Zhang, and Dejian Meng, "A dual sampling-rate active noise equalization algorithm for active sound quality control of vehicle interior noise," in *Proceedings of Meetings on Acoustics*. AIP Publishing, 2022, vol. 50.

[5] Haibo Huang, Teik C Lim, Jiuhui Wu, Weiping Ding, and Jian Pang, "Multitarget prediction and optimization of pure electric vehicle tire/road airborne noise sound quality based on a knowledge-and data-driven method," *Mechanical Systems and Signal Processing*, vol. 197, pp. 110361, 2023.

[6] Sen M Kuo and Dennis R Morgan, "Active noise control: a tutorial review," *Proc. of the IEEE*, vol. 87, no. 6, pp. 943–973, Jun. 1999.

[7] Yoshinobu Kajikawa, Woon-Seng Gan, and Sen Maw Kuo, "Recent applications and challenges on active noise control," in *Proc. 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, Sep. 2013, pp. 661–666.

[8] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, USA, second edition, 1991.

[9] John C Burgess, "Active adaptive sound control in a duct: A computer simulation," *J. Acoust. Soc. Am.*, vol. 70, no. 3, pp. 715–726, Sep. 1981.

[10] Yu Tian-ze, Xiao Yan, Luo Xiya, Li Wenyu, Yu Xingbo, and Su Jiaming, "Application of variable step feedback adaptive active control algorithm in underwater noise control," in *Proc. 11th International Conference on Computer Science and Network Technology (ICCSNT)*, Oct. 2023, pp. 315–320.

[11] Stephen J Elliott and Christopher C Boucher, "Interaction between multiple feedforward active control systems," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 521–530, Oct. 1994.

[12] Takumi Miyake, Kenta Iwai, and Yoshinobu Kajikawa, "Head-mounted multi-channel feedforward active noise control system for reducing noise arriving from various directions," *IEEE Access*, vol. 11, pp. 6935–6943, 2023.

[13] Sen M Kuo, Xuan Kong, and Woon S Gan, "Applications of adaptive feedback active noise control system," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 2, pp. 216–220, Mar. 2003.

[14] Eberhard Zwicker and Hugo Fastl, *Psychoacoustics: Facts and models*, vol. 22, Springer Science & Business Media, 2013.

[15] Frederic Marmel, Daniela Cortese, and Karolina Kluk, "The ongoing search for cochlear synaptopathy in humans: Masked thresholds for brief tones in threshold equalizing noise," *Hearing Research*, vol. 392, pp. 107960, 2020.

[16] James D Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 314–323, Feb. 1988.

[17] Brian CJ Moore, Brian R Glasberg, and Thomas Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240, Apr. 1997.

[18] Brian R Glasberg and Brian CJ Moore, "Development and evaluation of a model for predicting the audibility of time-varying sounds in the presence of background sounds," *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 906–918, Oct. 2005.

[19] Joan L Mitchell, "Introduction to digital audio coding and standards," *J. Electron. Imaging*, vol. 13, no. 2, pp. 399, 2004.

[20] Manfred R Schroeder, Bishnu S Atal, and JL Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Am.*, vol. 66, no. 6, pp. 1647–1652, Dec. 1979.

[21] A. Gilloire and V. Turbin, "Using Auditory Properties to Improve the Behaviour of Stereophonic Acoustic Echo Cancellers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, USA, May 1998, vol. 6, pp. 3681–3684.

[22] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation - An Overview of the Fundamental Problem," *IEEE Signal Process. Lett.*, vol. 2, pp. 148–151, Aug. 1995.

[23] Jussi Rämö, Vesa Välimäki, Mikko Alanko, and Miikka Tikander, "Perceptual frequency response simulator for music in noisy environments," in *Proc. Audio Eng. Soc. 45th*

Conference on Applications of Time-Frequency Processing in Audio, Mar. 2012.

[24] Jussi Rämö, Vesa Välimäki, and Miikka Tikander, "Perceptual headphone equalization for mitigation of ambient noise," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 724–728.

[25] Juan Estreder, Gema Piñero, Fabian Aguirre, Maria de Diego, and Alberto Gonzalez, "On perceptual audio equalization for multiple users in presence of ambient noise," in *Proc. IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jul. 2018, pp. 445–449.

[26] Juan Estreder, Gema Piñero, Miguel Ferrer, Maria de Diego, and Alberto Gonzalez, "Perceptual active equalization of multi-frequency noise," in *Proc. 18th International Conference on Signal Processing and Multimedia Applications (SIGMAP)*, Jul. 2021, pp. 39–47.

[27] Juan Estreder, Gema Piñero, Maria de Diego, Jussi Rämö, and Vesa Välimäki, "Improved Aures tonality metric for complex sounds," *Appl. Acoust.*, vol. 204, pp. 109238, 2023.

[28] Markus Christoph, "Noise dependent equalization control," in *Proc. Audio Eng. Soc. 48th International Conference on Automotive Audio*, Sep. 2012.

[29] Janne Riionheimo and Vesa Välimäki, "Parameter estimation of a plucked string synthesis model using a genetic algorithm with perceptual fitness calculation," *EURASIP J. Adv. Signal Process.*, vol. 2003, pp. 1–15, 2003.

[30] Eberhard Hänsler and Gerhard Schmidt, *Speech and audio processing in adverse environments*, Springer Science & Business Media, 2008.

[31] H Schöpp and H Hetze, "A linear-phase 512-band graphic equalizer using the fast-fourier transform," in *Proc. 96th Convention of the Audio Engineering Society*, Feb. 1994.

[32] Panagiotis D Hatziantoniou and John N Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, Apr. 2000.

[33] Stefania Cecchi, Alessandro Terenzi, Francesco Piazza, and Ferruccio Bettarelli, "Applying sound equalization to vibrating sound transducers mounted on rigid panels," in *Proc. 147th Convention of the Audio Engineering Society*, Oct. 2019.