Acoustic Environment Control : Implementation of a Reverberation Enhancement System

Clement S. J. Doire[1], Mike Brookes[1], Patrick A. Naylor[1], Enzo De Sena[2], Toon van Waterschoot[2], and Søren Holdt Jensen[3]

[1]*Electrical and Electronic Engineering Department, Imperial College London, U.K.*

[2]*Department of Electrical Engineering (ESAT-STADIUS/ETC), KU Leuven, Belgium*

[3]*Department of Electronic Systems, Aalborg University, Denmark*

Correspondence should be addressed to Clement S. J. Doire (`clement.doire11@imperial.ac.uk`)

ABSTRACT

Reverberation enhancement systems allow the active control of the acoustic environment. They are subject to instability issues due to acoustic feedback, and are often installed permanently in large halls, sometimes at great cost. In this paper, we explore the possibility of implementing a cost-effective reverberation enhancement system to control the acoustics of typical rooms using a combination of spatial filtering, automatic calibration, adaptive notch filters, howling detection and manual adjustments. The effectiveness of the system is then tested inside a small soundproof booth.

1. INTRODUCTION

Reverberation enhancement systems have been an active research topic for decades [1, 2] as they provide a convenient way to actively control the acoustics of a room. As with any other electroacoustic system using microphones and loudspeakers connected to each other, instability issues arising from acoustic feedback are a major implementation challenge [3]. Early systems used time-invariant techniques, controlling the gain of the electroacoustic forward path [4]. Time-varying systems are also a popular choice as they allow a greater acoustic gain; these include frequency shifting, phase modulation and delay modulation [5]. However, all of these techniques are tailored for expertly and permanently installed reverberation enhancement systems associated with high costs.

In this paper we detail possible solutions for implementing a smaller scale, inexpensive system aiming at actively controlling the acoustic environment without reinforcing the original signal. The system described in this paper consists of multiple reverberation panels each of which operates independently of the others. Therefore, each panel must be able to deal with the acoustic feedback problem by itself, and a user interface must be provided so that the acoustic properties of the room can be changed in real-time. In order to achieve this, each panel uses a 2-microphone array to provide spatial filtering and adaptive notch filters to prevent howling. The software was written in C++ and implemented as a VST plugin, enabling us to process the audio in real-time using a Digital Audio Workstation (DAW) and provide us with standard user interface tools.

The system described in this paper was implemented as part of a live demonstration involving human-robot interaction using Automatic Speech Recognition (ASR). By controlling the reverberation characteristics of the environment, it was possible to determine the effect of reverberation on the ASR performance. Therefore, the main goal was to modify the perceived reverberation for speech sources. For this reason, loudspeakers with a narrow frequency response, i.e. 500 to 5000 Hz were used.

To produce the additional artificial reverberation, the choice was made to convolve the input signal with recorded Room Impulse Responses (RIRs) in order to give the user the possibility to switch between easily identifiable room acoustics. Performing efficient low latency real-time convolution in a VST plugin is challenging as it requires complicated scheduling inside a single

computing thread [6]. Therefore, the real-time convolution engine of the WDL open source C++ library was used [7]. A diagram of the whole system is shown in Fig. 1.

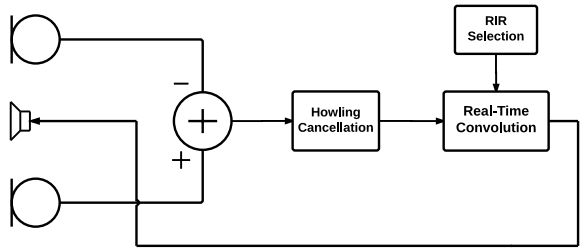

Fig. 1: Diagram of the reverberation enhancement system

The paper is organised as follows. Sec. 2 focuses on the spatial filtering component of the system, while Sec. 3 is concerned with the automatic calibration of the microphones. The notch filter-based howling cancellation system is described in Sec. 4. Sec. 5 presents the system implementation details and results. Finally, Sec. 6 concludes the paper.

2. DIFFERENTIAL MICROPHONE ARRAY

Acoustic feedback occurs when loudspeaker sound is picked up again by a microphone. Even when direct sound transmission from loudspeaker to microphone is avoided, sound is fed back due to reflections against walls and other objects. This generates a closed loop, which can give rise to system instability.

An effective way to reduce the impact of the closed loop is to use spatial filtering, i.e. to place a null at the location of the loudspeaker and a maximum in the direction of the sound source [3]. In order to keep both the cost and computational complexity to a minimum, it was decided to place a first-order differential microphone array symmetrically around the loudspeaker. This means 2 omnidirectional microphones are placed symmetrically on either side of the loudspeaker and set in opposite phase so as to cancel the contribution of the loudspeaker signal. However, this influences the spatial response of the microphone array.

A diagram of the system is presented on Fig. 2 in which the microphones lie on the x-axis with a spacing of d and the loudspeaker is modelled as a point source. This simplification is valid provided that the polar response of the loudspeaker is symmetric with respect to the microphones.

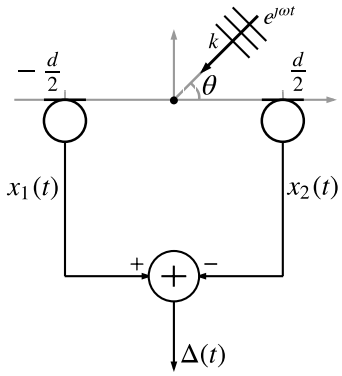

Fig. 2: Diagram of the differential microphone array. Loudspeaker is modelled as a point source at the origin.

Let $x_1(t)$ and $x_2(t)$ be the left and right microphone signals, respectively. Consider a unit-magnitude monochromatic plane sound wave with frequency f propagating in the direction of the wave vector $\mathbf{k} = \frac{2\pi f}{c}[\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi]$ where c is the speed of sound and θ and ϕ are the azimuth and elevation of the sound source direction. For clarity, and without loss of generalisation, we take $\phi = \pi/2$ in the description below. As detailed in [8], the acoustic pressure waveforms at the two microphones are

$$x_1(t) = e^{j2\pi ft} e^{-j\frac{2\pi f}{c}\frac{d}{2}\cos\theta} \tag{1}$$

$$x_2(t) = e^{j2\pi ft} e^{j\frac{2\pi f}{c}\frac{d}{2}\cos\theta}. \tag{2}$$

The difference signal is therefore

$$\Delta(t) = x_1(t) - x_2(t) = -2j\sin\left(\frac{\pi f d}{c}\cos\theta\right) e^{j2\pi ft}. \tag{3}$$

To illustrate the effect of the differential microphone array, Figs. 3 and 4 show polar plots of the $2\sin\left(\frac{\omega d}{2c}\cos\theta\right)$ term from (3) at selected frequencies, f, for microphone separations, d, of 10 and 15 cm respectively.

For low frequencies, $f \ll f_{\lim} = \frac{c}{2d}$, the argument of $\sin\left(\frac{\pi f d}{c}\cos\theta\right)$ in (3) has magnitude $\ll \pi/2$ and we can use a first order Taylor series to obtain the approximation

$$\Delta(t) \simeq -2j\frac{\pi f d}{c} e^{j2\pi ft}\cos\theta. \tag{4}$$

This corresponds to the figure-of-eight pattern visible for low frequencies on the polar plots of Figs. 3 and 4, with

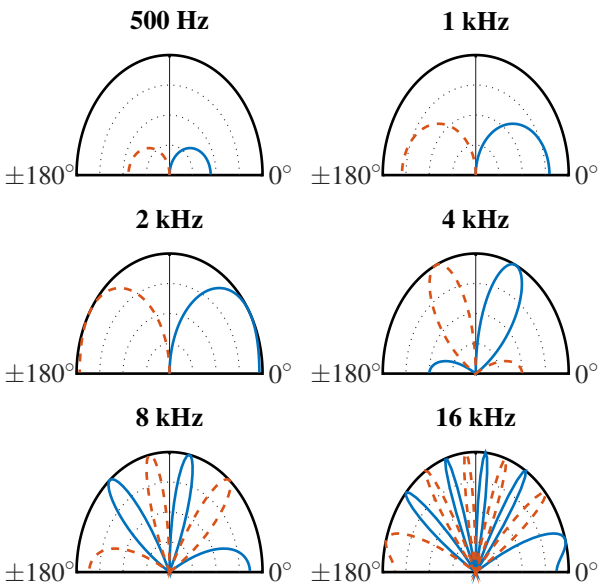

Fig. 3: Polar patterns of the differential microphone array for $d = 10$ cm ($f_{\text{lim}} = 1.7$ kHz). Positive and negative sign are denoted by solid and dashed line respectively.

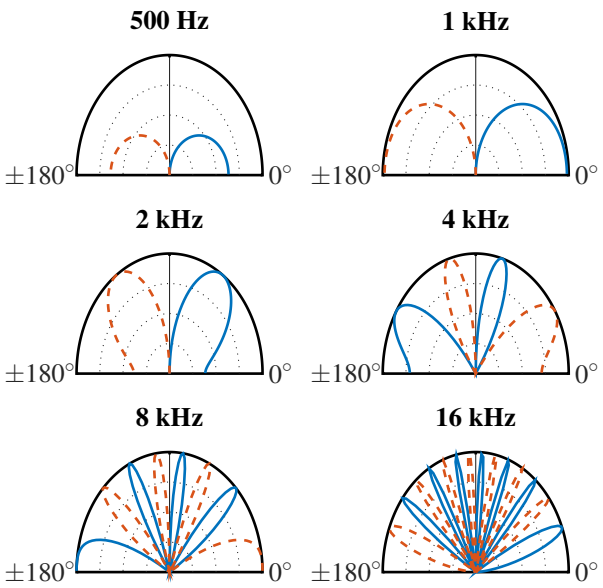

Fig. 4: Polar patterns of the differential microphone array for $d = 15$ cm ($f_{\text{lim}} = 1.13$ kHz). Positive and negative sign are denoted by solid and dashed line respectively.

a single null in the direction $\theta = 90°$. For $d = 10$ cm, we have $f_{\text{lim}} = 1.7$ kHz and for $d = 15$ cm we have $f_{\text{lim}} = 1.13$ kHz. In practice, the beam pattern remains close to a figure-of-eight even for frequencies slightly larger than f_{lim}. However, for $f > 2f_{\text{lim}}$, spatial aliasing causes additional nulls to appear in the polar response.

In the implementation for which this system was designed, the reverberation panels are inside a soundproof booth for interaction with a robot, and the position of the user (i.e. the sound source) corresponds to $0 \leq \theta < 90°$. Hence, in order to avoid unwanted cancellations of some directions of arrival at frequencies of interest, the smallest distance between microphones is preferable. This allows us to have a figure-of-eight polar response at the array for the biggest range of frequencies possible, including frequencies containing speech information.

Another way of looking at the influence of the term $2\sin\left(\frac{\omega}{c}\frac{d}{2}\cos\theta\right)$ on the response of the system is to consider the magnitude response as a function of frequency for different directions of arrival. It creates unwanted coloration of the output signal with a dependence on the angle of incidence θ. At low frequencies, for $f \ll f_{\text{lim}}$, and for all incidence angles, the response of the differential array is proportional to ω. This high-pass coloration could be compensated straightforwardly using a correction filter. However, as we are using a narrowband loudspeaker with a range of 500 Hz to 5000 Hz, the impact of the filter coloration is small and a correction filter was unnecessary.

3. MICROPHONE ARRAY CALIBRATION

The concept of cancelling the feedback path by using a differential microphone array with a loudspeaker in its centre of symmetry is only valid if the microphones are perfectly calibrated and all gains in the signal paths are identical. Since this is not strictly possible in practice, automatic software calibration of the microphone signals is necessary.

Consider the system diagram of figure 5. It is a discretized version of the system presented in section 2, where a delay has been added to the signal path of the left channel, and a calibration Finite Impulse Response (FIR) filter $\mathbf{h}_c$ added to the signal path of the right channel. The fixed delay of D samples was added in order to have an acausal calibration filter to cope with possible misalignments of the microphones and fractional sample delays.

We want to find $\mathbf{h}_c$ so that $x_2(n) * \mathbf{h}_c - x_1(n-D) = 0$ for all sound waves coming from the loudspeaker. To

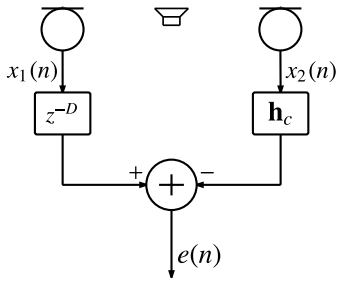

Fig. 5: Diagram of the calibration system. The calibration process aims at estimating filter $\mathbf{h}_c$ using a Normalized Least Mean Squares algorithm.

do so, a calibration button is added to the user interface to the VST plugin. Upon pressing of this button, white noise is forced at the output of the system through the loudspeaker for a period of several seconds, time during which an adaptive Normalized Least Mean Squares (NLMS) algorithm is used to converge to the correct $\mathbf{h}_c$ [9].

Let $\mathbf{h}_c(n) = [h_{c0}(n), h_{c1}(n), ..., h_{cp}(n)]^T$ with p the filter length. Each time a new sample is ready to be processed, we push values into the circular buffer $\mathbf{x}(n) = [x_2(n), x_2(n-1), ..., x_2(n-p+1)]^T$. Initialising $\mathbf{h}_c$ to an all-zero vector at first, we now have the following update equation:

$$\mathbf{h}_c(n+1) = \mathbf{h}_c(n) + \mu \frac{e(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n)} \tag{5}$$

with $e(n) = x_1(n-D) - \mathbf{x}^T(n)\mathbf{h}_c(n)$ and μ the learning rate parameter.

By having p corresponding to a length of a few milliseconds, the calibration should also compensate for immediate first-order reflections coming back to the differential microphone.

4. ADAPTIVE HOWLING CANCELLATION

As the environment in which the reverberation panels are used is constantly changing (e.g. people moving around creating time varying feedback paths through low-order reflection), even using a perfectly calibrated differential microphone array is not enough to prevent howling.

One way of dealing with this would be to use phase modulation, as it should not create audible artefacts when used with a speech source [10, 11]. However, the potential artefacts could be detrimental to the ASR used in conjunction with our system.

As we want minimum latency as well as minimum processing load, the choice of a purely time domain method based on adaptive notch filters was made. Indeed, performing the howling detection in the frequency domain would require additional overhead through the computation of a Fast Fourier Transform. There are many different transfer functions for adaptive notch filters [12, 13]. In the remainder of this section we use the one described in [14] as it always provides 0dB gain away from the notch frequency:

$$N(z) = \frac{1}{2}\left(1 + \frac{r + \alpha z^{-1} + z^{-2}}{1 + \alpha z^{-1} + r z^{-2}}\right) \tag{6}$$

with r determining the elimination bandwidth of the notch filter and α linked with the notch frequency through the relation

$$\alpha = -(1+r)cos\left(2\pi \frac{f_n}{f_s}\right) \tag{7}$$

where f_n is the notch frequency and f_s is the sampling frequency. The notch filter is implemented using the structure described on figure 6.

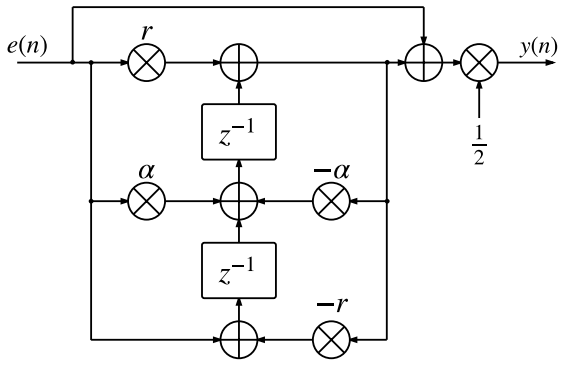

Fig. 6: Structure of the implemented notch filter

Following the adaptive howling cancellation method described in [15], it is possible to use a combination of notch filter and linear predictor to determine the value of α adaptively. The update is given by the following set of equations:

$$\alpha(n) = \frac{1+r}{2} h(n) \tag{8}$$

$$h(n+1) = h(n) - \rho \frac{\mathscr{E}(n)e(n-1)}{E\left[e^2(n-1)\right]} \tag{9}$$

with $\mathscr{E}(n) = e(n) + h(n)e(n-1) + e(n-2)$ and ρ the learning rate parameter.

In order to add a howling detection mechanism to this system, it was decided to use a method similar to [16] and have two regularised adaptive notch filters (RANF) running in parallel. The NLMS update equation of (9) is therefore modified into a leaky-NLMS:

$$h_i(n+1) = h_i(n) - \rho\left(\frac{\mathscr{E}_i(n)e(n-1)}{E\left[e^2(n-1)\right]} + \lambda_i h_i(n)\right) \quad (10)$$

for $i = 1,2$. The regularisation term is negligible when there is howling in the signal, and penalises the estimate of α when howling is not present. We choose $\lambda_1 = 0$ and λ_2 as [16]

$$\lambda_2 = \frac{1}{\rho}\left[\sqrt[2M]{1+\left(\frac{2\pi}{f_s}\Delta f\right)^2} - 1\right] \quad (11)$$

where Δf is the desired deviation of the notch frequency in Hz introduced over M samples by the regularisation term when howling is not present. As $\lambda_1 = 0$, this is equivalent to having a regularized and non-regularized adaptive notch filters running in parallel. We will therefore refer to their outputs and notch frequencies as $\{\tilde{y}(n), \tilde{f}_n\}$ and $\{y(n), f_n\}$ respectively.

The computed notch frequencies f_n and $\tilde{f}_n$ are stored in L samples buffers. The means of these two buffers, m_1 and m_2 are then computed, and the likelihood of howling being present is calculated using the following:

$$\mathscr{L}(m_1, m_2) = e^{-\frac{(m_1-m_2)^2}{2\sigma^2}} \quad (12)$$

with σ a fixed variance which can be viewed as a threshold on the difference $m_1 - m_2$ that is permitted before howling is considered a possibility.

The posterior probability of howling at discrete time n, $p(H|n)$, is then obtained using a sequential bayesian treatment, i.e. using the previous posterior probability as the new prior probability:

$$p(H|n) = \frac{p(H|n-1)\mathscr{L}(m_1,m_2)}{p(H|n-1)\mathscr{L}(m_1,m_2) + (1-p(H|n-1))(1-\mathscr{L}(m_1,m_2))} \quad (13)$$

The posterior probability can then be used as a soft weight on the output of the non-regularised notch filter, as the leaky version of the NLMS update might introduce some bias once convergence has been reached [17]. This gives:

$$s(n) = p(H|n)y(n) + (1 - p(H|n))\,e(n) \quad (14)$$

As it is possible for howling to occur at multiple frequencies simultaneously, a filterbank decomposition of the input signal is performed. Pairs of adaptive notch filters are thus working in parallel in each sub-band. We assume that howling can be present, in the worst case, in each octave. Therefore, in order to obtain an octave-band filterbank, eight 4th-order Butterworth filters were designed and implemented as cascaded bi-quad filters.

For each pair of adaptive notch filters running in parallel, the update equation (10) aimed at determining the notch frequency as well as the presence of howling is done using the sub-band input value $e_k(n)$. However, since the notch filters have unity gain away from their notch frequency, we can process the fullband signal as in (14) with the non-regularized sub-band notch filters placed in series, as shown on Fig. 7. We denote the output of the whole system by $s(n)$. Subscript indices for the system input $e_k(n)$ and probability of howling $p_k(H|n)$ indicate octave-band k.

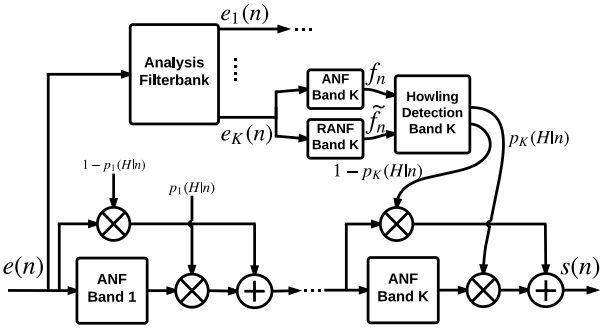

Fig. 7: Diagram of the notch filter based howling cancellation system

5. IMPLEMENTATION AND RESULTS

As mentioned in Sec. 1, this reverberation enhancement system was optimised and built to work in conjunction with a human-robot interaction demo at the Royal Society's Summer Science Exhibition which took place in London early July 2015 [18]. The demo was run inside a soundproof booth of inner dimensions $172 \times 110 \times 220$ cm, and providing us with an average 35 dB attenuation of the ambient noise present at the exhibition. Upon entering the booth, visitors were presented with the structure modelled on Fig. 8, with push buttons to select the

acoustic environment, a robot standing on the platform, and 3 reverberation panels surrounding the robot.

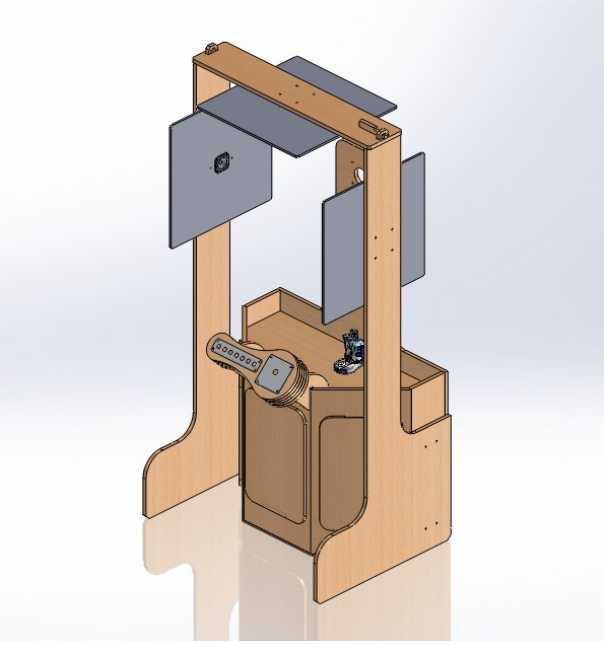

Fig. 8: 3D model of the human-robot interaction platform and reverberation panels installation

The aim of the demo was to illustrate to the visitors the influence of different room acoustics on the accuracy of ASR systems and how this can change how effective the human-machine interaction is. Users could also experiment with the acoustics of familiar environments such as a bathroom, classroom, concert hall or theatre. At any given time, the number of members of the public present inside the booth varied between 2 and 5.

The RIRs used for the real-time convolution process were a combination of impulse responses from [19] and personally recorded ones. Reverberation times (T_{60}) ranged from 0.1 s to 3 s.

The three panels were placed so that all the loudspeakers were located in the null direction of all three differential microphone arrays. It turned out, once the building was complete, that the material used for the panels on which the loudspeaker and differential microphones were mounted was acoustically reflective. With the left and right panels being parallel and only separated by one metre, a standing wave was created which started the howling process more easily than with the prototype panel used for development. For this reason, sound absorbing foam was placed on the left and right panels themselves, however placing the panels further from each other and in a random pattern would have avoided the need for sound absorbing foam.

We chose to run the system at a sampling rate of 32 kHz as the loudspeakers are narrowband and so as to avoid computational overload. For each panel, the distance between the pair of microphones was set to $d = 12$ cm. During the calibration process, because of their proximity and in order to take all three sources into account at each pair of receivers, the calibration process was run at the exact same time for the three panels, for a duration of 4 seconds and using $\mu = 0.23$.

For the howling cancellation part of the system, the notch elimination bandwidth was set using $r = 0.93$, and $\rho = 0.051$ was used in the NLMS update of the notch frequencies. The parameters λ_2 and σ were chosen to be different in each octave band, typically with small deviations of the leaky update and a small variance for the howling detection at low frequencies, and high deviations and variance at high frequencies. As an example, σ was set to 7 Hz for the lowest octave band and 80 Hz for the highest one. The parameter λ_2 was always set to a few Hz deviation over 30 samples.

Finally, the output gain of the amplifier controlling the loudness of the panels was manually set so that howling did not occur for any of the 6 acoustic conditions. The total latency introduced by the reverberation enhancement system was around 12 ms.

In order to test the effectiveness of the system, we recorded the impulse response inside the soundproof booth with the reverberation panels on, for all 6 acoustic environments offered to visitors. A Fostex 6301BX loudspeaker was placed at approximately mouth position (1.6 m height) and 2 omnidirectional DPA 4060 microphones were used: one placed at the centre of the platform where the robot was standing, at 1 m height, and one close to the entrance of the booth, behind the loudspeaker, at 1.7 m height, as shown on Fig. 9 & 11.

Impulse responses were recorded using a swept-sine technique. The squared impulse responses in the log domain for the microphone at the back of the booth are shown on Fig. 10 after smoothing was applied using a moving average filter of length 9 ms. The noise floor was at -120 dB.

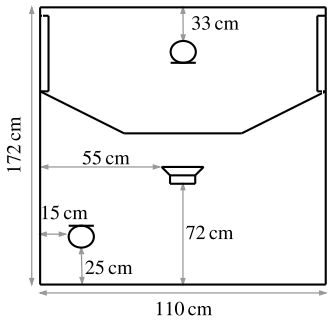

Fig. 9: Diagram of the measurement setup

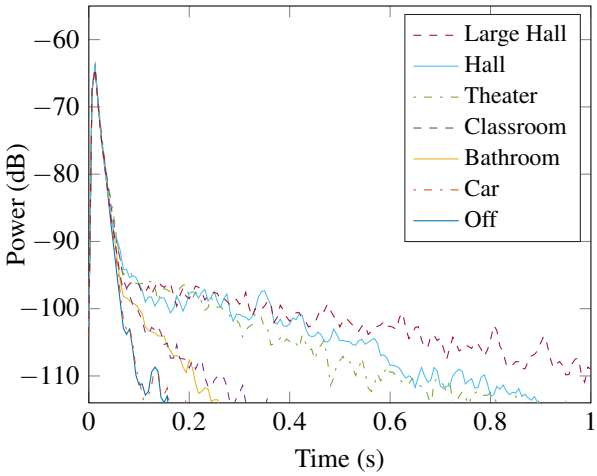

Fig. 10: Measured impulse responses, squared and smoothed, for all reverberant conditions offered to visitors.

Even though the latency of the system is around 12 ms, it is only possible to observe noticeable changes on the measured impulse response after 50 ms or so. This means that the natural impulse response inside the booth has much more energy in the first 50 ms, the early reflections pattern is slightly changed but remains mostly the natural one. This creates the "room in room" effect noticeable on all the plots of Fig. 10. However, one can clearly see the increased energy and length of the reverberation tail of the selected acoustic environments. Impulse responses recorded at the centre of the human-robot interaction structure, on the other side of the booth, are almost identical and therefore are not shown here.

In practice, as the room-in-room effect only occurred during the first 50 ms, it was barely noticeable. It did however have a perceivable impact on the frequency response of the artificial reverberation, as all the acoustic environments tended to be perceived as having the same coloration.



Fig. 11: Photograph of the measurement setup inside the soundproof booth

6. CONCLUSION

In this paper, we demonstrated the possibility of implementing a cost-effective reverberation enhancement system to actively control the acoustic environment. Using a combination of spatial filtering, automatic calibration, adaptive notch filters and manual adjustments, it is possible to implement a system of independently working panels to modify the amount of late reverberation present in a given room. The effects of the system were demonstrated in a small soundproof booth and can be generalised for larger rooms, provided more panels are available in order to build a sufficient sound field.

7. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the EU 7th Framework Programme (FP7/2007-2013) under grant agreement ITN-GA-2012-316969. The authors would like to thank Ray Thompson for his help in designing and building the human-robot interaction structure and panels.

8. REFERENCES

[1] R. W. Guelke and A. D. Broadhurst, "Reverberation time control by direct feedback," Acustica, Vol. 24, No. 1, 1971, pp. 33–41.

[2] Peter U. Svensson, "Computer simulations of periodically time-varying filters for acoustic feedback control," Journal of the Audio Engineering Society, Vol. 43, No. 9, 1995, pp. 667–677.

[3] Toon van Waterschoot and Marc Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," Proceedings of the IEEE, Vol. 99, No. 2, February 2011, pp. 288–327.

[4] Asbjörn Krokstad, "Electroacoustic means of controlling auditorium acoustics," Applied Acoustics, Vol. 24, 1988, pp. 275–288.

[5] Johan L. Nielsen and Peter U. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," Journal of the Acoustical Society of America, Vol. 106, No. 1, 1999, pp. 240–254.

[6] Jeffrey Hurchalla, "A time distributed FFT for efficient low latency convolution," 129th Audio Engineering Society Convention, San Francisco, 2010.

[7] Cockos Incorporated, WDL Open Source library, "http://www.cockos.com/wdl/" Online, accessed July 2015.

[8] Enzo de Sena, Hüseyin Hacihabiboğlu and Zoran Cvetković, "On the Design and Implementation of Higher Order Differential Microphones," IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 1, 2012, pp. 162–174.

[9] Simon S. Haykin, "Adaptive Filter Theory," Pearson Education, 2008.

[10] M. R. Schroeder, "Improvement of feedback stability of public address systems by frequency shifting", Journal of the Audio Engineering Society, Vol. 10, No. 2, 1962, pp. 108–109.

[11] Peter U. Svensson, "On reverberation enhancement in auditoria," Ph.D. dissertation, Dept. Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden, 1994.

[12] Soo-Chang Pei and Chien-Cheng Tseng, "IIR Multiple Notch Filter Design Based on Allpass Filter," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Vol. 44, No. 2, 1997, pp. 133–136.

[13] Arye Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP- 33, No. 4, 1985, pp. 983–996.

[14] Philip A. Regalia, "An Improved Lattice-Based Adaptive IIR Notch Filter," IEEE Transactions on Signal Processing, Vol. 39, No. 9, Sep. 1991, pp. 2124–2128.

[15] Akira Sogami, Yosuke Sugiura, Arata Kawamura, and Youji Iiguni, "An Adaptive Howling Canceller Using 2-Tap Linear Predictor," SciRes. Circuits and Systems, Vol. 4, No. 1, Jan. 2013, pp. 9–13.

[16] Pepe Gil-Cacho, Too van Waterschoot, Marc Moonen and Søren Holdt Jensen, "Regularized Adaptive Notch Filters for Acoustic Howling Suppression," 17th European Signal Processing Conference (EUSIPCO), Glasgow, Scotland, August 2009, pp. 2574–2577.

[17] Max Kamenetsky and Bernard Widrow, "A Variable Leaky LMS Adaptive Algorithm," Conference Record of the Thirty-Eighth Asilomar Conference on Signals Systems and Computers, 2004, pp. 125–128.

[18] Royal Society's Summer Science Exhibition 2015, Sound Interactions Exhibit, "http://sse.royalsociety.org/2015/sound-interactions/" Online, accessed July 2015.

[19] Samplicity, M7 Impulse Response Library v1.1, "http://www.samplicity.com/bricasti-m7-impulse-responses/" Online, accessed July 2015.