TEMPERATURE ROBUST ACTIVE-COMPENSATED SOUND FIELD REPRODUCTION USING IMPULSE RESPONSE SHAPING

Terence Betlehem

Seoul R&D Centre
Samsung Electronics
betl.terence@samsung.com

Lakshmi Krishnan and Paul Teal

Victoria University of Wellington
(lakshmi.krishnan, paul.teal)@ecs.vuw.ac.nz

ABSTRACT

Spatial audio may be performed in rooms using active-compensated sound field reproduction (AC-SFR), to cancel the reverberant reflections. Reproduction of a sound field may be performed using pre-calculated loudspeaker filters. However the robustness can be poor due to the temperature-induced perturbation of the room impulse responses (RIR). To compensate for reverberation, filters with significant non-causal components are designed. It is known that the requirements on the loudspeaker filters for room equalization may be reduced by applying the method of impulse response shaping (IRS). In this paper, we show that actively compensating paths with a long propagation distance are more heavily influenced by temperature than short paths. We hence impose a shaping penalty on the non-causal components of designed loudspeaker filters, to reduce the error caused by a change in temperature. We obtain AC-SFR designs that are more robust to changes in temperature.

I. INTRODUCTION

In sound field reproduction, a sound field is reproduced over an extended listening area using a number of loudspeakers [1]. In a reverberant room, accurate signal reproduction is possible using active-compensated sound field reproduction (AC-SFR), where room impulse responses are actively compensated. Yet such approaches are known to have poor robustness to temperature changes.

In AC-SFR, early reflections may be used to help reconstruct a sound field. A small number of higher order loudspeakers [2], that are capable of directing the sound [3–6], can work together to create a virtual source between loudspeakers. For example, directing the sound towards a wall may be used to create a virtual source in the direction of a first order reflection. However to actively compensate for the room reverberation requires pre-exciting the room. The designed filters may possess non-causal components which require adding significant modeling delay.

One way AC-SFR may be performed is by pre-calibrating a loudspeaker system. This has advantages over adaptive filter approaches [7–13] in which the loudspeaker filters must be constantly updated in real time using an in-situ array of microphones. Yet pre-calibration may suffer from robustness problems, due to temperature-induced perturbation of room impulse responses (RIRs). This may result in noticeable artifacts. Particularly disturbing is pre-reverberation [14], which is heard as a noticeable buildup of sound energy at the listener before the actual virtual source sound arrives.

The goal in this paper is to devise an AC-SFR approach with reduced dependence on non-causal components. One such approach is to simply perform a minimum-phase equalization. This approach however obliterates the phase response, which may degrade the fidelity of the acoustic image. A more powerful approach is the robust quadratic control approach of Brännmark et. al. [15–17]. Alternatively, planarity control [18] could be applied to control the direction in which the sound energy arrives at the listener, though the phase information is not controlled. Other methods have been proposed to reduce or eliminate the non-causal components [19] but the perceptual effects on spatial image quality are not clear.

The fine structure of an RIR is influenced by the speed of sound which is a function of air temperature and humidity and affects the wavefront propagation time. The perturbation in RIRs causes error in the room equalization, which results in pre-reverberation in AC-SFR systems [14, 19]. RIRs have been compensated by measuring temperature changes with thermometers and digitally dilating the RIRs [20, 21]. Yet perhaps at best, this approach is capable of correcting for the *average* temperature of the room, but not local fluctuations.

To improve robustness, the harmful effects of a temperature perturbation may be reduced using regularization, especially if applied to reduce the energy of the most deleterious components in the filters. Previously, regularization had been applied to SFR [22] and impulse response shaping [23, 24]. Regularization is relevant for systems perturbed by additive noise-like processes, but is not necessarily the best method of producing a design robust to temperature fluctuations.

Impulse response shaping (IRS) is an approach for directly manipulating the nett impulse responses (NIRs) to points within the listening zone resulting from the combined action of the loudspeakers. In this approach, the pre-echo, direct part and reverberant reflections of the NIRs may be individually shaped. IRS has been applied to equalizing a single loudspeaker-microphone channel [25, 26], for crosstalk cancelation in systems consisting of multiple microphones and loudspeakers [27, 28] and sound field reproduction [29, 30]. It results in short efficient filters that may reduce the sensitivity to RIR error [23, 26, 31].

In this paper, we show that a long propagation distance is more heavily influenced by temperature than a short one. We hence identify the components of the designed loudspeaker filters responsible for generating the pre-reverberation in the NIRs. Utilizing an IRS-based approach to SFR to impose a penalty on these components, we show that the error caused by a change in temperature is able to be reduced, thereby obtaining an AC-SFR design that is more robust to a changes in temperature.

II. TEMPERATURE ROBUSTNESS

A perturbation in room temperature causes a variation in the speed of sound. Consider an RIR between a loudspeaker and a microphone in a room. It can be well-modelled with the finite impulse response $h(t)$ of N propagation paths resulting from wall reflections:

$$h(t) = \sum_{n=1}^{N} a_n \delta(t - \frac{d_n}{c}), \tag{1}$$

where a_n is the accumulated reflection coefficient for a wall reflection arriving at the microphone via a particular path, d_n is the corresponding propagation distance and $c = c(T)$ is the speed of sound in air which is a function of temperature T. Taking the Fourier transform of $h(t)$ leads to the room transfer function (RTF):

$$H(f) = \sum_{n=1}^{N} a_n e^{i2\pi f d_n/c}, \tag{2}$$

for frequency f. When the speed of sound changes, the time $\tau_n = d_n/c$ taken to travel distance d_n changes. Let the original and perturbed speeds of sound be c_0 and c respectively, corresponding to temperatures T_0 and T. The change in propagation time is:

$$\Delta\tau_n = d_n \left(\frac{1}{c} - \frac{1}{c_0}\right). \tag{3}$$

For a temperature increase (decrease), the propagation time is shortened (lengthened) and $\Delta\tau_n$ is negative (positive). The change propagation time corresponds to a change in path length:

$$\Delta d_n = c\Delta\tau_n = d_n \left(1 - \frac{c}{c_0}\right). \tag{4}$$

Measured in terms of the number of wavelengths:

$$\frac{\Delta d_n}{\lambda} = f d_n \left(\frac{1}{c} - \frac{1}{c_0}\right). \tag{5}$$

The change in the length of the propagation path, when expressed in number of wavelengths, is hence proportional to both frequency and the original path length.

When equalizing a room impulse response, performance can be degraded significantly if the path length changes by only a tenth of a wavelength [32] or 36° phase shift. One method for canceling a reverberant reflection in a room is destructive interference. Here a reflection is canceled by anti-phasing the signal with a signal of equal amplitude generated by a loudspeaker located along the path of the propagation.

Consider canceling a single tone $\sin\omega t$ of frequency $\omega = 2\pi f$ at a single point by producing an anti-phase signal $\sin(\omega t + \pi)$. For a temperature-induced change Δt in a propagation delay of the original signal, the resultant is

$$\sin(\omega t + \omega\Delta t) + \sin(\omega t + \pi) = 2\sin\left(\frac{\omega\Delta t}{2}\right)\cos\left(\omega t - \frac{\omega\Delta t}{2}\right)$$

which has amplitude $2\sin\omega\Delta t/2 \approx \omega\Delta t$ for small Δt. For a phase shift in the original tone of 36°, the tone is only suppressed by 4.2 dB. A 20 dB tone suppression requires the phase shift to be 6° or less (i.e. $\lambda/60$). Rearranging (5), for a 20 dB suppression of an acoustic path:

$$f = \frac{1}{60 d_n \left|c^{-1} - c_0^{-1}\right|}. \tag{6}$$

For only a 1 m/s change in the speed of sound from $c_0 = 342$ m/s to $c = 343$ m/s and propagation distance of $d_n = 10$ m, which in a typical living room includes all first and some second order wall reflections, the 20 dB suppression of such a wall reflection is only possible up to 195 Hz. The change in path length due to a small change in the speed of sound is hence a significant factor in the performance of acoustic equalization systems.

Temperature is the most significant physical characteristic influencing the speed of sound in air. A prominent model of the speed of sound originating from the ideal gas law, for dry air is [20]:

$$c = 20.03\sqrt{K + T} \tag{7}$$

where T is temperature in degrees Celsius and $K = 273.15$ is the temperature in degrees Kelvin at 0°C. Substituting (7) into (5)

$$\frac{\Delta d_n}{\lambda} = \frac{f d_n}{20.03}\left(\frac{1}{\sqrt{K+T}} - \frac{1}{\sqrt{K+T_0}}\right) \tag{8}$$

where $T = T_0 + \Delta T$ and T_0 is the original temperature. Taking a first order Taylor approximation in ΔT, the change in the path length is

$$\frac{\Delta d_n}{\lambda} \approx -\kappa f d_n \Delta T, \tag{9}$$

where $\kappa = (20.03)^2/2c_0^3$. Temperature-induced perturbations to the path length are hence more extreme at higher frequencies and for longer propagation distances.

This temperature sensitivity poses a problem for any algorithm seeking to equalize the sound at one or more points in a room. In AC-SFR, sound is reflected from walls to achieve the reproduction of a spatial sound field. Here, certain sound propagations must be canceled over the spatial extent of the zone of reproduction. Each propagation could either be the sound coming directly from a loudspeaker, or a wall reflection. The cancelation is achieved by anti-phasing the sound propagation using a combination of other propagations.

In a room, higher order reflections are more distant from the listener and, by (9), are more susceptible to phase changes induced by a change in temperature. When performing AC-SFR then, a weak higher order reflection may be used to anti-phase a sound propagation. However, since the phase of a higher order reflection is more heavily influenced by temperature, the cancelation is less robust. It is preferable in AC-SFR then to penalize the use of higher order reflections.

We now seek a method of optimizing the designed filters to reduce the utilization of higher order reflections. We propose doing this by adding a penalty function which weights different parts of the designed filters. The greatest penalty is applied to the earliest parts of the filters, which utilize the least robust reflections. The penalty is relaxed for the latter parts of the filters.

III. SOUND FIELD REPRODUCTION

The task of sound field reproduction is to create an intended sound field around the listener using several surrounding loudspeakers. The most important type of reproductive sound field is that of a virtual sound source.

To compensate for room reverberation, the RIRs from each loudspeaker to each point in the sound reproduction region may be determined using a microphone array. To keep room simulations simple, in Section IV a pair of concentric arrays of microphones suspended in free space will be used.

III-A. Inverse Filter Solution

To reproduce a sound field over a region of space, one can either match the sound pressure at a number of points, or match the spatial modes. We do the former, leaving a modal formulation for future work. A solution is presented in terms of a matrix of RIRs.

For comparison with the IRS method, the pressure matching problem is formulated in the time domain. A sound field produced using L loudspeakers is measured at Q microphones. This MIMO problem is written as:

$$\mathbf{Ch} = \mathbf{d} \tag{10}$$

where $\mathbf{C}$ is the block Toeplitz convolution matrix of the RIRs between the microphones and loudspeakers, $[\mathbf{C}]_{ql} = \mathrm{convmatrix}(\mathbf{c}_{ql})$ is the (q,l)th submatrix of $\mathbf{C}$, the vector $\mathbf{c}_{ql}$ is the RIR from the lth loudspeaker to qth microphone. Vector $\mathbf{h}$ is a concatenated vector of loudspeaker filters, and vector $\mathbf{d}$ is the concatenated vector of desired impulse responses to be obtained at each of the microphones, defined $(\mathbf{d})_q = \mathbf{d}_q$ where vector $\mathbf{d}_q$ is the desired impulse response at microphone q.

The least squares objective, with an added $\mathcal{L}_2$ penalty weighting on the designed loudspeaker filters $\mathbf{h}$ is:

$$\mathcal{J}_{\mathrm{Inv}}(\mathbf{h}) = \|\mathbf{Ch} - \mathbf{d}\|^2 + \mu\|\mathbf{W}_{\mathrm{h}}\mathbf{h}\|^2$$

where $\mathbf{W}_{\mathrm{h}} = \mathrm{Diag}(\mathbf{w}_{\mathrm{h}})$, η is a Tikhonov regularization parameter to limit the energy in $\mathbf{h}$ and $\mathbf{w}_{\mathrm{h}}$ is a weighting vector on the loudspeaker filters. Minimizing the objective leads to the solution

$$\mathbf{h}_{\mathrm{Inv}} = (\mathbf{C}^T\mathbf{C} + \eta\mathbf{W}_{\mathrm{h}}^T\mathbf{W}_{\mathrm{h}})^{-1}\mathbf{C}^T\mathbf{d}. \tag{11}$$

The matrix $\mathbf{C}$ in this time-domain formulation is quite large. If each RIR is N_{c} taps long and each filter is N_{h} taps long, then each resultant impulse response is $N_{\mathrm{r}} \triangleq N_{\mathrm{c}} + N_{\mathrm{h}} - 1$. Matrix $\mathbf{C}$ then has $L\,N_{\mathrm{h}}$ columns and QN_{r} rows where, for typical audio sampling rates, N_{r} and N_{c} are typically in the order of thousands of taps. If $\mathbf{W}_{\mathrm{h}}$ is set to a scalar multiple of the identity matrix, this inverse problem reduces to a typical least squares problem which may be solved efficiently using a frequency-domain approach.

The virtual sound source is chosen here as an acoustic monopole at a position $\mathbf{y}$. The desired sound pressure at each point in the sound field can be expressed as follows. The sound pressure at each microphone q at time t is

$$p_{\mathrm{d}}(t;\mathbf{x}_q) = \int_{-\infty}^{\infty} \frac{e^{i2\pi f(t-d_q/c)}}{d_q}\,df = \frac{1}{d_q}\delta(t - d_q/c),$$

where $d_q \triangleq \|\mathbf{x}_q - \mathbf{y}\|$ and $\mathbf{x}_q$ is the position of microphone q. When signals are generated digitally, the inverse discrete Fourier transform (DFT) is used to construct the desired sound field at each point:

$$p_{\mathrm{d}}(mT_{\mathrm{s}};\mathbf{x}_q) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} \frac{1}{d_q} e^{i2\pi\left(\frac{nm}{N} - \frac{f_n d_q}{c}\right)} = \frac{1}{d_q}\frac{\sin\frac{\pi}{2}(m-\tau_q)}{\sin\frac{\pi N}{2}(m-\tau_q)}$$

where $F_s = 1/T_{\mathrm{s}}$ is the sampling rate, $f_n = nF_{\mathrm{s}}/N$ is the analog frequency of the DFT at index n and $\tau_q = F_{\mathrm{s}}d_q/c$ is the propagation delay of the virtual source to microphone q expressed in numbers of samples. The desired response at each microphone q is the vector $(\mathbf{d}_q)_m = p_{\mathrm{d}}(mT_{\mathrm{s}};\mathbf{x}_q)$.

III-B. Impulse Response Shaping

IRS is a more versatile approach than inverse filtering, enabling explicit control of pre-echo, early reflections and late reverberation. The objective function for the IRS solution to the pressure matching problem of sound reproduction is:

$$\mathcal{J}_{\mathrm{IRS}} = \|\mathbf{W}_{\mathrm{d}}(\mathbf{Ch} - \mathbf{d})\|^2 + \eta\|\mathbf{W}_{\mathrm{h}}\mathbf{h}\|^2 \tag{12}$$

where $\mathbf{W}_{\mathrm{d}} = \mathrm{diag}(\mathbf{w}_{\mathrm{d}})$ and $\mathbf{w}_{\mathrm{d}}$ is a weighting vector on the desired microphone responses that influence the manner in which they are shaped. The essence of IRS is to choose $\mathbf{w}_{\mathrm{d}}$ and $\mathbf{w}_{\mathrm{h}}$ so as to penalize the undesirable components of the resultant RIRs such as pre-echo and late reverberation whilst maintaining potentially desirable components such as early reflections.

The choice of the shaping weight used is based on pyscho-acoustic considerations such as the audibility of reverberation [25, 27]. The loudspeakers filters will be designed here with a modeling delay of N_{model} taps, a *don't care* duration of N_{early} taps for the region of early reflections and the direct part, followed by late reverberant duration of N_{late} taps. In the pressure matching problem, the modeling delay is a function of microphone position.

The filter weight vector $\mathbf{w}_{\mathrm{h}}$ provides shaping of the loudspeaker filters, to avoid placing high levels of energy in the early parts of the filters, which were shown above to cause pre-reverberation. The expression for the loudspeaker weight vector $\mathbf{w}_{\mathrm{h}}$ is the concatenation of L vectors $\mathbf{w}_{\mathrm{h}}^{(\ell)}$ each defined $(\mathbf{w}_{\mathrm{h}}^{(\ell)})_n = 1 + \alpha w_{\mathrm{h}}(n)$ where α is the relative weighting of the pre-reverberation penalty to the normal Tikhonov penalty, and $w_{\mathrm{h}}(n)$ is an exponentially-decaying penalty function for the early parts of the loudspeaker filters, but applies no weight past time index n_0:

$$w_{\mathrm{h}}(n) = \frac{e^{-(n-n_0)/\tau} - 1}{e^{n_0/\tau} - 1},$$

for $0 < n \leq n_0$ and is zero otherwise. Assume loudspeakers are approximately equidistant from the reproduction region. Define the average propagation time N_{prop} from the loudspeakers to the centre of the listening region. The choice of n_0 and τ are based upon considerations regarding pre-reverberation:

- The n_0 taps corresponds to the duration of time used to exploit room reflections. The components of the loudspeaker filters contributing to pre-reverberation occurs until a time equal to $N_{\mathrm{model}} + 1$ taps when the direct response arrives, minus the direct propagation time of N_{prop} taps. We apply weight $w_{\mathrm{h}}(n)$ to penalize the loudspeaker filter up to tap $n_0 = N_{\mathrm{model}} - N_{\mathrm{prop}}$.
- Parameter τ determines how aggressively the early energy in loudspeaker filters is penalized in relation to the late energy.

The solution minimizing $\mathcal{J}_{\mathrm{IRS}}$ is the same as the inverse filter solution in (11), after setting $\mathbf{C} \to \mathbf{W}_{\mathrm{d}}\mathbf{C}$ and $\mathbf{d} \to \mathbf{W}_{\mathrm{d}}\mathbf{d}$. Similar to the inverse filter solution, solving the IRS problem analytically requires construction of a large matrix which cannot be stored in the memory of most personal computers for channels of practical length.

A set of methods that does not require construction of matrix $\mathbf{C}$ are gradient methods. Here, the concatenated vector of filters is updated iteratively, e.g. through:

$$\mathbf{h}_{i+1} = \mathbf{h}_i - \mu\frac{\partial\mathcal{J}}{\partial\mathbf{h}_i},$$

where μ is the stepsize parameter and the gradient of the objective function $\partial\mathcal{J}/\partial\mathbf{h}$ is:

$$\begin{aligned}\frac{1}{2}\frac{\partial\mathcal{J}_{\mathrm{IRS}}}{\partial\mathbf{h}} &= (\mathbf{C}^T\mathbf{W}_{\mathrm{d}}^T\mathbf{W}_{\mathrm{d}}\mathbf{C} + \eta\mathbf{W}_{\mathrm{h}}^T\mathbf{W}_{\mathrm{h}})\mathbf{h} - \mathbf{C}^T\mathbf{W}_{\mathrm{d}}^T\mathbf{W}_{\mathrm{d}}\mathbf{d}\\ &= \mathbf{C}^T\mathbf{W}_{\mathrm{d}}^2(\mathbf{r} - \mathbf{d}) + \eta\mathbf{W}_{\mathrm{h}}^2\mathbf{h},\end{aligned} \tag{13}$$

where $\mathbf{r} = \mathbf{Ch}$ is the concatenated vector of resultant RIRs. We use a conjugate gradient method with a line search based on polynomial interpolation [33] for faster convergence.

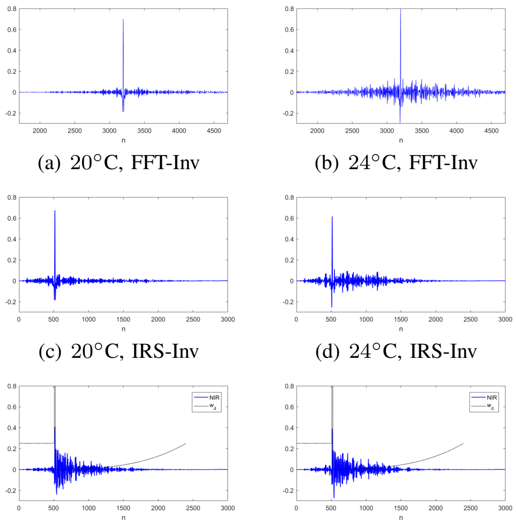
The parameters in (13) are computed using computationally-efficient means. The multiplication of block convolution matrix $\mathbf{C}$ with partitioned filter vector $\mathbf{h}$ is replaced by summing the convolutions of $\mathbf{c}_{ql}$ and $\mathbf{h}_l$. Each convolution is computed using the fast convolution method. Similarly $\mathbf{C}^T\mathbf{x}$ is computed by summing the autocorrelations of channels $\mathbf{c}_{ql}$ and vector partitions $\mathbf{x}_q$ of $\mathbf{x}$. Each autocorrelation is computed using the FFT.

IV. RESULTS

We evaluate the performance of inverse and IRS approaches, for the task of reproducing a virtual source over a spherical reproduction region located in the centre of the loudspeaker array. We explore the case that filters are designed for a temperature of 20°C are applied at a perturbed temperature of 24°C. To keep the computational requirements modest, the simulation is performed using a 16 kHz sampling rate.

The microphone array used to characterize the response between each loudspeaker and the reproduction region consists of a pair of concentric arrays: an inner array of radius 20 cm and an outer array of radius 30 cm, each using 100 sampling points arranged in a Fliege geometry [34]. The loudspeaker array consists of 40 monopole drivers arranged into five circular arrays of eight drivers each having a diameter of 30 cm. Each circular array was located 2m from the centre of listening region. The circular array configuration was intended to emulate a 2D higher order source capable of producing up to 3rd order directivity patterns [6]. Circular arrays are able to control the sound emitted in the horizontal plane (i.e. in azimuth) but are unable to control out-of-plane reverberation.

Reveberant RIRs are simulated by the image-source method, in a room of dimensions $6.4 \times 5 \times 4$ m with a wall reflection



(a) 20°C, FFT-Inv (b) 24°C, FFT-Inv

(c) 20°C, IRS-Inv (d) 24°C, IRS-Inv

(e) 20°C, IRS (f) 24°C, IRS

Fig. 1. Resultant RIR at microphone for (a) (b) FFT-Inv, (c) (d) IRS-Inv and (e) (f) an impulse response shaping (IRS) approaches, at original temperature 20°C and perturbed temperature 24°C.

coefficient of 0.7. The spherical reproduction region is located at $(-0.4, 0.2, -0.5)$ m relative to the centre of the room. The virtual source steering angle is 36°, positioned mid-way between circular array loudspeakers, at the same distance from the listening region.

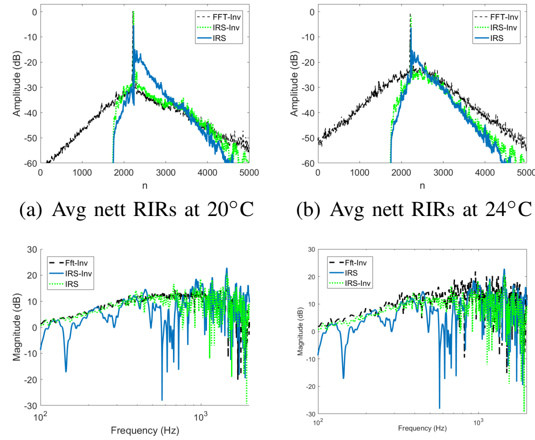
Three types of filters were compared: inverse filters designed using the FFT method, impulse response shaping filters designed with 30 ms of pre-echo for which the reshaping weight is shown for a microphone in Fig. 1(e) and (f), and inverse filters designed using the impulse response shaping approach with the same parameters but a flat weight vector $\mathbf{w}_\mathrm{d} = \mathbf{1}$. These three approaches are labelled *FFT-Inv*, *IRS* and *IRS-Inv* in the figures.

To reduce computational complexity in filter design, the RIRs were pre-processed by removing the direct component propagation time of 5.5 msec and applying a bandpass filter with a passband of 200 Hz to 2 kHz. RIRs were then truncated to $N_\mathrm{c} = 1200$ taps (75 msec). The *IRS* and *IRS-Inv* filter parameters were set to $N_\mathrm{h} = 1200$ taps, $\alpha = 3$ and $\tau = 12$. For inverse filtering, a longer filter was found to be effective; a length of 3600 taps was chosen. The inverse filter was designed with an 16x frequency oversampling factor, followed by applying a Hamming window in the time domain. All filters had 0.25% of their energy clipped off the start and end of their filter impulse responses. The IRS filters were computed with 240 iterations of the algorithm in [33].

The nett RIRs to a microphone, located at the top of the inner spherical array at $(0, 0, 0.2)$ m, are shown in Fig. 1 for the three filtering approaches at the designed and perturbed temperatures. The temperature change was seen to produce both pre-reveberation and post-reverberation in the NIRs. For a temperature perturbation, the IRS methods were seen to produce much less pre-reverberation. The post-reverberation levels of the *IRS* filter appeared unchanged by perturbation whilst for *IRS-Inv* the post-reverberation was significantly increased.

The time-aligned squared RIRs were averaged over the 200 microphones in Fig. 2(a) and (b) for original and perturbed temperatures. In the case of a temperature perturbation, the inverse approach is seen to have an early onset and slow rise in pre-reverberation, whilst the IRS methods are seen to have a much later, sharper rise in pre-reverberation.

In Table I the direct-to-prereverberation energy ratio (DPRR) is



(a) Avg nett RIRs at 20°C (b) Avg nett RIRs at 24°C

(c) Nett freq. resps at 20°C (d) Nett freq. resps at 24°C

Fig. 2. Aligned averaged energies of nett RIRs for (a) design and (b) perturbed room temperatures. Nett frequency responses at the top microphone, for (c) design and (d) perturbed temperatures.

Method	DPRR (dB)			
	Original		Perturbed	
	Inner	Outer	Inner	Outer
FFT-Inv	13.8	10.5	0.45	-0.71
IRS-Inv	14.2	11.8	11.4	9.4
IRS	13.4	11.2	9.3	7.9

Table I. Average direct-to-prereverberation-energy-ratio (DPRR) for impulse response shaping and inverse approaches at original and perturbed room temperatures.

calculated for the perturbed temperature. *IRS* and *IRS-Inv* are seen to produce 8.9 dB and 10.7 dB less pre-reverberant energy than *FFT-Inv* over the microphone locations. Some attenuation of the direct part of NIRs occurs for *IRS*, reducing DPRR performance slightly below that of *IRS-Inv*. The reason of this is as yet unclear.

The improved robustness is likely due to the shorter modeling delay chosen for the IRS designs. Shorter modeling delays were previously observed to be more robust in [35]. For the inverse filter, the modeling delay was approximately 175 msec but for *IRS* filters it was 30 msec. This way, the IRS filter approaches only utilize sound propagations up to a distance of 10 m. The inverse filter however attempts to use propagations up to 60 m away which is less temperature robust.

The different perturbed levels of post-reverberation in the IRS and inverse filtering methods can be explained intuitively. The *IRS* filter was designed to exploit room reflections for creating the phantom source, without de-reverberating the NIRs. The *Inv* and *IRS-Inv* filters in addition seek to de-reverberate the NIRs. This is arguably less robust than simply exploiting several low-order reflections to reproduce the phantom source.

V. CONCLUSION

Temperature perturbations were shown to significantly degrade the performance of AC-SFR. Impulse response shaping was more robust to a temperature perturbation when using a short modeling delay. For sound field reproduction, impulse response shaping has been shown as a flexible approach that can control the extent to which higher order reflections are exploited and the amount of reverberation that is removed.

VI. REFERENCES

[1] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[2] M. A. Poletti and T. Betlehem, "Higher order loudspeakers for improved sound reproduction in rooms," in *Proc. AES 52nd International Conference*, September 2013.

[3] T. Betlehem, C. Anderson, and M. A. Poletti, "A directional loudspeaker array for surround sound in reverberant rooms," in *Proc. Int. Conf. Acoust.(ICA'10)*, Sydney, 2010, pp. 1–6.

[4] M. A. Poletti, F. M. Fazi, and P. A. Nelson, "Sound reproduction systems using variable-directivity loudspeakers," *J. Acoust. Soc. Amer.*, vol. 129, pp. 1429–1438, 2011.

[5] M. A. Poletti and T. D. Abhayapala, "Spatial sound reproduction systems using higher order loudspeakers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 57–60.

[6] T. Betlehem and M.A. Poletti, "Two dimensional sound field reproduction using higher order sources to exploit room reflections," *J. Acoust. Soc. Amer.*, vol. 135, no. 4, pp. 1820–1833, 2014.

[7] S. Spors, H. Buchner, and R. Rabenstein, "Efficient active listening room compensation for wave field synthesis," in *Proc. 116th Convention of the Audio Engineering Society*, Berlin, May 2004.

[8] S. Spors, M. Renk, and R. Rabenstein, "Limiting effects of active room compensation using wave field synthesis," in *Proc. 118th Convention of the Audio Engineering Society*, Barcelona, May 2005.

[9] S. Petrausch, S. Spors, and R. Rabenstein, "Simulation and visualization of room compensation for wave field synthesis with the functional transformation method," in *Proc. 119th Convention of the Audio Engineering Society*, 2005.

[10] P.-A. Gauthier and A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound reproduction: theory," *J. Acoust. Soc. Amer.*, vol. 119, no. 5, pp. 2721–2737, 2006.

[11] P.-A. Gauthier and A. Berry, "Adaptive wave field synthesis for sound field reproduction: theory, experiments, and future perspectives," *Journal of the Audio Eng. Soc.*, vol. 55, no. 12, pp. 1107–1124, 2007.

[12] P.-A. Gauthier and A. Berry, "Adaptive wave field synthesis for active sound field reproduction: experimental results," *J. Acoust. Soc. Amer.*, vol. 123, no. 4, pp. 1991–2002, 2008.

[13] M. Schneider and W. Kellermann, "A wave-domain model for acoustic MIMO systems with reduced complexity," in *Proc. IEEE HSCMA*, Edinburgh, UK, May 2011, pp. 133 – 138.

[14] L. Bianch, F. Antonacci, A. Canclini, A. Sarti, and S. Tubaro, "A psychoacoustic-based analysis of the impact of pre-echoes and post-echoes in soundfield rendering applications," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Aachen, Sept. 2012.

[15] L.-J. Brännmark, "Robust audio precompensation with probabilistic modeling of transfer function variability," in *Proc. IEEE Workshop on Applications of Audio and Acoustics*, New Paltz, 2009, pp. 193–196.

[16] L.-J. Brännmark, A. Bahne, and A. Ahlén, "Improved loudspeaker-room equalization using multiple loudspeakers and MIMO feedforward control," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, 2012, pp. 237–240.

[17] L-J. Brännmark, A Bahne, and A. Ahlén, "Compensation of loudspeaker-room responses in a robust MIMO control framework," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 21, no. 6, pp. 1201 – 1215, 2013.

[18] P. Coleman, P. Jackson, M. Olik, and J.A. Pederson, "Optimizing the planarity of sound zones," in *Proc. 52nd Audio Engineering Society International Conference*, Guildford, Sep. 2013, pp. 1–10.

[19] S. Carini, A.and Cecchi and L. Romoli, "Multipoint room response equalization with group delay compensation," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aug. 2010.

[20] T. Hikichi and F. Itakura, "Time variation of room acoustic transfer functions and its effects on a multi-microphone dereverberation approach," in *Proc. Workshop on Microphone Arrays: Theory, Design and Application*, October 1994.

[21] Y. Yai, S. Miyabe, H. Saruwatari, K. Shikano, and Y. Tatekura, "Rapid compensation of temperature fluctuation effect for multichannel sound field reproduction system," *IEICE Trans. Fundamentals*, vol. E91-A, no. 6, pp. 1329–1336, 2008.

[22] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 20, no. 8, pp. 2388–2392, 2012.

[23] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Room impulse response shaping based on estimates of room impulse responses," in *DAGA2008*, Dresden, March 2008.

[24] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 21, no. 9, pp. 1879–1890, 2013.

[25] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening reshaping with infinity- and p-norm optimization," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 249–259, 2010.

[26] J.O. Jungmann, R. Mazur, and A. Mertins, "Perturbation of room impulse responses and its application in robust listening room compensation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, May 2013.

[27] J. O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. Mertins, "Combined acoustic MIMO channel crosstalk cancellation and room impulse response reshaping," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1829–1842, 2012.

[28] T. Betlehem, P. Teal, and Y. Hioka, "Efficient crosstalk canceler design with impulse response shortening filters," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, March 2012, pp. 393–396.

[29] T. Mei and A. Mertins, "On the robustness of room impulse response reshaping," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aug. 2010.

[30] H. Morgenstern and B. Rafaely, "Spatial reverberation and dereverberation using an acoustic multiple-input multiple-output system," *Journal of the Audio Eng. Soc.*, vol. 65, no. 1/2, pp. 42–55, 2017.

[31] L. Krishnan, P. D. Teal, and T. Betlehem, "A robust sparse approach to acoustic impulse response shaping," in *Proc. IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, April 2015, pp. 738–742.

[32] B. Radlovic, R.C. Williamson, and R.A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Trans. Audio, Speech and Language Processing*, vol. 8, pp. 311–319, 2000.

[33] C.E. Rasmussen, "Optimize," accessed 16-Feb-2018, http://learning.eng.cam.ac.uk/carl/code/minimize/.

[34] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding curbature formulae," *IMA J. Numer. Anal.*, vol. 19, pp. 317–334, 1999.

[35] T. Hikichi, M. Delcroix, and M. Miyoshi, "On robust inverse filter design for room transfer function fluctuations," in *14th European Signal Processing Conference*, Florence, Sep. 2006.