ACTIVE SPEECH CONTROL USING WAVE-DOMAIN PROCESSING WITH A LINEAR WALL OF DIPOLE SECONDARY SOURCES

Jacob Donley[⋆]*, Christian Ritz*[⋆] *and W. Bastiaan Kleijn*[†]

[⋆] School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Australia
[†] School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

ABSTRACT

In this paper, we investigate the effects of compensating for wave-domain filtering delay in an active speech control system. An active control system utilising wave-domain processed basis functions is evaluated for a linear array of dipole secondary sources. The target control soundfield is matched in a least squares sense using orthogonal wavefields to a predicted future target soundfield. Filtering is implemented using a block-based short-time signal processing approach which induces an inherent delay. We present an autoregressive method for predictively compensating for the filter delay. An approach to block-length choice that maximises the soundfield control is proposed for a trade-off between soundfield reproduction accuracy and prediction accuracy. Results show that block-length choice has a significant effect on the active suppression of speech.

***Index Terms*—** spatial audio, personal sound, active noise control, noise barrier, delay compensation, speech emission control.

1. INTRODUCTION

Personal sound [1] has been a topic of great interest to researchers in recent years. Spatial regions of controlled sound can be created using loudspeaker arrays and superposition of soundwaves can be used to actively control sound over space [2]. Active Noise Control (ANC) is a technique that allows secondary sources in electro-acoustic systems to reproduce destructive soundfields thus reducing energy levels of primary soundfields. The resultant suppressed soundfields have been successfully employed in several applications, including noise-cancelling headphones [3] and ANC in vehicle cabins [4, 5, 6]. Offices, libraries, teleconferencing rooms, restaurants and cafes may also benefit from ANC over broad spatial areas where physical partitions could be replaced with an active loudspeaker array.

ANC systems typically comprise a reference signal and/or error signal which are either fed forward and/or backward, respectively, to an algorithm for generating loudspeaker signals [2]. Hybrid systems exist that incorporate both feedforward and feedback techniques [7, 8]. Least Mean Squares (LMS) and Filtered-x LMS (FxLMS) control methods work by adaptively minimising the error signal in a least squares sense [9, 10]. Multichannel systems with numerous microphones inside, or near, the control space often use adaptive algorithms to minimise the error over the region [10, 11].

More recent techniques have been shown to be more accurate by measuring acoustic pressures on boundaries and using the Kirchhoff-Helmholtz integral to determine the soundfield [12, 13, 14]. Sampling the boundary that encloses the space, with microphones, allows the target soundfield to be estimated in the wave-domain. This extends the multipoint method by synthesising the entire spatial area and minimising the error over large spaces [13, 14].

In order to perform wave-domain analysis it is necessary to transform received signals into the (temporal) frequency domain where basis functions are a function of the wavenumber and spatial locations [12, 15]. This transformation induces a delay where numerous samples are required to analyse the signal with high resolution in the frequency domain. Adaptive algorithms overcome this issue by automatically compensating for any errors received at the error microphones [9, 13, 14]. In scenarios where microphones are not placed inside the control region, it is necessary to account for delay by other means. Linear prediction with pitch repetition has been shown to be viable for active speech cancellation with short predictions, up to $2\,\mathrm{ms}$, and at discrete points in a space [16]. However, the predictions do not predict a regular speech frame of length around $16\,\mathrm{ms}$ and cancellation occurs only in the vicinity of the control points.

The active control of sound over a linear array has been envisioned [17] using interconnected control units consisting of a microphone, directional loudspeaker and processing modules. However, the interconnection and modules do not model the received signals on the boundary in the wave-domain and perform only a phase inversion which is less robust to soundfield variation. Linear arrays [18] have also been investigated for improvement of noise barriers [19, 20] which aim to reduce diffraction of sound over a physical barrier by minimising the pressure at points in space, usually modelled on a plane spanning height and width. The use of linear arrays, without a physical barrier, for control over large spatial areas using recently advanced wave-domain processing is explored in this work.

As a baseline study, we analyse the delay caused by transforming reference ANC signals to the wave-domain using a block-based signal processing approach. We propose an autoregressive transform-delay compensator in conjunction with an inverse filter that together produce a virtual source soundfield used in wavefield decomposition to minimise energy residual of a control soundfield. Through analysis of the soundfield suppression we show that an optimal block-length can be chosen for active speech control using wave-domain filtering without error microphones in the control region. The optimal block-length is used in a simulated acoustic environment with dipole secondary sources in a linear array. Acting as an active wall, we show that the optimal block-length, along with the dipole sources, provide significant cancellation of traversing speech waves with minimal reproduction towards the primary source.

A description of the error minimised control soundfield synthesis using basis wavefields is given in section 2. An explanation of dipole modelled soundfield reproduction using synthesised loudspeaker weights is given in section 3. The short-time block-based signal processing approach with autoregressive and geometric delay compensation is presented in section 4 with results, analysis, discussion and conclusions in sections 5 and 6.

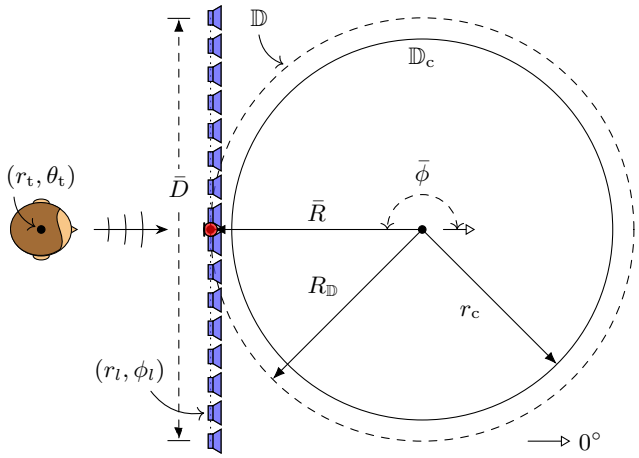

Fig. 1. Active control layout for a linear dipole array (blue) directed to the right. The microphone (red) is used to predict the unwanted speech source crossing the array.

2. WAVE-DOMAIN SOUNDFIELD SUPPRESSION

This section derives an expression for loudspeaker weights which reproduce a soundfield that minimises the residual energy over a control region, $\mathbb{D}_\mathrm{c}$. The active control layout and wave-domain solution to minimise residual energy are described.

2.1. Active Control Layout and Definitions

The proposed system using a linear dipole array is shown in Fig. 1 where the loudspeakers form an active wall between a talker and target quiet zone. The reproduction region for the soundfield, $\mathbb{D}$, with spatial sampling points $\mathbf{x} \in \mathbb{D}$, has a radius of $R_\mathbb{D}$ and contains a control subregion, $\mathbb{D}_\mathrm{c} \subseteq \mathbb{D}$, of radius r_c. The centre of the loudspeaker array is located at angle $\bar{\phi}$ and distance $\bar{R}$. The length of the loudspeaker array is $\bar{D}$ and is designed to reproduce a soundfield for a virtual point source located at $\mathbf{v}$. In this work we refer to the external source that is to be controlled as the talker with location $\mathbf{t} \equiv \mathbf{v} \equiv (r_\mathrm{t}, \theta_\mathrm{t})$. We assume $\mathbf{t}$ is known, or can be reliably estimated with multiple microphones, thus a single reference microphone suffices and is placed at the centre of the loudspeaker array with location $\mathbf{z} \equiv \left(\bar{R}, \bar{\phi}\right)$. Loudspeaker locations are $\mathbf{l}_l \equiv (r_l, \phi_l)$ for $l \in [\![\bar{L}]\!]$ where $\bar{L}$ is the number of loudspeakers, $k = 2\pi f/c$ is the wavenumber and c is the speed of sound in air. The euclidean norm is denoted using $\|\cdot\|$, $i = \sqrt{-1}$ and sets of indices are $[\![A]\!] \triangleq \{x : x \in \mathbb{N}_0, x < A\}$.

2.2. Soundfield Control Technique

The goal is to find coefficients for a set of basis functions that minimise the residual energy of a control soundfield, $S^\mathrm{c}(\mathbf{x}; k)$, and an arbitrary talker soundfield, $S^\mathrm{t}(\mathbf{x}; k)$. A simple solution is to perform an orthogonalisation on a set of plane-wave basis functions that produces a well-conditioned triangular matrix and a set of orthogonal basis functions. Expansion coefficients for the orthogonal basis functions can be easily solved with an inner product.

Any arbitrary soundfield can be completely defined by an orthogonal set of solutions of the *Helmholtz* equation [21]. An arbitrary 2D soundfield function that satisfies the wave equation, such as $S^\mathrm{c}(\mathbf{x}; k) : \mathbb{D} \times \mathbb{R} \to \mathbb{C}$, can be written as

$$S^\mathrm{c}(\mathbf{x}; k) = \sum_{g \in [\![G]\!]} E_{g,m} F_g(\mathbf{x}; k), \tag{1}$$

where $\{F_g\}_{g \in [\![G]\!]}$ is the set of orthogonal basis functions, $m \in [\![N]\!]$ are N frequency indices, the expansion coefficients for a particular frequency are $E_{g,m}$ and G is the number of basis functions [22].

Solving the inner product $E_{g,m} = \langle S^\mathrm{t}(\mathbf{x}; k), F_g(\mathbf{x}; k) \rangle$ yields the $E_{g,m}$ that minimise

$$\min_{E_{g \in [\![G]\!], m \in [\![N]\!]}} \| \sum_g E_{g,m} F_g(\mathbf{x}; k) + S^\mathrm{t}(\mathbf{x}; k) \|^2, \tag{2}$$

where $\|X\|^2 = \langle X, X \rangle$. The set of orthogonal basis functions, $\{F_g\}_{g \in [\![G]\!]}$, can be found by implementing an orthogonalisation on a set of planewaves, $P_h(\mathbf{x}; k) = e^{ik\mathbf{x} \cdot \rho_\mathbf{h}}$, where $\rho_\mathbf{h} \equiv (1, \rho_h)$, $\rho_h = (h-1)\Delta\rho$ and $\Delta\rho = 2\pi/G$. A Gram-Schmidt process gives the orthogonalised basis functions, which results in [22]

$$F_g(\mathbf{x}; k) = \sum_{h \in [\![G]\!]} \mathbf{R}_{hg,m} P_h(\mathbf{x}; k), \tag{3}$$

such that $\langle F_i(\mathbf{x}; k), F_j(\mathbf{x}; k) \rangle = \delta_{ij}$, where $\mathbf{R}_{hg}$ is the (h, g)th element of the lower triangular matrix, $\mathbf{R}$. Substituting (3) in (1), yields

$$S^\mathrm{c}(\mathbf{x}; k) = \sum_{h \in [\![G]\!]} \mathcal{Q}_{m,h} P_h(\mathbf{x}; k), \tag{4}$$

where $\mathcal{Q}_{h,m} = \sum_{g \in [\![G]\!]} E_{g,m} \mathbf{R}_{hg,m}$ are the plane-wave coefficients used to construct an approximation of the control soundfield.

3. LOUDSPEAKER WEIGHTS

In this section, the loudspeaker signals needed for soundfield reproduction with monopole and dipole sources are described.

3.1. Monopole Secondary Source Weights

To reproduce $S^\mathrm{c}(\mathbf{x}; k)$ with minimal error to $S^\mathrm{t}(\mathbf{x}; k)$, loudspeaker weights are found in the (temporal) frequency domain [23, 24, 25]

$$Q_l(k) = \frac{2\Delta\phi_\mathrm{s}}{i\pi} \sum_{\bar{m}=-\bar{M}}^{\bar{M}} \sum_{h \in [\![G]\!]} \frac{i^{\bar{m}} e^{i\bar{m}(\phi_l - \rho_h)}}{H_{\bar{m}}^{(1)}(r_l k)} \mathcal{Q}_{h,m}, \tag{5}$$

where $\Delta\phi_\mathrm{s} = 2\tan^{-1}(\bar{D}/2\bar{R})/\bar{L}$ approximates angular spacing of $\mathbf{l}_l$ for a linear array, $H_\nu^{(1)}(\cdot)$ is a νth-order Hankel function of the first kind and $\bar{M} = \lceil kR_\mathbb{D} \rceil$ is the modal truncation length [24]. However, monopole sources produce acoustic energy in all directions which may be undesirable as it would present an artificial echo towards $\mathbf{t}$.

3.2. Dipole Secondary Source Weights

To reproduce a soundfield with reduced acoustic energy presented towards the talker, dipole sources are modelled to reproduce predominantly over $\mathbb{D}$. The loudspeakers at $\mathbf{l}_l$ with weights $Q_l(k)$ are split into two point sources at $\mathbf{l}_{l,s}$ for $s \in [\![2]\!]$ with weights $Q_{l,s}(k)$. The dipole source pair locations are given by

$$\mathbf{l}_{l,s} = \mathbf{l}_l + (\ddot{d}/2, \bar{\phi} - s\pi), \tag{6}$$

where $\ddot{d}$ is the distance between the dipole point sources. The objective of each dipole source pair is to reproduce a wave which constructs in the direction $(1, \bar{\phi} - \pi)$ from $\mathbf{l}_l$ and de-constructs in the direction $(1, \bar{\phi})$ from $\mathbf{l}_l$ whilst maintaining the same amplitude and phase as a monopole source in the constructive direction. This can be accomplished by phase shifting and amplitude panning the monopole loudspeaker weights with the following [21, 26]

$$Q_{l,s}(k) \triangleq Q_l(k) \frac{e^{i(-1)^s (k\ddot{d} - \pi)/2}}{2k\ddot{d}}, \tag{7}$$

where as $\ddot{d}$ becomes small, $\mathbf{l}_{l,s}$ approach ideal dipole sources.

4. SHORT-TIME SIGNAL PROCESSING

In order to reproduce a control soundfield, a time-domain control signal is filtered using $Q_{l,s}(k)$ in the (temporal) frequency domain and inverse transformed back to the time-domain to yield the set of loudspeaker signals. Here, a block based approach is used. This section investigates the inherent time delay that is induced during the filtering process due to the wave-domain transformation used to compute the loudspeaker weights of (7).

4.1. Block Processing

An input signal, $v(n)$, broken into blocks (frames) using an analysis windowing function, $w(n)$, of length M, results in an ath windowed frame:

$$\widetilde{v}_a(n) \triangleq v(n+aR)w(n), \tag{8}$$

where $n \in \mathbb{Z}$ is the sample number in time, $a \in \mathbb{Z}$ is the frame index and $R \leq M$ is the step size in samples. The ath frame is transformed to the frequency domain to give the ath spectral frame as $\widetilde{V}_a(k_m) = \sum_{n\in[\![N]\!]} \widetilde{v}_a(n) e^{-icnk_m/2\dot{f}}$, where $k_m \triangleq 2\pi \dot{f} m/cN$ and the frame is oversampled with $N \geq M + L - 1$ for a filter length L.

Each spectral frame is filtered using $Q_{l,s}(k)$ from (7) up to the maximum frequency, $\dot{f}$, and inverse transformed to the time-domain

$$\widetilde{q}_{a,l,s}(n) = \Re\left\{ \frac{1}{N} \sum_{m\in[\![N]\!]} Q_{l,s}(k_m) \widetilde{V}_a(k_m) e^{icnk_m/2\dot{f}} \right\}, \tag{9}$$

$\forall n \in [\![N]\!]$, where $\Re\{\cdot\}$ returns the real part of its argument, after which a synthesis window, $w(n)$, equivalent to the analysis window, is applied to yield the weighted output

$$q^w_{a,l,s}(n) = \widetilde{q}_{a,l,s}(n - aR)w(n - aR). \tag{10}$$

The weighted output, $q^w_{a,l,s}(n)$, is added to the accumulated output signal, $q_{l,s}(n)$, for each dipole source. The analysis and synthesis windows are chosen so that $\sum_{a\in\mathbb{Z}} w(n-aR)^2 = 1$, $\forall n \in \mathbb{Z}$.

4.2. Autoregression Parameter Estimation

The soundfield filtering process induces a delay of M samples to build the current ath frame, $\widetilde{v}_a(n)$, from (8), essential for accurate reproduction. To perform active control, it is necessary to find R future samples of the accumulated $q_{l,s}(n)$ that estimate $v(n)$.

Forecasting the input signal's future values can be accomplished using an autoregressive (AR) linear predictive filter. Assuming the signal is unknown after the current time, n, the AR parameters, $\widehat{a}_j$, are estimated using $B > \mathcal{P}$ known past samples with

$$\epsilon(n + \grave{b} + 1) = v(n + \grave{b} + 1) + \sum_{j\in[\![\mathcal{P}]\!]} \widehat{a}_j v(n + \grave{b} - j), \tag{11}$$

$\forall \grave{b} \in \mathcal{B}$, where $\mathcal{B} = \{-B, \ldots, \mathcal{P}-1\}$, $\{\epsilon(n+\grave{b}+1)\}_{\grave{b}\in\mathcal{B}}$ are prediction errors, the predictor order is $\mathcal{P}$ and $j \in [\![\mathcal{P}]\!]$ are the coefficient indices. Stable AR coefficients, $\widehat{a}_j$, can be estimated using the *autocorrelation method* [27, 28] (equivalent to the *Yule-Walker method*) by approximating the minimisation of the expectation of $|\epsilon(n+\grave{b}+1)|^2, \forall \grave{b} \in \mathbb{Z}$ where, prior to minimisation, $v(n+\grave{b}+1)$ is windowed with $\bar{w}(\grave{b})$, assuming $\{\bar{w}(\grave{b})\}_{\grave{b}\notin\{-B,\ldots,-1\}} = 0$, to give $\bar{v}(\grave{b})$. Multiplying (11) by $v(n+\grave{b}-\breve{b}), \breve{b} \in [\![\mathcal{P}]\!]$ and taking the expectation gives the Yule-Walker (YW) equations, $\sum_{j\in[\![\mathcal{P}]\!]} r_{\breve{b}-j}\widehat{a}_j = -r_{\breve{b}}$. We estimate the jth autocorrelation, r_j, as $\widehat{r}_j \triangleq B^{-1}\sum_{\grave{b}=j}^{-1} \bar{v}(\grave{b})\bar{v}(\grave{b}-j)$. The YW equations can be written in matrix form as $\widehat{\mathbf{R}}\widehat{\mathbf{a}} = -\widehat{\mathbf{r}}$ where $\widehat{\mathbf{a}} = [\widehat{a}_0, \ldots, \widehat{a}_{\mathcal{P}-1}]^T$, $\widehat{\mathbf{r}} = [\widehat{r}_0, \ldots, \widehat{r}_{\mathcal{P}-1}]^T$ and the estimated autocorrelation matrix, $\widehat{\mathbf{R}}$, has a Toeplitz structure allowing for an efficient solution.

4.3. Filter-Delay Compensation

Once the $\widehat{a}_j$ are estimated following section 4.2, $v(n)$ can be extrapolated by

$$v(n + \acute{b} + 1) = -\sum_{j\in[\![\mathcal{P}]\!]} \widehat{a}_j v(n + \acute{b} - j), \quad \forall \acute{b} \in [\![\widehat{M}]\!] \tag{12}$$

where $\{v(n+\acute{b}+1)\}_{\acute{b}\in[\![\widehat{M}]\!]}$ are $\widehat{M}$ future estimates of $v(n)$. From (8), $\widetilde{v}_a(n)$ is an estimated future windowed frame when $\widehat{M} \geq M$. The estimated $\widetilde{v}_a(n)$ and partially estimated $\{\widetilde{v}_{a-\grave{a}-1}(n)\}_{\grave{a}\in[\![\frac{M}{R}-1]\!]}$ are transformed, filtered, inverse transformed and windowed through (9) and (10). Adding $q^w_{a,l,s}(n)$ to the previous frames obtains R future estimated samples for the output loudspeaker signals, $q_{l,s}(n)$. The procedures of section 4.2 and section 4.3 are repeated every R samples, including the estimation of $\widehat{a}_j$.

4.4. Geometric-Delay Compensation

The control soundfield modelling requires a virtual source location and signal. In this work, the reference microphone recording, $z(n)$, located at $\mathbf{z}$, is an attenuated and time delayed version of $v(n)$. Under the assumption of free-space and that the talker location, $\mathbf{t}$, is known, or can be reliably estimated, the talker signal is found by

$$v(n) = \Re\left\{ \frac{1}{N} \sum_{m\in[\![N]\!]} \frac{4\left\{\sum_{n\in[N]} z(n) e^{-icnk_m/2\dot{f}}\right\}}{iH_0^{(1)}\left(k_m \left\|\mathbf{v}-\mathbf{z}\right\|\right)} e^{icnk_m/2\dot{f}} \right\}, \tag{13}$$

where $z(n)$ is inverse filtered in the frequency domain with N sufficiently large compared to the time-delay. For the purpose of soundfield control, $\mathbf{t} \equiv \mathbf{v}$ and $v(n)$ is also the virtual source signal.

4.5. Loudspeaker Signals and Reproduction

Upon receiving the reference signal, $z(n)$, the final dipole loudspeaker signals, $q_{l,s}(n)$, are produced by firstly compensating for the geometric-delay with (13) to obtain $v(n)$. The virtual source signal is then extrapolated by $\widehat{M}$ future estimates computed with (12). The estimated $v(n)$ is transformed to the frequency domain after (8). The dipole loudspeaker weights, $Q_{l,s}(k)$, are computed with (7) through (5) after $\mathcal{Q}_{h,m}$ is found via (2) and (3).

For the reproduction, $Q_{l,s}(k)$ are used as filters via (9) to obtain $q_{l,s}(n)$. The actual reproduced control soundfield is given by

$$\mathcal{S}^{\mathrm{c}}(\mathbf{x};k) = \sum_{l\in[\![\bar{L}]\!],s\in[\![2]\!],n\in\mathbb{Z}} q_{l,s}(n) e^{-icnk/2\dot{f}} T(\mathbf{x}, \mathbf{l}_{l,s}; k), \tag{14}$$

$\forall \mathbf{x} \in \mathbb{D}_{\mathrm{c}}$, where the 2D acoustic transfer function for each source is $T(\mathbf{x},\mathbf{l};k) = \frac{i}{4}H_0^{(1)}\left(k\left\|\mathbf{l}-\mathbf{x}\right\|\right)$. Note, $\mathcal{S}^{\mathrm{c}}(\mathbf{x};k)$ depends on $v(n)$.

5. RESULTS AND DISCUSSION

5.1. Experimental Setup

For evaluation, the layout of Fig. 1 is used with $R_{\mathbb{D}} = \bar{R} = 1\,\mathrm{m}$, $r_{\mathrm{c}} = 0.9\,\mathrm{m}$, $\bar{\phi} = \pi$ and $\bar{D} = 2.1\,\mathrm{m}$. There are $\bar{L} = 18$ dipole speaker pairs with $\ddot{d} \ll 1/k_{\max} = 2.73\,\mathrm{cm}$ spacing [21, 26], where $k_{\max} = 2\pi(2\,\mathrm{kHz})/c$ and $c = 343\,\mathrm{m\,s^{-1}}$. Spatial aliasing in the soundfield reproduction begins to occur near 2 kHz which reduces the control capability. All signals are sampled at a rate of 16 kHz with a frame step of $R = 0.5M$ for 50% overlapping and $M = \{64, 128, 192, 256, 320, 384, 448, 512\}$ are window lengths in samples. A prediction of $\widehat{M} = M$ future samples is made using $B = 2M$ past samples with an order of $\mathcal{P} = M$. The window,

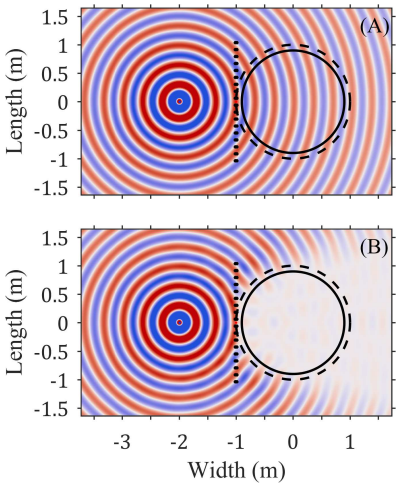

Fig. 2. The pressure field for an ideal periodic cancellation at 1kHz when the linear dipole array is inactive (A) and active (B).

$w(n)$, is a square root Hann window. The location of the talker is $\mathbf{t} = (2\,\mathrm{m}, \pi)$ and speech samples used to evaluate the performance were obtained from the TIMIT corpus [29]. Twenty files were randomly chosen such that the selection was constrained to have a male to female speaker ratio of $50:50$.

5.2. Soundfield Suppression

In order to evaluate the suppression of the control system, 32 virtual microphones are placed in random locations throughout $\mathbb{D}_\mathrm{c}$. The actual control and talker soundfields, $\mathcal{S}^\mathrm{c}(\mathbf{x};k)$ and $\mathcal{S}^\mathrm{t}(\mathbf{x};k)$, respectively, are approximated over $\mathbb{D}_\mathrm{c}$ using the 32 virtual recordings. To gauge the performance of the system, the normalised acoustic suppression between $\mathcal{S}^\mathrm{c}(\mathbf{x};k)$ and $\mathcal{S}^\mathrm{t}(\mathbf{x};k)$ is defined as

$$\zeta\left(k\right) \triangleq \frac{\int_{\mathbb{D}_\mathrm{c}} \left|\mathcal{S}^\mathrm{t}(\mathbf{x};k) + \mathcal{S}^\mathrm{c}(\mathbf{x};k)\right| \, d\mathbf{x}}{\int_{\mathbb{D}_\mathrm{c}} \left|\mathcal{S}^\mathrm{t}(\mathbf{x};k)\right| \, d\mathbf{x}}, \tag{15}$$

where $\mathcal{S}^\mathrm{c}(\mathbf{x};k)$ is from (14) and, in this work, for simplicity, $\mathcal{S}^\mathrm{t}(\mathbf{x};k) = \sum_{n\in\mathbb{Z}} v(n) e^{-icnk/2f} \frac{i}{4} H_0^{(1)}\left(k\left\|\mathbf{v}-\mathbf{x}\right\|\right)$. $\zeta\left(k\right)$ is found from (15) for a range of frequencies from 100 Hz to 8 kHz. The real part of $\mathcal{S}^\mathrm{t}(\mathbf{x};k)$ is shown in Fig. 2 at 1 kHz for when $\mathcal{S}^\mathrm{c}(\mathbf{x};k)$ is active and inactive, as an example. Fig. 2 clearly shows significant suppression on only one side of the linear dipole array providing a large quiet zone across the wall of loudspeakers. It is also apparent that by not strictly sampling the entire boundary of the control region for the Kirchhoff-Helmholtz integral, the loudspeaker array does not restrict the movement of a listener in and out of $\mathbb{D}$.

5.3. Synthesis and Prediction Accuracy Trade-off

A trade-off between soundfield reproduction accuracy and prediction accuracy is apparent in Fig. 3 which shows mean suppression from 156 Hz to 2 kHz. Assuming the signal is known (equivalent to a perfect prediction), as shown in blue in Fig. 3, the longer block length provides better control whereas a longer (and presumably therefore less accurate) prediction is required. A smaller block length is expected to perform worse as it results in fewer analysis frequencies in the wave domain and, hence, is filtered with less accuracy. Using a larger block length overcomes this issue and, assuming perfect prediction, is capable of -18.8 dB of suppression on average over $\mathbb{D}_\mathrm{c}$ with a 32 ms block length. However, with the necessary prediction

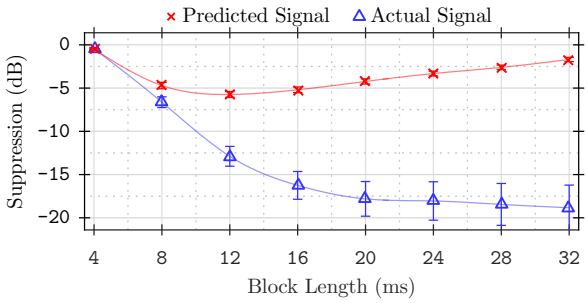

Fig. 3. The mean suppression, ζ, computed using $1/6$th octave band means from 156 Hz to 2 kHz over $2.54\,\mathrm{m}^2$ for an actual future block in blue and predicted in red. 95% confidence intervals are shown.

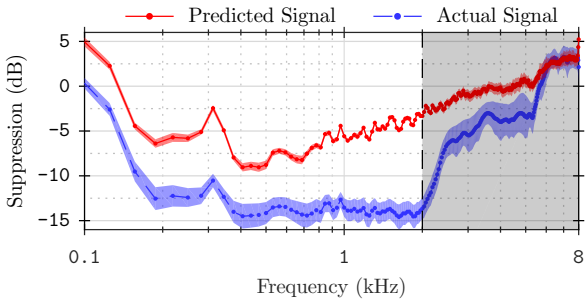

Fig. 4. The suppression, $\zeta\left(k\right)$, for a 12 ms block length from 100 Hz to 8 kHz over $2.54\,\mathrm{m}^2$. 95% confidence intervals are shaded red and blue. The bandwidth where spatial aliasing occurs is shaded grey.

to overcome the filtering delay, as shown in red in Fig. 3, the longer prediction results in less suppression. The peak suppression occurs with a 12 ms block length and -5.74 dB of suppression on average.

Choosing the block length which attains maximum suppression from Fig. 3 has the potential to provide the best suppression for wave-domain processed soundfield control. The optimal block length in this case is 12 ms and the suppression for this block length is shown per frequency in Fig. 4. The downward trend in Fig. 4 as frequency decreases from 2 kHz suggests that the control from the predicted block performs best for lower frequencies. The increase below 156 Hz and peak near 300 Hz is due to the finite length filter causing a loss of reproduction accuracy. It can be seen from Fig. 4 that the mean suppression reaches a peak of -9.1 dB near 400 Hz and maintains mean suppression below -7.5 dB from 365 Hz to 730 Hz. Future work could include investigating the control above the spatial Nyquist frequency by either increasing the loudspeaker density or using hybrid loudspeaker and ANC systems [30, 31].

6. CONCLUSIONS

We have investigated the effects of autoregressive delay compensation on active speech control when using wave-domain processing to improve active control over large spatial regions. A system has been proposed using a linear array of secondary dipole sources which uses autoregressive prediction with wavefield decompositions used to minimise residual soundfield energy. The proposed system is capable of a significant mean speech suppression of -18.8 dB with an ideally predicted 32 ms block over a large $2.54\,\mathrm{m}^2$ area. Through analysis of the proposed control system, a trade-off between reproduction accuracy and prediction accuracy has been shown to exist. A predicted block with an optimal length of 12 ms has shown to provide a mean suppression of -5.74 dB over a $2.54\,\mathrm{m}^2$ area.

7. REFERENCES

[1] T. Betlehem, W. Zhang, M. Poletti, and T. D. Abhayapala, "Personal Sound Zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, pp. 81–91, 2015.

[2] Y. Kajikawa, W.-S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Trans. Signal Inform. Process.*, vol. 1, pp. 1–21, 2012.

[3] S. M. Kuo, S. Mitra, and W.-S. Gan, "Active noise control system for headphone applications," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 2, pp. 331–335, 2006.

[4] T. J. Sutton, S. J. Elliott, A. M. McDonald, and T. J. Saunders, "Active control of road noise inside vehicles," *Noise Control Eng. J.*, vol. 42, no. 4, 1994.

[5] H. Sano, T. Inoue, A. Takahashi, K. Terai, and Y. Nakamura, "Active control system for low-frequency road noise combined with an audio system," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 7, pp. 755–763, 2001.

[6] J. Cheer and S. J. Elliott, "The design and performance of feedback controllers for the attenuation of road noise in vehicles," *Int. J. Acoust. Vibration*, vol. 19, no. 3, pp. 155–164, 2014.

[7] Y. Xiao and J. Wang, "A new feedforward hybrid active noise control system," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 591–594, 2011.

[8] N. V. George and G. Panda, "On the development of adaptive hybrid active noise control system for effective mitigation of nonlinear noise," *Signal Process.*, vol. 92, no. 2, pp. 509–516, 2012.

[9] O. J. Tobias and R. Seara, "Leaky delayed LMS algorithm: stochastic analysis for gaussian data and delay modeling error," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1596–1606, 2004.

[10] I. T. Ardekani and W. H. Abdulla, "Adaptive signal processing algorithms for creating spatial zones of quiet," *Digital Signal Process.*, vol. 27, pp. 129–139, 2014.

[11] S. Elliott, *Signal processing for active control*. Academic press, 2000.

[12] S. Spors and H. Buchner, "Efficient massive multichannel active noise control using wave-domain adaptive filtering," in *Int. Symp. Commun., Control Signal Process. (ISCCSP)*. IEEE, 2008, pp. 1480–1485.

[13] J. Zhang, W. Zhang, and T. D. Abhayapala, "Noise cancellation over spatial regions using adaptive wave domain processing," in *Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*. IEEE, 2015, pp. 1–5.

[14] J. Zhangg, T. D. Abhayapala, P. N. Samarasinghe, W. Zhang, and S. Jiang, "Sparse complex FxLMS for active noise cancellation over spatial regions," in *Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2016, pp. 524–528.

[15] W. Jin, "Adaptive reverberation cancelation for multizone soundfield reproduction using sparse methods," in *Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2016, pp. 509–513.

[16] K. Kondo and K. Nakagawa, "Speech emission control using active cancellation," *Speech Commun.*, vol. 49, no. 9, pp. 687–696, Sep. 2007.

[17] L. Athanas, "Open air noise cancellation," U.S. Patent 2011/0 274 283 A1, Nov. 10, 2011.

[18] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp. 2038–2050, 2010.

[19] C. R. Hart and S.-K. Lau, "Active noise control with linear control source and sensor arrays for a noise barrier," *Journal of Sound and Vibration*, vol. 331, no. 1, pp. 15–26, 2012.

[20] W. Chen, H. Min, and X. Qiu, "Noise reduction mechanisms of active noise barriers," *Noise Control Eng. J.*, vol. 61, no. 2, pp. 120–126, 2013.

[21] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[22] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, pp. 2343–2355, 2015.

[23] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, pp. 107–116, 2009.

[24] W. Jin, W. B. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," in *Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2013, pp. 311–315.

[25] J. Donley, C. Ritz, and W. B. Kleijn, "Improving speech privacy in personal sound zones," in *Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2016, pp. 311–315.

[26] F. Dunn, W. M. Hartmann, D. M. Campbell, N. H. Fletcher, and T. Rossing, *Springer handbook of acoustics*. Springer, 2015.

[27] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters," in *Speech Coding and Synthesis*. Elsevier Science Inc., 1995, ch. 12, pp. 433–466.

[28] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[29] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguistic Data Consortium*, 1993.

[30] J. Donley, C. Ritz, and W. B. Kleijn, "Reproducing Personal Sound Zones Using a Hybrid Synthesis of Dynamic and Parametric Loudspeakers," in *Asia-Pacific Signal & Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC)*. IEEE, Dec. 2016, pp. 1–5.

[31] K. Tanaka, C. Shi, and Y. Kajikawa, "Binaural active noise control using parametric array loudspeakers," *Applied Acoustics*, vol. 116, pp. 170–176, Jan. 2017.