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
jrd089@uowmail.edu.au, critz@uow.edu.au, bastiaan.kleijn@vuw.ac.nz

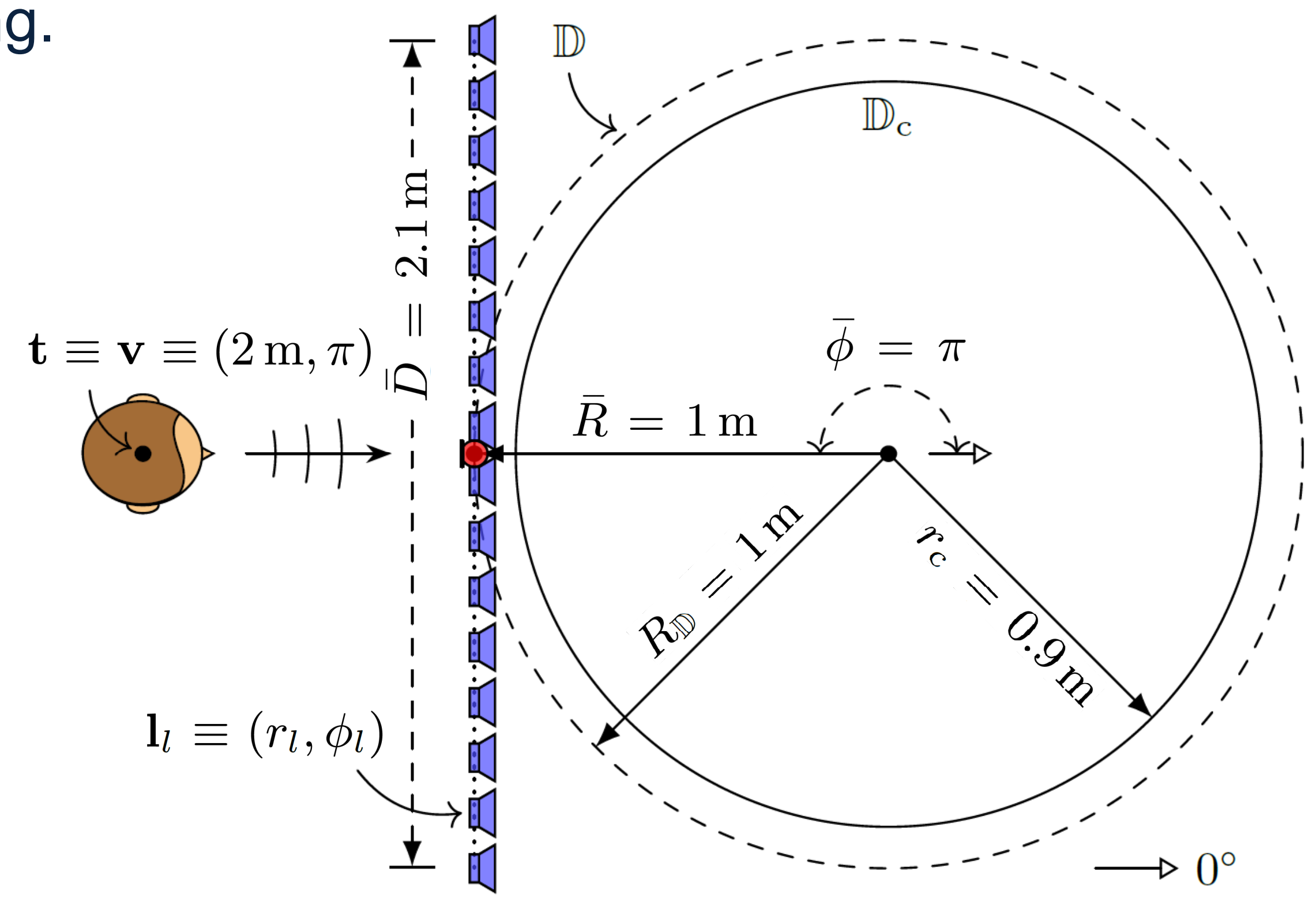
INTRODUCTION

Personal sound can be provided by spatial regions of controlled sound reproduced using loudspeaker arrays.

Active Speech Control (ASC) is a technique that allows secondary sources to reproduce destructive soundfields.

Physical partitions may be replaced by an active loudspeaker array to benefit offices, libraries, teleconferencing rooms, restaurants, cafes, stadiums and vehicle cabins.

In this work, we investigate the effects of compensating for wave-domain filtering delay in an ASC system using autoregressive (AR) modelling.



Active control layout for a linear dipole array directed to the right.

WAVE-DOMAIN SOUNDFIELD SUPPRESSION

- Soundfield function as an expansion of orthogonal basis wavefields

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

- Minimise the error between the talker soundfield and a control soundfield

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

- QR factorisation results in

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

- Sub (3) in (1), where $\mathcal{Q}_{m,h}$ **are control soundfield coefficients**, gives

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

LOUDSPEAKER WEIGHTS

- Monopole loudspeaker weights are found using cylindrical harmonic expansion

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

- Dipole loudspeaker positions are given by

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

- **Dipole loudspeaker weights** are found by converting monopole weights

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

Dipole weights reproduce a control soundfield on one side of the array

SHORT-TIME SIGNAL PROCESSING

- An input signal, $v(n)$, is broken into frames using a window, $w(n)$, of length M
- Each frame is filtered using the dipole weights, $Q_{l,s}(k)$, from (7)

$$\tilde{q}_{a,l,s}(n) = \Re\left\{ \frac{1}{N} \sum_{m\in[\![N]\!]} Q_{l,s}(k_m) \tilde{V}_a(k_m) e^{icnk_m/2\dot{f}} \right\}, \tag{8}$$

- A synthesis window, also $w(n)$, is applied to the weighted output from (8) and is added to the accumulated output signal, $q_{l,s}(n)$.
- The ***autocorrelation method*** estimates AR parameters, $\{\hat{a}_j\}_{j\in[\![\mathcal{P}]\!]}$.
- Soundfield filtering delay is compensated for using linear prediction
- Geometric delay compensation by inverse filtering

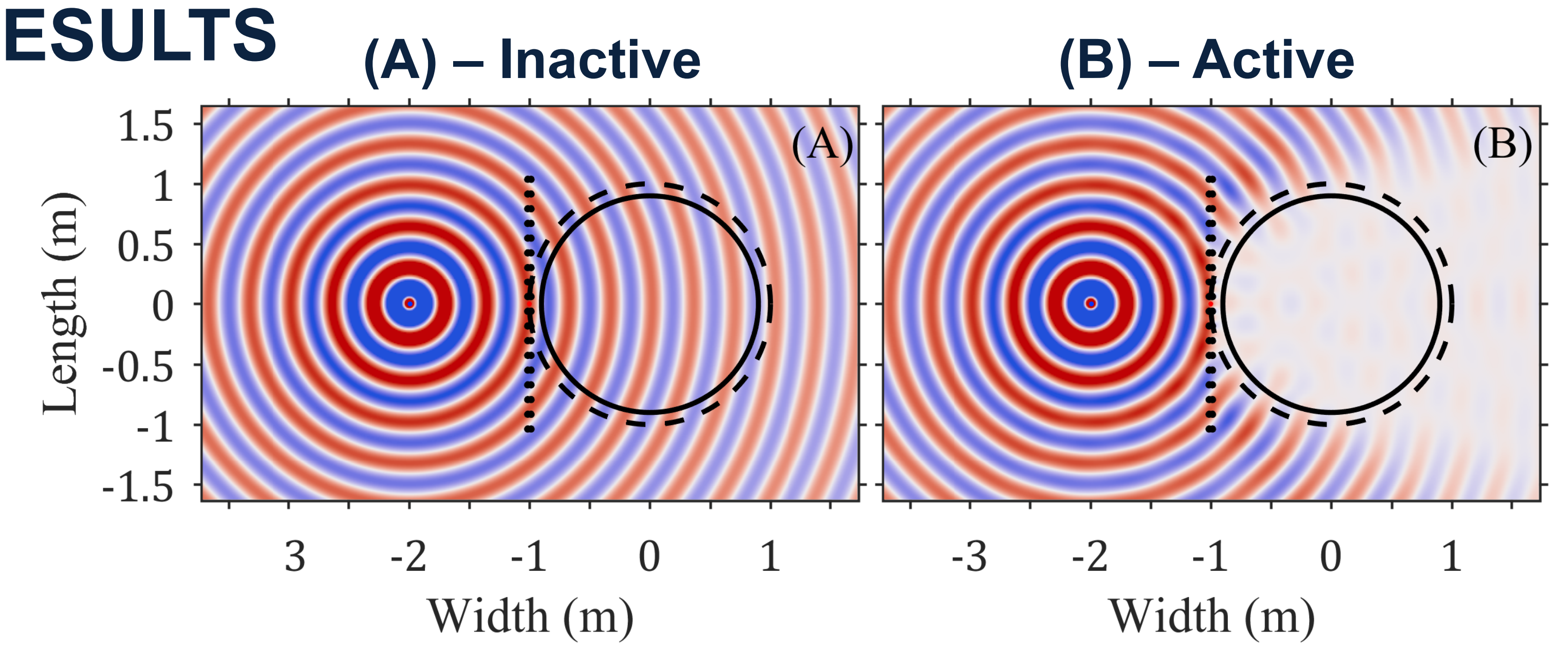
$$v(n) = \Re\left\{ \frac{1}{N} \sum_{m\in[\![N]\!]} \frac{4\left\{\sum_{n\in[\![N]\!]} z(n) e^{-icnk_m/2\dot{f}}\right\}}{iH^{(1)}_0(k_m\|\mathbf{v}-\mathbf{z}\|)} e^{icnk_m/2\dot{f}} \right\}, \tag{9}$$

- Reproduced control soundfield

$$\mathcal{S}^{\mathrm{c}}(\mathbf{x};k) = \sum_{l\in[\![\bar{L}]\!],s\in[\![2]\!],n\in\mathbb{Z}} q_{l,s}(n) e^{-icnk/2\dot{f}} T(\mathbf{x},\mathbf{l}_{l,s};k), \tag{10}$$

where the acoustic transfer function is $T(\mathbf{x},\mathbf{l};k) = \frac{i}{4} H^{(1)}_0(k\|\mathbf{l}-\mathbf{x}\|)$

RESULTS

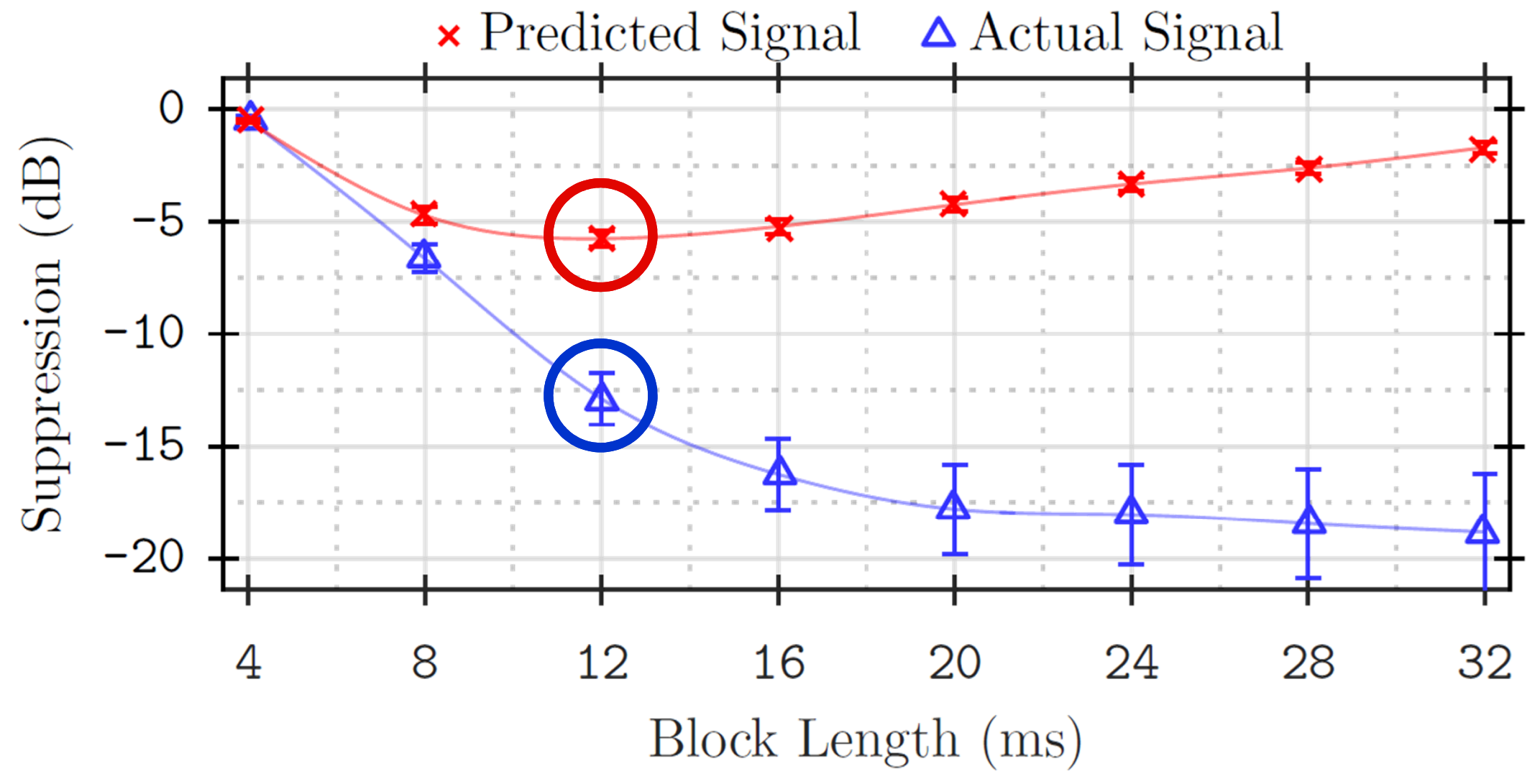


Using 18 dipole loudspeakers

Performance is gauged using the normalised acoustic suppression:

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

Mean suppression, ζ.



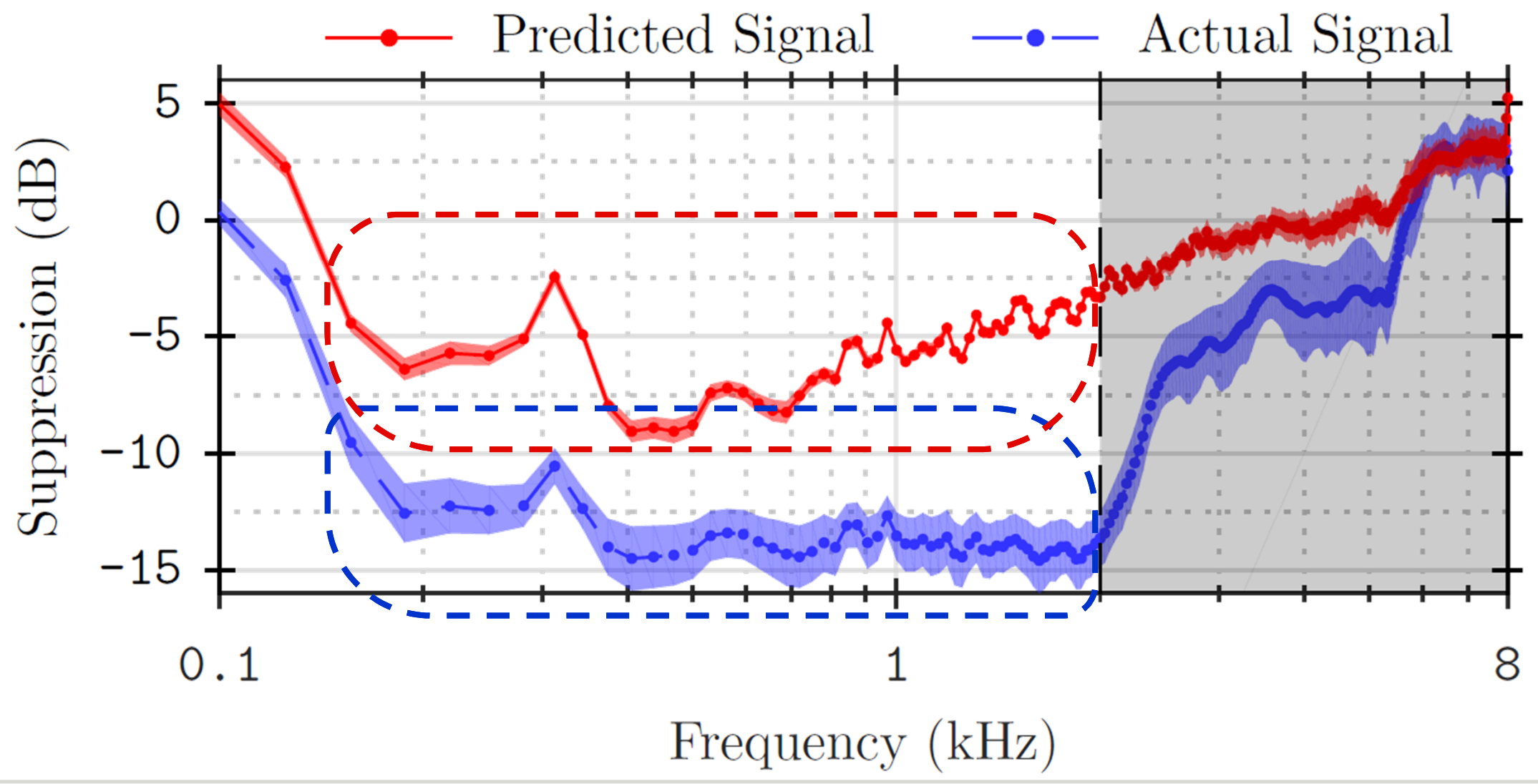
Best mean suppressions over 2.54 m²:

Actual Signal at $M = 32$ ms: −18.8 dB

Predicted Signal at $M = 12$ ms: −5.74 dB

Actual Signal at $M = 12$ ms: −12.9 dB

Suppression, $\zeta(k)$, for a 12 ms block length.



CONCLUSIONS

Investigated effects of AR delay-compensation on ASC using wave-domain processing over large spatial regions

Proposed and evaluated a linear dipole array system using AR prediction and wavefield decompositions to minimise residual energy

Analysis of proposed system finds **optimal block length of 12 ms**

Shown trade-off between reproduction and prediction accuracy

Proposed system capable of **suppression of -5.74 dB over 2.54 m²**

ICASSP 2017 — 5th-9th March, New Orleans, USA