Stefanakis, Nikolaos; Delikaris-Manias, Symeon; Mouchtaris, Athanasios

Acoustic beamforming in front of a reflective plane

Published in:
2018 26th European Signal Processing Conference, EUSIPCO 2018

DOI:
10.23919/EUSIPCO.2018.8553103

Published: 29/11/2018

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Stefanakis, N., Delikaris-Manias, S., & Mouchtaris, A. (2018). Acoustic beamforming in front of a reflective plane. In *2018 26th European Signal Processing Conference, EUSIPCO 2018* (Vol. 2018-September, pp. 26-30). Article 8553103 (European Signal Processing Conference). IEEE.
https://doi.org/10.23919/EUSIPCO.2018.8553103

Acoustic Beamforming in Front of a Reflective Plane

Nikolaos Stefanakis* ,‡, Symeon Delikaris-Manias†, Athanasios Mouchtaris* ,§
* FORTH-ICS, Heraklion, Crete, Greece, GR-70013
† Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland, FI-00076
‡ Technological Educational Institute of Crete, Department of Music Technology and Acoustics Engineering, Rethymno, Greece, GR-74100
§ University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-70013

***Abstract*—In this paper, we consider the problem of beamforming with a planar microphone array placed in front of a wall of the room, so that the microphone array plane is perpendicular to that of the wall. While this situation is very likely to occur in a real life problem, the reflections introduced by the adjacent wall can be the cause of a serious mismatch between the actual acoustic paths and the traditionally employed free-field propagation model. We present an adaptation from the free-field to the so-called reflection-aware propagation model, that exploits an in-situ estimation of the complex and frequency-dependent wall reflectivity. Results presented in a real environment demonstrate that the proposed approach may bring significant improvements to the beamforming process compared to the free-field propagation model, as well as compared to other reflection-aware models that have been recently proposed.**

I. Introduction

State-of-the-art devices such as mobile phones, laptops, tablets or home automation systems feature multiple microphones to reduce reverberation and noise or interference from simultaneous competing sound sources. Exploiting the spatial diversity of the sound scene, acoustic beamforming is an essential functionality for these applications. However, the deviation of many parameters from their ideal conditions, under which the theoretical performance of the system is predicted, causes degradation in the system performance by reducing the array gain and altering the beam pattern.

A regular cause for these phenomena is the fact that in real acoustic environments, the transmitted signal is often received via multiple paths due to reflection, diffraction and scattering by objects in the transmission medium. The idea to constructively combine one or more distinct reflections in the transmission path in order to improve the performance of applications related to microphone array signal processing has been expressed by several researchers in the past. Li et al [1] proposed the use of hemi-spherical array for sound capturing and beamforming in the half-space. In other works it was demonstrated that, accounting for first order reflections in reverberant rooms, may not only assist sound source localization [2]–[6], but may reveal additional spatial information regarding the sound sources, as for example their elevation angles and ranges [7] or their orientations [8], [9]. The improvement in the beamforming performance when taking into account multiple distinct reflections has been demonstrated along the concept of Acoustic Rake Receivers (ARRs) in [10], [11]. It should be noted however that most of these works are evaluated in simulated environments, considering rooms with simple geometries (mostly rectangular ones) and room boundaries with unrealistic reflectivity properties.

In this paper, we perform experiments in a real environment in order to study the problem of acoustic beamforming, considering a planar microphone array placed just in front of one of the walls of the room, so that the microphone array plane is perpendicular to that of the reflecting wall. This so-called half-space configuration may often appear in a real-life application due to practical constraints, as for example, small room size or limited space for the sound sources, and we believe that it deserves special attention. For each sound source present in the room, it is then expected that there will always be a secondary acoustic path which is coherent and comparable in strength to the direct path [12]. To improve the beamforming performance, we propose an adaptation from the free-field to the so-called Reflection-Aware propagation model, that accounts for the joint contribution of the direct path and the earliest reflection introduced by the adjacent wall. Doing so, we extend earlier work presented in [6] as well as along the concepts of ARRs, by employing an in-situ calibration phase to estimate a complex and frequency-dependent model of the wall reflectivity. It is demonstrated that the reflection-aware propagation model may achieve significant improvements when compared to a simpler approach that a fixed value for the wall reflectivity is used, as well as compared to the free-field steering vector. The advantages of the proposed propagation model are demonstrated in terms of the responses of two well known signal-independent beamformers, the matched-filter and the superdirective beamformer.

II. Propagation Model

Planar microphone arrays are among the most popular types of microphone arrays due to their ability to provide 360^o coverage in capturing, enhancing and separating acoustic sources

The project leading to this publication has received funding by the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 644283, Project LISTEN.

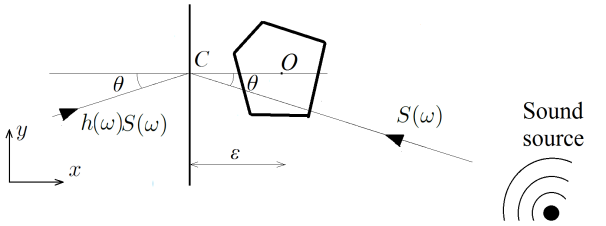

Fig. 1. A plane wave of strength $S(\omega)$ impinging on a planar array with incident angle θ, defined with respect to the projection of O on the reflective boundary, denoted with C.

[13]–[15]. The propagation model for a planar microphone array of M sensors is typically expressed through vector $\mathbf{a}(\omega,\theta) = [a_1(\omega,\theta),\cdots,a_M(\omega,\theta)]^T$ where $a_m(\omega,\theta)$ carries the transfer path characteristics from a sound source at angle θ to the mth microphone, relevant to a reference point O which typically coincides with the center of the planar array. Typically the transfer path characteristics involve the direct path from the sound source to each microphone assuming anechoic conditions and ignoring any distinct reflections that may occur due to the surrounding environment. Certainly, estimating all the secondary paths in a indoor environment is very difficult, as it would require detailed knowledge of the room geometry as well as of any objects in it. Assuming however that the microphone array is very close to a particular wall, it can be expected that the earliest reflection introduced by the adjacent wall will carry a relatively large portion of the energy of the Room Impulse Response (RIR). Moreover, assuming far field conditions and a specular reflection, the reflected signal component can be efficiently estimated in the frequency domain, as a function of the array's orientation and distance from the closest wall.

In Fig. 1 we present the proposed geometric model for the case of a planar array of arbitrary geometry placed just in front of a planar reflector so that an incident angle $\theta = 0$ is normal to the adjacent wall. Let point C coincide with the projection of the array center to the wall and let its distance from that wall be symbolized with ϵ. For the case of a specular reflection, a direct plane wave of amplitude $S(\omega)$ arriving at C with incident angle θ will be the cause of a reflected component of strength $h(\omega)S(\omega)$ travelling at an angle $\pi-\theta$. Here, $h(\omega)$ is the so-called Image Source Relative Gain (ISRG) that expresses the relative gain with which the image source contributes to the sound field. Using the free field steering vector $\mathbf{a}(\omega,\theta)$, we propose an adaptation to a so-called reflection-aware steering vector $\mathbf{a}'(\omega,\theta)$ through

$$\mathbf{a}'(\omega,\theta) = \mathbf{a}(\omega,\theta)e^{jk\epsilon\cos\theta} + h(\omega)\mathbf{a}(\omega,\pi-\theta)e^{-jk\epsilon\cos\theta}. \quad (1)$$

To be noticed that $h(\omega)$ does not account for the difference in the time of arrival between the primary sound wave and its reflection; this is explicitly taken into account by the two phasors $e^{jk\epsilon\cos\theta}$ and $e^{-jk\epsilon\cos\theta}$ in (1), which makes $h(\omega)$ to be independent on the array configuration. In a previous work [6], it was shown that a fixed and real value for h close to 1 represents a reasonable choice for constructing the reflection-aware steering vector. Such an approach is conceptually similar to the delay-and-sum rake receiver [10], in which case the secondary paths are modelled as delayed and attenuated versions of the direct path. However, as $h(\omega)$ relates to the wall acoustic impedance [16], it may be expected that a frequency-dependent and complex ISRG represents a more valid choice for deployment in a real environment. In a latter section in this paper we propose a procedure for in-situ estimation of $h(\omega)$.

III. Beamforming techniques

Signals are represented in the time-frequency (TF) domain with $\omega,\tau \in \mathbb{Z}$ denoting the angular frequency and the time-frame index respectively. We let $\mathbf{x}(\tau,\omega) = [X_1(\tau,\omega),...,X_M(\tau,\omega)]^T$ denote the STFTs of the observed microphone signals. We assume the presence of a single acoustic source at location θ_l which is assumed known, using, for example, the DOA estimation procedure presented in [6]. The output of the beamformer which is steered at θ_l in order to capture the sound source signal reads

$$Y(\tau,\omega) = \mathbf{w}(\omega,\theta_l)^H\mathbf{x}(\tau,\omega). \quad (2)$$

where $\mathbf{w}(\omega,\theta_l)$ is the $M\times 1$ complex vector with the beamformer weights.

Superdirective beamforming is well known for its ability to maximize array gain subject to a given model of the noise spatial coherence which is assumed to be stationary [17], [18]. The spherical isotropic model is a widely used model as it has been shown that it reflects the statistics of the reverberant part of the signals in rooms [19], [20]. Here, we employ that model in terms of the noise coherence matrix $\mathbf{Q}(\omega)$, used for constructing beamformer weights of the form

$$\mathbf{w}^{(\lambda)}(\omega,\theta_l) = \frac{(\mathbf{Q}(\omega)+\lambda\mathbf{I})^{-1}\mathbf{a}(\omega,\theta_l)}{\mathbf{a}^H(\omega,\theta_l)(\mathbf{Q}(\omega)+\lambda\mathbf{I})^{-1}\mathbf{a}(\omega,\theta_l)}, \quad (3)$$

where λ is a scalar which can be associated with the white noise gain constraint [21]. Using the two different propagation models, we consider in this paper two different superdirective (SD) beamformers and particularly

- the Free-Field Superdirective (FFSD) beamformer, derived by using the free-field propagation model and
- the Reflection-Aware Superdirective (RASD) beamformer, derived by using the reflection-aware propagation model of Eq. (1).

For reasons of comparison, we also consider the case of Matched Filter (MF) beamformers, which can be realized with weights of the form

$$\mathbf{w}(\omega,\theta_l) = \frac{\mathbf{a}(\omega,\theta_l)}{\mathbf{a}^H(\omega,\theta_l)\mathbf{a}(\omega,\theta_l)}. \quad (4)$$

Again, we can exploit the two different propagation models in order to deploy two different MF beamformers and particularly

- the Free-Field Matched Filter (FFMF) beamformer derived by using the free-field propagation model and

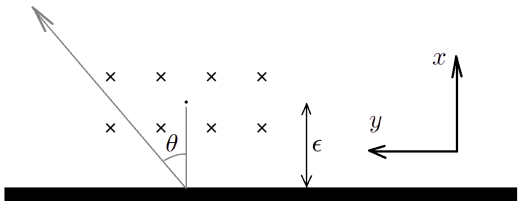

Fig. 2. Sketch of the microphone array and its arrangement inside the rectangular room.

- the Reflection-Aware Matched Filter (RAMF) beamformer derived by using the reflection-aware propagation model of Eq. (1).

IV. Estimation of ISRG

In-situ estimation of the specific acoustic impedance of materials is an active research area in acoustics [22]. Inspired by such attempts, we propose an off-line calibration process during which a complex model of the ISRG is estimated in-situ, considering a wideband sound source at a small distance from the microphone array and at a known location $\mathbf{r}_a = [x_a, y_a, z_a]$. By defining the coordinate system so that the plane $x = 0$ coincides with the nearest wall, the image source location will be at $\mathbf{r}_I = [-x_a, y_a, z_a]$. Assuming that the direct sound and the first reflection are significantly more dominant than other components of the RIR, then we can explain the observation signal $\mathbf{x}(\tau, \omega)$ as a result of the actual sound source signal $s_a(\tau, \omega)$ and the image source signal $s_i(\tau, \omega)$ through

$$\mathbf{x}(\tau, \omega) = \mathbf{G}(\omega)\mathbf{s}(\tau, \omega) \tag{5}$$

where $\mathbf{s}(\tau, \omega) = [s_a(\tau, \omega), s_i(\tau, \omega)]^T$, and $\mathbf{G}$ is a $M \times 2$ matrix such that $G_{m,1}(\omega) = e^{-jk_\omega \|\mathbf{r}_A - \mathbf{r}_m\|_2}$ and $G_{m,2}(\omega) = e^{-jk_\omega \|\mathbf{r}_I - \mathbf{r}_m\|_2}$ represent the free-field transfer path from the actual source and the image source to the mth microphone respectively and k_ω represents the wave number at frequency ω. The actual and image source signal can be recovered in a least square sense using Tikhonov regularization through

$$\mathbf{s}(\tau, \omega) = (\mathbf{G}(\omega)^H \mathbf{G}(\omega) + \mu \mathbf{I})^{-1} \mathbf{G}(\omega)^H \mathbf{x}(\tau, \omega). \tag{6}$$

We claim that a local estimation of the complex ISRG can be derived at TF point τ, ω from the ratio

$$\hat{h}(\tau, \omega) = \frac{s_a(\tau, \omega)}{s_i(\tau, \omega)}. \tag{7}$$

Having a large collection of local estimates from different time-frames at frequency ω, we expect that these will cluster around the actual $h(\omega)$ value. At each frequency ω we construct the set $\mathbb{H}(\omega)$ with the local estimates $\hat{h}(\tau, \omega)$ obtained from all time-frames and across a continuous range of frequencies from bin $\omega - L$ to bin $\omega + L$. For a given set, we then form a histogram with respect to the elements in the set $\mathbb{H}(\omega)$ in order to derive a final estimation the ISRG at each frequency as

$$h(\omega) = H_{Re}(\omega) + iH_{Im}(\omega), \tag{8}$$

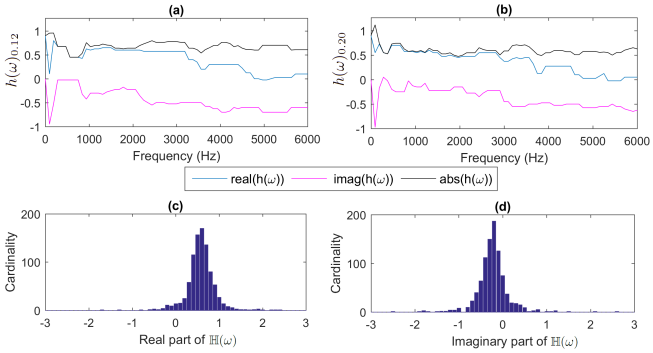

Fig. 3. Estimated ISRG as a function of frequency when the array is placed at a distance of $\epsilon = 0.12$ m from the wall in (a) and at a distance of $\epsilon = 0.20$ m in (b). Distribution of the $\mathbb{H}(\omega)$ coefficients at the frequency of 2720 Hz is shown for the real part in (c) and the imaginary part in (d).

where $H_{Re}(\omega)$ and $H_{Im}(\omega)$ correspond to the values of the histogram with the highest cardinality, referring to the real part and imaginary part respectively. The method allows for in-situ estimation of the ISRG without the need of a known excitation signal. Assuming that the reflecting wall is locally reacting [16] and that the reflections are always specular, the derived ISRG coefficients should be valid across a wide range of incident angles.

V. Evaluation

Experiments were performed in a reverberant room of approximately $\text{RT}_{60} = 0.7$ s with an open planar microphone array. The microphone array consisted of 8 omnidirectional microphones DPA4060 in a 2×4 grid with inter-microphone distance of 2 cm, as shown in Fig. 2. The microphones were connected to an RME Micstasy AD converter which was then connected to an RME Fireface400 through the ADAT connection. The signals were acquired at a sampling rate of 48 kHz while the time-frame length used for the TF analysis was equal to 512 samples. The microphone array was mounted on a microphone stand at a height of 1.5 m from the floor and was placed next to the wall at distances of $\epsilon = 0.12$ and $\epsilon = 0.20$ m. A two-way active Genelec loudspeaker was used for exciting the sound field at the same height and in different angles with respect to the array and the wall.

To estimate the ISRG, we followed the procedure explained in section IV. A Gaussian white noise signal of 5 s duration was used as input to the loudspeaker, which was placed at a distance of $d = 0.5$ m from the center of the microphone array each time and at an angle of $\theta = 0^o$. The value of L was set equal to 3 frequency bins. The first measurement was performed with the array at a distance of $\epsilon = 0.12$ m from the wall and a second measurement followed with the array at a distance of $\epsilon = 0.20$ m. We will use the notation $h(\omega)_{0.12}$ and $h(\omega)_{0.20}$ in order to refer to the values of ISRG retrieved at each distance. The real and imaginary part of the estimated ISRG, as well as the magnitude, can be seen as a function of frequency in Fig. 3(a) for $h(\omega)_{0.12}$ and in Fig. 3(b) for $h(\omega)_{0.20}$. Also, a histogram formed with the real and

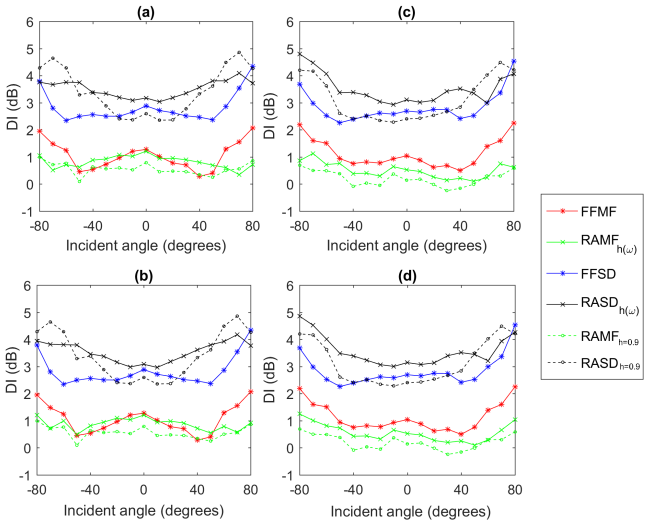

Fig. 4. DI as a function of the incident angle for $\epsilon = 0.12$ in (a) and (b) and $\epsilon = 0.20$ in (c) and (d). The values of ISRG are equal to those in $h(\omega)_{0.12}$ in (a) and (c) and equal to those in $h(\omega)_{0.20}$ in (b) and (d).

imaginary part of the local ISRG estimates for $h(\omega)_{0.12}$ at the frequency of 2720 Hz is shown in (c) and (d) respectively.

It can be generally observed from (a) and (b) that in both experiments the retrieved ISRG values have a non-trivial imaginary part. Considering that the boundary is a rigid wall, we believe that the magnitude of ISRG is underestimated in both $h(\omega)_{0.12}$ and $h(\omega)_{0.20}$. This can be attributed to the fact that attenuation is neglected in the actual and image source transfer path in section IV. From the same figure it can be observed that the retrieved ISRG values exhibit strong fluctuations in the low frequency range. We believe that below a cut-off frequency f_0, the proposed procedure does not present a viable choice for estimating the reflective properties of the boundary. This can be attributed to the fact that sound waves behave as rays only for wavelengths which are small compared to the dimensions of the room and that at lower frequencies, sound propagation is governed by the room modal behavior. A rough approximation regarding the frequency limit above which sound waves behave as rays is $f_0 = 4f_c$ where f_c is the Schroeder frequency [16]. Taking this into account, we decided to use a fixed value of $h(\omega) = 0.9$ for frequencies below the value of 844 Hz, while $h(\omega)$ was set to the retrieved values of $h(\omega)_{0.12}$ or $h(\omega)_{0.20}$ above this frequency.

The ability to define different values for the ISRG at each frequency point inserts an additional degree of freedom in the beamforming process. Essentially, using the propagation model of Eq. (1) and the retrieved ISRG coefficients, we can design a steering vector which is better suited to the particularities of the reflective boundary. As a means for evaluating the advantage gained from the proposed calibration phase, we consider an additional reflection-aware propagation model which is based on fixed-with-frequency ISRG values of $h = 0.9$.

Among the various ways to quantify directivity of a beamformer, the polar directivity pattern and the Directivity Index (DI) are those most widely used in the literature and in industry [16]. Due to particularities of the half-space arrangement, it is meaningful to calculate the DI as a function of the look direction θ_l. To calculate this metric, we created a spatial grid of 17 different locations on a semicircle of radious R=1.5 m with center defined as point C (see Fig. 2). These 17 locations span the azimuth angle range from $\theta = -80^o$ to $\theta = 80^o$ with a step of 10^o. A Gaussian white noise sequence of 5 s duration was emitted from the loudspeaker at each location on the spatial grid. The DI at each look direction θ_l was calculated across a continuous frequency range from $\omega_{min} = 60$ Hz to $\omega_{max} = 6$ kHz using the formula [23], [24]

$$\mathrm{DI}(\theta_l) = 10\log_{10}\left(\frac{p(\theta_l|\theta_l)}{\frac{1}{17}\sum_{n=1}^{17} p(\theta_l|\theta_n)}\right) \tag{9}$$

where θ_n, $n = 1, ..., 17$ is the set of angular points on the orbit of investigation and

$$p(\theta_l|\theta_n) = \sum_{\tau} \sum_{\omega=\omega_{min}}^{\omega_{max}} |Y(\tau,\omega)|^2 \tag{10}$$

is the total beafmormer output energy at look direction θ_l when the sound source is located at angle θ_n.

In general, it was observed that the value of λ used in Eq. (3) may significantly affect the performance of the SD beamformers. The values of λ that guaranteed the best performance with respect to DI were found to be the same for the RASD and the FFSD beamformers and were equal to $\lambda = 3.16$. Based on these values, the DI as a function of the look direction can be seen for all different beamformers in Fig. 4. The DI is shown for a distance of $\epsilon = 0.12$ from the wall in (a) and (b) and for a distance of $\epsilon = 0.20$ in (c) and (d). The performance for the RAMF and RASD beamformers implemented with a fixed ISRG value of $h = 0.9$ are shown with the dashed-green and dashed-black line respectively in each plot. On the other hand, subfigures (a) and (c) illustrate the performance for RAMF and RASD beamformer using $h(\omega)_{0.12}$ while subfigures (b) and (d) illustrate the performance when using $h(\omega)_{0.20}$. Note that we use the subscript $_{h=0.9}$ and $_{h(\omega)}$ to discriminate between a beamformer which is based on a fixed and on frequency-dependent ISRG coefficients respectively.

Several interesting facts become apparent from these plots. A first thing to observe is that the SD beamformers obtain a much larger DI compared to the MF beamformers. This can be attributed to the ability of the spherical isotropic coherence model to sufficiently represent the second-order statistics of the reverberant part of the sound field. Now, comparing the different SD beamformer implementations, it can be seen that the RASD$_{h=0.9}$ provides an improvement compared to the FFSD beamformer but only at relatively large incident angles of $\theta_l > 40^o$ and $\theta_l < -40^o$. At angles approaching $\theta_l = 0^o$, the RASD$_{h=0.9}$ beamformer seems to deteriorate compared to the FFSD beamformer. Evidently, this deterioration is avoided when using the RASD$_{h(\omega)}$ beamformers, the latter ones providing higher DI values compared to all

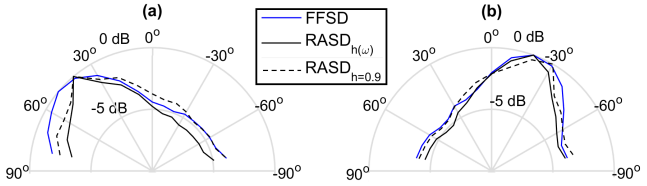

Fig. 5. Beampattern for $\theta_l = 40^o$ in (a) and $\theta_l = -20^o$ in (b).

other beamformers in the range from -60^o to 60^o. The four graphs demonstrate that usage of the estimated ISRG values may provide a substantial advantage as opposed to a fixed and real ISRG value, especially in the angle range that these values were estimated during the calibration phase (in our case, close to $\theta = 0$). Furthermore, subfigures (b) and (c) demonstrate that the estimated ISRG values are usable even when the array is at a different distance from the wall during the calibration and the operation phase. Finally, similar to the SD beamforming implementation, Fig. 4 illustrates that the estimated ISRG coefficients provide an improvement compared to the fixed ISRG value also for the case of MF beamforming.

In Fig. 5, we demonstrate the SD beamformers' polar patterns for $\theta_l = 40^o$, in (a), and $\theta_l = -20^o$, in (b), when the array is placed $\epsilon = 0.20$ m. The results for RASD$_{h(\omega)}$ beamformer are in both cases obtained using $h(\omega)_{20}$. In accordance to the results shown in Fig. 4, a higher spatial selectivity can be achieved with the RASD$_{h(\omega)}$ beamformer compared to the RASD$_{h=0.9}$ and the FFSD beamformer.

VI. Conclusion

The idea to constructively combine one or more distinct reflections in the transmission path in order to improve the performance of acoustic beamforming has been expressed by several researchers in the past. In this paper, we considered a reflection-aware propagation model which takes into account the earliest reflection introduced by the adjacent wall. It was shown that a complex and frequency-dependent model of the wall reflectivity may significantly improve the performance of the RASD and RAMF beamformers, and we presented a simple and intuitive approach for estimating this model in-situ. The performance seems to improve especially in a range of incident angles that are close to those used during the calibration phase. In the future, we intend to experiment with a frequency- and angle-dependent model of the ISRG, using an interpolation process that exploits estimations of the ISRG derived from more than one sound source locations.

References

[1] Z. Li and R. Duraiswami, "Hemispherical microphone arrays for sound capture and beamforming," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, 2005, pp. 106–109.

[2] P. Bergamo, S. Asgari, H. Wang, and D. Maniezzo, "Collaborative sensor networking towards real-time acoustical beamforming in free-space and limited reverberance," *IEEE Transaction on Mobile Computing*, vol. 3, no. 3, pp. 211–224, 2004.

[3] T. Korhonen, "Acoustic localization using reverberation with virtual microphones," in *Proc. of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2008.

[4] O. Öçal, I. Dokmanić, and M. Vetterli, "Source localization and tracking in non-convex rooms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2014, pp. 1443–1447.

[5] A. Alexandridis, N. Stefanakis, and A. Mouchtaris, "Towards wireless acoustic sensor networks for location estimation and counting of multiple speakers in real life conditions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2017, pp. 6140–6144.

[6] N. Stefanakis and A. Mouchtaris, "Direction of arrival estimation in front of a reflective plane using a circular microphone array," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2016, pp. 622–626.

[7] F. Ribeiro, D. Ba, C. Zhang, and D. Florêncio, "Turning enemies into friends: using reflections to improve sound source localization," in *IEEE Int. Conf. Multimedia and Expo (ICME)*, 2010, pp. 731–736.

[8] P. Svaizer, A. Brutti, and M. Omologo, "Environment aware estimation of the orientation of acoustic sources using a line array," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2012, pp. 1024–1028.

[9] K. Niwa, Y. Hioka, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimation of sound source orientation using eigenspace of spatial correlation matrix," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2010, pp. 129–132.

[10] I. Dokmanić, R. Scheibler, and M. Vetterli, "Raking the cocktail party," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 825–836, 2015.

[11] A. Javed, A. Moore, and P. Naylor, "Spherical microphone array acoustic rake receivers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, 2016, pp. 111–115.

[12] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[13] J. Maganti, D. Gatica-Perez, and I. McCowan, "Speech enhancement and recognition in meetings with an audio-visual sensor array," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 15, no. 8, pp. 2257–2269, 2007.

[14] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 10, pp. 2193–2206, 2013.

[15] J. Benesty and J. Chen, "Differential beamforming with circular microphone arrays," *J. Acoust. Soc. Am.*, vol. 139, no. 4, pp. 2049–2049, 2016.

[16] L. Beranek, *Acoustics*, American Institute of Physics, 1986.

[17] H. Coh, R. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acous. Speech, Signal Process.*, vol. 34, no. 3, pp. 393–398, 1986.

[18] M. Wang, X. Ma, P. Yang, C. Hao, X. Feng, and Y. Zhang, "Design of robust high-order superdirectivity for circular arrays with sensor gain and phase errors," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, pp. 19, 2017.

[19] N. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 18, no. 7, pp. 1830–1840, 2010.

[20] N. Ito, N. Ono, E. Vincent, and S. Sagayama, "Designing the Wiener post-filter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2010, pp. 2818–2821.

[21] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 35, pp. 1365–1376, 1987.

[22] Y. Zhang, W. Lin, and C. Bi, "A technique based on the equivalent source method for measuring the surface impedance and reflection coefficient of a locally reacting material," in *Proc. of International Congress on Noise Control Engineering*, 2014.

[23] I. Thomas, M.Tashev, F. Lim, and P. Naylor, "Optimal beamforming as a time domain equalization problem with application to room acoustics," in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on*, 2014, pp. 75–79.

[24] J. Tylka and E. Choueiri, "On the calculation of full and partial directivity indices," *3D Audio and Applied Acoustics Laboratory, Princeton University, Technical Report*, 2014.