

ACOUSTIC REFLECTOR LOCALIZATION AND CLASSIFICATION

Luca Remaggi¹, Hansung Kim¹, Philip J. B. Jackson¹, Filippo Maria Fazi² and Adrian Hilton¹

¹Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

²Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, UK

ABSTRACT

The process of understanding acoustic properties of environments is important for several applications, such as spatial audio, augmented reality and source separation. In this paper, multichannel room impulse responses are recorded and transformed into their direction of arrival (DOA)-time domain, by employing a superdirective beamformer. This domain can be represented as a 2D image. Hence, a novel image processing method is proposed to analyze the DOA-time domain, and estimate the reflection times of arrival and DOAs. The main acoustically reflective objects are then localized. Recent studies in acoustic reflector localization usually assume the room to be free from furniture. Here, by analyzing the scattered reflections, an algorithm is also proposed to binary classify reflectors into room boundaries and interior furniture. Experiments were conducted in four rooms. The classification algorithm showed high quality performance, also improving the localization accuracy, for non-static listener scenarios.

Index Terms— Reflector Localization, Reflector Size Classification, Room Impulse Response, Beamforming

1. INTRODUCTION

In audio processing, several areas can benefit from an accurate estimation of the environmental acoustic properties. Information about the main reflectors can aid audio forensic [1] and simultaneous localization and mapping methods [2]. Moreover, knowledge of the reflected signal characteristics can improve systems in source separation [3], dereverberation [4], and microphone array raking [5]. Another application, that this paper targets, is the active manipulation of reverberant spatial audio objects, for next generation audio production [6]. Furthermore, augmented reality (AR) systems [7] can be improved by the auditory scene analysis provided here. The main contributions of this paper are: a novel method to localize acoustic reflectors given room impulse responses (RIRs); a novel classification algorithm, that analyzed the scattered reflections to classify reflectors into large (i.e. room boundaries) and small (e.g. furniture).

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and BBC as part of the BBC Audio Research Partnership.

One of the first attempts to localize acoustic reflectors was proposed in [8]. There, the time-space domain of multichannel RIRs was employed. This was done by assuming a linear microphone array to be perpendicular to the analyzed direction. Later, in [9], the direction of arrival (DOA)-time domain was analyzed, by employing the plane-wave decomposition method. Nevertheless, a large space was required to be sampled, to correctly capture the sound wave. Spherical harmonics were employed in [10, 11], to localize reflections. As in [12], they employed continuous signals, however, having issues with low SNRs. In [13], reflectors were localized by first estimating the related image source positions [14], having knowledge of the reflections' times of arrival (TOAs). However, microphones were assumed to be sparse in space. A 2D geometric approach was presented in [15]. Its full 3D extension was later proposed in [16], where reflectors were localized by constructing ellipsoids, having axes proportional to the reflection TOAs. Despite the good performance, it had a high computational cost. In this paper, we propose a novel approach, that exploits the reflection TOAs and DOAs. These are estimated by a new 2D morphological image processing method, which detects meaningful peaks in noisy spectral images, such as the DOA-time domain of multichannel RIRs.

Every state-of-the-art method that has been mentioned above aimed to localize acoustic reflectors by observing their specular reflections [17]. However, a correct analysis of the scattered reflections can help achieving several goals [18]. For instance, it can lead to a discrimination between walls and furniture, in a room. This kind of classification allows production of synthetic RIRs [19], and estimation of reverberant spatial audio object parameters [20]. To our knowledge, in the literature, there is yet no method to classify acoustic reflectors in terms of their spatial extension. To fill this gap, in this paper, we propose a binary classification algorithm, that uses the scattered reflections' energy as feature.

In this paper, Sec. 2 presents novel methods to localize and classify acoustic reflectors; in Sec. 3, experiments and results are discussed; Sec. 4 draws the overall conclusion.

2. THE PROPOSED METHODS

We propose a new pipeline (see Fig. 1), to localize and classify acoustic reflectors. RIRs are employed, by defining them

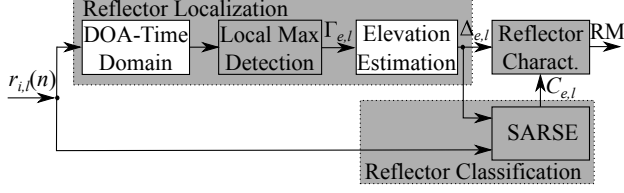


Fig. 1. The proposed pipeline, with the novelties in grey. $\Gamma_{e,l}$ contains the estimated TOAs and azimuth DOAs; $\Delta_{e,l}$ contains $\Gamma_{e,l}$ and the elevations. RM stands for room model.

as $r_{i,l}(n) = \sum_{e=0}^{T_m} h_{e,i,l}(n - n_{e,i,l}) + g_{i,l}(n)$, where i and l are the microphone and source indexes, respectively; n is the discrete time index, T_m is the last early reflection, and $g_{i,l}(n)$ the late reverberation. $n_{e,i,l}$ is the early reflection TOA, with e being the reflection index ($e = 0$ indicates the direct sound).

2.1. The DOA-Time Domain

Multichannel RIRs can be transformed to the joint DOA-time domain, where the acoustic energy is function of both azimuth DOA and time. This domain transformation is done by using a superdirective array (SDA) beamformer [21]. Before employing the SDA, RIRs are band-pass filtered between $\Omega_1 = 2$ kHz and $\Omega_2 = 8$ kHz, to avoid the poor low frequency directivity, and the spatial aliasing at high frequencies.

The SDA is our 3D version [21] of [22], and it can be applied to any kind of microphone array. Points in the 3D Cartesian coordinates can be defined as $\mathbf{X} = [u_x; u_y; u_z]^T = [\cos \theta \sin \phi; \sin \theta \sin \phi; \cos \phi]^T$, where $[\cdot]^T$ is the matrix transpose operator, θ the azimuth, and ϕ the elevation. The array manifold can be then written as $\mathbf{a}(\omega, \theta, \phi) = \exp(j \frac{\omega}{c_0} \mathbf{A} \mathbf{X})$, where $j = \sqrt{-1}$, \mathbf{A} contains the M microphones' coordinates, ω is the angular frequency, and c_0 the sound speed. The elements of the noise cross-spectral matrix \mathbf{S} are: $s_{i,i'} = \text{sinc}(\frac{\omega}{c_0} \|\mathbf{A}_i - \mathbf{A}_{i'}\|)$, where \mathbf{A}_i and $\mathbf{A}_{i'}$ contain the i -th and i' -th microphone coordinates, respectively, and $\|\cdot\|$ is the ℓ_2 norm operator. The regularization matrix is $\mathbf{B} = [\mathbf{S} + q\mathbf{I}]^{-1}$, with $q = 0.01$ (empirically found) and \mathbf{I} the identity matrix. The SDA weights are: $\mathbf{w}(\omega, \theta, \phi) = \frac{\mathbf{B}\mathbf{a}}{\mathbf{a}^* \mathbf{B}\mathbf{a}}$, where $[\cdot]^*$ is the complex conjugate transpose operator. $\mathbf{w}(\omega, \theta, \phi)$ contains the M microphone signals' weights $w_i(\omega, \theta, \phi)$.

Beamformed RIRs are produced by using the SDA as:

$$r_l^B(n, \theta, \phi) = \mathcal{F}^{-1} \left\{ \sum_{i=1}^M \sum_{\omega=\Omega_1}^{\Omega_2} w_i(\omega, \theta, \phi) R_{i,l}(\omega) \right\}. \quad (1)$$

$R_{i,l}(\omega)$ is the $r_{i,l}(n)$'s Fourier transform and $\mathcal{F}^{-1}\{\cdot\}$ the inverse Fourier transform operator. By setting $\phi = 0$, beamformed RIRs are produced for each θ , and placed one next to the others, to obtain the DOA-time domain (see Fig. 2).

2.2. Local Maxima Detection

In the DOA-time domain, the energy maximum coordinates are the reflection TOAs and azimuth DOAs. A peak-picking

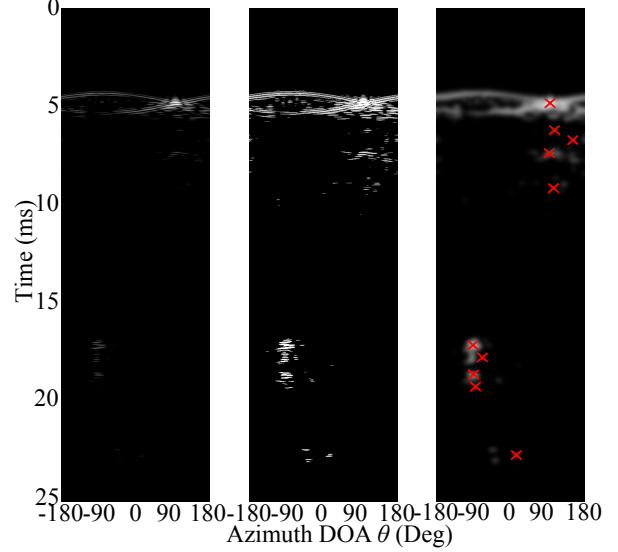


Fig. 2. DOA-time domain example (left); local maxima detector pre-processor output (center); local maxima detector output (right), with estimated maxima (red crosses). Figures are in greyscale (white means signal power is 1, black 0).

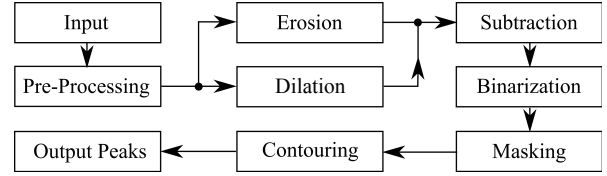


Fig. 3. Block diagram for local maxima detection.

algorithm can be exploited to detect them. Automated peak detection for 2D spectral images have been actively researched for more than 30 years, in image processing and bioinformatics [23, 24, 25]. The major challenges of 2D peak detection algorithms is the distinction of artifacts from real peaks, and also detecting peaks with overlapping resonances. We propose a 2D peak detection algorithm modified from the morphology-based method in [25], for meaningful peak detection in the DOA-time domain images. Fig. 3 shows the block diagram of the proposed 2D peak detection algorithm.

In the pre-processing stage, the input image is filtered by a Gaussian kernel to reduce noise. A sound wave decreases its energy by 6 dB every double of the traveled distance [26]. To compensate this bias, the original values in the DOA-time domain are updated by $y = y + 6 \log_2(x)$, where y is the energy level in dB and x is time in ms. The middle image of Fig. 2 shows the result of this bias compensation. This image is converted into two images by two morphological process: erosion and dilation, then subtracted from the original image. Two subtraction images are binarized and the erosion image is used as a mask to filter out the peak image from the dilation image. Finally, contours of the images are found, and their centers are defined as the detected peak locations. The right image of Fig. 2 shows the proposed peak detector output.

2.3. Elevation Estimation

From the position of the maxima in the DOA-time domain, the reflection TOAs $n_{e,l}$ and azimuth DOAs $\theta_{e,l}$ are obtained. However, to localize reflections in the 3D space, elevations $\phi_{e,l}$ are also needed. By fixing θ to the estimated values $\theta_{e,l}$, the SDA weights $\mathbf{w}(\omega, \theta = \theta_{e,l}, \phi)$ are calculated considering elevation angles between 0° and 90° with a resolution of 10° . The corresponding beamformed signals $r_l^B(n, \theta = \theta_{e,l}, \phi)$ are then obtained by following Eq. (1). By using a Hamming window of length $D = 2.7$ ms (value heuristically found), centered at $n_{e,l}$, the beamformed reflections are segmented. The elevations $\phi_{e,l}$ are selected as the ones producing the reflection segments having the greatest energy: $\phi_{e,l} = \operatorname{argmax}_\phi \sum_{n=n_{e,l}-D}^{n_{e,l}+D} |r_l^B(n, \theta = \theta_{e,l}, \phi)|^2$.

2.4. The SARSE Algorithm

The novel algorithm, to discriminate between large and small reflectors is named as scattering analysis based reflector size estimator (SARSE). Different from the well-known scattering coefficient [18], we define here the partial reflection scattering coefficient K . This measures the partial amount of sound that is scattered away from the specular reflection direction, by observing it from a single listening position. SARSE's assumptions are: reflectors are not perfectly smooth; small objects are not closely spaced.

Reflections are segmented out from the beamformed signals, through a Hamming window of length D centered at $n_{e,l}$. The specular component energy is calculated by using the beamformed reflection segment associated to $\theta_{e,l}$ and $\phi_{e,l}$: $E_{e,l}^{\text{SP}} = \sum_{n=n_{e,l}-D}^{n_{e,l}+D} r_{e,l}^B(n, \theta = \theta_{e,l}, \phi = \phi_{e,l})$. The energy of the partial scattered reflections is obtained by using the beamformed reflection segments $\pm\alpha$ around the specular component: $E_{e,l}^{\text{sc}} = \sum_{n=n_{e,l}-D}^{n_{e,l}+D} r_{e,l}^B(n, \theta_{e,l} - \alpha \leq \theta \leq \theta_{e,l} + \alpha, \phi_{e,l} - \alpha \leq \phi \leq \phi_{e,l} + \alpha)$, with $\theta \neq \theta_{e,l}$ and $\phi \neq \phi_{e,l}$. α has to be narrow enough to avoid the inclusion of other reflection energies. Since the SDA beam size, at 2 kHz, is 5° , we suggest $\alpha = 20^\circ$, i.e. four times it. Partial scattered energies, for a large and a small reflector, are shown Fig. 4. The partial reflection scattering coefficient is $K_{e,l} = 1 - (E_{e,l}^{\text{SP}}/E_{e,l}^{\text{sc}})$. By setting a threshold K^{th} , a binary classification is performed, with output $C_{e,l}$: if a reflector is large, then $K_{e,l} > K^{\text{th}}$ ($C_{e,l} = 1$), since scattered components are received at the listener from around the specular reflection direction; if a reflector is small, this does not happen, since no surface exists around the specular reflective area, thus $K_{e,l} < K^{\text{th}}$ ($C_{e,l} = 0$).

2.5. Reflector Characterization

By combining the localization and classification outputs, reflectors can be finally characterized. From the estimated $n_{e,l}$, the distance between the reflection image sources and the mi-

crophone array are obtained as $\rho_{e,l} = n_{e,l}c_0$. By also having $\theta_{e,l}$ and $\phi_{e,l}$, the image sources are localized in spherical coordinates, as it was done in our image source direction and ranging (ISDAR) [16]. The loudspeaker image bisection (LIB) [27] is then used to calculate $\mathbf{M}_{e,l}$, that is the midpoint between the image source position, in Cartesian coordinates, $\mathbf{B}_{e,l}$ and the loudspeaker $\mathbf{B}_{0,l}$: $\mathbf{M}_{e,l} = (\mathbf{B}_{e,l} + \mathbf{B}_{0,l})/2$.

Midpoints that are related to reflections classified as large $\mathbf{M}_{e,l, C_{e,l}=1}$ are converted to planes by employing our ISDAR-LIB method (for a detailed description, please, refer to [16]). Looking towards the six spatial directions, the closest planes to the microphone array are selected to generate the room shoebox model. If L loudspeakers are used, the shoebox estimation is refined, by applying our mean-ISDAR-LIB [16], which averages each room boundary over its L estimations. Finally, the remaining midpoints $\mathbf{M}_{e,l, C_{e,l}=0}$, that are related to small reflectors, are modeled as cuboids of dimensions ten times smaller than the shoebox in which they are in. This size was chosen to approximate a typical furniture dimension.

3. EXPERIMENTS AND RESULTS

SARSE and its application to the new localization method are evaluated. Experiments aim to demonstrate the reflector classification importance, for dynamic listener applications.

3.1. The Datasets

RIRs were recorded in four rooms. A bi-circular array of microphones was employed. Each circle, of radii 85 cm and 106 cm, respectively, is composed of 24 microphones, for a total of 48. Prior knowledge about the reflector positions was used to solve the up-down ambiguity generated by planar arrays [16]. The room dimensions, reverberation times (RT60s) averaged between 500 Hz and 4 kHz, and loudspeaker positions (with respect to the microphone array) are reported on the left side of Tab. 1. All the recorded RIRs, together with relative detailed documentation, can be downloaded from [28, 29]. The first room, named as ‘‘BBC LR’’, is a listening environment at the BBC MediaCity, in Salford, UK, where the interior objects were limited to the recording devices [29]. The second room is named as ‘‘BBC UL’’. It is a lab at the BBC MediaCity, furnished to reproduce a living room-like environment [29]. ‘‘Vislab’’ was an acoustically treated lab at the University of Surrey, where the Surrey Sound Sphere was built (i.e. the main interior object) [28]. The last dataset is ‘‘VML’’. It was a mock room built in a lab at the University of Surrey, with one wall and ceiling missing, like a film set [28]. The interior objects were the recording devices.

3.2. Evaluation Metrics

The first part of the experiments evaluates SARSE's classification performance. Four metrics are used [30]: the precision,

Table 1. Dataset characteristics (left), classification results (center) and reflector localization results (right).

	Room Size (m ³)	RT60	Loudsp. Pos.	Precis.	Recall	Accur.	F ₁ Score	ε _{SARSE}	ε
BBC LR	5.6 × 5.0 × 2.9	222 ms	0°, 180°	100 %	100 %	100 %	1.00	118 cm	141 cm
BBC UL	5.6 × 5.2 × 2.9	275 ms	0°, -110°	83 %	100 %	89 %	0.91	150 cm	197 cm
Vislab	7.8 × 6.1 × 4.0	326 ms	0°, 180°	100 %	100 %	100 %	1.00	178 cm	388 cm
VML	2.4 × 4.0 × 2.4	445 ms	90°, -90°	83 %	83 %	71 %	0.83	148 cm	128 cm
AVG	–	–	–	92 %	96 %	90 %	0.93	149 cm	214 cm

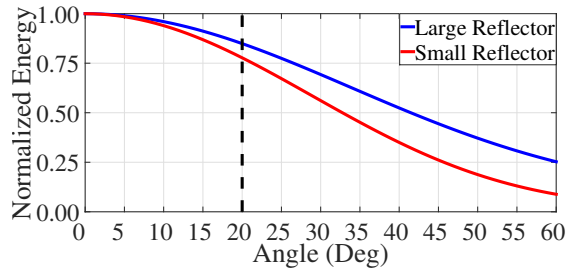


Fig. 4. 8 kHz octave band energy, around the specular reflection (at 0°), for a large (i.e. the ceiling) and a small (i.e. a sofa) reflector in BBC UL. The vertical line shows $\alpha = 20^\circ$.

that is calculated as $\frac{TP}{TP+FP}$, where TP stands for true positive and FP for false positive; the recall that is $\frac{TP}{TP+FN}$, where FN means false negative; the accuracy $\frac{TP+TN}{TP+TN+FP+FN}$, with TN being true negative; the F₁ Score, that is $2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$.

For the acoustic reflector localization evaluation, the Euclidean distances between the image source positions (related to the estimated reflectors), and their groundtruth are calculated to obtain the errors ϵ_e . The first evaluated error is the average over the E room boundaries, that were found without using SARSE: $\epsilon = \frac{1}{E} \sum_{e=1}^E \epsilon_e$. The same kind of error is calculated by introducing SARSE, and named as ϵ_{SARSE} .

3.3. Classification and Localization Performance

Results at the center of Table 1 shows SARSE’s high performance, with an average, over the four datasets, that is always above 90 %. It works perfectly in both BBC LR and Vislab. SARSE classifies as small objects the recording devices in BBC LR, and the Surrey Sound Sphere bars, in Vislab. This suggests robustness of SARSE also with non-flat objects. In BBC UL, although having a recall of 100 %, the other metrics show some errors. These are caused by a coffee table, placed between one loudspeaker and the microphone array, which occludes the floor. Therefore, SARSE recognizes this table as being large. VML shows decent performance for most of the metrics. However, the accuracy is about 70 %. This is caused by the floor, i.e. the furthest planar reflector. There, SARSE mistakenly classifies first order reflections as belonging to small objects, whereas higher order ones as planes.

The advantage of using SARSE is underlined by non-static listener scenarios, for applications such as AR, or next generation spatial audio production. By classifying a reflector, the image source position can be correctly dynamically



Fig. 5. BBC UL 360 image with large (square) and small (circle) reflectors correctly (green) and wrongly (red) classified.

readjusted with the listener movements. In fact, small reflectors are modeled as small cuboids instead of planes. To evaluate this performance gain, the errors made by localizing the reflector, with (i.e. ϵ_{SARSE}) and without (i.e. ϵ) SARSE, are compared on the right side of Table 1. At first glance, these errors may look greater than the ones in the literature [16, 31]. However, there, just the first arriving reflections were analyzed. Here, the error is averaged over all the estimated reflectors composing the room shoebox. Furthermore, here, gross and fine errors are not discriminated. In BBC LR, BBC UL and Vislab, the improvement is remarkable. Without SARSE, in BBC UL, three planes would be generated at the position of two sofas and a computer screen (see Fig. 5). In Vislab, several planes would surround the listener at the position of the Surrey Sound Sphere. However, SARSE classifies a second-order reflection as the VML’s floor. The related plane lies far from the groundtruth, increasing ϵ_{SARSE} . In preliminary tests on a large room (RT60 = 910 ms), SARSE’s classification performance was similar to that in the VML (precision = 50 %, recall = 100 %, accuracy = 71 %, F1 score = 0.67). However, the error in reflector localization, ϵ , was large. Further work is needed to investigate reflector localization in large reverberant rooms for application to concert halls.

4. CONCLUSION

A novel pipeline, that localizes and classifies acoustic reflectors, has been proposed. Experiments were performed by employing four different rooms, showing the high performance of the novel classification algorithm SARSE and the importance of its integration within the novel reflector localization method. Future work may focus on extending the analysis to a greater room number. Moreover, alternative reflector classification methods may be studied. Investigations about combinations with vision based methods may also be carried out.

5. REFERENCES

- [1] H. Malik, "Acoustic environment identification and its applications to audio forensics," *IEEE Trans. on Information Forensics and Security*, vol. 8, no. 11, pp. 1827–1837, 2013.
- [2] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to SLAM," in *the IEEE ICASSP*, Shanghai, China, 2016.
- [3] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, "From blind to guided audio source separation," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 107–115, 2014.
- [4] P. A. Naylor and N. D. Gaubitch, *Speech dereverberation*, Springer Science & Business Media, 2010.
- [5] I. Dokmanić, R. Scheibler, and M. Vetterli, "Raking the cocktail party," *IEEE J. of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 825–836, 2015.
- [6] P. Coleman, A. Franck, P. J. B. Jackson, R. Hughes, L. Remaggi, and F. Melchior, "Object-based reverberation for spatial audio," *J. Audio Engineering Society*, vol. 65, no. 1/2, pp. 66–77, 2017.
- [7] Y. Vazquez-Alvarez, I. Oakley, and S. A. Brewster, "Auditory display design for exploration in mobile audio-augmented reality," *Personal and Ubiquitous Computing*, vol. 16, no. 8, pp. 987–999, 2012.
- [8] M. Kuster, D. de Vries, E. M. Hulsebos, and A. Gisolf, "Acoustic imaging in enclosed spaces: analysis of room geometry modifications on the impulse response," *J. Acoustical Society of America*, vol. 116, no. 4, pp. 2126–2137, 2004.
- [9] A. M. Torres, J. J. Lopez, B. Pueo, and M. Cobos, "Room acoustics analysis using circular arrays: an experimental study based on sound field plane-wave decomposition," *J. Acoustical Society of America*, vol. 133, no. 4, pp. 2146–2156, 2013.
- [10] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," in *the IEEE ICASSP*, Prague, Czech Republic, 2011.
- [11] E. Mabande, K. Kowalczyk, H. Sun, and W. Kellermann, "Room geometry inference based on spherical microphone array eigenbeam processing," *J. Acoustical Society of America*, vol. 134, no. 4, pp. 2773–2789, 2013.
- [12] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *Proc. of the IEEE ICASSP*, Dallas, USA, 2010.
- [13] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *PNAS*, vol. 110, no. 30, pp. 12186–12191, 2013.
- [14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Society of America*, vol. 4, no. 65, pp. 943–950, 1979.
- [15] F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 20, no. 10, pp. 2683–2695, 2012.
- [16] L. Remaggi, P. J. B. Jackson, P. Coleman, and W. Wang, "Acoustic reflector localization: novel image source reversion and direct localization methods," *IEEE/ACM Trans. on Audio, Speech and Lang. Proc.*, vol. 25, no. 2, pp. 296–309, 2017.
- [17] H. Kuttruff, *Room acoustics - Fifth edition*, Spon press, 2009.
- [18] T. J. Cox, B.-I. L. Dalenback, P. D'Antonio, J. J. Embrechts, J. Y. Jeon, E. Mommertz, and M. Vorländer, "A tutorial on scattering and diffuse coefficients for room acoustic surfaces," *ACTA Acustica United with Acustica*, vol. 92, no. 1, pp. 1–15, 2005.
- [19] N. Tsingos, C. Dachsbacher, S. Lefebvre, and M. Dellepiane, "Instant sound scattering," in *the 18th Eurographics Conf. on Rendering Techniques (EGSR)*, Grenoble, France, 2007.
- [20] L. Remaggi, P. J. B. Jackson, and P. Coleman, "Estimation of room reflection parameters for a reverberant spatial audio object," in *the 138th AES Conv.*, Warsaw, Poland, 2015.
- [21] H. Kim, L. Remaggi, P. J. B. Jackson, F. M. Fazi, and A. Hilton, "3D room geometry reconstruction using audio-visual sensors," in *the Conf. on 3D Vision (3DV)*, Qingdao, China, 2017.
- [22] M. R. Bai and C.-C. Chen, "Application of convex optimization to acoustical array signal processing," *J. of Sound and Vibration*, vol. 332, no. 25, pp. 6596–6616, 2013.
- [23] D. S. Garrett, R. Powers, A. M. Gronenborn, and G. M. Clore, "A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams," *J. of Magnetic Resonance*, vol. 95, pp. 214–220, 1991.
- [24] P. Klukowski, M. J. Walczak, A. Gonczarek, J. Boudet, and G. Wider, "Computer vision-based automated peak picking applied to protein NMR spectra," *Bioinformatics*, vol. 31, no. 18, pp. 2981–2988, 2015.
- [25] T. Banelli, M. Vuano, F. Fogolari, A. Fusiello, G. Esposito, and A. Corazza, "Automation of peak-tracking analysis of stepwise perturbed NMR spectra," *J. of Biomolecular NMR*, vol. 67, no. 2, pp. 121–134, 2017.
- [26] J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, "Review of noise propagation in the atmosphere," *J. Acoustical Society of America*, vol. 61, no. 6, pp. 1403–1418, 1977.
- [27] S. Tervo and T. Tossavainen, "3D room geometry estimation from measured impulse responses," in *the IEEE ICASSP*, Kyoto, Japan, 2012.
- [28] P. Coleman, L. Remaggi, and P. J. B. Jackson, "S3A room impulse responses," 2015, <http://dx.doi.org/10.15126/surreydata.00808465>.
- [29] H. Kim, L. Remaggi, P. J. B. Jackson, and A. Hilton, "S3A audio-visual captures," 2017, <https://doi.org/10.15126/surreydata.00812228>.
- [30] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [31] Y. El Baba, A. Walther, and E. A. P. Habets, "Time of arrival disambiguation using the linear Radon transform," in *the IEEE ICASSP*, New Orleans, USA, 2017.