

LOCALIZATION OF VIRTUAL ACOUSTIC SOURCES BASED ON THE HOUGH TRANSFORM FOR SOUND FIELD RENDERING APPLICATIONS

L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, S. Tubaro

Dipartimento di Elettronica ed Informazione, Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

ABSTRACT

In this paper we propose a methodology for the localization of virtual acoustic sources for sound field rendering applications. After the reconstruction of the sound field in the listening area by means of circular harmonic decomposition, the virtual source location is found through the Hough transform. We prove the accuracy of the proposed methodology by comparing the source locations estimates with those of a subjective test campaign.

Index Terms— Microphone array, loudspeaker array, sound field rendering

1. INTRODUCTION

This paper concerns the problem of localizing virtual acoustic sources in sound field rendering applications. The delivery of an accurate spatial impression of the location of the sound source is crucial for the effectiveness of rendering techniques. We can roughly categorize assessment solutions in two classes: based on subjective listening tests ([1, 2, 3, 4, 5, 6]) and based on objective criteria, possibly adopting psychoacoustics considerations ([7, 8, 9, 10]).

The first class of solutions presents the clear advantage of testing the accuracy directly on human beings. As an example, authors in [6] synthesize the soundfield to be reproduced by binaural synthesis and apply this stimulus to a panel of listeners. If samples are not accurately selected and their number is not sufficiently high, some bias could be introduced in the test. In order to prevent such inaccuracies, [4] gives also rules for the selection of the listening panel and for an appropriate statistical analysis. These recommendations are taken from [5], which is focused on the evaluation of small impairments. Subjective tests are, however, an expensive solution for many applications. In [8] the author compares the spatial responses of real world and simulative data. The spatial response is acquired by means of a linear microphone array. Even if cues related to source location are contained in the spatial response, localization of the virtual source is accomplished only through subjective tests.

In [9] the authors localize the virtual acoustic source by means of Vector-Based Amplitude Panning for Wave Field Synthesis installations. It is worth noticing that in certain conditions the rendered sound field could exhibit relevant distortions in portions of the listening area, thus delivering an incoherent spatial impression. It is therefore important that the estimate of the virtual source location is accomplished for different positions of the listener. Authors in [7] analyze the impact of reverberations in small auditoria and concert halls on the location of the sound source. For this purpose the authors employ binaural measurements acquired using a dummy-head at different locations inside the listening area. Motivated by the high costs of listening tests, authors in [10] perform a Inter-aural Time Difference-based virtual source localization. Methods described in

[9] and [10] present the advantage of enabling a position dependent localization of the virtual source, but, in order to attain this desirable feature, acquisitions must be repeated for every position of interest.

While keeping the advantage of a localization dependent on the listener position, in this paper we adopt a system that enables a single acquisition. For this purpose we adopt a measurement setup based on a single cardioid microphone that is moved at uniformly spaced positions on a circle that encloses the listening area. The sound field within the listening area is then reconstructed using circular harmonics decomposition, as described in [11]. The same setup has been used also in [12] and [13] for different purposes. More specifically, in [12] the authors aim at evaluating the impact of reverberations on the quality of the rendered sound field. In [13] artifacts such as pre-echoes and post-echoes are analyzed using a psychoacoustic metric. We localize virtual acoustic sources through a Hough transform analysis [14, 15] on the reconstructed sound field, searching for the centers of the wavefronts in the sound field. The localization can be conducted over the whole listening area or also on portions of it. This feature is useful in all the situations where the sound field exhibits deviations from the ideality in some parts of the listening area. In order to validate the presented technique, we accomplished a listening test. Results confirm that a good correlation between objective and subjective measurements exists.

The rest of the paper is organized as follows: Section 2 formulates the problem and introduces the notation. Section 3 describes the localization technique based on the Hough transform. Section 4 describes the simulative and experimental setup and discusses the results. Section 5 draws some conclusions. Finally, section 6 discusses the relation existing between the presented technique and prior work.

2. BACKGROUND AND PROBLEM STATEMENT

Consider the sound field in Figure 1, which is a snapshot of a simulated sound field. The reference frame, as shown, is at the center of the circle corresponding to the listening area. The virtual source V is on the right of the listening area and it forms an angle α with the x axis. The soundfield is observed at $N_{im} \times N_{im}$

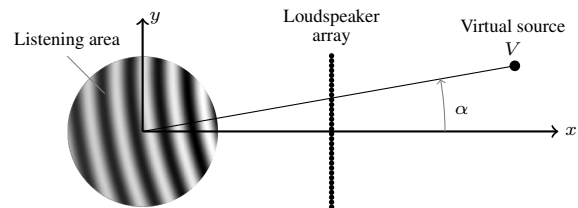


Fig. 1: Example of soundfield

points within the listening area. Our goal is to estimate the location of the virtual source from the measured sound field. Beforehand, therefore, we need a technique for the measurement of the sound field in the listening area. A brute-force technique could be based on sampling the listening area using microphones located on a regular grid and implementing a localization/tracking methodology, such as [16]. However, microphones would interact with the propagating sound field, thus corrupting the measurement. For this reason we use a less invasive technique, based on sequential measurements. A cardioid microphone is placed at one end of a rotating rig. The rig uniformly samples a circle by means of a stepper motor. The sound field, which is stationary during the measurement procedure, is sampled for each position of the microphone on the circle. The sound field is then reconstructed through an interpolation and extrapolation procedure based on circular harmonics decomposition, as described in [11, 12]. The reconstructed sound field is composed by multiple wavefronts. Under the assumption of homogeneous and isotropic propagation, the wavefronts are arcs of circles, whose centers are, ideally, at the virtual source location V . Due to non-idealities in the rendered sound field the estimated position $\hat{V}$ could be different from V . We propose to compute $\hat{V}$ through a generalized Hough transform, which can also work on portions of the sound field image, thus enabling a position dependent localization. Fig. 2 summarizes the workflow of the proposed technique.

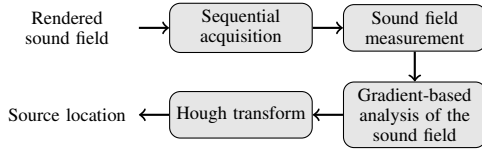


Fig. 2: Workflow of the localization technique.

3. VIRTUAL SOURCE LOCALIZATION

In this section we describe an approach to compute $\hat{V}$ based on the Hough transform, a feature extraction technique originally proposed for the detection of straight lines [14] and then extended to circle detection in [15]. The Hough transform operates a mapping of points in the image of the sound field (*image space*), to the search space of the potential source locations (*parameter space*). The image space is a square portion of the Cartesian plane centered in the origin of the reference frame. The sound field is measured on a grid of $N_{im} \times N_{im}$ locations. The resolution of the grid is Δ_{im} m/pt. The sound field image is processed with the edge detection algorithm in [17], which returns a binary image containing the points with maximum intensity of the gradient of the sound field. Fig. 3a shows the position of the image space with respect to the Cartesian plane.

The Cartesian coordinates of a point with i and j row and column indices are

$$x_i = \frac{(i - \lfloor (N_{im} + 1)/2 \rfloor)}{(N_{im} + 1)/2}, \quad y_j = \frac{(j - \lfloor (N_{im} + 1)/2 \rfloor)}{(N_{im} + 1)/2}.$$

We adopt the same parameter space described in [15]. The equation of a circle on a plane is $(x - a)^2 + (y - b)^2 = r^2$, where (x, y) are the Cartesian coordinates of points in the image space; a and b are the coordinates of the center of the circle; and r its radius.

As defined in [15], the parameter space $\mathcal{P} \subset \mathbb{R}^3$ is a grid that samples the 3D space with coordinate axes a, b, r and centered at (x_0, y_0, r_0) . The grid on the (a, b) axes has $N_p \times N_p$ points with a

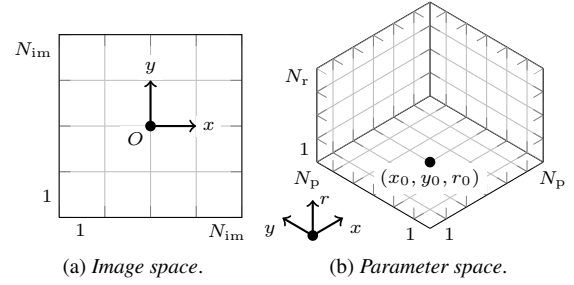


Fig. 3: Image space and parameter space mapped on the Cartesian plane.

resolution Δ_p m/pt. The r -axis is sampled at N_r points with resolution Δ_r m/pt between a minimum radius r_{min} and a maximum radius r_{max} . Fig. 3b shows the position of the parameter space with respect to the Cartesian plane.

The point on the grid with indices (l, m, k) , corresponds to a point in $\mathcal{P}$ with coordinates

$$x_l = l \cdot \Delta_p - \frac{N_p \cdot \Delta_p}{2} + x_0, \quad (1)$$

$$y_m = m \cdot \Delta_p - \frac{N_p \cdot \Delta_p}{2} + y_0, \quad (2)$$

$$r_k = k \cdot \Delta_r + r_{min}. \quad (3)$$

A point (x_i, y_j) in the image space is mapped into the cone

$$(x_i - a)^2 + (y_j - b)^2 - r^2 = 0 \quad (4)$$

in the parameter space. Let us now consider a set $\mathcal{W} = \{(x_i, y_j) : (x_i - \hat{a})^2 + (y_j - \hat{b})^2 = \hat{r}^2\}$, of points that belong to a circle. The cones generated by the points in $\mathcal{W}$ intersect at $(\hat{a}, \hat{b}, \hat{r}) \in \mathcal{P}$, whose coordinates are the parameters of the considered circle [15].

In order to compute $(\hat{a}, \hat{b}, \hat{r})$ we introduce on the parameter space the accumulation function $A : \mathcal{P} \rightarrow \mathbb{R}$, which initially is zero for $l = 1 \dots N_p, m = 1 \dots N_p, k = 1 \dots N_r$. Consider now the distance

$$d = \sqrt{(x_i - x_l)^2 + (y_j - y_m)^2}$$

for $i = 1 \dots N_{im}, j = 1 \dots N_{im}, l = 1 \dots N_p, m = 1 \dots N_p$. We select the radius $r_{\hat{k}}$ closest to d

$$r_{\hat{k}} = \arg \min_{r_k} (|d - r_k|), \quad k = 1 \dots N_r. \quad (5)$$

If $|d - r_{\hat{k}}| < \rho$, where ρ is an appropriate threshold, we increment the accumulation function, i.e. $A(l, m, \hat{k}) = A(l, m, \hat{k}) + 1$.

Consider now the case of an image containing two concentric circles. All the points in the image space belonging to one circle are mapped in the parameter space into cones intersecting at $(\bar{l}, \bar{m}, \bar{k}')$, whereas the points belonging to the other circle will be mapped into cones intersecting at $(\bar{l}, \bar{m}, \bar{k}'')$. As the circles are concentric, they share the same parameters $\bar{l}, \bar{m}$, but they have different radii. We take advantage of this consideration in order to reduce the dimensionality of the parameter space.

First, in order to increase the robustness of the proposed method against noise, we hard-limit the accumulation function adopting a threshold ε

$$A(l, m, k) = \begin{cases} 0, & \text{if } a(l, m, k) < \varepsilon \\ A(l, m, k), & \text{if } a(l, m, k) \geq \varepsilon \end{cases} \quad (6)$$

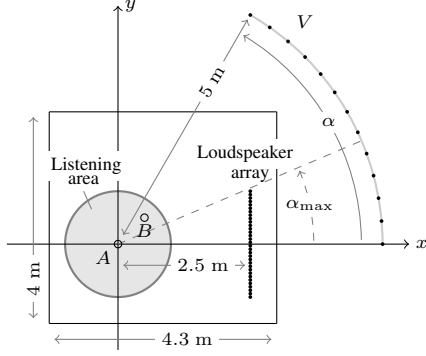


Fig. 4: Setup of the loudspeaker array and the listening area.

We then define a reduced accumulation function

$$A'(l, m) = \sum_{k=1}^{N_r} A(l, m, k).$$

The virtual source location is finally found as

$$(\hat{l}, \hat{m}) = \arg \max_{(l, m)} (A'(l, m)). \quad (7)$$

Algorithm 1 summarizes the proposed localization method. The localization technique presented here can work also on portions of the image space. Next section, with the aid of an illustrative example, will make clear the advantages of this generalization.

Algorithm 1 Estimation of the virtual acoustic source location.

```

for  $l, m = 1 \dots N_p, k = 1 \dots N_r$  do
   $a(l, m, k) \leftarrow 0$ 
end for
for  $i, j = 1 \dots N_{im}$  do
  for  $l, m = 1 \dots N_p$  do
     $d \leftarrow \sqrt{(x_i - x_l)^2 + (y_j - y_m)^2}$ 
     $r_{\hat{k}} \leftarrow \arg \min_k (|d - r_k|)$ 
    if  $|d - r_{\hat{k}}| < \rho$  then
       $a(l, m, \hat{k}) \leftarrow a(l, m, \hat{k}) + 1$ 
    end if
  end for
end for
for  $l, m = 1 \dots N_p, k = 1 \dots N_r$  do
  if  $a(l, m, k) < \varepsilon$  then
     $a(l, m, k) \leftarrow 0$ 
  end if
end for
for  $l, m = 1 \dots N_p$  do
   $a'(l, m) \leftarrow 0$ 
end for
for  $n = 1 \dots N_r$  do
   $a'(l, m) \leftarrow a'(l, m) + a(l, m, n)$ 
end for
 $(\hat{l}, \hat{m}) \leftarrow \arg \max_{(l, m)} (a'(l, m))$ 

```

4. RESULTS

In this section we present some simulative and experimental results to validate the proposed localization methodology. Moreover, in order to test the accuracy, we compare these results with those obtained in a subjective listening test.

The sound reproduction system is composed by a linear array of $M = 32$ equally spaced loudspeakers with an extension of $l =$

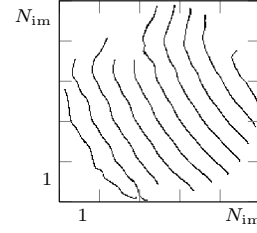


Fig. 5: Edge image of a sound field which exhibits spatial distortion.

2.035 m. The middle point of the array is distant 2.5 m from the center of the circular listening area, which is also the origin of the reference frame. The radius of the listening area is 0.86 m and coincides with the radius of the circle described by the rotating microphone. Fig. 4 depicts the setup of the loudspeaker array and the area. The spatial Nyquist frequency of the loudspeaker array is $f_{alias} = c/2d \approx 2.7\text{kHz}$, where $c = 343\text{ m/s}$ is the speed of sound at 20°C and $d = l/M = 6.3\text{ cm}$ is the distance between adjacent loudspeakers.

4.1. Simulations

To perform the simulations and the experiments we have employed two rendering techniques: *Geometric Rendering (GR)* ([18]) and *Wave Field Synthesis (WFS)* ([19]). Both techniques render an omnidirectional virtual source distant 5 m from the center of the listening area and with a variable *Direction Of Arrival (DOA)* α , as shown in Fig. 4. We define a grid of $N = 201 \times 201$ control points c_n , $n = 1 \dots N$ inside the listening area. The sound field $\mathbf{p}(\omega)$ is simulated on c_n by applying the filter coefficients $\mathbf{h}_m(\omega)$, $m = 1 \dots M$ to the $N \times M$ propagation matrix $\mathbf{G}(\omega)$, i.e.

$$\mathbf{p}(\omega) = \mathbf{G}(\omega)\mathbf{h}(\omega). \quad (8)$$

With reference to Fig. 4, α_{max} is the maximum DOA for which the line connecting the virtual source and the center of the reference frame intersects the array. It is known in the literature [8, 20] that for $\alpha > \alpha_{max}$ the accuracy of the rendering decreases and some artifacts appear in the sound field. More specifically, the curvature of the wavefronts is not consistent with the position of the desired virtual source. For instance, consider the illustrative sound field depicted in Fig. 5: here the wavefronts in the upper part of the figure exhibit a different curvature than in the rest of the listening area. In order to quantify position-dependent distortions of the wavefront, the localization procedure can be repeated for arbitrary zones in the listening area. For the sake of example, we show results for two zones.

We center the parameter space in $(x_0 = 3.75\text{ m}, y_0 = 2.15\text{ m})$ and sample it on $N_p = 305$ pt with a resolution $\Delta_p = 1.64\text{ cm/pt}$. The r -axis of the parameter space is sampled at $N_r = 61$ pt with a resolution $\Delta_r = 5\text{ cm/pt}$ between $r_{min} = 3.5\text{ m}$ and $r_{max} = 6.5\text{ m}$. We consider two square zones of the image space: one centered in point A and one centered in point B, with dimension $101\text{ pt} \times 101\text{ pt}$ and resolution $\Delta_{im} = 1\text{ cm/pt}$. The reference frame, for all evaluations, remains the same, i.e. the estimated DOA $\hat{\alpha}$ is consistent for all analysis positions. The localization algorithm introduced in Section 3 is tuned with $\rho = 2.5\text{ cm}$ and $\varepsilon = 70\%$ of the maximum value of $a(l, m, k)$.

We simulated the sound fields of GR and WFS for $\alpha = 0^\circ \div 60^\circ$ and $f = 900\text{ Hz}$. Fig. 6 shows the DOA $\hat{\alpha}$ estimated from the simulations as a function of the actual DOA α . Both GR and WFS

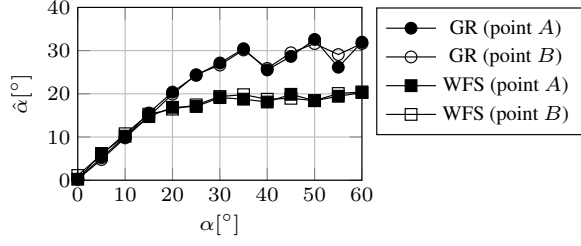


Fig. 6: DOA $\hat{\alpha}$ versus the actual DOA α for GR and WFS in positions A and B, as from Fig. 4.

are able to correctly render the sound field only for $\alpha < \alpha_{\max}$, as expected. In this simulative scenario there is no noticeable difference between localizations in points A and B.

4.2. Experiments

In order to apply the localization procedure on real data, we employ the sound field measurement methodology described in Section 2. For this purpose the sound field is sampled at $N_{\text{mic}} = 180$ points on a circumference of radius $\rho_0 = 86$ cm using a condenser cardioid microphone *AKG C1000s*. The experiments are performed with the same setup depicted in Fig. 4, hosted in a semi-anechoic room with reverberation time $T_{60} \approx 50$ ms. The excitation signal is a sinusoid with frequency $f = 900$ Hz, rendered by both GR and WFS for $\alpha = 0^\circ \div 50^\circ$. We also carried out a set of listening tests with the same setup adopted for simulations and sound field measurements. The goal is to assess the relationship existing between the localization accuracy of the proposed methodology and that of a panel of expert listeners, who were asked to assess the DOA of a virtual source emitting an excerpt from Suzanne Vega's *Tom's Diner*, converted into a monophonic signal and rendered with both GR and WFS. The virtual source is omnidirectional, distant 5 m for $\alpha = 0^\circ \div 50^\circ$. We presented 12 stimuli to the listeners (6 rendered with GR and 6 with WFS) and asked them to annotate the DOA $\hat{\alpha}_s$ of each stimulus. An angular scale ranging from -40° to 40° with a resolution of 10° was provided as an hint.

Fig. 7 shows the results. $\hat{\alpha}_m$ and $\hat{\alpha}_s$ are, respectively, the measured DOA and the average DOA obtained from subjective tests. The 95% confidence interval, according to the recommendation in [5], is also shown for the subjective listening tests. In all the cases where the real rendering system is able to reproduce the wavefronts $\hat{\alpha}_m$ well approximates $\hat{\alpha}_s$. Fig. 7a and 7b do not show $\hat{\alpha}_m$ for $\alpha = 30^\circ, 40^\circ$, respectively, because the sound field does not exhibit regular wavefronts. For those angles the confidence intervals of the subjective listening tests are larger, meaning a less accurate estimate. As expected, accurate rendering is possible only for $\alpha < \alpha_{\max}$. Due to non-idealities in the loudspeaker array, subjective results provide different estimates in points A and B. This difference was not evident with simulative data. Notice that, though subtle, these differences are well captured by the proposed localization methodology, thus confirming its accuracy.

5. CONCLUSIONS

In this paper we have presented a technique for the localization of virtual acoustic sources for sound field rendering applications. We first measure the sound field in the listening area, which is composed by multiple concentric wavefronts originating from the virtual

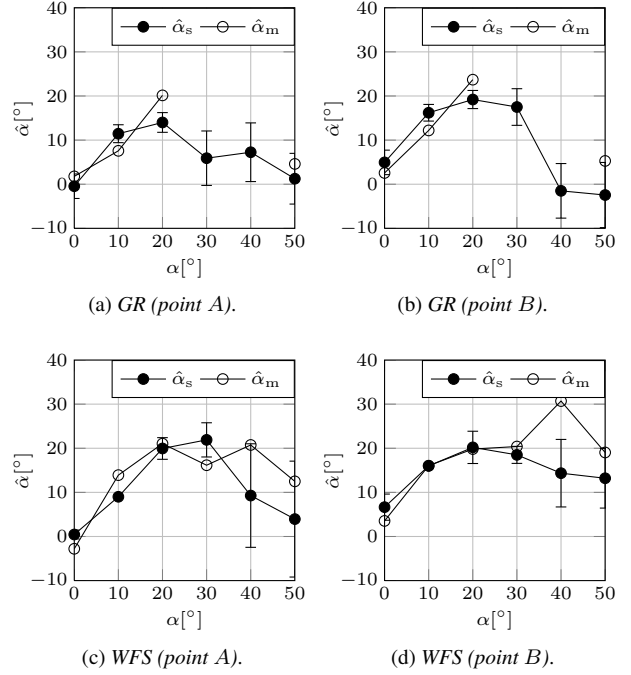


Fig. 7: DOAs $\hat{\alpha}_m$ and $\hat{\alpha}_s$ versus the actual DOA α for GR and WFS in positions A and B, as from Fig. 4.

source location. By means of a generalized Hough transform, we localize the common center of the wavefronts. With a single acquisition of the sound field, localization can be accomplished in an arbitrary number of points within the listening area. The presented technique turns out to be accurate. In fact, results show that results of subjective listening tests can be predicted with a good accuracy, capturing also small localization impairments.

6. RELATION TO PRIOR WORK

The work presented here enables a position-dependent objective localization of the virtual sound source using a single acquisition and with an accuracy comparable to that of listening tests that follow recommendations in [5]. We mutated the measurement methodology in [11], but with different goals with respect to previous works [12, 13]. As in [7, 8, 9, 10], localization is accomplished through objective measurements. In [9] localization is performed for sound field coding purposes exploiting the inverse operation of vector-based amplitude panning. In [10] the authors exploit the precedence effect and inter-aural time difference to perform localization analysis of the sound field acquired by means of a parallel circular microphone array. In [7] the authors perform localization on binaural measurements using the inter-aural time difference. In [8] the analysis is limited to a qualitative comparison of real world and simulated spatial responses. For all the mentioned references, acquisitions must be repeated if position-dependent localization is in order.

In order to validate the proposed methodology, we performed a subjective test campaign adopting the methodology described in [4], which, at its turn, mutated recommendations for the evaluation of small impairments from [5].

7. REFERENCES

- [1] H. Wittek, *Perceptual Differences between Wavefield Synthesis and Stereophony*, Ph.D. thesis, Dep. of Music and Sound Recording, University of Surrey, Guildford, UK, Oct. 2007.
- [2] B. Klehs and T. Sporer, "Wave Field Synthesis in the Real World: Part 1 - In the Living Room," in *114th Audio Engineering Society (AES) Conv.*, Amsterdam, NL, Mar. 2003.
- [3] T. Sporer and B. Klehs, "Wave Field Synthesis in the Real World: Part 2 - In the Movie Theatre," in *116th Audio Engineering Society (AES) Conv.*, Berlin, DE, May 2004.
- [4] International Telecommunication Union - Radiocommunication Assembly, *BS.1284 - General Methods for the Subjective Assessment of Sound Quality*, 2003.
- [5] International Telecommunication Union - Radiocommunication Assembly, *BS.1116 - Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, 1997.
- [6] H. Wierstorf, A. Raake, and S. Spors, "Localization of a Virtual Point Source within the Listening Area for Wave Field Synthesis," in *133rd Audio Engineering Society (AES) Conv.*, San Francisco, CA, USA, Oct. 2012.
- [7] E. Bates, G. Kearney, D. Furlong, and F. Boland, "Localization Accuracy of Advanced Spatialisation Techniques in Small Concert Halls," *The Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. 3069–3070, May 2007.
- [8] E. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, Delft, NL, 1998.
- [9] N. Kamado, M. Hirata, H. Saruwatari, and K. Shikano, "Object-Based Stereo Up-Mixer for Wave Field Synthesis Based on Spatial Information Clustering," in *Proc. of 20th European Signal Processing Conference*, Bucharest, RO, Aug. 2012.
- [10] T. Kimura, Y. Yamakata, M. Katsumoto, and K. Takehi, "Localization Model of Synthesized Sound Image Using Precedence Effect in Sound Field Reproduction Based on Wave Field Synthesis," *IEICE Transactions of Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 6, pp. 1310–1319, Jun. 2008.
- [11] A. Kuntz and R. Rabenstein, "Cardioid Pattern Optimization for a Virtual Circular Microphone Array," in *EAA Symposium on Auralization*, Helsinki, FI, Jun. 2009.
- [12] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, "A Methodology for Evaluating the Accuracy of Wave Field Rendering Techniques," in *Proc. of 2011 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Prague, CZ, May 2011.
- [13] L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, and S. Tubaro, "A Psychoacoustic-Based Analysis of the Impact of Pre-Echoes and Post-Echoes in Soundfield Rendering Applications," in *Proc. of IWAENC 2012, Int. Workshop on Acoustic Signal Enhancement*, Aachen, DE, Sept. 2012.
- [14] P. V. C. Hough, "Method and Means for Recognizing Complex Patterns," Dec. 1962.
- [15] R. O. Duda and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.
- [16] F. Antonacci, M. Matteucci, D. Migliore, D. Riva, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking Multiple Acoustic Sources in Reverberant Environments using Regularized Particle Filter," in *Proc. of 2007 15th Int. Conf. on Digital Signal Processing*, Cardiff, Wales, UK, July 2007.
- [17] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, Nov. 1986.
- [18] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield Rendering with Loudspeaker Arrays through Multiple Beam Shaping," in *Proc. of 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2009.
- [19] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic Control by Wave Field Synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.
- [20] S. Spors, R. Rabenstein, and J. Ahrens, "The Theory of Wave Field Synthesis Revisited," in *124th Audio Engineering Society (AES) Conv.*, Amsterdam, NE, May 2008.