TOWARDS OPTIMAL LOUDSPEAKER PLACEMENT FOR SOUND FIELD REPRODUCTION

Hanieh Khalilian, Ivan V. Bajić, and Rodney G. Vaughan

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

With a view to the suppression of unwanted sound, a planar array of loudspeakers is used to recreate a sound field in a nearby cubic listening area. Using free-space propagation, the formulation for selecting optimal locations of loudspeakers is developed so that numerical experiments can give a feel for the best possible suppression. First, to provide a benchmark, a target Acoustic Transfer Function (ATF) matrix is found that minimizes the reproduction error for a number of uniformly placed loudspeakers. Then the same number of loudspeakers is positioned by selecting from densely placed candidates so that their ATF matrix best approximates the target sound field. For recreating the field of a point source located in the direction of the array, the loudspeaker array with selected locations is shown to improve the reproduction accuracy over a reasonable bandwidth.

Index Terms— Active sound cancellation, sound field reproduction, noise cancellation, loudspeaker arrays

1. INTRODUCTION

Sound field reproduction is the process of synthesizing a sound field in a region of interest - the listening area - using an array of loudspeakers. This synthesis can be either for creating a spatial sound effect, or for creating a sound-suppressed volume by canceling the sound field. In the context of noise cancellation and adaptive algorithms, the term "desired field" is used for the negative of the undesired field which is to be canceled, so it is also referred to here as a target field.

Previous work on sound field reproduction falls under three terminologies [1]: wave field synthesis (WFS), higher order ambisonics (HOA), and direct approximation methods. WFS is based on the Kirchoff-Helmholtz integral equation, which states that the pressure in an enclosed volume is completely determined by the velocity and pressure on the surface enclosing that volume. In a limiting form, this principle predicts that a sound field sourced from a half space can be recreated (and therefore also canceled) by the pressure and velocity on the half space planar boundary plane [2, 3, 4]. In HOA methods, the desired, or target field, is expressed in spherical harmonics to facilitate finding the phase and magnitude for the loudspeakers [4, 5]. In [6], HOA was used to recreate a plane wave in a sphere of radius 0.2 m using an array of loudspeakers located on a sphere of radius of 1 m. In [7], a plane wave was recreated in a sphere of radius 0.11 m using loudspeakers on three surrounding rings, the largest of which was 2 m. Direct approximation methods minimize the energy of the approximation error between the synthesized field and the desired field at a set of sampling points. In [8] the Least Squares (LS) criterion is employed to recreate a plane wave within a square of length of 1 m by placing loudspeakers along the perimeter of a 7.5 m $\times 6.4$ m rectangle. In [1], the Least Absolute Shrinkage and Selection Operator (LASSO) was employed to recreate a spherical wave within circular and amphitheatrical surfaces. In [9], first, the regularization factor is found using singular value decomposition (SVD) of the ATF matrix, and then a spherical wave is recreated within a circular surface using a circular array of loudspeakers.

Most of the above papers use modeling based on free-space propagation, and omnidirectional loudspeakers and field samplers. In this paper the same situation applies because there is interest in the best possible performance available from idealized configurations, except for being spatially under-sampled. In real-world conditions, noise cancellation cannot be expected to perform as well as that modeled for free space.

The contribution of this paper is to use direct approximation for optimizing the discretized placement of loudspeakers, which has not been considered in previous work. First, the SVD of the ATF matrix is performed on uniform loudspeaker placement. Then, the entries of the ATF matrix are modified in order to decrease the approximation error and the total loudspeaker power simultaneously. This offers a benchmark performance for the number of loudspeakers and the configuration. Finally, the loudspeakers are relocated to best match the modified ATF matrix, and the LS solution of the new structure is found. The relocated loudspeakers outperform the uniform placement in both the approximation error and power.

2. PROPOSED METHOD

Let s be the vector containing complex amplitudes of N loudspeakers ($\mathbf{s} = [s_1, s_2, ...s_N]^T$), $p_{\max} \geq ||\mathbf{s}||^2$ be the maximum normalized power of the loudspeakers, $\mathbf{p}^{des}$ be the vec-

This work was supported in part by the NSERC grant EGP 445356-12.

tor containing the complex values of the desired sound field at $M > N$ sampling points, and $\mathbf{p}$ be the corresponding values of the synthesized field at the same sampling points. Direct least squares approximation attempts to solve $(\min ||\mathbf{p}^{des} - \mathbf{p}||^2$ s.t. $||\mathbf{s}||^2 \leq p_{\max})$, usually by minimizing

$$J = ||\mathbf{p}^{des} - \mathbf{p}||^2 + \gamma||\mathbf{s}||^2 \tag{1}$$

where γ is the regularization parameter. The monochromatic field from the loudspeakers at the sampling points is given by

$$\mathbf{p} = \mathbf{G}\mathbf{s} \tag{2}$$

where the ATF matrix $\mathbf{G}$ is $M \times N$, and its (m, n)-th entry, $G_{m,n}$, is the ATF of the n-th loudspeaker at the m-th sampling point. For example, assuming free space propagation, the ATF at point $\mathbf{y}_m$ from an omnidirectional loudspeaker located at $\mathbf{x}_n$ is

$$G_{m,n} = \frac{1}{4\pi} \cdot \frac{e^{-jk||\mathbf{x}_n - \mathbf{y}_m||}}{||\mathbf{x}_n - \mathbf{y}_m||}. \tag{3}$$

The major assumption of free space propagation allows simplistic reverberation modeling, and incorporation of the patterns of the loudspeakers and sampling microphones in (3), in turn allowing expressions such as (2), for fields, to also represent the signals for processing.

The SVD of the ATF matrix is $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are, respectively, $M \times M$ and $N \times N$ unitary matrices, and $\mathbf{\Sigma}$ is a $M \times N$ diagonal matrix containing N ordered singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_N$ of $\mathbf{G}$. According to [9], the approximation error and the loudspeaker power resulting from solving (1) are, in terms of the entries of SVD matrices,

$$||\mathbf{p}^{des} - \mathbf{p}||^2 = \sum_{n=1}^{N} \frac{\gamma^2}{(\sigma_n^2 + \gamma)^2}|c_n|^2 + \sum_{n=N+1}^{M} |c_n|^2, \tag{4}$$

$$||\mathbf{s}||^2 = \sum_{n=1}^{N} \frac{\sigma_n^2}{(\sigma_n^2 + \gamma)^2}|c_n|^2, \tag{5}$$

where $c_n = \mathbf{u}_n^H \mathbf{p}^{des}$ is the projection of $\mathbf{p}^{des}$ onto the n-th column of $\mathbf{U}$. Both the approximation error and the loudspeaker power can be reduced if the $|c_n|$'s are decreased. Since $|c_n|$ is the magnitude of the projection of $\mathbf{p}^{des}$ onto the n-th column of $\mathbf{U}$, we will attempt to minimize the $|c_n|$'s by designing the matrix $\mathbf{U}$ appropriately. This new matrix ($\mathbf{U}^{ideal}$) will in turn generate new ATF matrix $\mathbf{G}$, which becomes $\mathbf{G}^{ideal}$, below.

Suppose we pick the m'-th column of $\mathbf{U}^{ideal}$, denoted $\mathbf{u}_{m'}^{ideal}$, and make it parallel with $\mathbf{p}^{des}$, then

$$\mathbf{u}_{m'}^{ideal} = \mathbf{p}^{des}/||\mathbf{p}^{des}||,$$

so $|c_{m'}| = |(\mathbf{u}_{m'}^{ideal})^H \mathbf{p}^{des}| = ||\mathbf{p}^{des}||$. Other columns of $\mathbf{U}^{ideal}$ must be unit vectors orthogonal to $\mathbf{u}_{m'}^{ideal}$, so they lie in the $(M-1)$-dimensional nullspace of $\mathbf{u}_{m'}^{ideal}$. With all $\mathbf{u}_q^{ideal}$ with $q \neq m'$ orthogonal to $\mathbf{p}^{des}$, we get $|c_q| = 0$ for $q \neq m'$, so (4) and (5) simplify to

$$||\mathbf{p}^{des} - \mathbf{p}||^2 = \begin{cases} \frac{\gamma^2}{(\sigma_{m'}^2 + \gamma)^2}||\mathbf{p}^{des}||^2 & \text{if } m' \leq N, \\ ||\mathbf{p}^{des}||^2 & \text{if } m' > N. \end{cases} \tag{6}$$

$$||\mathbf{s}||^2 = \begin{cases} \frac{\sigma_{m'}^2}{(\sigma_{m'}^2 + \gamma)^2}||\mathbf{p}^{des}||^2 & \text{if } m' \leq N, \\ 0 & \text{if } m' > N. \end{cases} \tag{7}$$

The question now is – which m' should we pick? Since $\gamma^2/(\sigma_{m'}^2 + \gamma)^2 < 1$, we see that the approximation error in (6) will be minimized if we pick $m' \leq N$. Further, since σ_1 is the largest singular value, then $||\mathbf{p}^{des}||^2\gamma^2/(\sigma_1^2 + \gamma)^2$ is the smallest error we can achieve. Hence, we pick $m' = 1$, i.e., $\mathbf{u}_1^{ideal} = \mathbf{p}^{des}/||\mathbf{p}^{des}||$. This defines $\mathbf{u}_q^{ideal}$ for $q > 1$ as an orthonormal basis of the nullspace of $\mathbf{u}_1^{ideal}$. This arrangement of $\mathbf{u}^{ideal}$'s make a new matrix $\mathbf{U}^{ideal}$ with a new ATF matrix

$$\mathbf{G}^{ideal} = \mathbf{U}^{ideal}\mathbf{\Sigma}\mathbf{V}^H. \tag{8}$$

$\mathbf{G}^{ideal}$ has the same singular values as $\mathbf{G}$.

Using $\mathbf{G}^{ideal}$, the approximation error and the total loudspeaker power are given by (6) and (7), respectively, with $m' = 1$. There are options for attempting to realize this ATF matrix, and the choice of interest is to fix the sampling points and approximate the "ideal" ATF matrix in (8) by adjusting the locations of loudspeakers. In practice, this may be from a screen of reconfigurable, densely spaced transducers (where power consumption is important), or set up as a fixed configuration for a fixed, targeted sound field.

3. LOUDSPEAKER PLACEMENT

N loudspeaker locations $\mathbf{x}_n$ are sought that minimize the approximation error at M sampling points $\mathbf{y}_m$. The following algorithm selects discetized locations.

Step 1: Distribute N loudspeakers uniformly across the planar square region (loudspeaker region) shown in Fig. 1.

Step 2: Compute the ATF matrix $\mathbf{G}$ for the structure in Step 1, and modify it as explained in the previous section to create $\mathbf{G}^{ideal}$. Let $\mathbf{g}_n^{ideal}$ be the n-th column of $\mathbf{G}^{ideal}$. This M-dimensional vector contains the ATF of the n-th loudspeaker at M sampling points.

Step 3: Choose a number $N_v \gg N$, distribute N_v virtual loudspeakers uniformly across the loudspeaker region (l-th virtual loudspeaker located at $\mathbf{x}_l^v$), and compute the ATF of each virtual loudspeakers at M sampling points. Store the results in an $M \times N_v$ matrix $\mathbf{H}$. The l-th column of $\mathbf{H}$, $\mathbf{h}_l$, contains the ATF of the l-th virtual loudspeaker at each of the M sampling points.

Step 4: For each $n = 1, 2, ..., N$, find $\mathbf{h}_l$ that best matches $\mathbf{g}_n^{ideal}$,

$$l^*(n) = \underset{l}{\operatorname{argmin}}\ d(\mathbf{h}_l, \mathbf{g}_n^{ideal}),$$

and place the n-th loudspeaker at $\mathbf{x}^v_{l^*(n)}$, the location of the $l^*(n)$-th virtual loudspeaker. In our experiments, as a measure of matching between $\mathbf{h}_l$ and $\mathbf{g}_n^{ideal}$, we used the negative of the absolute value of the inner product between $\mathbf{h}_l$ and $\mathbf{g}_n^{ideal}$, that is $d(\mathbf{h}_l, \mathbf{g}_n^{ideal}) = -|\mathbf{h}_l^H \mathbf{g}_n^{ideal}|$. Although $d(\cdot, \cdot)$ is not a formal matrix metric, it has several nice features in the context of our problem: (1) orthogonal vectors are most "distant" from each other ($d = 0$ for orthogonal vectors, which is the highest it can be), and (2) unlike Euclidean distance, d is invariant to vector inversion (multiplication by -1), which is important in our case because SVD is unique only up to a unit-phase factor of left- and right-singular vectors.

Step 5: After finding the locations, calculate the ATF matrix of the new arrangement, and find the complex amplitude of the loudspeakers ($\mathbf{s}$) by solving (1).

4. EXPERIMENTAL RESULTS

The formulation allows parametric study, but is not yet tied to any particular configuration. The interest here is on the structure of Fig. 1. The loudspeakers occupy a 3 m $\times$3 m square centered at the origin in the $x - y$ plane, while the listening region is a 1m cube, located 1 m away from the $x - y$ plane in the direction of positive z axis. The sampling points are uniform throughout the cube, with Cartesian spacing of 0.2 m, so the number of sampling points is $M = 125$. (The sample locations are a half sample spacing from the edges of the cube.) The number of loudspeakers is $N = 25$, with a Cartesian spacing of 0.5 m. The number of virtual loudspeakers in Step 3 of the placement algorithm was $N_v = 400$ (Cartesian spacing of 0.15 m). The desired field is from a point source at $(x, y, z) = (0, 0, -8)$, with complex amplitude equal to 8. The regularization factor is taken from [9]. Spot frequencies of interest, elaborated below, (and their wavelengths) are 200 Hz (1.7 m), 600 Hz (0.56 m) and 2000 Hz (0.17 m). So it is clear that for the higher frequencies the planar aperture is under-sampled; for example at 2000 Hz, the loudspeaker spacing of 1.17 wavelengths is under-sampled, although the directional coverage (directions encompassing the cube) being smaller than π helps somewhat. The angular extent of the planar aperture as viewed from the cube affects the spatial sampling requirement in the cube. From these considerations, achieving perfect performance is unlikely, depending on the target field, even at single frequencies. The experiments, reported below with the free space conditions, offer a feel for the best possible performance.

In the first experiment, the frequency of the desired wave is 600 Hz, and the maximum normalized loudspeaker power is $p_{\max} = 0.1$. Fig. 2 shows the real parts of the desired

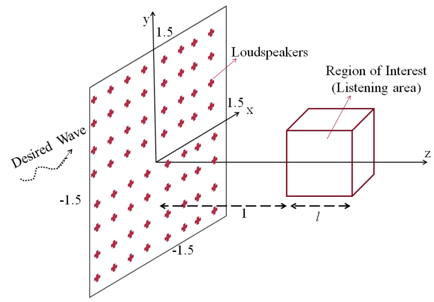
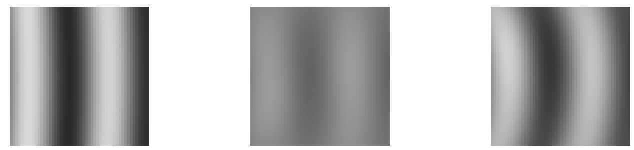


Fig. 1. Sound field reproduction in a cube by a planar array.



(a) (b) (c)

Fig. 2. Pressure at a plane in the cube at 600 Hz, indicating the wave structure. Real parts of (a) the desired field, (b) the field produced by uniform loudspeaker placement and (c) the field produced by the modified placement.

field (left column), the field produced by the uniform loudspeaker distribution (middle column), and the field produced by the modified loudspeaker placement (right column), on a plane that passes through $(0, 0, 1)$ and $(1, 0, 1)$ in the cubic listening region. By eyeballing the figures, it is evident that the modified loudspeaker placement offers a better reproduction of the desired field, compared to a uniform loudspeaker distribution. To quantify the improvement, the relative error is used,

$$\text{Error (dB)} = 10 \log_{10} \left(\frac{||\mathbf{p}^{des} - \mathbf{p}||^2}{||\mathbf{p}^{des}||^2} \right). \tag{9}$$

For this calculation, the desired or target field $\mathbf{p}^{des}$ and the reproduced field $\mathbf{p}$ are evaluated at more than the $M = 125$ sampling points used to generate loudspeaker placement. Here, $125,000$ uniformly spaced points are used in (9). This ensures that the cube is comfortably over-sampled - even at 2000 Hz the Cartesian spacing is about a tenth of a wavelength.

Fig. 3 shows the error obtained for various frequencies of the desired wave between 200 Hz (spatially oversampled) and 2000 Hz (undersampled). The modified placement reduces the error by 6 dB at mid-frequencies, and by up to 2 dB at high frequencies, relative to uniform loudspeaker placement.

In the next experiment, the frequency is 600 Hz, and the maximum normalized loudspeaker power $p_{\max}$ is varied. Fig. 4 shows that the error of the modified placement is less than that of the uniform placement for the same maximum power. Equivalently, for the same approximation error, the

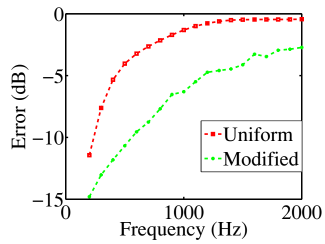


Fig. 3. Relative error for the uniform and modified loudspeaker placement for various frequencies of the desired field.

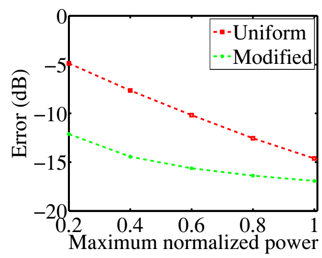


Fig. 4. Relative error for the uniform and modified loudspeaker placement for various values of $p_{\max}$, at 600Hz

modified placement consumes less power.

In Fig. 4, the gap between the errors produced by uniform and modified placement reduces as the maximum power increases. However, this gap increases at higher frequencies. For example, in Fig. 4, for 600 Hz, the gap between error curves at $p_{\max} = 1$ is about 2 dB. At 1200 Hz, this error gap is about 10 dB. The desired and reproduced fields at 1200 Hz are shown in Fig. 5 on a plane that passes through $(0, 0, 1)$ and $(1, 0, 1)$ in the cubic listening region.

Next we seek a feel for the bandwidth limitations of the configuration. The placement is optimized for 300 Hz and 1000 Hz, denoted P300 and P1000, respectively, and their error plotted from 300 Hz to 1000 Hz in Fig. 6, along with the error for the uniform placement. The maximum normalized power is 0.1. The placement, optimized at one frequency, say 300 Hz, is seen to work well over a wide bandwidth.

In the next experiment, the sensitivity of the system to the source location is sought. The target field's point source is placed at $(0, 0, -4)$, $(0, 0, -12)$, $(0, 8, -8)$, $(4, 0, -8)$, and $(4, 4, -4)$, and for each case, the loudspeaker locations are optimized separately. The frequency is 600 Hz, and the maximum normalized power is 0.2. The error values are in Table 1, showing that the modified placement reduces the error relative to uniform placement by 3–6 dB. Fig. 7 shows the selected locations (blue) when the source is at $(0, 8, -8)$, with red dots indicating the examined locations of the virtual loudspeakers. Note that due to the location of the source ($x = 0, y > 0$), the selected loudspeaker locations cluster around the plane $x = 0$ and in the upper part of the array.

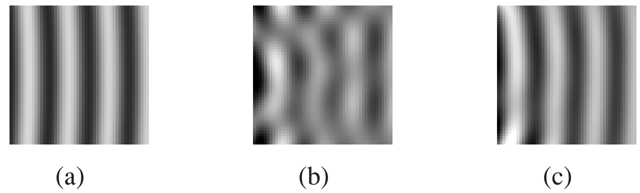


Fig. 5. Same as Fig. 2, but for 1200 Hz.

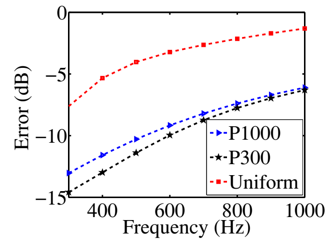


Fig. 6. Relative error for the uniform and modified placement optimized at 300 Hz (P300) and 1000 Hz (P1000).

Table 1. Relative error in dB of the uniform placement and the modified one, for various locations of the desired source

	(0,0,−4)	(0,0,−12)	(0,8,−8)	(4,0,−8)	(4,4,−4)
Uniform	−1.70	−4.91	−2.93	−3.81	−2.79
Modified	−5.55	−11.41	−9.63	−9.24	−5.64

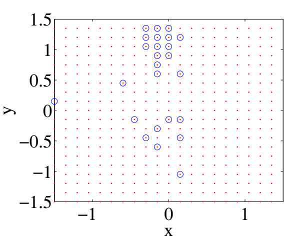


Fig. 7. Selected locations of the loudspeakers.

5. CONCLUSION

The formulation for optimizing loudspeaker placement for sound field reproduction has been presented, allowing parametric study of the configuration. The physical basis is the ATF matrix of the loudspeakers to the sample points, and here, idealized (no reverberation, dispersion-free) free space propagation was used. Densely spaced candidate loudspeaker locations offer a selection pool for finding the best-performing set of placements. The modified placements, relative to a uniform loudspeaker arrangement, result in a lower reproduction error and lower power consumption. Numerical experiments demonstrate that the modified spacing improves the relative approximation error between 3 dB and 10 dB, depending on the configuration.

6. REFERENCES

[1] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using lasso," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, pp. 1902–1921, Nov. 2010.

[2] A. J. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Amer.*, vol. 93, pp. 27642778, May 1993.

[3] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, pp. 2038–2050, Nov. 2010.

[4] J. Ahrens and S. Spors, "A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," in *In 125th Conv. of the AES*, San Francisco, CA, Oct. 2008.

[5] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *In AES 116th Convention Preprints*, Berlin, Germany, May 2004.

[6] D. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 9, pp. 697–707, Sep. 2001.

[7] A. Gupta and T. Abhayapala, "Three-dimensional sound field reproduction using multiple circular loudspeaker arrays," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, pp. 1149–1159, July 2011.

[8] P. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 2, pp. 662–678, Feb. 2005.

[9] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, pp. 2388–2392, Oct. 2012.