JOINT SOURCE LOCALIZATION AND DEREVERBERATION BY SOUND FIELD INTERPOLATION USING SPARSE REGULARIZATION

Niccolò Antonello[1], Enzo De Sena[2], Marc Moonen[1], Patrick A. Naylor[3] and Toon van Waterschoot[1,4]

[1] KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), 3001 Leuven, Belgium
[2] University of Surrey, Institute of Sound Recording, GU2 7XH, Guilford, UK
[3] Imperial College London, Dept. of Electrical Engineering, SW7 2AZ, London, UK
[4] KU Leuven, Dept. of Electrical Engineering (ESAT-ETC), 3000 Leuven, Belgium

ABSTRACT

In this paper, source localization and dereverberation are formulated jointly as an inverse problem. The inverse problem consists in the interpolation of the sound field measured by a set of microphones by matching the recorded sound pressure with that of a particular acoustic model. This model is based on a collection of equivalent sources creating either spherical or plane waves. In order to achieve meaningful results, spatial, spatio-temporal and spatio-spectral sparsity can be promoted in the signals originating from the equivalent sources. The inverse problem consists of a large-scale optimization problem that is solved using a first order matrix-free optimization algorithm. It is shown that once the equivalent source signals capable of effectively interpolating the sound field are obtained, they can be readily used to localize a speech sound source in terms of Direction of Arrival (DOA) and to perform dereverberation in a highly reverberant environment.

***Index Terms*—** Dereverberation, Source localization, Sparse sensing, Inverse problems, Large-scale optimization

1. INTRODUCTION

While there are many source localization methods that work well in free-field acoustic scenarios, source localization in highly reverberant environments is challenging [1, 2]. Reverberant environments are also problematic for speech intelligibility and significant research efforts have been focusing on dereverberation [3]. Dereverberation and source localization are often connected: for example many dereverberation methods require the knowledge of the Direction of Arrival (DOA) of the sound source [4, 5]. Instead, other methods rely either on channel equalization which requires estimation of the Room Impulse Responses (RIRs) [6] or on Multi-Channel Linear Prediction (MCLP) which requires no a priori knowledge of the acoustics but is non-robust to additive noise [7, 8]. These methods are either data-driven or make use of parametric acoustic models. Recently source localization has been posed as an inverse problem where physical acoustic models are used to reconstruct and localize the sound source [9–11]. Such methods allow precise localization of the source position inside the room but require detailed knowledge of the room geometry and boundary conditions. Alternatively, the Plane Wave Decomposition Model (PWDM) has been shown to approximate well any sound field in source-free volumes [12]. This allows sound sources to be localized without knowledge of the room geometry but requires a large number of microphone measurements scattered in a large volume [13, 14].

In this paper, a recently proposed RIR interpolation method [15], is reformulated and modified to be able to perform joint source localization and dereverberation. The aim of this paper is not to demonstrate that this method is competitive with other state-of-the art methods but rather to present a novel approach which is substantially different from the traditional localization and dereverberation methods. The proposed method relies on the interpolation of the sound field recorded by a set of microphones which is formulated as a regularized inverse problem. This consists of an optimization problem that matches the sound pressure measured by microphones with the sound pressure predicted by an acoustic model. Here two acoustic models are compared: the PWDM and the Time-domain Equivalent Source Model (TESM) which both are capable of approximating the sound field in a source-free volume [15]. In both methods the modeling is based on a collection of equivalent sources producing either spherical (TESM) or plane waves (PWDM) controlled by signals that are estimated through the optimization problem. It is shown that, by imposing a specific sparsity-inducing regular-

This research work was carried out at the ESAT Laboratory of KU Leuven, the frame of the FP7-PEOPLE Marie Curie Initial Training Network "Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)", funded by the European Commission under Grant Agreement no. 316969, KU Leuven Research Council CoE PFV/10/002 (OPTEC), the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office IUAP P7/19 "Dynamical systems control and optimization" (DYSCO) 2012-2017, KU Leuven Impulsfonds IMP/14/037 and KU Leuven C2-16-00449 "Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking". The scientific responsibility is assumed by its authors.

ization with a specific model, three kinds of sparse priors can be imposed in these signals: spatial sparsity, spatio-temporal sparsity and spatio-spectral sparsity. A novel procedure for tuning the level of regularization is presented that requires an additional reference microphone. The resulting optimization problem is of large scale and has a non-smooth cost function: this is solved using a matrix-free accelerated version of the Proximal Gradient (PG) algorithm [16] and combined with a Weighted Overlap-Add (WOLA) strategy. Once the interpolation step is achieved, the equivalent source signals can be used to estimate the DOA of the sound source. Additionally, a dereverberated audio signal can be readily obtained by selecting the equivalent source signal corresponding to the estimated DOA. Simulation results show that in a sound field generated by a speech source, spatio-spectral and spatial sparsity have similar performances and outperform spatio-temporal sparsity both in terms of sound field interpolation and dereverberation.

2. ACOUSTIC MODELS

2.1. Time-domain equivalent source method

The time-domain Green's function for a point-like source in free-field is defined as:

$$\phi_{l,m}(t) = 1/(4\pi d_{l,m})\delta\left(t - d_{l,m}/c\right), \tag{1}$$

where c is the speed of sound, δ is the Dirac delta function and $d_{l,m} = \|\mathbf{x}_l - \mathbf{x}_m\|_2$ is the distance between the m-th microphone position $\mathbf{x}_m$ and the l-th equivalent source position $\mathbf{x}_l$. Equation (1) represents a particular solution of the free-field wave equation with null initial conditions and describes a *spherical wave*. Equation (1) can be discretized over time at a sampling frequency F_s using a *fractional delay* filter with Impulse Response (IR) $h_{l,m}$ [17]. The TESM can be described by the following equation:

$$p(\mathbf{x}, n)|_{\mathbf{x}=\mathbf{x}_m} \approx \sum_{l=0}^{N_w-1} \frac{1}{4\pi d_{l,m}} h_{l,m}(n) * w_l(n), \tag{2}$$

for $\mathbf{x}_m \in \Omega$, where $*$ represents convolution, $p(\mathbf{x}, n)|_{\mathbf{x}=\mathbf{x}_m}$ is the sound pressure at $\mathbf{x}_m$ at a discrete time n and $\Omega \subset \mathbb{R}^3$ is a source-free volume where any sound field can be well approximated using a collection of equivalent sources [15] controlled by the signals $w_l(n)$ referred to here as *weight signals*. Equation (2) can be generalized for N_m discrete positions $\mathbf{x}_m \in \Omega$ and N_t discrete times: $\mathbf{P} = D_t(\mathbf{W})$, where $\mathbf{P} \in \mathbb{R}^{N_t \times N_m}$ is a matrix in which the m-th column is the sound pressure signal $p(\mathbf{x}, n)|_{\mathbf{x}=\mathbf{x}_m}$ for $n = 1, ..., N_t$, and $\mathbf{W} \in \mathbb{R}^{N_t \times N_w}$ is a matrix in which the l-th column is the weight signal $w_l(n)$. The linear operator $D_t : \mathbb{R}^{N_t \times N_w} \rightarrow \mathbb{R}^{N_t \times N_m}$ maps the weight signals to the sound pressures and represents a dictionary of spherical waves.

2.2. Plane wave decomposition method

A plane wave is defined as

$$\hat{\phi}_{l,m}(f) = e^{-ik_f d_{l,m}} \tag{3}$$

and is the homogeneous solution of the Helmholtz equation, i.e. the frequency domain counterpart of the free-field wave equation. Here f is the frequency index and k_f is the wave number defined as $k_f = \omega_f/c$. A sound field in a source-free volume Ω can be as well represented by a finite weighted sum of plane waves coming from N_w different directions [12]:

$$\hat{p}(\mathbf{x}, f)|_{\mathbf{x}=\mathbf{x}_m} \approx \sum_{l=0}^{N_w-1} \hat{\phi}_{l,m}(f)\hat{w}_l(f) \text{ for } \mathbf{x}_m \in \Omega, \tag{4}$$

where the *weight* $\hat{w}_l(f)$ is a complex scalar that weights the l-th plane wave at the frequency index f. Equation (4) describes the PWDM: this equation can be generalized as well for N_m discrete positions $\mathbf{x}_m \in \Omega$ and N_f discrete frequencies $\hat{\mathbf{P}} = D_p(\hat{\mathbf{W}})$ where $\hat{\mathbf{P}} \in \mathbb{C}^{N_f \times N_m}$ is a matrix in which the m-th column is the Discrete Fourier Transform (DFT) of the sound pressure signal $p(\mathbf{x}, n)|_{\mathbf{x}=\mathbf{x}_m}$ and $\hat{\mathbf{W}} \in \mathbb{C}^{N_f \times N_w}$ is a matrix containing the weights $\hat{w}_l(f)$.

3. THE INVERSE PROBLEM

Consider a sound source in the far field and a set of of N_m microphones positioned inside a source-free volume $\Omega \in \mathbb{R}^3$. The aim is to interpolate the sound field inside this volume in order to jointly localize and dereverberate a sound source. What is sought by the inverse problem is to extrapolate out of the microphone measurements the optimal weight signals that lead to the optimal sound field approximation in the least-squares sense. The following optimization problem can be used to solve this inverse problem:

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\text{argmin}} \quad f(\mathbf{W}) = \frac{1}{2}\|D(\mathbf{W}) - \tilde{\mathbf{P}}\|_F^2, \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm $\|\mathbf{A}\|_F = \|\text{vec}(\mathbf{A})\|_2$, with $\text{vec}(\cdot)$ representing vectorization and $D(\cdot)$ the acoustic model of choice, i.e. either $D_t(\cdot)$ or $D_p(\cdot)$. The columns of the matrix $\tilde{\mathbf{P}}$ contain the microphone measurements, i.e. either the N_t-long measured sound pressure signals or their DFT. Problem (5) is heavily *ill-posed*: if many sound waves are used to construct $D(\cdot)$, multiple solutions for $\mathbf{W}$ can minimize the cost function effectively. This will in general lead to *over-fitting*: the measured sound pressure will coincide with the sound pressure of the acoustic model but only at the microphone positions, leading to a poor sound field interpolation. To avoid this, it is necessary to regularize problem (5) by adding a regularization term to its cost function:

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\text{argmin}} \quad f(\mathbf{W}) + g(\mathbf{W}). \tag{6}$$

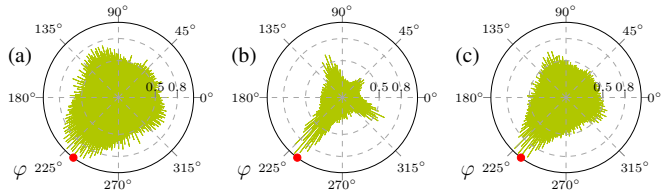

Fig. 1. Visualization of the normalized energy of the weight signals, i.e. $\|\mathbf{w}_l^\star\|_2^2$, as a function of the azimuthal angle φ. The red dots represent the true source position. (a) TESM with l_1-norm regularization, (b) TESM with sum of l_2-norms regularization, (c) PWDM with l_1-norm regularization.

A possible choice is the *sum of l_2-norms regularization* corresponding to $g(\mathbf{W}) = \lambda \sum_{l=0}^{N_w-1} \|\mathbf{W}_{:,l}\|_2$ where $\mathbf{W}_{:,l}$ indicates the lth column of $\mathbf{W}$ and λ is a scalar that balances the level of regularization. This regularization promotes only few columns of $\mathbf{W}$ to have non-zero coefficients, in practice encouraging *spatial sparsity*. Another very common regularization is the *l_1-norm regularization* corresponding to $g(\mathbf{W}) = \lambda\|\text{vec}(\mathbf{W})\|_1$ which promotes sparsity in $\mathbf{W}$, that is only few elements of the matrix are non-zero. If a time domain acoustic model (TESM) is used, *spatio-temporal sparsity* is promoted while with a frequency domain acoustic model (PWDM), *spatio-spectral sparsity* is promoted. The equivalent sources are positioned in a Fibonacci lattice, providing a nearly uniform sampling of the surface of a sphere [18]. In order to achieve accurate interpolation, a large number of equivalent sources must be used: here $N_w = 500$. Once a solution is obtained, the DOA of the sound source can be inferred by finding the weight signal with the strongest energy. Fig. 1 shows the energy of the weight signals as a function of the azimuthal angle for the simulation results presented in Section 5. A clear maximum is visible towards the direction of the sound source shown by the red dot. Finally, a dereverberated audio signal can be obtained by selecting the weight signal corresponding to the estimated DOA.

4. OPTIMIZATION ALGORITHM

Problem (6) is non-smooth and can easily become of large-scale. A well known algorithm that can deal with this type of problems is the PG algorithm which is a first order optimization algorithm suitable for non-smooth cost functions and having minimal memory requirements [19]. The PG algorithm generalizes the gradient descent algorithm to a class of non-smooth cost functions and solves optimization problems such as the problem in (6) where $f(\cdot)$ is convex and smooth, and $g(\cdot)$ is non-smooth, such as for instance the regularization terms described in Section 3. The PG algorithm consists of iterating

$$\mathbf{W}^{k+1} = \text{prox}_{\gamma g}\left(\mathbf{W}^k - \gamma \nabla f(\mathbf{W}^k)\right), \tag{7}$$

starting from an initial guess $\mathbf{W}^0$. Here $\nabla f(\cdot)$ is the Jacobian operator of $f(\cdot)$, γ is the step-size, and $\text{prox}_{\gamma g}(\cdot)$ is the *proximal mapping* of the function $g(\cdot)$ [19]. For the regularization terms described in Section 3 the proximal mapping consists of a cheap operation [19], e.g. the l_1-norm regularization reduces to a soft-thresholding of the elements of $\mathbf{W}$. The Jacobian can be computed using the adjoint operator of $D(\cdot)$ [15]. Both $D(\cdot)$ and its adjoint can be evaluated without the usage of matrices, which for this problem can become unfeasible to store, leading to *matrix-free optimization*. Finally, an accelerated variant of the PG algorithm is used: this algorithm uses a quasi-Newton method to accelerate the PG algorithm substantially [16]. An implementation of the algorithm is also available online [20].

4.1. Weighted Overlap-add

Solving the optimization problem in (6) using microphone signals with a duration of the order of seconds is not feasible since evaluating the linear operator $D(\cdot)$ and its adjoint becomes too costly. For example choosing $N_w = 500$ equivalent sources and $N_t = 16000$ (2 s with a sampling frequency $F_s = 8$ kHz) would result in having $8 \cdot 10^6$ optimization variables. To overcome this issue, the optimization problem is split into several smaller sub-problems. A WOLA procedure is used: the microphone signals are split into frames of $N_{\bar{t}}$ samples. Here, $N_{\bar{t}} = 512$, resulting in a sub-problem having $256 \cdot 10^3$ optimization variables when $N_w = 500$. A square-rooted Hanning window is used with an overlap of 50%. If a frequency domain model is used, an additional DFT is applied to the sound pressure frames and the solution is converted back to the time domain before and after solving the optimization sub-problem respectively.

4.2. Tuning of parameter λ

The parameter λ appearing in the regularization terms $g(\cdot)$ controls the level of regularization. In order to obtain meaningful results it is essential to tune λ properly. The following strategy is used: an additional microphone, positioned at the center of the microphone array is used to validate the quality of the interpolation. For each frame, the optimization sub-problem is solved multiple times using different values of λ. Initially, a low level of regularization is used: λ_0 is chosen to be $10^{-6}\lambda_{max}$, where λ_{max} is the value for which $\mathbf{W}^\star = \mathbf{0}$ [15]. Once a solution is obtained, the Normalized Mean Squared Error (NMSE) of the interpolation error $\epsilon_{in} = \|\mathbf{p}_{v,\lambda_z} - \tilde{\mathbf{p}}_v\|_2^2 / \|\tilde{\mathbf{p}}_v\|_2^2$ is computed, namely the distance between the reference microphone signal $\tilde{\mathbf{p}}_v$ of the current frame and $\mathbf{p}_{v,\lambda_z}$, the reference microphone sound pressure predicted by the acoustic model at the z-th iteration. For small values of λ the prediction error is expected to be large due to over-fitting. The optimization sub-problem is then solved once more by increasing λ_z logarithmically: this is warm-started using the previous solution. The procedure is

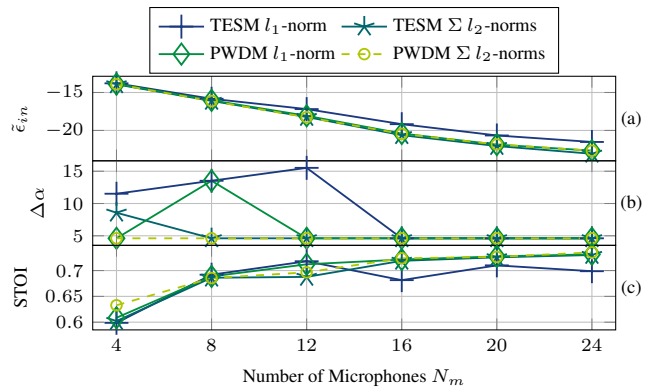

Fig. 2. Mean of the NMSE interpolation error in dB (a), angular distance in degrees (b) and STOI scores (reverberant microphone score is 0.5) (c) for different types of acoustic models and regularizations as a function of the number of microphones (excluding the reference microphone).

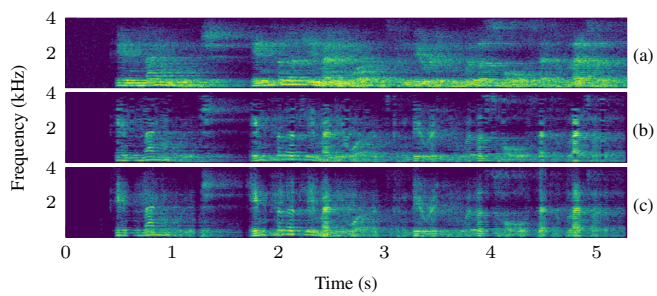

Fig. 3. Spectrogram of reverberant microphone signal (a), dereverberated signal obtained through PWDM with l_1-norm regularization (b), and through TESM with sum of l_2-norms regularization (c), using $N_m = 12$ microphones.

stopped once the prediction error stops decreasing, $\epsilon_{in,z} > \epsilon_{in,z-1} + 10^{-4}$, namely when the regularization ceases to be beneficial in terms of interpolation error. Finally the solution with optimal lambda, λ_{z-1} is added to $\mathbf{W}^\star$.

5. SIMULATION RESULTS

In this Section, results of simulations using the proposed method are presented. A reverberant shoebox room with dimensions $[L_x, L_y, L_z] = [7.34, 8.09, 2.87]$ m and reverberation time of $T_{60} = 1$ s is modeled using the Randomized Image Method (RIM) [21]. The sound source is placed in the front left corner of the room ($\mathbf{x}_s = [L_x/8,\ L_x/8,\ 1.6]$ m), and a sampling frequency of $F_s = 8$ kHz is used. An anechoic audio sample of 5.3 s of male speech from [22] is convolved with the RIRs to simulate the microphone signals. White noise is added with a SNR of 40 dB to simulate sensor noise. The microphones are positioned in a spherical microphone array with a radius of 10 cm, centered at $\mathbf{x}_c = [4.4,\ 5.7,\ 1.4]$. The equivalent sources are also centered at $\mathbf{x}_c$ with a radius of 2.9 m.

Fig. 2(a) shows the mean of the interpolation error obtained at each frame. Almost identical performances are achieved between the PWDM with either sum of l_2-norms or l_1-norm. Slightly better results are obtained using the TESM with sum of l_2-norms. The worse results are obtained using the TESM with l_1-norm. This method, which was shown to have good performance for the task of RIR interpolation [15], has poorer performance in this context because the sound field is not generated by a temporally sparse source. All methods achieve reasonable localization even when only 4 microphones are used with spatial sparsity outperforming the other regularizations as it can be seen in Fig. 2(b). Here the minimum angular error is 4.5, which is due to the finite number of equivalent sources, and, as a consequence, of directions. Fig. 3 compares the spectrogram of the weight signals corresponding to the estimated DOAs with a microphone recording: in the latter the speech components are smeared out by the reverberation while in the former these are clearly more visible. Fig. 2(c) shows speech intelligibility scores obtained using the STOI measure [23]: these are in line with the results of Fig. 2(a). Informal listening tests indicate that the spatio-temporal sparsity audio samples have many more artifacts than the ones obtained with spatial sparsity or spatio-spectral, with the latter having less audible distortions and that the dereverberation effect increases as more microphone are used. Audio samples can be found in [24].

6. CONCLUSIONS

This paper proposes a novel method for joint source localization and dereverberation. This is achieved by interpolating the sound field using the measurements of a set of microphones and by solving an inverse problem that relies on a particular acoustic model. The inverse problem is solved using an accelerated version of the PG algorithm using matrix-free optimization and a WOLA strategy in order to obtain the weight signals that control the sound waves which effectively are able to interpolate the sound field. Here, two acoustic models are compared: the TESM and the PWDM. The inverse problem is regularized using sparsity promoting regularization and depending on the choice of the acoustic model, spatial, spatio-temporal and spatio-spectral sparsity can be promoted in the weight signals. The level of regularization is tuned by comparing the interpolated sound field with the one recorded by an additional microphone. It is shown that, by finding the weight signal with strongest energy, the sound source can be localized in terms of DOA. The same weight signal can then also be used for a dereverberation task. Simulations shows that DOA estimation can be achieved using relatively few microphones ($N_m \geq 4$) when a speech source generates the sound field and that spatial and spatio-spectral sparsity outperform spatio-temporal sparsity in terms of both interpolation quality and dereverberation.

7. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer, 2001.

[2] S. Argentieri, P. Danes, and P. Souères, "A survey on sound source localization in robotics: From binaural to array processing methods," *Computer Speech & Language*, vol. 34, no. 1, pp. 87–112, 2015.

[3] P. A. Naylor and N. D. Gaubitch, *Speech dereverberation*. Springer, 2010.

[4] E. A. P. Habets and S. Gannot, "Dual-microphone speech dereverberation using a reference signal," in *Proc. 2007 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '07)*, 2007, pp. 901–904.

[5] A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 6, pp. 1006–1018, 2015.

[6] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 9, pp. 1879–1890, 2013.

[7] T. Nakatani, B.-H. Juang, T. Yoshioka, K. Kinoshita, M. Delcroix, and M. Miyoshi, "Speech dereverberation based on maximum-likelihood estimation with time-varying gaussian source model," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 8, pp. 1512–1527, 2008.

[8] A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 9, pp. 1509–1520, 2015.

[9] I. Dokmanić and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12)*, 2012, pp. 2617–2620.

[10] N. Antonello, T. van Waterschoot, M. Moonen, and P. A. Naylor, "Source localization and signal reconstruction in a reverberant field using the FDTD method," in *Proc. 22nd European Signal Process. Conf. (EUSIPCO-14)*. IEEE, 2014, pp. 301–305.

[11] S. Kitić, L. Albera, N. Bertin, and R. Gribonval, "Physics-driven inverse problems made tractable with cosparse regularization," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 335–348, 2016.

[12] A. Moiola, R. Hiptmair, and I. Perugia, "Vekua theory for the Helmholtz operator," *Zeitschrift für angewandte Mathematik und Physik*, vol. 62, no. 5, pp. 779–807, 2011.

[13] G. Chardon, T. Nowakowski, J. De Rosny, and L. Daudet, "A blind dereverberation method for narrowband source localization," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 815–824, 2015.

[14] T. Nowakowski, J. de Rosny, and L. Daudet, "Robust source localization from wavefield separation including prior information," *J. Acoust. Soc. Amer.*, vol. 141, no. 4, pp. 2375–2386, 2017.

[15] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, 2017.

[16] A. Themelis, L. Stella, and P. Patrinos, "Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone line-search algorithms," *arXiv preprint arXiv:1606.06256*, 2016.

[17] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay," *IEEE Signal Process. Mag.*, vol. 13, no. 1, pp. 30–60, 1996.

[18] Á. González, "Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices," *Mathematical Geosciences*, vol. 42, no. 1, pp. 49–64, 2010.

[19] N. Parikh and S. P. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[20] L. Stella and N. Antonello. (2017) Proximal Algorithms. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/587243

[21] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.

[22] Bang and Olufsen, "Music for Archimedes," CD B&O 101, 1992.

[23] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.

[24] N. Antonello. (2017) Audio samples. [Online]. Available: https://nantonel.github.io/jsld/