Efficient Relative Transfer Function Estimation Framework in the Spherical Harmonics Domain

Yoav Biderman and Boaz Rafaely
Department of Electrical and
Computer Engineering
Ben-Gurion University of the Negev

Sharon Gannot
Faculty of Engineering
Bar Ilan University
Ramat-Gan, Israel

Simon Doclo
Dept. of Medical Physics and Acoustics
Cluster of Excellence Hearing4All
University of Oldenburg, Germany

***Abstract*—In acoustic conditions with reverberation and coherent sources, various spatial filtering techniques, such as the linearly constrained minimum variance (LCMV) beamformer, require accurate estimates of the relative transfer functions (RTFs) between the sensors with respect to the desired speech source. However, the time-domain support of these RTFs may affect the estimation accuracy in several ways. First, short RTFs justify the multiplicative transfer function (MTF) assumption when the length of the signal time frames is limited. Second, they require fewer parameters to be estimated, hence reducing the effect of noise and model errors. In this paper, a spherical microphone array based framework for RTF estimation is presented, where the signals are transformed to the spherical harmonics (SH)-domain. The RTF time-domain supports are studied under different acoustic conditions, showing that SH-domain RTFs are shorter compared to conventional space-domain RTFs.**

I. Introduction

Microphone arrays, used for speech enhancement, motivated research of spatial filters, also known as beamformers. First, fixed beamformers, such as the delay-and-sum beamformer [1], were developed. Later, data-dependent beamformers that include noise minimization, such as the linearly constrained minimum variance (LCMV) beamformer [2] with its well-known special case, the minimum variance distortionless response (MVDR) beamformer [3], were introduced. The generalized sidelobe canceler (GSC) [4] was developed as an effective way of implementing the MVDR beamformer. The original GSC was developed under the assumption of a simple anechoic acoustic environment, where the received signals are delayed versions of the desired signal. However, in real rooms with reflections from the room boundaries this assumption is not valid. To address degradation due to coherent reflections, in [5] the TF-GSC was proposed by adapting the GSC to the general acoustic transfer function (ATF) case. It was found that the TF-GSC can be implemented using the relative transfer functions (RTFs) within the linear constraints of the beamformer.

Due to its importance in beamformer design, RTF estimation has been widely explored. In [6], Cohen developed a method based on spectral subtraction. Under stricter assumptions regarding time segmentation of noise and the desired sources, Markovich *et al.* [8] proposed an eigenspace method for RTF estimation in a multi-speaker problem, which was further analyzed in [9] to demonstrate its superiority over covariance subtraction based methods [6].

These methods estimate the space-domain RTFs. Their performance depends on the complexity of the RTFs and, more particularly, on the time-domain support of their corresponding impulse responses. This support affects the required length of the time frames employed when processing the speech signal using the short-time Fourier transform (STFT). Standard speech processing applications may impose limits on the frame length, such that shorter RTFs may be desirable. Furthermore, RTFs corresponding to short impulse responses justify the use of the multiplicative transfer function (MTF) approximation [10], and reduce the analytical and computational complexity imposed by algorithms which use space-domain RTFs, as in [11].

A noise reduction algorithm for spherical arrays, using an LCMV framework that consists of a fixed beamformer and a blocking matrix, based on the spherical Fourier transform (SFT) is introduced in [12]. The spherical array enables spatial separation of the acoustic reflections, simplifying the RTFs and improving the noise reduction performance of the LCMV beamformer. The current contribution provides a comprehensive analysis of the RTFs in the SH-domain, comparing the support of space-domain RTFs and SH-domain RTFs under various acoustic conditions. It is shown that SH-domain RTFs are significantly shorter, which motivates the application of SH-domain array processing for improved performance.

In section II the system model is presented, and in sections III and IV the RTFs in the SH-domain will be defined and discussed. Section V outlines the results of a simulation study and, finally, conclusions are drawn in section VI.

II. LCMV framework

The LCMV framework, developed in [12], is briefly reviewed in this section. This framework is an example for the use of RTFs in the SH-domain, and the results may also be useful for other algorithms using RTFs.

For the measurement model, consider a spherical array of radius r, consisting of J microphones receiving a speech signal emitted by a point source in a noisy and reverberant room. The pressure recorded at the array elements consists of the desired signal and its reflections from the room boundaries.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

The pressure at the j^{th} microphone, in the frequency-domain, can be expressed by:

$$p_j(k) = h_j(k)s(k) + u_j(k) = \sum_{q=1}^{Q} v_j(k,\psi_q)s(k) + u_j(k)$$
$$j = 1, \cdots, J \quad (1)$$

where k denotes the frequency index, Q is the number of reflections (including the direct path), $h_j(k)$ is the space-domain ATF of the source to the j^{th} microphone, $v_j(k,\psi_q)$ is the ATF of the q^{th} reflection to the j^{th} microphone, with ψ_q the direction of arrival (DOA) of the q^{th} reflection. $s(k)$ is the speech signal and $u_j(k)$ represents the noise and interference signals at the j^{th} microphone.

Figure 1 shows a block diagram of the LCMV framework [12] divided into several stages, detailed in the following.

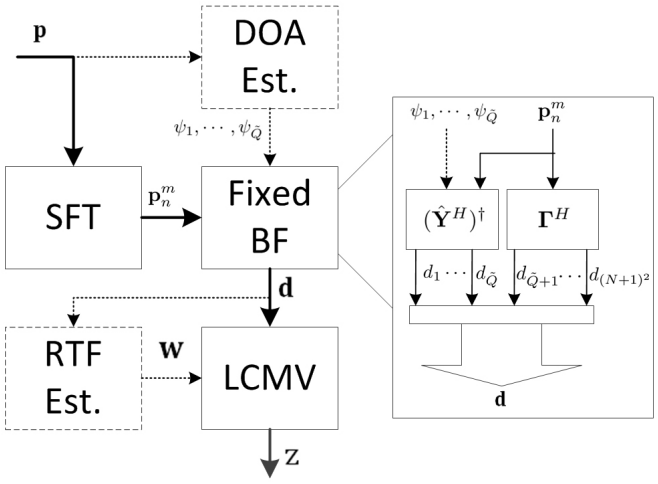

Fig. 1. LCMV block diagram

A. Spherical Fourier transform (SFT)

In the first stage, the spherical functions are represented by their SH decomposition [13]. Under the necessary condition that $J \geq (N+1)^2$, the SFT up to the array order N for $kr < N$ can be approximated using a weighted summation over the microphone signals [14]. Using the representation of plane waves in the SH-domain, the spherical Fourier coefficients of the pressure are given by:

$$p_n^m(k,r) = \sum_{q=1}^{Q} a_q(k) b_n(kr) Y_n^m(\psi_q)^* s(k) + u_n^m(k) \quad (2)$$

where $a_q(k)$ is the complex amplitude of the q^{th} reflection, $b_n(kr)$ is determined by the array configuration [14], $Y_n^m(\psi_q)$ is the SH function of order n and degree m, and $u_n^m(k)$ are the spherical Fourier coefficients of the noise. Equation (2) is valid under the far-field assumption that approximates the reflections as plane waves. Dropping k, r for brevity, the vector of coefficients up to order N is:

$$\mathbf{p}_n^m = \mathbf{B}_n \mathbf{Y}^H \mathbf{a} s + \mathbf{u}_n^m \quad (3)$$

with $(\cdot)^H$ the conjugate transpose operator and:

$$\mathbf{p}_n^m = [p_0^0, p_1^{(-1)}, p_1^0, p_1^1, \ldots, p_N^N]^T$$
$$\mathbf{u}_n^m = [u_0^0, u_1^{(-1)}, u_1^0, u_1^1, \ldots, u_N^N]^T \quad (4)$$

the $(N+1)^2 \times 1$ pressure and noise spherical Fourier coefficient vectors, respectively. $\mathbf{B}_n = \mathrm{diag}(b_0, b_1, b_1, b_1, ..., b_N)$ is an $(N+1)^2 \times (N+1)^2$ matrix and the amplitude vector is $\mathbf{a} = [a_1(k), \cdots, a_Q(k)]^T$. The $Q \times (N+1)^2$ SH matrix $\mathbf{Y}$ is defined as:

$$\mathbf{Y} = \begin{pmatrix} Y_0^0(\psi_1) & Y_1^{(-1)}(\psi_1) & Y_1^0(\psi_1) & \cdots & Y_N^N(\psi_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\psi_Q) & Y_1^{(-1)}(\psi_Q) & Y_1^0(\psi_Q) & \cdots & Y_N^N(\psi_Q) \end{pmatrix}. \quad (5)$$

B. Fixed Beamformer

From the spherical Fourier coefficients, the pressure field plane wave density (PWD) function [14] is:

$$\mathbf{y} = \mathbf{B}_n^{-1} \mathbf{p}_n^m = \mathbf{Y}^H \mathbf{a} s + \tilde{\mathbf{u}}_n^m \quad (6)$$

where $\tilde{\mathbf{u}}_n^m = \mathbf{B}_n^{-1} \mathbf{u}_n^m$. Circumventing the matrix $\mathbf{B}_n$ to be singular at some frequencies can be achieved by using specific array configurations, and by limiting the order at low frequencies [14].

For the fixed beamformer stage, it is assumed that the DOAs of $\tilde{Q}$ reflections are known, with $\tilde{Q} \leq Q$. While this does not necessarily hold in practice, it enables analysis of the algorithm building blocks. It is also assumed that the array is capable of decomposing all known reflections, requiring that:

$$(N+1)^2 \geq \tilde{Q}. \quad (7)$$

The following beamformer and blocking matrix are then applied to the PWD function:

$$\mathbf{d} = \begin{pmatrix} (\tilde{\mathbf{Y}}^H)^\dagger \\ \mathbf{\Gamma}^H \end{pmatrix} \mathbf{y} \quad (8)$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo-inverse operator. $\tilde{\mathbf{Y}}$ is a $\tilde{Q} \times (N+1)^2$ matrix with a structure similar to (5), based on the known DOAs of the $\tilde{Q}$ reflections.

Each row of $(\tilde{\mathbf{Y}}^H)^\dagger$ is actually a beamformer directed to the DOA of the q^{th} reflection, with nulls pointed at the other $\tilde{Q} - 1$ reflections [15]. The columns of the blocking matrix $\mathbf{\Gamma}$ span the null space of $\tilde{\mathbf{Y}}$, and can be constructed from the first $(N+1)^2 - \tilde{Q}$ columns of $\mathbf{I} - \tilde{\mathbf{Y}}^H (\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^H)^{-1} \tilde{\mathbf{Y}}$.

C. LCMV

The final stage is to apply an LCMV beamformer to the fixed beamformer output signals $\mathbf{d}$, using as constraints estimated RTFs in the SH-domain with respect to the desired source. The first channel in (8) is always used as the reference channel for the RTF estimation. A noise reference signal is required for estimating the noise covariance matrix. The reader is referred to [12] for more details.

III. Space-domain and SH-domain RTFs

The goal of this section is to define the RTFs in the space-domain and the SH-domain and to define a measure for their time-domain support.

The j^{th} space-domain RTF can be calculated according to the corresponding ATFs defined in (1) as:

$$w_j(k) = \frac{h_j(k)}{h_1(k)} \qquad j = 2, \cdots, J. \tag{9}$$

In the SH-domain, we first define the corresponding ATFs relating to the processing in (6) and (8) (i.e. PWD and fixed beamformer) as:

$$\mathbf{g}(k) = \begin{pmatrix} (\tilde{\mathbf{Y}}^H)^\dagger \\ \mathbf{\Gamma}^H \end{pmatrix} \mathbf{B}_n^{-1}(k)\mathbf{h}_n^m(k) \tag{10}$$

with the $(N+1)^2 \times 1$ vector of spherical Fourier coefficients of the space-domain ATF vector $\mathbf{h}(k) = [h_1(k), \cdots, h_J(k)]^T$:

$$\mathbf{h}_n^m(k) = [h_0^0(k), h_1^{(-1)}(k), h_1^0(k), h_1^1(k), \ldots, h_N^N(k)]^T. \tag{11}$$

Similarly, the RTFs in the SH-domain are defined as:

$$w_i(k) = \frac{g_i(k)}{g_1(k)} \qquad i = 2, \cdots, (N+1)^2. \tag{12}$$

The time-domain support of the RTFs is calculated using Schroeder integration [16]. However, since the RTFs are not necessarily causal, i.e. the non-causal parts appear at the end of their impulse responses, the RTFs are centralized by a reverse circular shift before applying the inverse Discrete Fourier transform (IDFT), i.e.:

$$\begin{aligned} w_i^c(m) = \text{IDFT}\{w_i(k)e^{-jk\pi}\} \qquad & i = 2, \cdots, (N+1)^2 \\ & m = 0, \cdots, M-1 \end{aligned} \tag{13}$$

where $w_i^c(m)$ is the centralized time-domain RTF, and M denotes the number of STFT frequencies (and thus the length of $w_i^c(m)$). Due to the two-sided nature of the centralized RTFs (see Figure 3), the Schroeder integral is applied twice, i.e. once from each side. The discrete time versions of the Schroeder integrals are given by:

$$\begin{aligned} S_i^1(m) &= \sum_{l=m}^{M-1} (w_i^c(l))^2, && m = 0, \cdots, M-1 \\ S_i^2(m) &= \sum_{l=m}^{M-1} (w_i^c(M-1-l))^2, && m = 0, \cdots, M-1 \end{aligned} \tag{14}$$

Finally, the time-domain support of the i^{th} RTF is defined as:

$$\Delta t_i = \frac{m_{25}(S_i^1) - (M - 1 - m_{25}(S_i^2))}{f_s} \tag{15}$$

where f_s is the sampling frequency and $m_{25}(S_i^1)$ and $m_{25}(S_i^2)$ are the sample indices for which $S_i^1(m)$ and $S_i^2(m)$ decay by 25dB relative to their value at $m = 0$, corresponding to the total energy of the RTF. This measure provides the time period during which most of the energy of the RTF resides.

IV. Analysis of SH-domain RTFs

The main purpose of this paper is to analyze the properties of the RTFs in the SH-domain. The analysis is divided into two cases: one in which a complete separation of the reflections is obtained in the SH-domain ($\tilde{Q} = Q$), and the other in which the separation is only partial ($\tilde{Q} < Q$). For each case, the advantages and limitations of the analytic approach will be examined. The separation capabilities of the array depend on its SH order and, therefore, the number of microphones. This is contrary to the space-domain, where the number of microphones does not affect the RTF support.

1) Full separation ($\tilde{Q} = Q$): When the DOAs of all reflections are known with high accuracy and the array order is high enough to enable their separation, each of the first Q channels of $\mathbf{d}$ comprises a single source reflection (where the first channel comprises the direct signal), i.e.:

$$d_i(k) = \begin{cases} a_i(k)s(k) & i = 1, \cdots, Q \\ 0 & i = Q+1, \cdots, (N+1)^2 \end{cases}. \tag{16}$$

Therefore, the ATFs are given by $g_i(k) = a_i(k)$ for $i = 1, \cdots, (N+1)^2$, and the RTFs in the SH-domain are (12):

$$w_i(k) = \begin{cases} \frac{a_i(k)}{a_1(k)} & i = 2, \cdots, Q \\ 0 & i = Q+1, \cdots, (N+1)^2 \end{cases}. \tag{17}$$

Assuming that the direct path ATF $a_1(k)$ is constant over frequency, the first Q RTF supports are dictated by the time-domain support of the corresponding SH-domain ATF $a_i(k)$.

2) Partial separation ($\tilde{Q} < Q$): Due to the $Q - \tilde{Q}$ unknown DOAs, the fixed beamformer in (8) does not have nulls pointed in the directions of the unknown reflections. Thus, all channels in $\mathbf{d}$ will contain residual reflections. Let $\mathbf{r}$ be the $(N+1)^2 \times 1$ residual vector whose elements are:

$$r_i(k) = \sum_{j=1}^{Q} \gamma_{ij} a_j(k) \qquad i = 1, \cdots, (N+1)^2 \tag{18}$$

where γ_{ij} is a gain factor inserted by the fixed beamformer for the j^{th} reflection. Elements of $\mathbf{d}$ can be expressed as:

$$d_i(k) = \begin{cases} a_i(k)s(k) + r_i(k)s(k) & i = 1, \cdots, \tilde{Q} \\ r_i(k)s(k) & i = \tilde{Q}+1, \cdots, (N+1)^2 \end{cases} \tag{19}$$

The RTFs in the SH-domain are then given by:

$$w_i(k) = \begin{cases} \frac{a_i(k)+r_i(k)}{a_1(k)+r_1(k)} & i = 2, \cdots, \tilde{Q} \\ \frac{r_i(k)}{a_1(k)+r_1(k)} & i = \tilde{Q}+1, \cdots, (N+1)^2 \end{cases}. \tag{20}$$

The residual reflections will typically result in a longer support in the time-domain. However, for the first channel, obtained from a beamformer directed to the direct path of the desired source, it is reasonable to assume that the direct sound is dominant, i.e. $|a_1(k)| >> |\gamma_{1j}a_j(k)|$ for $j = 1, \cdots, Q$, such that the residual term can be neglected. The RTFs in the SH-domain can then be approximated by:

$$w_i(k) \approxeq \begin{cases} \frac{a_i(k)+r_i(k)}{a_1(k)} & i = 2, \cdots, \tilde{Q} \\ \frac{r_i(k)}{a_1(k)} & i = \tilde{Q}+1, \cdots, (N+1)^2 \end{cases}. \tag{21}$$

Under this assumption, the only influence of the denominator is a time-shift, which is not reflected in a longer support. This is definitely an advantage over space-domain RTFs, where a complex ATF in the denominator results in long impulse responses in the time-domain.

V. Simulation Study: SH-domain RTFs

The simulations in this section compare the RTFs' time-domain supports in the space-domain and the SH-domain for different acoustic environments. A room with dimensions $8 \times 6 \times 3\,\mathrm{m}$ was simulated by an in-house developed simulator, implementing the image method [17], to create an environment with various reverberation times. A spherical microphone array with a radius of $4.2\,\mathrm{cm}$ was placed at $(4.5, 2.5, 1.5)\,\mathrm{m}$ and a point source at 16 uniformly spaced positions around the array, each at a distance of 2 m from the array center. The sampling frequency was set to $8\,\mathrm{kHz}$. In the simulations we considered arrays of orders $N = 4$ and $N = 10$ with $J = 64$ and $J = 242$ microphones, respectively. The room boundary reflection coefficients were chosen such that T_{60} ranges from $42\,\mathrm{ms}$ to $645\,\mathrm{ms}$.

For the considered reflection coefficients, the simulated room had many reflections, such that the given array order does not enable full separation of the reflections, i.e. $\tilde{Q} < Q$. Since the simulation engine provided more reflection DOAs than the analysis can handle, a DOA subset was chosen. Two criteria were taken into account in the selection of this subset:

1) Avoiding ill-posed distributions of reflection DOAs, manifested as a high condition number of matrix $\tilde{\mathbf{Y}}$ in (8).
2) The reflection power.

Adjacent DOAs produce linearly dependent columns in $\tilde{\mathbf{Y}}$, such that taking the pseudo-inverse in (8) becomes a highly non-stable operation, causing unstable beam patterns. The chosen approach for the DOA selection was to take the $(N+1)^2 \cdot 1.5$ reflections with the highest amplitudes from the set given by the simulation engine, and to sort them by the corresponding condition number of $\tilde{\mathbf{Y}}$. Although the DOA sorting is supposed to solve the robustness issue, the matrix is still unstable when $\tilde{Q} \approx (N+1)^2$. For this reason, diagonal loading by a factor of a tenth of the largest singular value is used as another means to improve the robustness.

In practice, the information of multiple reflection DOAs is not likely to be available. However, the analysis in this section uses this information to study RTFs in the SH-domain for a wide range of conditions, aiming to provide recommendations and insights regarding performance under real conditions. The following simulations also investigate scenarios where only the source DOA is known, i.e. $\tilde{Q} = 1$. Note that in this case, all $(N+1)^2$ RTFs are relevant, because all channels contain residual reflections.

For 6 different settings, Figure 2 depicts the average RTF supports, calculated using (14), as a function of the reverberation time. It is evident that the SH-domain RTFs are generally shorter than the space-domain RTFs. Furthermore, as anticipated, the array order greatly affects the support of the SH-domain RTFs, due to the improved spatial resolution of arrays of higher order, but hardly makes any difference in the space-domain.

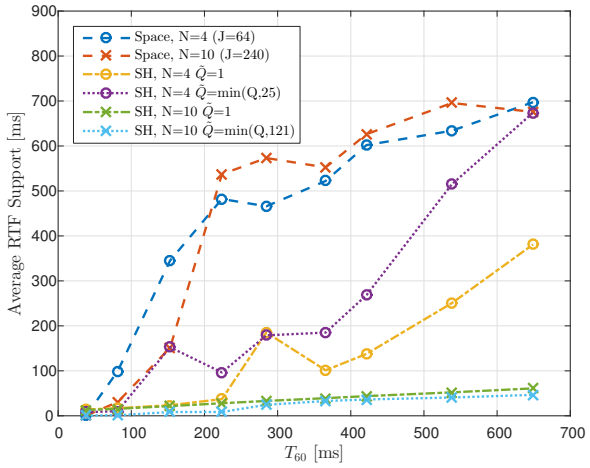

Fig. 2. RTF supports for different array configurations as a function of the reverberation time, averaged over 16 source positions and on the $(N+1)^2$ RTF supports per position.

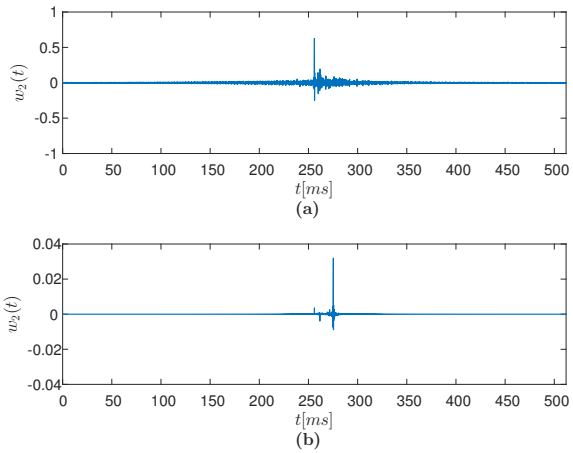

Fig. 3. Examples of (a) space-domain and (b) SH-domain RTFs, array order $N = 10$, $T_{60} = 153\,\mathrm{ms}$, $Q = \tilde{Q} = 54$, after circular shift correction.

Another observation regards the array separation capability. For small reflection coefficients, where $\tilde{Q} = Q$ can be achieved, all reflections are decomposed by the array, such that the RTF support is close to 0. For $\tilde{Q} = 1$, it is to be expected that the RTFs will be longer, but they are still quite short because the first channel contains only the direct source (see Section IV). When the number of reflections surpasses the array separation capability, the results for $N = 4$ show that separating only the direct source and using it for RTF estimation gives better results than separating many reflections, such that there is no advantage in choosing $\tilde{Q} > 1$. This is a reasonable outcome, because as we try to separate a larger number of reflections, the pseudo-inverse of the matrix $\tilde{\mathbf{Y}}$ is less stable, causing amplification of undesired reflections. As a result, the first channel, which greatly affects the RTF support,

contains significant reflections other than the direct signal.

An example of a space-domain and a SH-domain RTF is depicted in Figure 3, demonstrating that, indeed, SH-domain RTFs have a shorter support.

Figure 4 depicts the influence of the number of known reflection DOAs, $\tilde{Q}$, on the RTF support. For the larger reverberation times, the average RTF support increases as more reflections are taken into account in the fixed beamformer. This observation confirms the statement that for reverberant scenarios it may be better to use only the direct source channel to keep the fixed beamformer robust. However, if the total number of reflections, Q, is smaller than the number of separable reflections $(N+1)^2$ (in this simulation 25), or only slightly larger, the RTF support is maintained or decreased when a larger $\tilde{Q}$ is selected. This is due to the successful separation in these low reverberation cases. However, even in these cases the condition number of $\tilde{\mathbf{Y}}$ still increases significantly above 1, and, in practice, this might raise an issue of noise amplification.

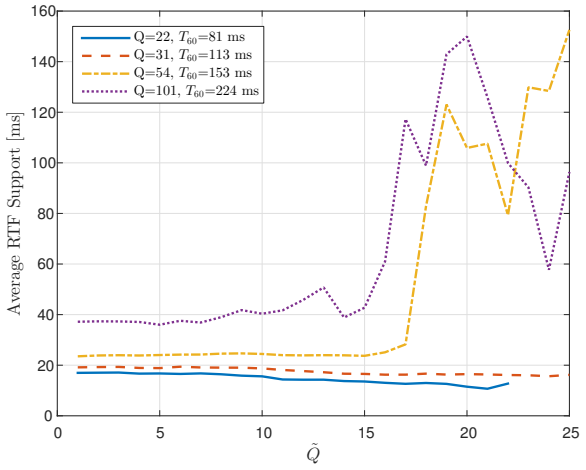

Fig. 4. Average RTF supports as a function of the known reflections $\tilde{Q}$, for an array of order $N = 4$ and varying reverberation levels.

VI. Conclusion

In this paper, a previously developed framework for spherical microphone arrays that transforms the space-domain signals to the SH-domain was considered. Under a set of assumptions, an analysis of the RTFs in the SH-domain was carried out and tested under different acoustic environments and array configurations. It was shown that the time-domain support of the SH-domain RTFs, which influences the performance of subsequent localization and beamforming algorithms, was shorter compared to the support of conventional space-domain RTFs. Simulations have examined the dependence on array order, level of room reverberation, and number of known reflections. It was shown that, when using a high order array in low reverberation conditions, it is better to use all the known reflections in the SH-domain. However, if the room response consists of many reflections, $\tilde{Q}$ should be limited.

Future work will include the development of improved RTF estimation methods in the SH-domain, and a comparison of LCMV performance using these SH-domain RTFs with LCMV with space-domain RTFs.

Acknowledgment

The research leading to these results has received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465.

References

[1] W. Hansen and J. Woodyard, "A new principle in directional antenna design," *Proceedings of the Institute of Radio Engineers*, vol. 26, no. 3, pp. 333–345, 1938.

[2] H. L. Van-Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*, 1st ed. Wiley-Interscience, 2002.

[3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug 1969.

[4] L. J. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[5] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Sig. Proc.*, vol. 49, no. 8, pp. 1614–1626, 2001.

[6] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. on Speech and Audio Proc.*, vol. 12, no. 5, pp. 451–459, Sept 2004.

[7] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 8, pp. 2055–2063, Aug. 1996.

[8] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. on Audio, Speech, and Language Proc.*, vol. 17, no. 6, pp. 1071–1086, Aug 2009.

[9] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 544–548.

[10] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Proc. Letters*, vol. 14, no. 5, pp. 337 –340, May 2007.

[11] E. Hadad, S. Doclo, and S. Gannot, "The binaural lcmv beamformer and its performance analysis," *IEEE/ACM Tran. on Audio, Speech, and Language Proc.*, vol. PP, no. 99, pp. 1–1, 2016.

[12] Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE Trans. on Audio, Speech, and Language Proc.*, vol. 21, no. 12, pp. 2532–2540, Dec. 2013.

[13] J. R. Driscoll and D. M. H. Jr., "Computing Fourier transforms and convolutions on the 2-sphere," *Adv. App. Math.*, vol. 15, pp. 202–250, 1994.

[14] B. Rafaely, *Fundamentals of Spherical Array Processing*. Berlin: Springer-Verlag, February 2015.

[15] ——, "Spherical microphone array with multiple nulls for analysis of directional room impulse responses," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, USA, Mar. 2008, pp. 281–284.

[16] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412, 1965.

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