Generalizing the Relative Transfer Function to a Matrix for Multiple Sources and Multichannel Microphones

Thushara D. Abhayapala*, Lachlan Birnie, Manish Kumar, Daniel Grixti-Cheng, Prasanga N. Samarasinghe
Audio and Acoustic Signal Processing Group, School of Engineering
The Australian National University
Canberra, Australia
*Email: Thushara.Abhayapala@anu.edu.au

***Abstract*—The Relative Transfer Function (ReTF) describing the acoustic channel between two receivers in response to a single source is a useful tool in many acoustic applications like source localization, separation, and remote microphone methods. In this paper, we propose an approach to generalize the ReTFs of multiple simultaneous sound sources. We allocate receivers into two multichannel groups and formulate the Relative Transfer Matrix (ReTM) to describe the spatial acoustic channel between them. We show that the ReTM is signal independent, a spatial property of the acoustic environment, and that it can be blindly estimated from the observed signals with covariance matrices. We provide preliminary validation on simulated and experiment recordings.**

***Index Terms*—Relative Transfer Function, Spatial Audio, Microphone Array**

I. Introduction

Solutions to acoustic problems have found success in exploiting the spatial mapping provided by ReTFs [1]. The ReTF is a spatial function that describes the position of the receivers and the position of the sound source within an acoustic environment. It is given by the ratio of acoustic transfer functions between a source and two receivers [2], [3]. In a sense, the ReTF represents the coupling or acoustic channel between two microphones in response to a single sound source. Most notably, the ReTF is a property of an acoustic system with three key attributes:

i) It is independent of the source's emitted signal.
ii) It provides a unique signature of the source-microphone positions and the acoustic environment such as room size and reverberation [2].
iii) It can be reliably and robustly estimated blindly from the received signals when only the single sound source is active [4]–[6].

Its unique properties make the ReTF easily estimated and exploited as a tool in various signal processing applications, such as blind source separation [7]–[9], beamforming [4], [10], [11], sound source localization [2], [12], [13], acoustic echo cancellation [14], microphone array calibration [3], speech enhancement [15], and in hearing aids [16]. Moreover, the ReTF has also shown usefulness as a spatial feature in learning algorithms for source localization [17] and low SNR spatial-filtering-based speech enhancement [18]. The direct-path ReTF has also been defined and used in many applications [19]. Moreover, the ReTF has been extended to the spherical harmonic domain as relative harmonic coefficients [1], [20] with successful use in source localization [21], [22].

The ReTF is limited in that it is only valid when there is a single active sound source in the acoustic environment. The preceding acoustic applications which employ the ReTF are often more challenged in lively settings with multiple sound sources, as a result. There is often a compromise or assumption made to subvert this challenge. One answer is to assume that the multiple sound sources alternate in activity [23]. However, this does not always match reality, especially in the presence of sources that are continuously active.

Ultimately, a good answer is likely a generalization to the ReTF for multiple simultaneous sound sources. Deleforge, Gannot, and Kellermann [24], however, have shown that such a generalization is not possible when using a single spectro-temporal observation (which we briefly review in Sec. II). Instead, Deleforg et al. [24] proposed a ReTF generalization by using a plucker spectrogram transform to turn the observation of multiple sources for multiple timeframes into an observation of a single compound source. Good results in a source localization application are achieved in a multi-source scenario. But this approach still assumes that the number of active sources in the environment is known and that all these sources are at least partially active over the estimation period.

In this paper, we propose another extension toward generalizing the ReTF for multiple simultaneous sound sources. We consider two multichannel groups of receivers to formulate the spatial mapping of the acoustic environment as a *Relative Transfer Matrix* (ReTM) (Sec. III). The ReTM describes the coupling or acoustic channel between the two multichannel groups of receivers, similar to the ReTF. Moreover, the ReTM exhibits the same three key properties as the ReTF. The ReTM:

i) is independent of the signals emitted by multiple sources;
ii) provides a unique signature of the source-microphone positions and the acoustic environment;
iii) can be estimated blindly from the received signals through a covariance-based method (Sec. III-A).

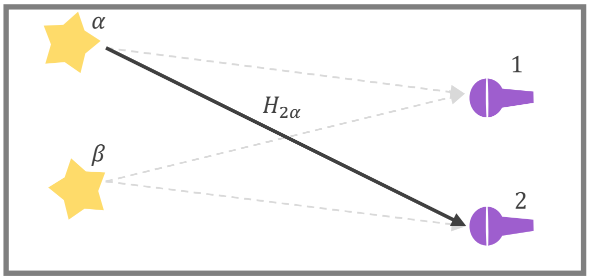

Fig. 1. Drawing of receivers, sources, and transfer functions.

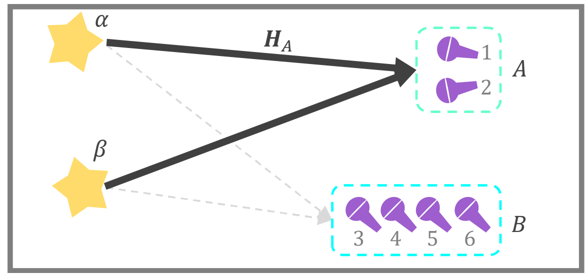

Fig. 2. Illustration of grouped microphones.

For the remainder of this paper, we provide a preliminary and theoretical introduction to the ReTM. We do not consider a specific acoustic application here, as we want to emphasize that the ReTM is essentially applicable to any approach that utilizes the ReTF. As such, we demonstrate estimating the ReTM channel for a simple remote microphone application in Sec. IV-A for both a simulated and live recording. While not examined explicitly here, such remote microphone channels are useful tools in spatial audio applications like active noise control [25], [26]. We leave more thorough examinations of the ReTM to future work, as its effectiveness is expected to depend on the specific use case whether it be employed in source localization, speech enhancement, or other applications.

II. Problem Formulation

Consider two microphones indexed by integers $\{1, 2\}$ and a single sound source denoted by $\{\alpha\}$, as shown in Fig. 1. We will consider the second source, $\{\beta\}$, later on. Let us denote the acoustic transfer functions between the source and the microphones as $H_{1\alpha}$ and $H_{2\alpha}$. The short-time frequency domain signals received by the microphones, $M_{\{1,2\}}$, are described as

$$M_1(f,t){=}H_{1\alpha}(f)S_\alpha(f,t), \quad \text{and} \quad M_2{=}H_{2\alpha}S_\alpha, \tag{1}$$

where (f, t) denotes frequency and time, and S is the source's signal. Note that we often drop the (f, t) notation for brevity.

A. The Relative Transfer Function

In theory, the ReTF is given by the ratio of the acoustic transfer functions between the source and two receivers. The ReTFs for the source α are given by

$$\mathcal{R}^{\alpha}_{1,2}(f) = H_{1\alpha}(f)/H_{2\alpha}(f), \text{ and } \mathcal{R}^{\alpha}_{2,1} = H_{2\alpha}/H_{1\alpha}, \tag{2}$$

depending on whether microphone-$\{2\}$ or -$\{1\}$ is taken as the reference, respectively. The ReTF can be estimated using (2) when the acoustic transfer functions are known. However, this typically requires impulse response (IR) measurements of the source and a stable acoustic environment.

In another sense, the ReTF of $\mathcal{R}^{\alpha}_{1,2}$ can be considered to be a spatial mapping between the received signal at microphone-$\{2\}$ to the received signal at microphone-$\{1\}$, due to the single sound source $\{\alpha\}$. This mapping can be illustrated by

$$M_1(f,t) = \mathcal{R}^{\alpha}_{1,2}(f) M_2(f,t) = H_{1\alpha}(f)S_\alpha(f,t). \tag{3}$$

B. Estimating the ReTF from Received Signals

In practice, the ReTF can be estimated directly from signals measured in a live environment. Consider the simple example of taking the ratio of the two received signals in (1) as

$$\frac{M_1(f,t)}{M_2(f,t)} = \frac{H_{1\alpha}(f)S_\alpha(f,t)}{H_{2\alpha}(f)S_\alpha(f,t)} = \frac{H_{1\alpha}(f)}{H_{2\alpha}(f)} = \mathcal{R}^{\alpha}_{1,2}(f). \tag{4}$$

Observe that the source's signal elegantly cancels off leaving behind only spatial information about the acoustic environment. We note that the ratio of (4) is susceptible to noise. Therefore, often the ReTF is estimated with the cross power spectral density, $\mathcal{P}$, of the received signals, [27]

$$\mathcal{R}^{\alpha}_{1,2}(f) \approx \mathcal{P}_{1,2}(f,t)/\mathcal{P}_{2,2}(f,t). \tag{5}$$

C. The ReTF with Multiple Sources

Consider now a second sound source denoted by $\{\beta\}$. The received signals analogous to (1) for these two sources are

$$M_1 = H_{1\alpha}S_\alpha{+}H_{1\beta}S_\beta, \text{ and } M_2 = H_{2\alpha}S_\alpha{+}H_{2\beta}S_\beta. \tag{6}$$

Estimating the ReTF by substituting (6) into (4) gives us

$$\frac{M_1(f,t)}{M_2(f,t)} = \frac{H_{1\alpha}S_\alpha + H_{1\beta}S_\beta}{H_{2\alpha}S_\alpha + H_{2\beta}S_\beta} = \mathcal{F}(f,t,S_\alpha,S_\beta), \tag{7}$$

which is some unknown function $\mathcal{F}$ that is no longer independent of the source's signal. In this simplified sense, the ReTF is not generalizable to multiple simultaneous sound sources. Therefore, the ReTF becomes an ineffective spatial feature for acoustic processing (e.g. localization) in multi-source environments. This issue remains even when increasing the number of receivers [24].

III. The Relative Transfer Matrix

In this section, we propose a spatial feature analogous to the ReTF in the form of a matrix that is generalizable to multiple simultaneous sound sources. To this end, consider Q receivers indexed $q = \{1, \cdots, Q\}$ and $\mathcal{L}$ sound sources indexed $\ell = \{1, \cdots, \mathcal{L}\}$. Let us separate the receivers into two multichannel subgroups denoted by $\{A\}$ and $\{B\}$ assigned with Q_A and Q_B receivers, respectively. We illustrate one example in Fig. 2 with $Q_A = 2$, $Q_B = 4$, and $\mathcal{L} = 2$ where $\ell = \{\alpha, \beta\}$.

We express the signals received by each microphone group in matrix form as

$$\boldsymbol{M}_A(f,t) = \boldsymbol{H}_A(f)\boldsymbol{S}(f,t), \tag{8}$$
$$\boldsymbol{M}_B(f,t) = \boldsymbol{H}_B(f)\boldsymbol{S}(f,t), \tag{9}$$

where $\boldsymbol{M}_A = [M_1, \cdots, M_{Q_A}]^T$, $\boldsymbol{S} = [S_1, \cdots, S_{\mathcal{L}}]^T$, $[\cdot]^T$ is matrix transpose, and $\boldsymbol{H}_A \in \mathbb{C}^{Q_A \times \mathcal{L}}$ is a matrix with elements defined by the acoustic transfer functions. The vector $\boldsymbol{M}_B \in \mathbb{C}^{Q_B \times 1}$ and matrix $\boldsymbol{H}_B \in \mathbb{C}^{Q_B \times \mathcal{L}}$ are similar.

Following the ReTF's mapping property in (3), we intend to find a matrix denoted $\boldsymbol{\mathcal{R}}_{A,B}(f)$ which defines the spatial mapping between receiver groups- $\{A\}$ and $\{B\}$, such that

$$\boldsymbol{M}_A(f,t) = \boldsymbol{\mathcal{R}}_{A,B}(f)\boldsymbol{M}_B(f,t). \tag{10}$$

We term $\boldsymbol{\mathcal{R}}_{A,B}(f)$ the ReTM (*Relative Transfer Matrix*). A theoretical definition of the ReTM is found by left-side multiplying (9) with a suitable pseudo-inverse of $\boldsymbol{H}_B$, and then substituting the result for $\boldsymbol{S}$ in (8), to get

$$\boldsymbol{\mathcal{R}}_{A,B}(f) = \boldsymbol{H}_A(f)\boldsymbol{H}_B^{\dagger}(f), \tag{11}$$

where $(\cdot)^{\dagger}$ denotes Moore–Penrose inverse, assuming validity (i.e., $Q_B \geq \mathcal{L}$).

Just like the ReTF, we observe that the ReTM (11) is independent of the source signals. Furthermore, the ReTM is seen to be a matrix described solely by the spatial properties (the acoustic transfer functions) of the acoustic environment, the source positions, and the receiver positions. These are two of the three key properties of the ReTF we discussed in the introduction. We show that the ReTM has the third property of blind estimation next.

A. Blind Estimation of the ReTM from Received Signals

There are likely multiple approaches to estimate the ReTM directly from received signals. In this work, we explore a method using covariance matrices of

$$\boldsymbol{\mathcal{P}}_{AA}(f) \triangleq E\{\boldsymbol{M}_A\boldsymbol{M}_A^*\}, \text{ and } \boldsymbol{\mathcal{P}}_{BA}(f) \triangleq E\{\boldsymbol{M}_B\boldsymbol{M}_A^*\}, \tag{12}$$

where $[\cdot]^*$ is conjugate transpose, and $E\{\cdot\}$ denotes the expectation which can be found from averaged time frames,

$$\boldsymbol{\mathcal{P}}_{XY}(f) \cong \tfrac{1}{T}\textstyle\sum_{t=1}^{T} \boldsymbol{M}_X(f,t)\,\boldsymbol{M}_Y^*(f,t). \tag{13}$$

Using (8) and (9) in (12), we write

$$\boldsymbol{\mathcal{P}}_{AA}(f) = \boldsymbol{H}_A\boldsymbol{\mathcal{P}}_S\boldsymbol{H}_A^* \tag{14}$$
$$\boldsymbol{\mathcal{P}}_{BA}(f) = \boldsymbol{H}_B\boldsymbol{\mathcal{P}}_S\boldsymbol{H}_A^* \tag{15}$$

where $\boldsymbol{\mathcal{P}}_S \triangleq E\{\boldsymbol{S}\boldsymbol{S}^*\}$ is the expectation of the source signals. Multiplying (15) with $\boldsymbol{H}_B^{\dagger}$ and substituting into (14) gives

$$\boldsymbol{\mathcal{P}}_{AA}(f) = \boldsymbol{H}_A\boldsymbol{H}_B^{\dagger}\boldsymbol{\mathcal{P}}_{BA}(f) = \boldsymbol{\mathcal{R}}_{A,B}(f)\boldsymbol{\mathcal{P}}_{BA}(f). \tag{16}$$

Therefore, the ReTM can be estimated by applying the pseudo-inverse of $\boldsymbol{\mathcal{P}}_{BA}(f)$ the right side of $\boldsymbol{\mathcal{P}}_{AA}(f)$ in (16), giving

$$\boldsymbol{\mathcal{R}}_{A,B}(f) \approx \boldsymbol{\mathcal{P}}_{AA}(f)\boldsymbol{\mathcal{P}}_{BA}^{\dagger}(f). \tag{17}$$

Like the ReTF of (5), (17) is an approximation of the ReTM.

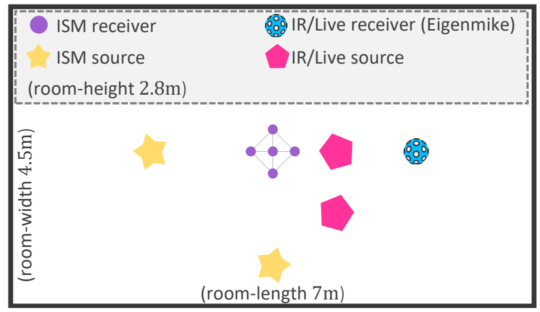

Fig. 3. Approximate setup of the receiver and source positions in the numerical simulation and experiment recordings.

Compared to the generalization of the ReTF in [24], the ReTM is similar in that it also works with the signal subspace of the acoustic system. The ReTM, however, does not require any prior knowledge of the number of sound sources. Accordingly, this independence from a pre-defined source count makes the ReTM adaptable to inactive sources unlike [24]. The main disadvantage of the ReTM is that it relies on the estimation of covariance matrices, whereas [24] does not. We do not consider our ReTM approach to be futile, however, as the covariance estimation is similar to that used in the popular MUSIC source localization approach [28] which has seen great success across many practical acoustic applications.

IV. Preliminary Analysis

There are many independent variables of the ReTM to validate such as the number of sources and receivers. The influence of each variable may differ depending on the application of the ReTM, whether it be used in a source localization, signal enhancement, or active noise control process. Therefore, we leave a more thorough analysis of the ReTM for future implementations in these specific applications.

In this section, we provide a broad preliminary analysis of the ReTM spatial mapping ability for a single scenario of two simultaneous sources and six microphones as shown in Fig. 2. We evaluate the ReTM in three cases:

1) *ISM*: Numerical simulations using the image source method;
2) *IR*: Experimental recordings from impulse responses;
3) *Live*: Experimental recordings from live loudspeakers.

Figure 3 provides an approximate illustration of the setup in each case. We inspect the condition number of covariance matrix $\boldsymbol{\mathcal{P}}_{BA}$, as well as the average magnitude spectrum error of the first (q=1) microphone in group A given by

$$\mathcal{E}_1(f) = \operatorname*{mean}_t 10\log_{10}|\hat{M}_1(f,t) - M_1(f,t)|^2/|M_1(f,t)|^2,$$

where $\hat{M}_1(f,t)$ is the estimated signal at the q=1 microphone found using the ReTM in $\hat{\boldsymbol{M}}_A \approx \mathcal{R}_{A,B}\boldsymbol{M}_B$. In this sense, $\hat{M}_1$ is a remote microphone signal estimated from the ReTM and B-microphone signals. The ReTM is considered to be a spatial map analogous to the ReTFs for multiple simultaneous sources if $\hat{M}_1$ matches M_1. Furthermore, if $\hat{M}_1$ remains accurate while the source signals change then the ReTM is shown to be independent of the emitted signals.

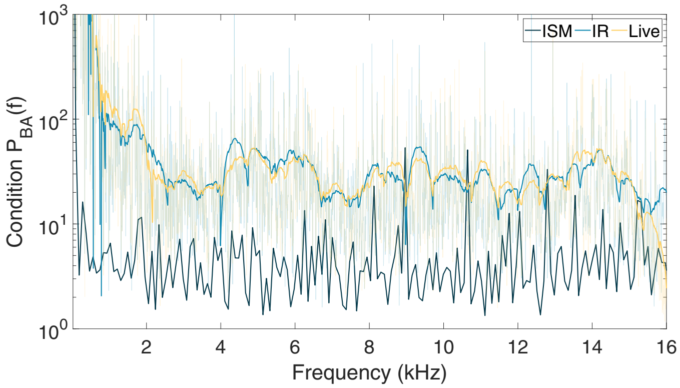

Fig. 4. Condition number of the $\boldsymbol{\mathcal{P}}_{BA}(f)$ covariance matrix.

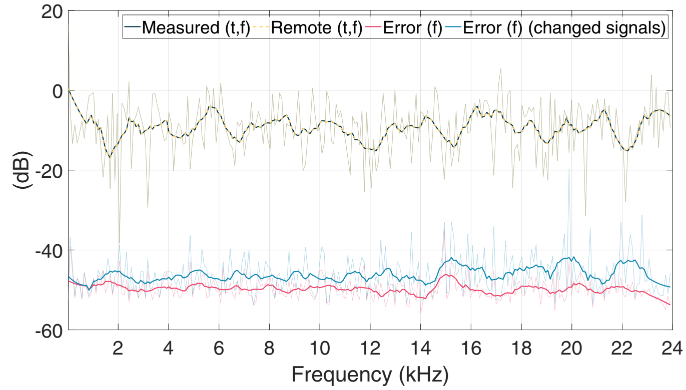

Fig. 5. Magnitude spectrum and error averaged over time for the *ISM*.

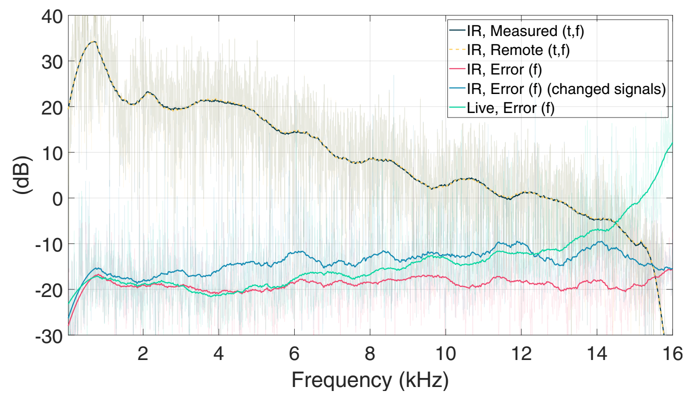

Fig. 6. Magnitude spectrum and error averaged over time for *IR* and *live*.

A. Numerical Simulations

In brief, we modeled the acoustic transfer functions between the sources and receivers with the image source method [29]. An 8^{th} order image depth and 0.9 reflection coefficient on all walls were used. The six receivers were configured as an octahedron ± 0.2 m along each axis. The two receivers on the z-axis comprised group A, and the remaining were group B. The signals were processed directly in the short-time Fourier domain for a 512 window size, 48 kHz sampling, and 10 second duration. Both sources emitted white noise.

B. Experimental Recordings

Both the *IR* and *live* recordings shared the same setup. We used an em32 Eigenmike [30] for recording with 48 kHz sampling. We assigned two receivers on the back (channels $\{18, 20\}$) to group A, and 4 receivers on the front (channels $\{$1-4$\}$) to group B. We note here that the Eigenmike is a spatially small and symmetric device. More spatially diverse microphone configurations are expected to be more useful for a ReTM application. The sources were two loudspeakers playing vacuum cleaner noise and music, placed ~1 m in front of the Eigenmike. We note that these stimulus signals were compressed by a 16 kHz low-pass, which limits our analysis. The room was a large office with minor acoustic treatment on the walls and a $T_{60} \approx 200$ ms reverberation time. The recording had noteworthy background air conditioning noise (~45 dBA noise floor). The 10 second recordings were short-time-Fourier-transformed with a 2^{14} window size that was long enough to satisfy the multiplicative transfer function [31].

C. Results

Figure 4 provides the condition number of $\boldsymbol{\mathcal{P}}_{BA}(f)$ estimated by (13). The low condition numbers indicate that ReTM estimation via (17) is robust to erroneous noise. The *ISM* simulation has a good condition of mostly <10 throughout the wide frequency band. Whereas, the *IR* and *live* experiment recordings are conditioned worse, but similar to each other with values between 10-100 above ~2 kHz. Ill-conditioning at low frequencies may be due to the Eigenmike's size.

Figure 5 gives the magnitude spectrum of the recorded ("Measured") M_1 and estimated ("Remote") $\hat{M}_1$ signals for a single time-frame, as well as their error averaged over all time frames for the *ISM* simulation. We observe an almost perfect match between M_1 and $\hat{M}_1$ which indicates the ReTM accurately maps the multiple source acoustic environment. Below -40 dB error occurs throughout the full (smoothed) 20-24k Hz band. To evaluate the ReTM's signal independence, we change the second source's signal while keeping the original ReTM estimate in (10). We denote this as "changed signals" in Fig. 5. We observe that the magnitude spectrum error of $\hat{M}_1$ remains low when the source signal changes, indicating that the ReTM is correctly mapping the source's spatial properties while being signal independent.

Figure 6 shows the magnitude spectrum error for the *IR* and *live* experiment recordings. We still see a close match between M_1 and $\hat{M}_1$ despite being real-world recordings. The *IR* error remains below -10 dB for the full (smoothed) 20-16k Hz band. We see a low error below 2 kHz despite the ill-conditioning observed in Fig. 4. The *live* recording error is also similarly good, remaining below -10 dB for most frequencies but becoming worse above 12 khz. This may be due to the source's energy being low at high frequencies compared to the non-stationary room noise. As before, we changed the source signals in the "changed signals" error while keeping the estimated ReTM constant for the *IR* recording. The error of $\hat{M}_1$ slightly worsens from the changed signals. However, it remains below -10 dB, supporting that the ReTM is independent of the multiple source signals.

V. Conclusion

We have proposed an approach to generalize the ReTM for multiple simultaneous sound sources. By separating the receivers into two multichannel groups we are able to model their coupled response due to multiple sound sources as a *relative transfer matrix*. We showed that this ReTM has the key properties of being: independent of the source signals, a useful spatial feature of the environment described by acoustic transfer functions, and that it can be blindly estimated from received signals through a covariance matrix approach. In this regard, the ReTM behaves analogously to the ReTF while being adaptable to multiple sources.

Thus far, we only provided a preliminary proof-of-concept examination of the ReTM. We expect that the ReTM can naturally extend the numerous acoustic applications of the ReTF toward multi-source environments. As such, we leave a deeper analysis of the ReTM as future work to be tailored for its implementation in each specific acoustic technique like source localization, source separation, remote microphone, etc. An analysis to provide rules-of-thumb for the aspects like the required number of receivers and their grouping is also left for the future. Nevertheless, we have proposed a novel and straightforward approach toward generalizing the ReTF to multiple sources which shows substantial promise.

References

[1] Y. Hu and S. Gannot, "Closed-form single source direction-of-arrival estimator using first-order relative harmonic coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 726–730.

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

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

[4] G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 4, pp. 711–727, 2008.

[5] R. Varzandeh, M. Taseska, and E. A. P. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *Hands-free Speech Communications and Microphone Arrays (HSCMA)*, 2017, pp. 11–15.

[6] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of relative transfer function in the presence of stationary noise based on segmental power spectral density matrix subtraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2015, pp. 320–324.

[7] N. Ito, S. Araki, and T. Nakatani, "Permutation-free clustering of relative transfer function features for blind source separation," in *Eur. Signal Process. Conf. (EUSIPCO)*. IEEE, 2015, pp. 409–413.

[8] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 3, pp. 516–527, 2010.

[9] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, 2000.

[10] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, 1997.

[11] N. Gößling, W. Middelberg, and S. Doclo, "Rtf-steered binaural mvdr beamforming incorporating multiple external microphones," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 373–377.

[12] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," *Signal Processing*, vol. 85, no. 1, pp. 177–204, 2005.

[13] S. Braun, W. Zhou, and E. A. P. Habets, "Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2015, pp. 1–5.

[14] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "Advances in network and acoustic echo cancellation," *Springer*, 2001.

[15] A. Sofer, T. Kounovskỳ, J. Čmejla, Z. Koldovskỳ, and S. Gannot, "Robust relative transfer function identification on manifolds for speech enhancement," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 401–405.

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Incorporating relative transfer function preservation into the binaural multi-channel wiener filter for hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 6500–6504.

[17] B. Yang, H. Liu, and X. Li, "Learning deep direct-path relative transfer function for binaural sound source localization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 3491–3503, 2021.

[18] L. Birnie, P. Samarasinghe, T. Abhayapala, and D. Grixti-Cheng, "Noise retf estimation and removal for low snr speech enhancement," in *Proc. IEEE Workshop Mach. Learning Signal Process. (MLSP)*, 2021, pp. 1–6.

[19] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of the direct-path relative transfer function for supervised sound-source localization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 11, pp. 2171–2186, 2016.

[20] Y. Hu, P. N. Samarasinghe, and T. D. Abhayapala, "Sound source localization using relative harmonic coefficients in modal domain," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2019, pp. 348–352.

[21] Y. Hu, P. N. Samarasinghe, T. D. Abhayapala, and S. Gannot, "Unsupervised multiple source localization using relative harmonic coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 571–575.

[22] Y. Hu, P. N. Samarasinghe, S. Gannot, and T. D. Abhayapala, "Decoupled multiple speaker direction-of-arrival estimator under reverberant environments," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 30, pp. 3120–3133, 2022.

[23] M. Taseska and E. A. P. Habets, "Relative transfer function estimation exploiting instantaneous signals and the signal subspace," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, 2015, pp. 404–408.

[24] A. Deleforge, S. Gannot, and W. Kellermann, "Towards a generalization of relative transfer functions to more than one source," in *Eur. Signal Process. Conf. (EUSIPCO)*. IEEE, 2015, pp. 419–423.

[25] A. Roure and A. Albarrazin, "The remote microphone technique for active noise control," in *INTER-NOISE and NOISE-CON congress and conference proceedings*, vol. 1999, no. 5. Institute of Noise Control Engineering, 1999, pp. 1233–1244.

[26] H. Sun, J. Zhang, T. D. Abhayapala, and P. N. Samarasinghe, "Active noise control over 3d space with remote microphone technique in the wave domain," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2021, pp. 301–305.

[27] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2055–2063, 1996.

[28] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

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

[30] M. Acoustics, "Em32 eigenmike microphone array release notes (v17. 0)," *25 Summit Ave, Summit, NJ 07901, USA*, 2013.

[31] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, 2007.