

On Beamforming with the Single-Sideband Transform

Vitor Probst Curtarelli *  and Israel Cohen * 

Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 3200003, Israel

* Correspondence: vitor.c@campus.technion.ac.il (V.P.C.); icohen@ee.technion.ac.il (I.C.)

Abstract: In this paper, we examine the use of the Single-Sideband Transform (SSBT) for convolutive beamformers. We explore its unique properties and implications for beamformer design. Our study sheds light on the tradeoffs involved in using the SSBT in beamforming applications, offering insights into both its strengths and limitations. Despite the advantage of having real-valued coefficients, we show that the convolution handling of the transform presents challenges that impact fundamental beamforming principles. When compared to the Short-Time Fourier Transform (STFT), the SSBT displays lower robustness, especially in scenarios involving mismatch and modeling noise. Notably, we establish a direct equivalence between the SSBT and STFT when using identical transform parameters, enabling their seamless interchangeability and joint use in time–frequency signal enhancements. We validate our theoretical findings through realistic simulations using the Minimum-Power Distortionless Response beamformer. These simulations illustrate that although the STFT performs marginally better than the SSBT under optimal conditions, it outperforms significantly in non-ideal scenarios.

Keywords: single-sideband transform; time–frequency transform; beamforming; array signal processing; signal enhancement



Citation: Curtarelli, V.P.; Cohen, I. On Beamforming with the Single-Sideband Transform. *Appl. Sci.* **2024**, *14*, 7514. <https://doi.org/10.3390/app14177514>

Academic Editors: Jie Zhang, Douglas O’Shaughnessy and Lamberto Tronchin

Received: 23 June 2024

Revised: 1 August 2024

Accepted: 22 August 2024

Published: 25 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Filtering is an essential tool available in the modern world, being necessary in fields such as communications [1], biomedical applications [2–4], and system control [5,6], among other areas [7–9]. At its core, filtering mitigates the effects of undesired sources and can use temporal and spatial information regarding the signals, sources, sensors, and environment to enhance the signals. Given time samples, these filtering processes can be implemented in the time-, frequency-, and time–frequency domains [10], each offering distinct advantages. In particular, time–frequency methods exploit frequency-related information while dynamically adapting to signal and environmental changes in time, providing a tradeoff between strictly time or frequency information. Generically, transforms are the primary tool for achieving desired time–frequency domain data, with the Short-Time Fourier Transform (STFT) [11,12] being prominent in the literature due to its widespread use. However, alternative transforms can also be employed [13–15], each offering a unique perspective and insight regarding the signal, leading to different performances depending on specific application requirements.

Among these alternatives, an approach based on single-sideband modulation has occasionally been proposed [16] for various applications in signal enhancement [17,18], from acoustic echo cancellation [19] to speech de-reverberation [20] and machine learning signal enhancement [21], showing varying levels of success. The Single-Sideband Transform (SSBT) holds promise in the field due to its real-valued representation. This characteristic simplifies mathematical developments by avoiding complex-valued coefficients or matrices, and it also leads to a more straightforward hardware implementation, resulting in more cost-efficient devices. Previous research shows that this transform works best with short time–frequency frames [20,21]. Despite this, the research on applying this transform for beamforming is limited. In particular, its potential for spatio-temporal multisensor

beamforming in reverberant scenarios [22] under a convolutive transfer function (CTF) model [23] is still open to be explored. However, it is necessary to understand that the transform's properties is crucial to ensure accurate and meaningful outputs. The SSBT lacks comprehensive examination in the literature, resulting in a limited understanding of its features and limitations compared to other methodologies. This knowledge gap complicates the appropriate application and utilization of this transform. This paper aims to explore the SSBT properties, how they interact with traditional beamforming assumptions, and how to properly implement a beamformer, specifically the Minimum-Power Distortionless Response, under this new transform.

We begin by examining a continuous-frequency version of the SSBT, comparing its properties to those of the Fourier Transform (FT) and the STFT. We study how these properties impact basic beamforming concepts such as the convolution theorem and relative frequency response estimation for the SSBT. While the SSBT is more error-prone and restrictive than the STFT regarding beamforming design, we also demonstrate a bijective interchangeability between them. This allows for their joint usage, potentially enhancing beamformer design within the SSBT domain by converting into the STFT and back without depending on inverse transforms, which are computationally intensive. We also test its direct implementation for beamforming in a real-life-like reverberant scenario, comparing it to the STFT in this regard. Our theoretical findings matched the experimental results: the beamformers based on SSBT were slightly less effective than the ones based on STFT in ideal conditions, and significantly less effective in non-ideal scenarios. This emphasizes the limitations of the SSBT compared to the STFT in practical beamforming.

The remainder of this paper is organized as follows: In Section 2, we introduce the proposed frequency and time–frequency transforms, explain their relationship, and elaborate on their relevant properties. These properties are fully developed in Appendix A where proofs are necessary. Section 3 presents the considered signal model and how to incorporate the desired constraints while considering the time–frequency transforms at hand, considering their features and peculiarities. In Section 4, we present and discuss the results, comparing the studied filtering approaches for various metrics and situations, each exposing different types of information regarding their performances. Finally, Section 5 concludes this paper.

2. Frequency and Time–Frequency Transforms

Hereafter, we assume that all time-domain signals are real-valued, which allows for shortcuts to some transforms and enables the use of others.

For continuous time and frequency domains, the Fourier Transform (FT) of a time signal $x(t)$ is defined as

$$\begin{aligned} X_{\mathcal{F}}(f) &\equiv \mathcal{F}\{x(t)\}(f) \\ &= \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \end{aligned} \quad (1)$$

where $X_{\mathcal{F}}(f) = X_{\mathcal{F}}^*(-f)$ if $x(t)$ is real-valued, with $(\cdot)^*$ being the complex-conjugate operation.

We define the Real Fourier Transform (RFT) similarly to the FT, constructed such that its frequency spectrum is real-valued without loss of information. The RFT is given by

$$\begin{aligned} X_{\mathcal{R}}(f) &\equiv \mathcal{R}\{x(t)\}(f) \\ &= \sqrt{2}\mathcal{R}\left\{\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft + j\frac{3\pi}{4}} dt\right\} \\ &= \int_{-\infty}^{\infty} x(t)[- \cos(2\pi ft) + \sin(2\pi ft)] dt \end{aligned} \quad (2)$$

and the Inverse Real Fourier Transform (IRFT) as below (see Property A2 in Appendix A):

$$\begin{aligned}
 x(t) &\equiv \mathcal{R}^{-1}\{X_{\mathcal{R}}(f)\}(t) \\
 &= \sqrt{2}\mathcal{R}\left\{\int_{-\infty}^{\infty} X_{\mathcal{R}}(f)e^{j2\pi ft-j\frac{3\pi}{4}}df\right\}.
 \end{aligned}
 \tag{3}$$

The RFT can also be defined via the FT through a simple substitution of Equation (1) in Equation (2) (see Property A1 in Appendix A), resulting in

$$X_{\mathcal{R}}(f) = -X_{\mathcal{F}}^{\mathcal{R}}(f) - X_{\mathcal{F}}^{\mathcal{I}}(f).
 \tag{4}$$

Using the fact that $x(t)$ is a real signal, it is possible to achieve the following (Property A1, Appendix A):

$$X_{\mathcal{R}}(f) = \frac{1}{\sqrt{2}}\left(e^{j\frac{3\pi}{4}}X_{\mathcal{F}}(f) + e^{-j\frac{3\pi}{4}}X_{\mathcal{F}}(-f)\right)
 \tag{5a}$$

$$X_{\mathcal{F}}(f) = \frac{1}{\sqrt{2}}\left(e^{-j\frac{3\pi}{4}}X_{\mathcal{R}}(f) + e^{j\frac{3\pi}{4}}X_{\mathcal{R}}(-f)\right).
 \tag{5b}$$

This means that the RFT can be defined in terms of the FT and vice-versa, forming a bijective relationship between the two transforms. This is only true if the original time-domain signal is real-valued, which is a requirement for the RFT to be invertible in the first place. A similar result was obtained previously (Equations (4) and (8) in [21]), where it was shown that the SSB representation depends on the complex-conjugate of the single-sideband modulated signal. Here, this complex-conjugate is represented by the negative frequency component, these being the same under the assumption of $x(t)$ being real-valued.

2.1. Convolution

Given an LTI system with an impulse response $h(t)$, the convolution theorem for the Fourier transform states that

$$h(t) * x(t) \stackrel{\mathcal{F}}{\rightleftharpoons} H_{\mathcal{F}}(f)X_{\mathcal{F}}(f)
 \tag{6}$$

where $\stackrel{\mathcal{F}}{\rightleftharpoons}$ indicates a Fourier transform pair. This theorem is not strictly valid for the RFT (see Property A3 in Appendix A). However, it is possible to prove that there is an equivalent of the convolution theorem for the RFT (see Property A4 in Appendix A), with

$$h(t) * x(t) \stackrel{\mathcal{R}}{\rightleftharpoons} X_{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{R}}(f) + X_{\mathcal{R}}(-f)H_{\mathcal{F}}^{\mathcal{I}}(f)
 \tag{7}$$

where $\stackrel{\mathcal{R}}{\rightleftharpoons}$ indicates an RFT pair. For a given frequency f , the convolution's output on the RFT domain depends on both it and its dual frequency $-f$. This makes intuitive sense, as the real-valued spectrum of the RFT impedes a correct phase representation using only the given frequency, making its conjugate frequency also necessary.

2.2. Relative Frequency Responses

Given two systems with an input $x(t)$, each with an impulse response $h_1(t)$ and $h_2(t)$, we can calculate their relative frequency responses (RFRs) respective to the output of one of the systems (assumed to be the first without compromise), these being denoted $A_1(f)$ and $A_2(f)$. Let $X_1(f)$ be the first system's output, given by

$$X_1(f) = H_1(f)X(f)
 \tag{8}$$

and similarly for $Y_2(f)$. Clearly, $X_1(f) = A_1(f)X_1(f)$ only if $A_1(f) = 1$. We can obtain $A_2(f)$ as

$$A_2(f) = \frac{H_2(f)}{H_1(f)} \tag{9}$$

which satisfies $A_2(f)X_1(f) = H_2(f)X(f)$. These RFRs can also be calculated as

$$A_m(f) = \frac{\mathbb{E}\{X_m(f)X_1^*(f)\}}{\mathbb{E}\{X_1(f)X_1^*(f)\}} \tag{10}$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator. It is easy to see that Equations (9) and (10) are equivalent, at least in an ideal scenario.

The RFR in the RFT domain is more intricate since each frequency depends on its conjugate after the convolution. We hereby have frequency responses $H'_m(f)$ and $H''_m(f)$ for each system, as well as two inputs $X'(f) = X(f)$ and $X''(f) = X(-f)$. Then, our outputs can be described as

$$X_1(f) = H'_1(f)X'(f) + H''_1(f)X''(f) \tag{11a}$$

$$X_2(f) = H'_2(f)X'(f) + H''_2(f)X''(f). \tag{11b}$$

From Equation (7), we easily see that

$$H'_m(f) = H'_m(-f) = H^{\mathbb{R}}_{\mathcal{F};m}(f) \tag{12a}$$

$$H''_m(f) = -H''_m(-f) = H^{\mathbb{I}}_{\mathcal{F};m}(f). \tag{12b}$$

We now let $X'_1(f) = X_1(f)$ and $X''_1(f) = X_1(-f)$, as our new system inputs. We write

$$X_1(f) = A'_1(f)X'_1(f) + A''_1(f)X''_1(f) \tag{13a}$$

$$X_2(f) = A'_2(f)X'_1(f) + A''_2(f)X''_1(f) \tag{13b}$$

where $A'_m(f)$ and $A''_m(f)$ are the relative frequency responses between the new inputs $X'_1(f)$ and $X''_1(f)$ and the output $X_m(f)$. Similarly to the FT, we have $A'_1(f) = 1$ and $A''_1(f) = 0$. Using Equations (11a) and (12) in Equation (13a), we obtain

$$X_2(f) = [A'_2(f)H'_1(f) - A''_2(f)H''_1(f)]X'(f) + [A'_2(f)H''_1(f) + A''_2(f)H'_1(f)]X''(f) \tag{14}$$

and, by comparing with $X_2(f)$ from Equation (11b), we form a system of equations

$$A'_2(f)H'_1(f) - A''_2(f)H''_1(f) = H'_2(f) \tag{15a}$$

$$A'_2(f)H''_1(f) + A''_2(f)H'_1(f) = H''_2(f) \tag{15b}$$

whose solution is

$$A'_2(f) = \frac{H'_1(f)H'_2(f) + H''_1(f)H''_2(f)}{H_1'^2(f) + H_1''^2(f)} \tag{16a}$$

$$A''_2(f) = \frac{H'_1(f)H''_2(f) - H''_1(f)H'_2(f)}{H_1'^2(f) + H_1''^2(f)}. \tag{16b}$$

Through the properties of the RFT in Appendix A, we obtain

$$\mathbb{E}\{X'_1(f)X'_1(f)\} = (H_1'^2(f) + H_1''^2(f))\sigma_X^2(f) \tag{17a}$$

$$\mathbb{E}\{X'_1(f)X_2(f)\} = (H'_1(f)H'_2(f) + H''_1(f)H''_2(f))\sigma_X^2(f) \tag{17b}$$

$$\mathbb{E}\{X''_1(f)X_2(f)\} = (H'_1(f)H''_2(f) - H''_1(f)H'_2(f))\sigma_X^2(f) \tag{17c}$$

where $\sigma_X^2(f) = \mathbb{E}\{X_{\mathcal{F}}(f)^2\} = \mathbb{E}\{X_{\mathcal{R}}(f)^2\}$. From this, our RFRs take a similar form to Equation (9), resulting in

$$A_2'(f) = \frac{\mathbb{E}\{X_1'(f)X_2(f)\}}{\mathbb{E}\{X_1'(f)X_1'(f)\}} \tag{18a}$$

$$A_2''(f) = \frac{\mathbb{E}\{X_1''(f)X_2(f)\}}{\mathbb{E}\{X_1'(f)X_1'(f)\}}. \tag{18b}$$

The RFRs in Equation (18) make intuitive sense: $A_2'(f)$ uses the correlation between $X_2(f)$ and $X_1'(f)$ (same frequency), and $A_2''(f)$ the correlation between $X_2(f)$ and $X_1''(f) = X_1(-f)$ (conjugate frequency), which is in line with Equation (13).

Generalizing to a situation with M sensors, one can follow the same steps and achieve that for each m -th sensor, we have

$$A_m'(f) = \frac{\mathbb{E}\{X_1'(f)X_m(f)\}}{\mathbb{E}\{X_1^2(f)\}} \tag{19a}$$

$$A_m''(f) = \frac{\mathbb{E}\{X_1''(f)X_m(f)\}}{\mathbb{E}\{X_1^2(f)\}}. \tag{19b}$$

Notably, Equation (19a) reduces to $A_1'(f) = 1$ and $A_1''(f) = 0$, using the fact that $X_1(f)$ and $X_1(-f)$ are uncorrelated (see Property A6).

Observing strictly the mathematical structures of Equation (8) and Equation (11), the FT can be treated as a particular case of the RFT formulation, where $A_m'(f) = A_{m;\mathcal{F}}(f)$, and $A_m''(f) = 0$. Knowing this, we use the SSBT formulation of Equation (26) for both the SSBT and the STFT, as it is a more general model. The necessary considerations are taken when particularizing the equations for the STFT.

2.3. Discrete Time–Frequency Transforms

The Short-Time Fourier Transform (STFT) [11,12] of a discrete-time signal $x[n]$ is

$$X_{\mathcal{F}}[l, k] = \sum_{n=0}^{K-1} w[n]x[n + l \cdot O]e^{-j2\pi k \frac{(n+l \cdot O)}{K}} \tag{20}$$

where $w[n]$ is an analysis window of length K , and O is the hop size between successive windows of the transform, usually $O = \lfloor K/2 \rfloor$. l is the frame index, and k is the frequency bin index. The STFT is generally seen as a discretization of the FT, being applied over sequential time “snippets”.

The Single-Sideband Transform (SSBT) [16,24] is similarly defined, being the RFT’s windowed-discrete-time equivalent. The SSBT of $x[n]$ is

$$X_S[l, k] = \sqrt{2}\mathbb{R}\left\{\sum_{n=0}^{K-1} w[n]x[n + l \cdot O]e^{-j2\pi k \frac{(n+l \cdot O)}{K} + j\frac{3\pi}{4}}\right\}. \tag{21}$$

One advantage of using the STFT is the need for only $\lfloor (K+1)/2 \rfloor + 1$ frequency bins, given its complex-conjugate behavior for real-time-domain signals. Meanwhile, the SSBT requires all K bins to correctly capture all the information of $x[n]$, but its coefficients are real.

Assuming that all K bins of the STFT are available (even though they are not all necessary), similarly to Equations (4), (5a) and (5b) we now have

$$\begin{aligned} X_S[l, k] &= \sqrt{2}\mathbb{R}\left\{X_{\mathcal{F}}[l, k]e^{j\frac{3\pi}{4}}\right\} \\ &= -\mathbb{R}\{X_{\mathcal{F}}[l, k]\} - \mathbb{I}\{X_{\mathcal{F}}[l, k]\}. \end{aligned} \tag{22}$$

For the abuse of notation, we let $X_S[l, K] \equiv X_S[l, 0]$, and equally for $X_{\mathcal{F}}[l, K]$.

As is the case for the RFT, the SSBT also does not hold the convolution theorem in the same way as the STFT. However, similarly to what is shown in Equation (7) (and in Property A4), the convolution on the SSBT domain through the MTF model [25] can be given by

$$h[n] * x[n] \stackrel{\text{S}}{\rightleftharpoons} X_S[l, k]H'_S[k] + X_S[l, K - k]H''_S[k] \tag{23}$$

or, with the CTF model,

$$h[n] * x[n] \stackrel{\text{S}}{\rightleftharpoons} X_S[l, k] * H'_S[l, k] + X_S[l, K - k] * H''_S[l, k] \tag{24}$$

in which this convolution is performed over l , with $H'_S[l, k] = H^R_{\mathcal{F}}[l, k]$ and $H''_S[l, k] = H^I_{\mathcal{F}}[l, k]$. Note that this is an approximation, as cross-band interference [26,27] is necessary to model the convolution perfectly; however, this effect is not considered here for either transform. The conjugate frequency's presence is a byproduct distinct to the cross-band interference, the latter happening due to aliasing and the windowing processes, and the former coming from the continuous-time convolution theorem for the SSBT (see Section 2.1 and Property A3).

Similarly to the relationship presented in Equation (5), we have

$$X_S[l, k] = \frac{1}{\sqrt{2}} \left(e^{j\frac{3\pi}{4}} X_{\mathcal{F}}[l, k] + e^{-j\frac{3\pi}{4}} X_{\mathcal{F}}[l, K - k] \right), \tag{25a}$$

$$X_{\mathcal{F}}[l, k] = \frac{1}{\sqrt{2}} \left(e^{-j\frac{3\pi}{4}} X_S[l, k] + e^{j\frac{3\pi}{4}} X_S[l, K - k] \right). \tag{25b}$$

Equations (25a) and (25b) give us a bijective relationship between time–frequency signals in the SSBT and STFT domains. That is, for the same transform parameters (window type and size, overlap, etc.), it is possible to convert from one to the other without using their inverses and going into the time-domain, which is resource-consuming. This interchangeability means they can be used together, transforming from one to the other for situations where each is more advantageous.

2.4. RFR Estimation for Time–Frequency Transforms

Under the same assumptions as in Section 2.2, we can write the output of our systems with the CTF model in the SSBT domain as

$$X_{m;S}[l, k] = X'_{1;S}[l, k] * A'_{m;S}[l, k] + X''_{1;S}[l, k] * A''_{m;S}[l, k] \tag{26}$$

where $X'_{1;S}[l, k]$, $X''_{1;S}[l, k]$, $A'_{m;S}[l, k]$ and $A''_{m;S}[l, k]$ are defined similarly to their counterparts from Section 2.2. In particular, given $X_{m;\mathcal{F}}[l, k]$ and $X_{1;\mathcal{F}}[l, k]$ under the STFT, we have

$$\begin{aligned} \mathbb{E} \left\{ X_{m;\mathcal{F}}[l + \lambda, k] X_{1;\mathcal{F}}^*[l, k] \right\} &= \mathbb{E} \left\{ \sum_{\tau} A_{m;\mathcal{F}}[\tau, k] X_{1;\mathcal{F}}[l + \lambda - \tau, k] X_{1;\mathcal{F}}^*[l, k] \right\} \\ &= \sum_{\tau} A_{m;\mathcal{F}}[\tau, k] \mathbb{E} \left\{ X_{1;\mathcal{F}}[l + \lambda - \tau, k] X_{1;\mathcal{F}}^*[l, k] \right\} \\ &= \sum_{\tau} A_{m;\mathcal{F}}[\tau, k] \sigma_{X_{1,\lambda-\tau}}^2[k] \\ &= A_{m;\mathcal{F}}[\lambda, k] \sigma_{X_{1,0}}^2[k] + \sum_{\tau \neq \lambda} A_{m;\mathcal{F}}[\tau, k] \sigma_{X_{1,\lambda-\tau}}^2[k]. \end{aligned} \tag{27}$$

This is equal to $A_{m;\mathcal{F}}[\lambda, k] \sigma_{X_{1,0}}^2$ iff $\sigma_{X_{1,\lambda-\tau}}^2[k] = 0 \forall \lambda \neq \tau$, which is equivalent to different samples of $X_{1;\mathcal{F}}[l, k]$ being uncorrelated. This contradicts the fact that $X_{1;\mathcal{F}}[l, k]$ is the output of a convolution, thus having correlated samples. The overlap between windows in the transform also contributes to the correlation of different windows. With this, the summation over $\tau \neq \lambda$ is an error term in the RFR estimation.

Applying this same process to the SSBT, we have that

$$\begin{aligned} & \mathbb{E}\{X_{1;s}[l, k]X_{m;s}[l + \lambda, k]\} \\ & = A_{m;s}[\lambda, k]\sigma_{X_{1;0}}^2[k] + \sum_{\tau \neq \lambda} A'_{m;s}[\tau, k]\sigma_{X'_{\lambda-\tau}}^2[k] + A''_{m;s}[\tau, k]\sigma_{X''_{\lambda-\tau}}^2[k]. \end{aligned} \tag{28}$$

In this scenario, there is an error term for the same-frequency component but also one for the cross-frequency interference. Expanding $\sigma_{X''_{\lambda-\tau}}^2[k]$ leads to

$$\begin{aligned} \sigma_{X''_{\lambda-\tau}}^2[k] & = \mathbb{E}\{X'_{1;s}[l + \tau - \lambda, k]X''_{1;s}[l, k]\} \\ & = \sigma_X^2[k] \sum_{\mu \geq 0} (H'_{1,\mu} - H'_{1,-\mu})(H''_{1,\mu+(\lambda-\tau)} - H''_{1,\mu-(\lambda-\tau)}). \end{aligned} \tag{29}$$

Three variables are of interest: the frame delay between different sensors λ , the RFR index τ , and the delay between same- and cross-frequency μ . In all cases where $\mu \neq 0$ and $\tau \neq \lambda$, the summation term of Equation (29) is not trivially null, and, therefore, the summation over τ in Equation (28) is also not identically zero. Comparing Equations (27) and (28) reveals that the SSBT estimation introduces error terms due to the cross-frequency cross-frame correlation, an issue absent in the STFT. This difference diminishes the robustness and performance of beamformers designed in the SSBT domain, potentially increasing output distortion due to inaccuracies in estimating relative frequency responses.

3. Signal Model and Beamforming

Let a device in a reverberant environment consist of M sensors and a loudspeaker. We assume the presence of a desired source and undesired uncorrelated noise at each sensor of the device. For simplicity, we assume that the environment and sources are spatially stationary, although this condition can be relaxed. Let $y_m[n]$ be the observed signal at the m -th sensor, given by

$$y_m[n] = x_m[n] + s_m[n] + r_m[n] \tag{30}$$

where $x_m[n]$ is a desired component, $s_m[n]$ is the undesired interfering signal (from the loudspeaker) captured by the sensors, and $r_m[n]$ is uncorrelated noise present in the sensors. The index m ($1 \leq m \leq M$) refers to the different sensors. Regarding the environment as an LTI system, then

$$\begin{aligned} x_m[n] & = h_m[n] * x[n] \\ & = a_m[n] * x_1[n] \end{aligned} \tag{31}$$

with $h_m[n]$ being the frequency response between the desired source $x[n]$ and the m -th sensor, and $a_m[n]$ is the relative impulse response between the reference (assumed to be $m = 1$) and the m -th sensors. In the time–frequency domain, Equation (30) becomes

$$Y_{m,k}[l] = X_{m,k}[l] + S_{m,k}[l] + R_{m,k}[l]. \tag{32}$$

Hereafter, the notation is different than it was previously to ease reading and emphasize the frame index l .

From Equation (31), using Equation (26) and assuming the CTF model, we can write $X_{m,k}[l]$ as

$$X_{m,k}[l] \approx A'_{m,k}[l] * X'_{1,k}[l] + A''_{m,k}[l] * X''_{1,k}[l] \tag{33}$$

with $X'_{1,k}[l] = X_{1,k}[l]$ and $X''_{1,k}[l] = X_{1,K-k}[l]$, and $A^n_{m,k}[l]$ being the RFRs between each sensor and the reference for same- and conjugate-frequencies. Note that $A^n_{m,k}[l]$ is not strictly causal, depending on the direction of arrival and features of the reverberant environment, as well as relative delays between the sources at each sensor. It is trivial to see that, for Equation (33) to be respected, $A'_{1,k}[l] = A''_{1,k}[l] = \delta_{0,l}$, which is a Kronecker delta at $l = 0$.

By approximating $A_{m,k}^n[l]$ through a truncated RFR $D_{m,k}^n[l]$ with L_D samples, we obtain

$$\begin{aligned} X_{m,k}^n[l] &= A_{m,k}^n[l] * X_{1,k}^n[l] \\ &\approx D_{m,k}^n[l] * X_{1,k}^n[l] \end{aligned} \tag{34}$$

in which $D_{m,k}^n[l]$ has Δ non-causal and Λ causal samples, with $L_D = \Delta + \Lambda + 1$. By defining $L_D \times 1$ vectors $\mathbf{x}_{1,k}^n[l]$ and $\mathbf{d}_{m,k}^n$

$$\mathbf{x}_{1,k}^n[l] = \left[X_{1,k}^n[l + \Delta], \dots, X_{1,k}^n[l], \dots, X_{1,k}^n[l - \Lambda] \right]^T \tag{35a}$$

$$\mathbf{d}_{m,k}^n = \left[A_{m,k}^n[-\Delta], \dots, A_{m,k}^n[0], \dots, A_{m,k}^n[\Lambda] \right]^T, \tag{35b}$$

we can rewrite the convolution of Equation (34) as a vector multiplication, given by

$$D_{m,k}^n[l] * X_{1,k}^n[l] = \mathbf{d}_{m,k}^{n\top} \mathbf{x}_{1,k}^n[l] \tag{36}$$

with $(\cdot)^T$ being the transpose operator. Then, our observed signal $Y_{m,k}[l]$ becomes

$$Y_{m,k}[l] \approx \mathbf{d}_{m,k}'^T \mathbf{x}_{1,k}'[l] + \mathbf{d}_{m,k}''^T \mathbf{x}_{1,k}''[l] + S_{m,k}[l] + R_{m,k}[l]. \tag{37}$$

The signal vectors $\mathbf{x}_{1,k}'[l]$ and $\mathbf{x}_{1,k}''[l]$ are still frame-dependent, while the RFR vectors $\mathbf{d}_{m,k}'$ and $\mathbf{d}_{m,k}''$ are not due to the assumption of spatial stationarity. Taking the L_Y most recent samples of $Y_{m,k}[l]$ implies

$$\begin{aligned} \mathbf{Y}_{m,k}[l] &= \left[Y_{m,k}[l], Y_{m,k}[l-1], \dots, Y_{m,k}[l-L_Y+1] \right]^T \\ &= \bar{\mathbf{D}}_{m,k}' \bar{\mathbf{x}}_{1,k}'[l] + \bar{\mathbf{D}}_{m,k}'' \bar{\mathbf{x}}_{1,k}''[l] + \mathbf{S}_{m,k}[l] + \mathbf{R}_{m,k}[l] \end{aligned} \tag{38}$$

in which $\bar{\mathbf{D}}_{m,k}^n$ is an $L_Y \times L$ Toeplitz matrix with $L = L_D + L_Y - 1$, and $\bar{\mathbf{x}}_{1,k}^n[l]$ is an $L \times 1$ vector of our desired signal

$$\bar{\mathbf{D}}_{m,k}^n = \begin{bmatrix} \mathbf{d}_{m,k}^{n\top} & 0 & \dots & 0 \\ 0 & \mathbf{d}_{m,k}^{n\top} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{d}_{m,k}^{n\top} \end{bmatrix} \tag{39a}$$

$$\bar{\mathbf{x}}_{1,k}^n[l] = \left[X_{1,k}^n[l + \Delta], \dots, X_{1,k}^n[l], \dots, X_{1,k}^n[l - (L_Y + \Lambda - 1)] \right]^T. \tag{39b}$$

Concatenating the observed signals sensor-wise yields $\mathbf{y}_k[l]$, defined by

$$\mathbf{y}_k[l] = \left[\mathbf{Y}_{1,k}^T, \dots, \mathbf{Y}_{M,k}^T \right]^T \tag{40a}$$

$$= \mathbf{D}_k' \bar{\mathbf{x}}_{1,k}'[l] + \mathbf{D}_k'' \bar{\mathbf{x}}_{1,k}''[l] + \mathbf{s}_k[l] + \mathbf{r}_k[l] \tag{40b}$$

and

$$\mathbf{D}_k^n = \begin{bmatrix} \bar{\mathbf{D}}_{1,k}^n \\ \vdots \\ \bar{\mathbf{D}}_{M,k}^n \end{bmatrix} \tag{41}$$

where $\mathbf{y}_k[l]$ is an $(ML_Y) \times 1$ vector, and \mathbf{D}_k^n is an $(ML_Y) \times L$ matrix. $\mathbf{s}_k[l]$ and $\mathbf{r}_k[l]$ are defined similarly to Equation (40a).

3.1. Filtering and the MPDR Beamformer

Our objective is to recover the desired signal at the reference sensor $X_{1,k}[l]$ without any distortion while minimizing the output signal's power. For this, a linear filter $\mathbf{f}_k[l]$ is employed, yielding an estimate $Z_k[l]$ of our desired signal:

$$\begin{aligned} Z_k^n[l] &\approx X_{1,k}[l] \\ &= \mathbf{f}_k^H[l] \mathbf{y}_k[l] \end{aligned} \tag{42}$$

with $(\cdot)^H$ being the transposed-complex-conjugate operator. Using Equation (40b) in Equation (42) yields

$$\begin{aligned} Z_k[l] &= \mathbf{f}_k^H[l] \mathbf{D}'_k \bar{\mathbf{x}}'_{1,k}[l] + \mathbf{f}_k^H[l] \mathbf{D}''_k \bar{\mathbf{x}}''_{1,k}[l] + \mathbf{f}_k^H[l] \mathbf{s}_k[l] + \mathbf{f}_k^H[l] \mathbf{r}_k[l] \\ &= X_{f,k}[l] + S_{f,k}[l] + R_{f,k}[l] \end{aligned} \tag{43}$$

where $X_{f,k}[l]$ is the filtered desired signal, $S_{f,k}[l]$ is the filtered interference signal, and $R_{f,k}[l]$ is the filtered noise signal. In particular, the filtered desired signal is

$$X_{f,k}[l] = \mathbf{f}_k^H[l] \mathbf{D}'_k \bar{\mathbf{x}}'_{1,k}[l] + \mathbf{f}_k^H[l] \mathbf{D}''_k \bar{\mathbf{x}}''_{1,k}[l] \tag{44}$$

where each component of the desired signal is exposed. The distortionless constraint on the desired signal is translated as

$$X_{f,k}[l] = X_{1,k}[l] \tag{45}$$

which is equivalent to requiring that each component of $X_{1,k}[l]$ is perfectly recovered. This means maintaining the same-frequency component while nulling the cross-frequency parcel that appears in Equation (44). From Equation (39b), we have that the desired signal for the current index l is the $(\Delta + 1)$ -th element of $\bar{\mathbf{x}}''_{1,k}[l]$; thus, the constraints are

$$\mathbf{f}_k^H[l] \mathbf{D}'_k = \mathbf{i}_\Delta^T \tag{46a}$$

$$\mathbf{f}_k^H[l] \mathbf{D}''_k = \mathbf{0}^T \tag{46b}$$

where \mathbf{i}_Δ is an $L \times 1$ vector of zeros except for the $(\Delta + 1)$ -th entry, which is 1, and $\mathbf{0}$ is an $L \times 1$ vector of zeros. For the STFT, only the first constraint of Equation (46) is considered since \mathbf{D}''_k is identically zero by definition, and, therefore, the second condition is trivially satisfied. With this, we write our constraint matrix as

$$\mathbf{f}_k^H[l] \mathbf{C}_k = \mathbf{i}^T \tag{47}$$

where for the STFT, $\mathbf{C}_k = \mathbf{D}'_k$ and $\mathbf{i} = \mathbf{i}_\Delta$; and for the SSBT, $\mathbf{C}_k = [\mathbf{D}'_k, \mathbf{D}''_k]$, and $\mathbf{i} = \begin{bmatrix} \mathbf{i}_\Delta \\ \mathbf{0} \end{bmatrix}$.

To minimize the output signal's power while satisfying the distortionless constraint, a Minimum-Power Distortionless Response (MPDR) beamformer is used, defined by

$$\mathbf{f}_{\text{mpdr};k}[l] = \min_{\mathbf{f}_k[l]} \mathbf{f}_k^H[l] \mathbf{\Gamma}_{\alpha;k}[l] \mathbf{f}_k[l] \text{ s.t. } \mathbf{f}_k^H[l] \mathbf{C}_k = \mathbf{i}^T \tag{48}$$

where $\mathbf{\Gamma}_{\alpha;k}[l] = (1 - \alpha) \mathbf{\Gamma}_{\mathbf{y}_k}[l] + \alpha \mathbf{I}$, with $\mathbf{\Gamma}_{\mathbf{y}_k}[l]$ being the pseudo-correlation matrix of $\mathbf{y}_k[l]$, and \mathbf{I} is the identity matrix, both of size $ML_Y \times ML_Y$. α is a regularization parameter for white noise gain control [28]. The solution to the minimization problem in Equation (48) is given by

$$\mathbf{f}_{\text{mpdr};k}[l] = \mathbf{\Phi}_{\mathbf{y}_k}^{-1}[l] \mathbf{C}_k \left[\mathbf{C}_k^H \mathbf{\Phi}_{\mathbf{y}_k}^{-1}[l] \mathbf{C}_k \right]^{-1} \mathbf{i}. \tag{49}$$

All conjugate-transpose operations are replaced with simple transposes for the SSBT, as all signals and matrices are real-valued in this transform.

3.2. Conjugate-Frequency Filtering with the SSBT

It is useful to bring Property A7 to light. From there, we have that RFRs with the SSBT are even functions of the frequency for the same-frequency portion and odd for the conjugate frequency. That is, $A'_{m,k}[l] = A'_{m,(K-k)}[l]$ and $A''_{m,k}[l] = -A''_{m,(K-k)}[l]$. Using this, from the constraint in Equation (46b), we have

$$\begin{aligned} \mathbf{f}_{K-k}^H[l] \mathbf{D}_{K-k}'' &= \mathbf{f}_{K-k}^H[l] (-\mathbf{D}_k'') \\ &= 0. \end{aligned} \quad (50)$$

It is also easy to see that $\Phi_{y_k}[l] = \Phi_{y_{K-k}}[l]$ (for the SSBT), given all the properties from Appendix A. Therefore, we have that $\mathbf{f}_k[l]$ achieves the distortionless constraint for the bin $K - k$, given that $A'_{m,k}[l] = A'_{m,(K-k)}[l]$. It achieves the null of the conjugate-frequency portion, given the results from Equation (50), and also minimizes the power of the output signal, given that $\Phi_{y_k}[l] = \Phi_{y_{K-k}}[l]$. Consequently, it is unnecessary to calculate $\mathbf{f}_{K-k}[l]$, given that $\mathbf{f}_k[l]$ fulfills the minimization problem from Equation (48) for the conjugate bin $K - k$ as well. Although with the STFT, only half the spectrum is needed (given its complex-conjugate properties), with the SSBT, the filter only needs to be calculated for half the spectrum (even though it needs to be applied to the whole spectrum), putting them on a similar footing in this regard.

3.3. Theoretical Disadvantages

There are two apparent weaknesses with the SSBT: the first is a byproduct of working with a real-valued transform, where each frequency is influenced by its conjugate when dealing with convolution. This is inherent to any time–frequency transform that operates on a real-valued frequency domain, assuming a correct model that preserves the phase. The need to work with two frequencies simultaneously implies that, for each constraint in the problem, two constraints are needed in the mathematical model, even in an ideal scenario. This adds further load to the minimization problem, limiting how much noise it can minimize.

The second disadvantage is a direct consequence of the first. As explained in Section 2.4, the SSBT transform is less robust than the STFT in terms of RFR estimation errors, this being caused by correlation between conjugate frequencies on different frames (see Section 2.4), having more error terms when compared to the STFT. While it is impossible to bypass the first one since it is a modeling obstacle, the second can be worked around as it is an outcome of non-ideal considerations. For example, these effects are lessened by minimizing the impact different windows of the CTF model have on each other (or by using the MTF model).

4. Comparisons and Simulations

In the simulations (code available at <https://github.com/VCurtarelli/py-ssb-ctf-bf> (accessed on 21 August 2024)), we employed room impulse responses that were generated using the RIR generator [29] and signals that were selected from the SMARD database [30]. In all cases, we used $L_D = L_Y$ and $\Delta = 0$; with this, we disregarded any non-causal frames and considered as many frame samples as the RIR samples. The room's dimensions were $4 \times 6 \times 3$ m (width \times length \times height), with a reverberation time of 0.3 s. The device composed of the loudspeaker plus sensors was centered at (3, 4, 1) m, comprising $M = 8$ sensors arranged in a circular array with a radius of 8 cm. All sensors were omnidirectional with a flat frequency response, with the reference being the sensor at (3, 1.92, 1) m. The positions and signals used for the sources are shown in Table 1. The room's layout is described in Figure 1, where in green we have the desired source (assumed to be omnidirectional), and in red the device, with the 8 sensors and the loudspeaker in the center.

We set the input Signal-to-Noise Ratio (iSNR) for the white Gaussian source to $\text{gSNR} = 30$ dB, and initially, the input Signal-to-Echo Ratio (iSER) for the interfering loudspeaker source to $\text{iSER} = -15$ dB. The loudspeaker interference was labeled as echo,

given that it was a feedback path between the source and the sensors, which were all mounted on the same device.

Table 1. Source information for the simulations.

Source	Position	Signal
$x[n]$	(2, 5, 1.8) m	50_male_speech_english_ch8_OmniPower4296.flac
$s[n]$	(3, 4, 1) m	69_abba_ch8_OmniPower4296.flac
$r[n]$	~	wgn_48kHz_ch8_OmniPower4296.flac

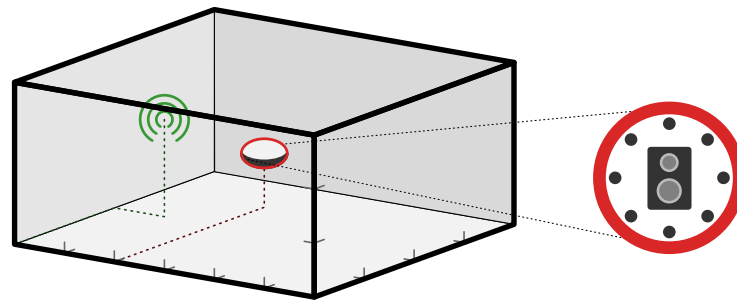


Figure 1. Room layout for simulations.

Hamming windows were used for the transform with an overlap of 50%, and all signals were resampled to the desired sampling frequency of 16kHz. Unless stated otherwise, $N = 32$ samples per window were used in the windowing processes. The regularization parameter was empirically set to $\alpha = 1 \times 10^{-4}$, just enough to control the gain in SNR. Although the developments allowed for a time-variant beamformer, we designed a single filter for the whole signal, favoring a faster processing time and ease of comparing the results.

We compare filters obtained via the two transforms for varying conditions of the signals and variables considered. In all plots, the STFT results are presented in red lines with squares, and the SSBT results in green lines with triangles. The results for an accurate *a priori* RFR are in lighter continuous lines, and those for the estimated RFR via Equations (10) and (18) are in darker dotted lines. For simplicity, the STFT for an accurate RFR is labeled STFT-A, and for an estimated RFR, it is STFT-E. The same notation applies to SSBT-A and SSBT-E.

4.1. Metrics of Interest

The compared filters aim to reduce the loudspeaker’s signal while preserving the desired signal. The minimal enhancement for white noise is also interesting, given the regularization parameter α added to the problem. These are measured by the desired signal reduction factor (DSRF, or ξ_x), gain in SER (gSER), and gain in SNR (gSNR), respectively. We also observe the directivity index (DI, or \mathcal{D}), which measures the beamformer’s behavior when employed in a spherically anisotropic noise field. Their time-dependent broadband formulations are given by

$$\xi_x[l] = \frac{\sum_k |X_{1,k}[l]|^2}{\sum_k |X_{f,k}[l]|^2} \tag{51a}$$

$$\text{gSER}[l] = \frac{\sum_k |S_{1,k}[l]|^2}{\sum_k |S_{f,k}[l]|^2} \cdot \frac{1}{\xi_x[l]} \tag{51b}$$

$$\text{gSNR}[l] = \frac{\sum_k |R_{1,k}[l]|^2}{\sum_k |R_{f,k}[l]|^2} \cdot \frac{1}{\xi_x[l]} \tag{51c}$$

$$\mathcal{D}[l] = \frac{\sum_k |\mathbf{f}_k^H[l] \mathbf{d}_k[l]|^2}{\sum_k \mathbf{f}_k^H[l] \bar{\Gamma}_k[l] \mathbf{f}_k[l]} \tag{51d}$$

in which $\bar{\Gamma}_k[l]$ is the spherical anisotropic noise field correlation matrix [31], and $\mathbf{d}_k[l]$ is the steering vector between the desired source and the sensor array, both assuming a far-end free field environment. We are also interested in a time-average broadband formulation for these metrics, these being given by

$$\xi_x = \frac{\sum_{l,k} |X_{1,k}[l]|^2}{\sum_{l,k} |X_{f,k}[l]|^2} \tag{52a}$$

$$gSER = \frac{\sum_{l,k} |S_{1,k}[l]|^2}{\sum_{l,k} |S_{f,k}[l]|^2} \cdot \frac{1}{\xi_x} \tag{52b}$$

$$gSNR = \frac{\sum_{l,k} |R_{1,k}[l]|^2}{\sum_{l,k} |R_{f,k}[l]|^2} \cdot \frac{1}{\xi_x} \tag{52c}$$

$$\mathcal{D} = \frac{\sum_{l,k} |\mathbf{f}_k^H[l] \mathbf{d}_k[l]|^2}{\sum_{l,k} \mathbf{f}_k^H[l] \bar{\Gamma}_k[l] \mathbf{f}_k[l]} \tag{52d}$$

We chose the gains in SER and SNR metrics rather than the more common ERLE and WNG [32], given an *a priori* knowledge of distortion on the desired signal. Therefore, the gains become more representative of the results.

4.2. Comparison for Different Observed Frames

In this simulation, we compare our filters for the accurate and estimated RFR for a range of considered observed frames L_Y , in which $L_Y = 1$ reduces to the MTF model. The results are in Figure 2. For all metrics except DI, the accurate results are consistently better than those achieved for an estimated RFR. Also, the SSBT-A results are, overall, worse than the accurate STFT-A ones, but not by too large a margin (around 3–4 dB for all metrics). This is also the case between the SSBT-E and STFT-E, but to a higher degree since the SSBT-E beamformer also led to a higher desired DSDI, which considerably adds to its performance loss of the former compared to the STFT-E for all metrics.

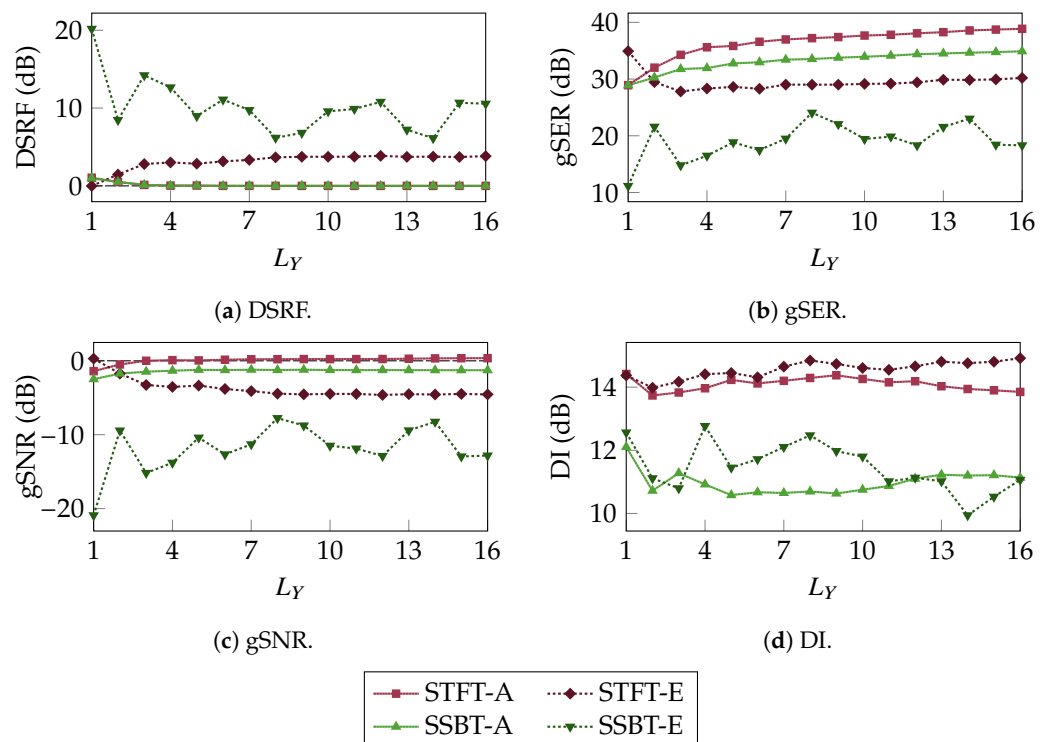


Figure 2. Output metrics for the beamformers over time for varying values of L_Y .

The best STFT-E results are obtained for $L_Y = 1$ in this scenario. With $L_Y = 2$, there is a notable increase in DSRF and a sharp decrease in gSER, with the latter increasing again for larger values of L_Y but never achieving the same performance. This is due to errors in the RFR estimation for frames other than $l = 0$, given that these windows carry less information, and these errors interfere more than adding new frames helps. Also, the estimate with $L_Y = 1$ already takes into account some information about the different windows' correlations with regard to the desired signal (see Equation (27)), explaining the STFT-E results being better than the STFT-A results in this case.

4.3. Comparison for Different Numbers of Samples per Frame

We now compare the beamformers in a scenario with the MTF model and change the number of samples per window (Figure 3). This circumvents the problems addressed in Section 2.4 for both transforms since we do not consider a convolutive filter. By increasing the number of samples per window, we minimize the frequency aliasing effects caused by the windowing process of the time–frequency transform. This allows us to capture more of the desired signal on each window.

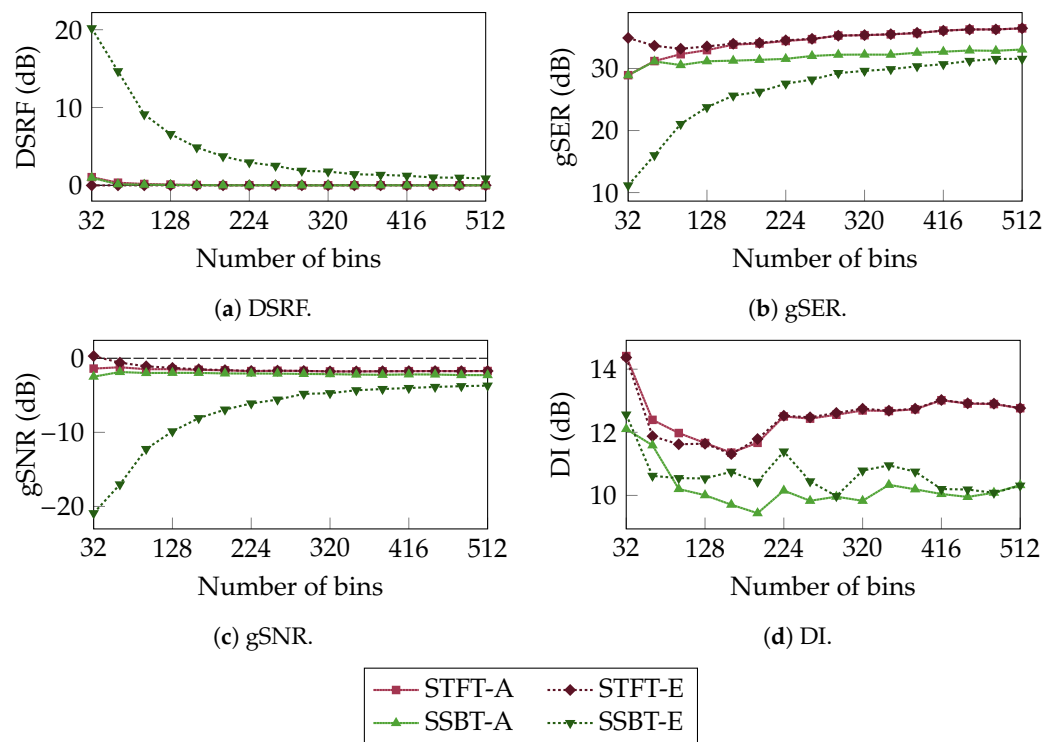


Figure 3. Output metrics for the beamformers over time, for varying N .

The first effect is that the desired signal's distortion for the SSBT-E decreases as more samples are considered. Also, for all other metrics, the SSBT-E performance approaches that of the SSBT-A as we increase N . The same happens for the STFT filters; however, the difference between the STFT-A and STFT-E is negligible with far fewer samples per frame. It is also observed that for a high N , the SSBT results are only slightly worse than the STFT ones for both the accurate and estimated RFR cases. This supports the previous theoretical claims, where we state that increasing the number of samples per frame reduces the RFR estimation errors.

4.4. Comparison for Different iSERs

We now compare the beamformers for different values of the input SER (iSER); that is, we change the signal's power in the source within the device, right next to the sensors. This simulation's results are presented in Figure 4. Overall, the accurate results for both

transforms are comparable, with the SSBT-A again underperforming marginally except for a high iSER and $L_Y = 1$. The same results as obtained previously are observed for the estimated RFR outputs. For $L_Y = 1$, the SSBT-E grossly underperforms compared to the STFT-E, and although their difference for $L_Y = 16$ is not as substantial, it is still relevant, except for a higher iSER.

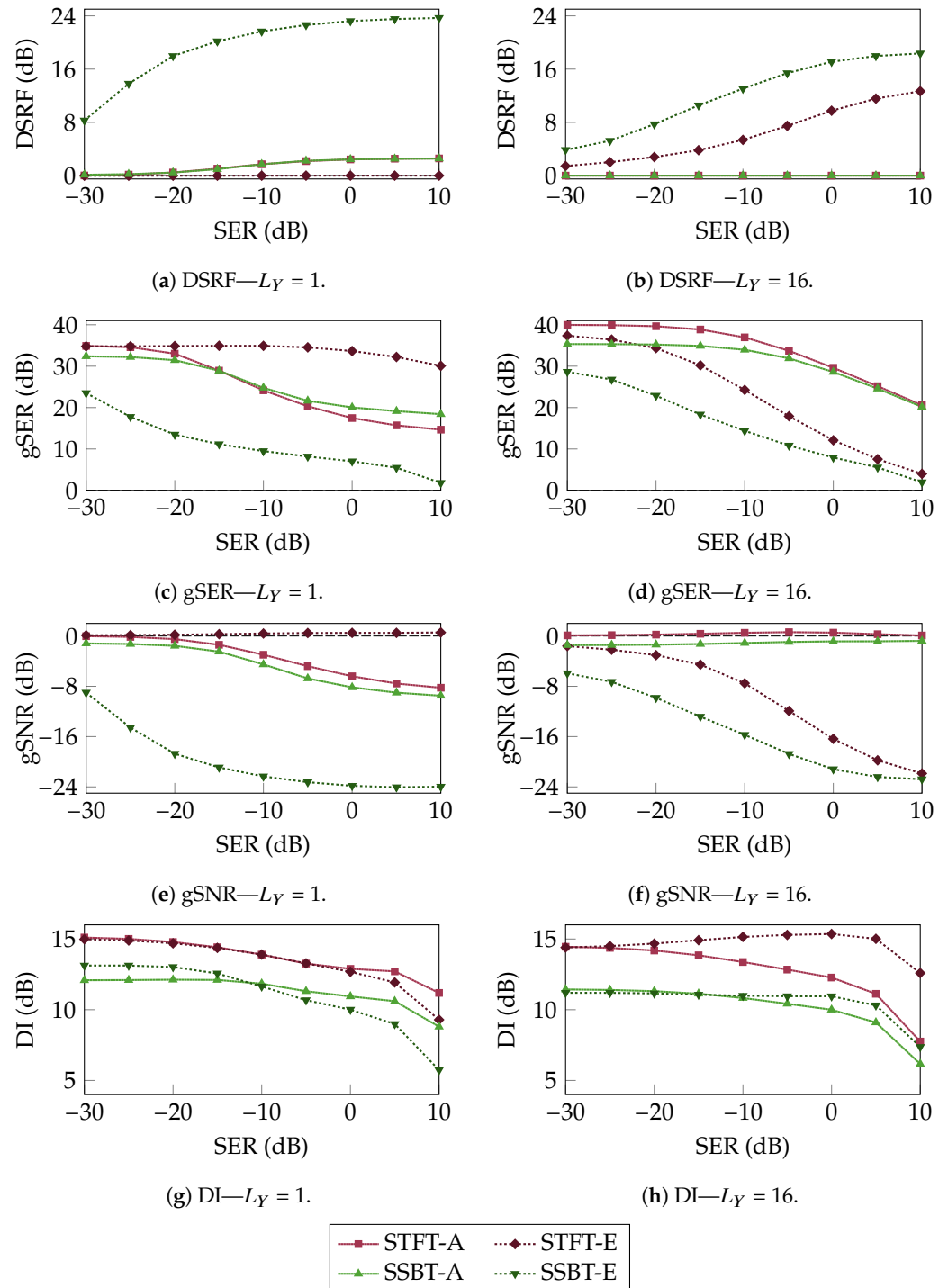


Figure 4. Output metrics for the beamformers for varying iSERs, with $N = 32$ samples, and two cases of L_Y .

We repeated these simulations, but instead of different values of L_Y , we changed the number of samples per frame to $N = 512$ and set $L_Y = 1$, with results in Figure 5. This was

chosen using the knowledge from Section 2.4 and Figure 3, where we both theoretically and practically have that a higher N minimizes the effects of estimation error. Notably, the results for the STFT-A and STFT-E are identical in this scenario, as in Figure 3 for a higher N . Meanwhile, the SSBT-A and SSBT-E results are similar only for very low iSERs (below -20 dB), and for higher input SERs, the estimated SSBT results deteriorate drastically. Although a high N value was used, which theoretically would lead to a better SSBT-E result (following Figure 3), we see that increasing the iSER led to a worse performance, similar to the results seen in Figure 4.

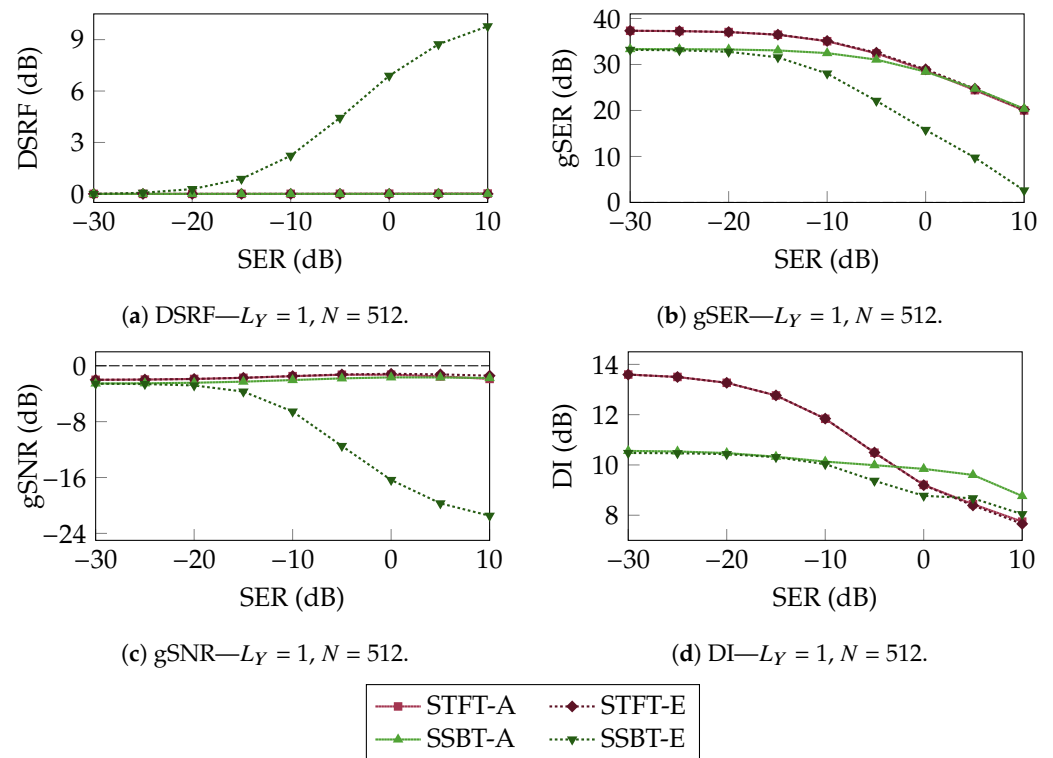


Figure 5. Output metrics for the beamformers for varying iSERs, with $N = 512$ samples.

4.5. General Simulation Results

For all simulations, the ideal SSBT-A results are similar (although mostly slightly worse) to the STFT-A results. Given that the SSBT beamformer is a strictly real-valued filter, this could be a worthwhile tradeoff of a slight performance loss for a faster beamforming algorithm, leading to a cheaper implementation. However, for a useful implementation of the SSBT in beamforming, these results should be followed in the estimated RFR case. However, in this scenario, the SSBT results were drastically worse when compared to those through the STFT, except for the specific case of a multiplicative filter, with a large number of samples per frame and a low iSER. This suggests that employing the SSBT with an inherently more robust beamformer with the MTF and high N could lead to viable applications.

5. Conclusions

We conducted a comprehensive investigation into the Single-Sideband Transform within the context of beamforming, examining its mathematical properties and interaction with key processes such as convolution, relative frequency response estimation, and the distortionless constraint. Our theoretical study reveals that despite its interesting real-valued representation, the SSBT exhibits higher susceptibility to errors in RFR estimation, requiring stricter constraints for its proper application. We found that in scenarios with longer time windows and under the multiplicative transfer function model, the estimation errors with the SSBT are reduced. Furthermore, we established that the SSBT and the Short-Time Fourier Transform are interchangeable in the time–frequency domain without

the need for their respective inverse transforms, enabling a seamless conversion between the two.

To validate our theoretical findings, we employed both transforms in the design of a convolutive Minimum-Power Distortionless Response beamformer within a reverberant environment across various scenarios. These practical results support our theoretical claims, showing that the SSBT-based filter slightly underperforms the STFT-based one in optimal conditions and significantly underperforms in non-ideal situations. While these findings highlight challenges in directly applying the SSBT for beamforming, the interchangeability of these transforms allows for filtering in the STFT domain—even when signals are initially in the SSBT domain. This allows for the combined use of the transformations, taking advantage of their respective strengths as applicable. Future research could explore integrating the SSBT into more robust beamformers, further comparing it to the STFT, and studying the cooperative integration of the two transforms in greater depth.

Author Contributions: Conceptualization, I.C. and V.P.C.; Methodology, V.P.C.; Formal analysis, V.P.C.; Software, V.P.C.; Writing—original draft: V.P.C.; Writing—review and editing, I.C. and V.P.C.; Supervision, V.P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Pazy Research Foundation and the Israel Science Foundation (grant no. 1449/23).

Data Availability Statement: The source-code for the simulations developed here is available at <https://github.com/VCurtarelli/py-ssb-ctf-bf> (accessed on 21 August 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CTF	Convolutive Transfer Function
DI	Directivity Index
DSRF	Desired Signal Reduction Factor
FT	Fourier Transform
MPDR	Minimum-Power Distortionless-Response
MTF	Multiplicative Transfer Function
RFR	Relative Frequency Response
RFT	Real Fourier Transform
SER	Signal-to-Echo Ratio
SNR	Signal-to-Noise Ratio
SSBT	Single-Sideband Transform
STFT	Short-Time Fourier Transform

Appendix A. Properties of the Real Fourier Transform

The Fourier Transform (FT) is defined as

$$X_{\mathcal{F}}(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (\text{A1})$$

with an inverse

$$x(t) = \int_{-\infty}^{\infty} X_{\mathcal{F}}(f)e^{j2\pi ft} dt. \quad (\text{A2})$$

The Real Fourier Transform (RFT) is defined as

$$X_{\mathcal{R}}(f) = \sqrt{2}\mathbb{R}\left\{\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft + j\frac{3\pi}{4}} dt\right\} \quad (\text{A3})$$

with an inverse

$$x(t) = \sqrt{2}\mathbb{R}\left\{\int_{-\infty}^{\infty} X_{\mathcal{R}}(f)e^{j2\pi ft - j\frac{3\pi}{4}} df\right\}. \tag{A4}$$

We assume that $x(t)$ is a real-valued time-domain signal unless stated otherwise.

Property A1. *The FT and RFT are bijective transformations of one another.*

Proof. By manipulating Equation (A3), we can obtain

$$X_{\mathcal{R}}(f) = \sqrt{2}\mathbb{R}\left\{\left(\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt\right)e^{j\frac{3\pi}{4}}\right\} \tag{A5}$$

where the term in parenthesis is trivially the FT of $x(t)$ from Equation (A1). Thus, we obtain that

$$\begin{aligned} X_{\mathcal{R}}(f) &= \mathbb{R}\left\{\left(X_{\mathcal{F}}^{\mathcal{R}}(f) + jX_{\mathcal{F}}^{\mathcal{I}}(f)\right) \cdot (-1 + j)\right\} \\ &= -X_{\mathcal{F}}^{\mathcal{R}}(f) - X_{\mathcal{F}}^{\mathcal{I}}(f). \end{aligned} \tag{A6}$$

Likewise, considering that the FT of a real signal is complex-conjugate, such that $X_{\mathcal{F}}(f) = X_{\mathcal{F}}^*(-f)$, it is easy to see that

$$X_{\mathcal{R}}(-f) = -X_{\mathcal{F}}^{\mathcal{R}}(f) + X_{\mathcal{F}}^{\mathcal{I}}(f). \tag{A7}$$

With this, we have that

$$\begin{aligned} \sqrt{2}X_{\mathcal{R}}(f)e^{-j\frac{3\pi}{4}} &= X_{\mathcal{F}}^{\mathcal{R}}(f) + X_{\mathcal{F}}^{\mathcal{I}}(f) + jX_{\mathcal{F}}^{\mathcal{R}}(f) + jX_{\mathcal{F}}^{\mathcal{I}}(f) \\ \sqrt{2}X_{\mathcal{R}}(-f)e^{j\frac{3\pi}{4}} &= X_{\mathcal{F}}^{\mathcal{R}}(f) - X_{\mathcal{F}}^{\mathcal{I}}(f) - jX_{\mathcal{F}}^{\mathcal{R}}(f) + jX_{\mathcal{F}}^{\mathcal{I}}(f) \end{aligned} \tag{A8}$$

and, therefore, we have

$$\begin{aligned} \frac{X_{\mathcal{R}}(f)e^{-j\frac{3\pi}{4}} + X_{\mathcal{R}}(-f)e^{j\frac{3\pi}{4}}}{\sqrt{2}} &= X_{\mathcal{F}}^{\mathcal{R}}(f) + jX_{\mathcal{F}}^{\mathcal{I}}(f) \\ &= X_{\mathcal{F}}(f). \end{aligned} \tag{A9}$$

Similarly, using Equation (A5), we have that

$$X_{\mathcal{R}}(f) = \sqrt{2}\mathbb{R}\left\{X_{\mathcal{F}}(f)e^{j\frac{3\pi}{4}}\right\}. \tag{A10}$$

Using the property of complex numbers that $\mathbb{R}\{a\} = \frac{a+a^*}{2}$, then

$$X_{\mathcal{R}}(f) = \frac{X_{\mathcal{F}}(f)e^{j\frac{3\pi}{4}} + X_{\mathcal{F}}(-f)e^{-j\frac{3\pi}{4}}}{\sqrt{2}}. \tag{A11}$$

Since for each $X_{\mathcal{F}}(f)$ there exists one and only one $X_{\mathcal{R}}(f)$, and vice-versa, they form a bijective relationship. \square

Property A2. *The IRFT is the inverse of the RFT.*

Proof. From Property A1, we have that

$$\mathcal{R}^{-1}\{X_{\mathcal{R}}(f)\}(t) = \frac{X_{\mathcal{F}}(f)e^{j\frac{3\pi}{4}} + X_{\mathcal{F}}(-f)e^{-j\frac{3\pi}{4}}}{\sqrt{2}}. \tag{A12}$$

Substituting this in Equation (A4), we have

$$\begin{aligned} \mathcal{R}^{-1}\{X_{\mathcal{R}}(f)\}(t) &= \sqrt{2}\mathbb{R}\left\{\int_{-\infty}^{\infty}\left(\frac{X_{\mathcal{F}}(f)e^{j\frac{3\pi}{4}}+X_{\mathcal{F}}(-f)e^{-j\frac{3\pi}{4}}}{\sqrt{2}}\right)e^{j2\pi ft-j\frac{3\pi}{4}}df\right\} \\ &= \mathbb{R}\left\{\int_{-\infty}^{\infty}\left(X_{\mathcal{F}}(f)e^{j2\pi ft}+X_{\mathcal{F}}(-f)e^{j2\pi ft-j\frac{3\pi}{2}}\right)df\right\}. \end{aligned} \tag{A13}$$

The first term expands to the inverse Fourier transform of $X_{\mathcal{F}}(f)$, which is trivially $x(t)$; the second term is the inverse Fourier transform of $X_{\mathcal{F}}(-f)$, which is from the time reversal property and is $x(-t)$. Therefore,

$$\begin{aligned} \mathcal{R}^{-1}\{X_{\mathcal{R}}(f)\}(t) &= \mathbb{R}\left\{x(t)+x(-t)e^{-j\frac{3\pi}{2}}\right\} \\ &= \mathbb{R}\{x(t)+jx(-t)\} \\ &= x(t). \end{aligned} \tag{A14}$$

Note that the invertibility of the RFT is a direct consequence of $x(t)$ being a real signal. Assuming the contrary, then, obviously, $\mathcal{R}^{-1}\{\mathcal{R}\{x(t)\}(f)\}(t) \neq x(t)$, given that the inverse RFT involves taking the real part of the inverse, and, therefore, it maps the original signal to a different one. \square

Property A3. *The FT convolution theorem does not apply for the RFT.*

Proof. Let $h(t)$ be the impulse response of an LTI system, with input $x(t)$. It is trivial that the system’s output, $y(t)$, is given by

$$y(t) = h(t) * x(t) \tag{A15}$$

with $*$ being the convolution operator. For the Fourier transform, through the convolution theorem, it is trivial that

$$Y_{\mathcal{F}}(f) = H_{\mathcal{F}}(f)X_{\mathcal{F}}(f). \tag{A16}$$

Expanding these in terms of real and imaginary parts (omitting the frequency index for clarity),

$$Y_{\mathcal{F}} = H_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{R}} + jH_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{I}} + jH_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{R}} - H_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{I}}. \tag{A17}$$

Now in the RFT domain, with Equation (A6), we have that

$$\begin{aligned} X_{\mathcal{R}}(f) &= -X_{\mathcal{F}}^{\mathbb{R}}(f) - X_{\mathcal{F}}^{\mathbb{I}}(f) \\ H_{\mathcal{R}}(f) &= -H_{\mathcal{F}}^{\mathbb{R}}(f) - H_{\mathcal{F}}^{\mathbb{I}}(f). \end{aligned} \tag{A18}$$

Assuming the convolution theorem is true for the RFT,

$$Y_{\mathcal{R}} = H_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{R}} + H_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{I}} + H_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{R}} + H_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{I}}. \tag{A19}$$

Now, by applying Equation (A6) on Equation (A17), we have

$$\tilde{Y}_{\mathcal{R}} = -H_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{R}} - H_{\mathcal{F}}^{\mathbb{R}}X_{\mathcal{F}}^{\mathbb{I}} - H_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{R}} + H_{\mathcal{F}}^{\mathbb{I}}X_{\mathcal{F}}^{\mathbb{I}} \tag{A20}$$

where it is explicit that $Y_{\mathcal{R}}(f) \neq \tilde{Y}_{\mathcal{R}}(f)$. Therefore, the RFT of the convolution (Equation (A20)) is not the product of the RFTs of the signals (Equation (A19)), and thus, the convolution theorem does not hold for the RFT. \square

Property A4. *There is an equivalent of the convolution theorem for the RFT.*

Proof. From Equation (A20), we have our objective for the “convolution theorem”-equivalent for the RFT. From both Equations (A6) and (A7), we have

$$\begin{aligned} X_{\mathcal{R}}(f) &= -X_{\mathcal{F}}^{\mathcal{R}}(f) - X_{\mathcal{F}}^{\mathcal{I}}(f) \\ X_{\mathcal{R}}(-f) &= -X_{\mathcal{F}}^{\mathcal{R}}(f) + X_{\mathcal{F}}^{\mathcal{I}}(f) \\ H_{\mathcal{R}}(f) &= -H_{\mathcal{F}}^{\mathcal{R}}(f) - H_{\mathcal{F}}^{\mathcal{I}}(f) \\ H_{\mathcal{R}}(-f) &= -H_{\mathcal{F}}^{\mathcal{R}}(f) + H_{\mathcal{F}}^{\mathcal{I}}(f). \end{aligned} \tag{A21}$$

We omit the frequency dependency in the FT values. Taking the possible combinations, we have

$$X_{\mathcal{R}}(f)H_{\mathcal{R}}(f) = H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{R}} + H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{I}} + H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{R}} + H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{I}} \tag{A22a}$$

$$X_{\mathcal{R}}(f)H_{\mathcal{R}}(-f) = H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{R}} + H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{I}} - H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{R}} - H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{I}} \tag{A22b}$$

$$X_{\mathcal{R}}(-f)H_{\mathcal{R}}(f) = H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{R}} - H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{I}} + H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{R}} - H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{I}} \tag{A22c}$$

$$X_{\mathcal{R}}(-f)H_{\mathcal{R}}(-f) = H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{R}} - H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{I}} - H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{R}} + H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{I}}. \tag{A22d}$$

Taking the difference between Equation (A22a) and Equation (A22d), and the sum of Equation (A22b) and Equation (A22c), we have

$$X_{\mathcal{R}}(f)H_{\mathcal{R}}(f) - X_{\mathcal{R}}(-f)H_{\mathcal{R}}(-f) = 2\left(H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{I}} + H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{R}}\right) \tag{A23a}$$

$$X_{\mathcal{R}}(f)H_{\mathcal{R}}(-f) + X_{\mathcal{R}}(-f)H_{\mathcal{R}}(f) = 2\left(H_{\mathcal{F}}^{\mathcal{R}}X_{\mathcal{F}}^{\mathcal{R}} - H_{\mathcal{F}}^{\mathcal{I}}X_{\mathcal{F}}^{\mathcal{I}}\right) \tag{A23b}$$

and, therefore, to achieve Equation (A20), we let

$$\begin{aligned} Y_{\mathcal{R}}(f) &= \frac{-X_{\mathcal{R}}(f)H_{\mathcal{R}}(f) + X_{\mathcal{R}}(-f)H_{\mathcal{R}}(-f) - X_{\mathcal{R}}(f)H_{\mathcal{R}}(-f) - X_{\mathcal{R}}(-f)H_{\mathcal{R}}(f)}{2} \\ &= X_{\mathcal{R}}(f)\frac{-H_{\mathcal{R}}(f) - H_{\mathcal{R}}(-f)}{2} + X_{\mathcal{R}}(-f)\frac{-H_{\mathcal{R}}(f) + H_{\mathcal{R}}(-f)}{2}. \end{aligned} \tag{A24}$$

Finally, from Equation (A21), we achieve

$$Y_{\mathcal{R}}(f) = X_{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{R}}(f) + X_{\mathcal{R}}(-f)H_{\mathcal{F}}^{\mathcal{I}}(f) \tag{A25}$$

and, for its conjugate frequency (that is, for $-f$), we have

$$\begin{aligned} Y_{\mathcal{R}}(-f) &= X_{\mathcal{R}}(-f)H_{\mathcal{F}}^{\mathcal{R}}(-f) + X_{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{I}}(-f) \\ &= X_{\mathcal{R}}(-f)H_{\mathcal{F}}^{\mathcal{R}}(f) - X_{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{I}}(f). \end{aligned} \tag{A26}$$

□

Property A5. *Frequencies in the RFT have the same variance as their FT counterpart.*

Proof. We now assume that $X_{\mathcal{F}}(f)$ is the transform of a random process, such that its real and imaginary parts are independent and identically distributed with zero means. Taking the complex correlation of a given frequency in the FT domain,

$$\mathbb{E}\{X_{\mathcal{F}}(f)X_{\mathcal{F}}^*(f)\} = \mathbb{E}\{X_{\mathcal{F}}^{\mathcal{R}}(f)^2 + X_{\mathcal{F}}^{\mathcal{I}}(f)^2\}. \tag{A27}$$

Since they are identically distributed, we denote

$$\mathbb{E}\{X_{\mathcal{F}}^{\mathcal{R}}(f)^2\} = \mathbb{E}\{X_{\mathcal{F}}^{\mathcal{I}}(f)^2\} = \sigma_f^2 \tag{A28}$$

and, therefore, we have

$$\mathbb{E}\{X_{\mathcal{F}}(f)X_{\mathcal{F}}^*(f)\} = 2\sigma_f^2. \tag{A29}$$

Now, in the RFT domain, we take the correlation of a given frequency Equation (A6),

$$\mathbb{E}\{X_{\mathcal{R}}(f)^2\} = \mathbb{E}\{X_{\mathcal{F}}^{\mathcal{R}}(f)^2 + 2X_{\mathcal{F}}^{\mathcal{R}}(f)X_{\mathcal{F}}^{\mathcal{I}}(f) + X_{\mathcal{F}}^{\mathcal{I}}(f)^2\}. \quad (\text{A30})$$

Using the fact that $X_{\mathcal{F}}^{\mathcal{R}}(f)$ and $X_{\mathcal{F}}^{\mathcal{I}}(f)$ are i.i.d. and zero-mean, the cross terms are zero, and thus,

$$\begin{aligned} \mathbb{E}\{X_{\mathcal{R}}(f)^2\} &= 2\sigma_f^2 \\ &= \mathbb{E}\{X_{\mathcal{F}}(f)X_{\mathcal{F}}^*(f)\}. \end{aligned} \quad (\text{A31})$$

It is trivial to see that the same applies for $X_{\mathcal{R}}(-f)$. \square

Property A6. *Conjugate frequencies in the RFT domain are independent.*

Proof. We take the same assumptions as those in Property A5. Taking the complex correlation between the two conjugate frequencies,

$$\mathbb{E}\{X_{\mathcal{F}}(f)X_{\mathcal{F}}^*(-f)\} = \mathbb{E}\{X_{\mathcal{F}}^{\mathcal{R}}(f)^2 + 2jX_{\mathcal{F}}^{\mathcal{R}}(f)X_{\mathcal{F}}^{\mathcal{I}}(f) - X_{\mathcal{F}}^{\mathcal{I}}(f)^2\}. \quad (\text{A32})$$

Using the fact that the $X_{\mathcal{F}}^{\mathcal{R}}(f)$ and $X_{\mathcal{F}}^{\mathcal{I}}(f)$ are independent and zero-mean, the cross terms are zero, and with Equation (A28), then

$$\mathbb{E}\{X_{\mathcal{F}}(f)X_{\mathcal{F}}^*(-f)\} = 0. \quad (\text{A33})$$

This result is known, but it is useful to show it since this same procedure is used for the RFT.

We now consider Equations (A6) and (A7). Taking the correlation between the two conjugate-frequencies yields

$$\mathbb{E}\{X_{\mathcal{R}}(f)X_{\mathcal{R}}(-f)\} = \mathbb{E}\{X_{\mathcal{F}}^{\mathcal{R}}(f)^2 - X_{\mathcal{F}}^{\mathcal{I}}(f)^2\}. \quad (\text{A34})$$

Under the same assumptions that the real and imaginary parts of $X_{\mathcal{F}}(f)$ are identically distributed, we obtain the same result as before, where

$$\mathbb{E}\{X_{\mathcal{R}}(f)X_{\mathcal{R}}(-f)\} = 0. \quad (\text{A35})$$

Note that, with the RFT, we did not use the complex correlation since it is real-valued. Lastly, we take the correlation between conjugate frequencies of the output of a system according to Equations (A25) and (A26):

$$\begin{aligned} \mathbb{E}\{Y_{\mathcal{R}}(f)Y_{\mathcal{R}}(-f)\} &= H_{\mathcal{F}}^{\mathcal{R}}(f)^2\mathbb{E}\{X_{\mathcal{R}}(f)X_{\mathcal{R}}(-f)\} - H_{\mathcal{F}}^{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{I}}(f)\mathbb{E}\{X_{\mathcal{R}}(f)^2\} \\ &\quad + H_{\mathcal{F}}^{\mathcal{R}}(f)H_{\mathcal{F}}^{\mathcal{I}}(f)\mathbb{E}\{X_{\mathcal{R}}(-f)^2\} + H_{\mathcal{F}}^{\mathcal{I}}(f)^2\mathbb{E}\{X_{\mathcal{R}}(f)X_{\mathcal{R}}(-f)\}. \end{aligned} \quad (\text{A36})$$

Since $X_{\mathcal{R}}(f)$ and $X_{\mathcal{R}}(-f)$ are independent and zero-mean, the first and last terms are zero, and the other two also cancel each other out since $\mathbb{E}\{X_{\mathcal{R}}(f)^2\} = \mathbb{E}\{X_{\mathcal{R}}(-f)^2\} = 2\sigma_f^2$. Therefore,

$$\mathbb{E}\{Y_{\mathcal{R}}(f)Y_{\mathcal{R}}(-f)\} = 0. \quad (\text{A37})$$

\square

Property A7. *Relative frequency responses with the RFT are even for frequency-to-frequency and odd for conjugate frequency.*

Proof. First, given two real systems with a shared input $x(t)$, each with an impulse response $h_m(t)$ and an output $y_m(t)$, such that

$$\begin{aligned} Y_1(f) &= X(f)H_{1,\mathcal{F}}^{\mathcal{R}}(f) + X(-f)H_{1,\mathcal{F}}^{\mathcal{I}}(f) \\ Y_2(f) &= X(f)H_{2,\mathcal{F}}^{\mathcal{R}}(f) + X(-f)H_{2,\mathcal{F}}^{\mathcal{I}}(f) \end{aligned} \quad (\text{A38})$$

and, for their conjugate frequencies,

$$\begin{aligned} Y_1(-f) &= X(-f)H_{1,\mathcal{F}}^{\text{R}}(-f) + X(f)H_{1,\mathcal{F}}^{\text{I}}(-f) \\ Y_2(-f) &= X(-f)H_{2,\mathcal{F}}^{\text{R}}(-f) + X(f)H_{2,\mathcal{F}}^{\text{I}}(-f) \end{aligned} \quad (\text{A39})$$

where we omit the transform index when it is an RFT's domain value for compactness. If we define $X'_1(f)$ and $X''_1(f)$ as

$$\begin{aligned} X'_1(f) &= Y_1(f) \\ X''_1(f) &= Y_1(-f) \end{aligned} \quad (\text{A40})$$

then

$$\begin{aligned} Y_m(f) &= A'_m(f)X'_1(f) + A''_m(f)X''_1(f) \\ Y_m(-f) &= A'_m(-f)X'_1(-f) + A''_m(-f)X''_1(-f) \end{aligned} \quad (\text{A41})$$

where $A'_m(f)$ and $A''_m(f)$ are given by

$$\begin{aligned} A'_m(f) &= \frac{H'_1(f)H'_m(f) + H''_1(f)H''_m(f)}{H_1'^2(f) + H_1''^2(f)} \\ A''_m(f) &= \frac{H'_1(f)H''_m(f) - H''_1(f)H'_m(f)}{H_1'^2(f) + H_1''^2(f)} \end{aligned} \quad (\text{A42})$$

and, for their conjugate frequencies,

$$\begin{aligned} A'_m(-f) &= \frac{H'_1(-f)H'_m(-f) + H''_1(-f)H''_m(-f)}{H_1'^2(-f) + H_1''^2(-f)} \\ A''_m(-f) &= \frac{H'_1(-f)H''_m(-f) - H''_1(-f)H'_m(-f)}{H_1'^2(-f) + H_1''^2(-f)}. \end{aligned} \quad (\text{A43})$$

Assuming that $h(t)$ is real-valued, then $H_{m,\mathcal{F}}(f) = H_{m,\mathcal{F}}^*(-f)$, and, therefore, $H'_m(f) = H'_m(-f)$, and $H''_m(f) = -H''_m(-f)$. With this,

$$\begin{aligned} A'_m(-f) &= \frac{H'_1(f)H'_m(f) + [-H''_1(f)][-H''_m(f)]}{H_1'^2(f) + [-H''_1(f)]^2(f)} \\ &= \frac{H'_1(f)H'_m(f) + H''_1(f)H''_m(f)}{H_1'^2(f) + H_1''^2(f)} \\ &= A'_m(f) \end{aligned} \quad (\text{A44})$$

$$\begin{aligned} A''_m(-f) &= \frac{H'_1(f)[-H''_m(f)] - [-H''_1(f)]H'_m(f)}{H_1'^2(f) + [-H''_1(f)]^2(f)} \\ &= \frac{-H'_1(f)H''_m(f) + H''_1(f)H'_m(f)}{H_1'^2(f) + H_1''^2(f)} \\ &= -A''_m(f). \end{aligned}$$

Therefore, our frequency-to-frequency RFR $A'_m(f)$ is even, since $A'_m(f) = A'_m(-f)$, and our conjugate-frequency RFR is odd, since $A''_m(f) = -A''_m(-f)$.

□

References

1. Chen, J.; Yao, K.; Hudson, R. Source localization and beamforming. *IEEE Signal Process. Mag.* **2002**, *19*, 30–39. [[CrossRef](#)]
2. Lobato, W.; Costa, M.H. Worst-Case-Optimization Robust-MVDR Beamformer for Stereo Noise Reduction in Hearing Aids. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2224–2237. [[CrossRef](#)]
3. Lu, J.Y.; Zou, H.; Greenleaf, J.F. Biomedical ultrasound beam forming. *Ultrasound Med. Biol.* **1994**, *20*, 403–428. [[CrossRef](#)]

4. Nguyen, N.Q.; Prager, R.W. Minimum Variance Approaches to Ultrasound Pixel-Based Beamforming. *IEEE Trans. Med. Imaging* **2017**, *36*, 374–384. [CrossRef]
5. Han, Y.; Luo, M.; Zhao, X.; Guerrero, J.M.; Xu, L. Comparative Performance Evaluation of Orthogonal-Signal-Generators-Based Single-Phase PLL Algorithms—A Survey. *IEEE Trans. Power Electron.* **2016**, *31*, 3932–3944. [CrossRef]
6. Häggglund, T. Signal Filtering in PID Control. *IFAC Proc. Vol.* **2012**, *45*, 1–10. [CrossRef]
7. Hathcock, D.; Sheehy, J.; Weisenberger, C.; Ilker, E.; Hinczewski, M. Noise Filtering and Prediction in Biological Signaling Networks. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2016**, *2*, 16–30. [CrossRef]
8. Lee, J.; Shin, S.Y. General construction of time-domain filters for orientation data. *IEEE Trans. Vis. Comput. Graph.* **2002**, *8*, 119–128. [CrossRef]
9. Shi, Z.; Butt, G. New enhancement filters for geological mapping. *ASEG Ext. Abstr.* **2004**, *2004*, 1–4. [CrossRef]
10. Benesty, J.; Cohen, I.; Chen, J. *Fundamentals of Signal Enhancement and Array Signal Processing*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
11. Kymık, M.; Güler, İ.; Dizibüyük, A.; Akin, M. Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. *Comput. Biol. Med.* **2005**, *35*, 603–616. [CrossRef]
12. Pan, C.; Chen, J.; Shi, G.; Benesty, J. On microphone array beamforming and insights into the underlying signal models in the short-time-Fourier-transform domain. *J. Acoust. Soc. Am.* **2021**, *149*, 660–672. [CrossRef] [PubMed]
13. Chen, W.; Huang, X. Wavelet-Based Beamforming for High-Speed Rotating Acoustic Source. *IEEE Access* **2018**, *6*, 10231–10239. [CrossRef]
14. Yang, Y.; Peng, Z.K.; Dong, X.J.; Zhang, W.M.; Meng, G. General Parameterized Time-Frequency Transform. *IEEE Trans. Signal Process.* **2014**, *62*, 2751–2764. [CrossRef]
15. Almeida, L. The fractional Fourier transform and time-frequency representations. *IEEE Trans. Signal Process.* **1994**, *42*, 3084–3091. [CrossRef]
16. Crochiere, R.E.; Rabiner, L.R. *Multirate Digital Signal Processing*; Prentice-Hall Signal Processing Series; Prentice-Hall: Englewood Cliffs, NJ, USA, 1983.
17. Wackersreuther, G. Some new aspects of filters for filter banks. *IEEE Trans. Acoust. Speech Signal Process.* **1986**, *34*, 1182–1200. [CrossRef]
18. Harteneck, M.; Weiss, S.; Stewart, R. Design of near perfect reconstruction oversampled filter banks for subband adaptive filters. *IEEE Trans. Circuits Syst. II Analog. Digit. Signal Process.* **1999**, *46*, 1081–1085. [CrossRef]
19. Chin, W.; Farhang-Boroujeny, B. Subband adaptive filtering with real-valued subband signals for acoustic echo cancellation. *IEE Proc. Vis. Image Signal Process.* **2001**, *148*, 283. [CrossRef]
20. Ozyerman, A.; Cohen, I. System identification and dereverberation of speech signals in the Single-Side-Band transform domain. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, 27–31 August 2012; pp. 360–364.
21. Okamoto, T.; Tachibana, K.; Toda, T.; Shiga, Y.; Kawai, H. Subband wavenet with overlapped single-sideband filterbanks. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Okinawa, Japan, 16–20 December 2017; pp. 698–704. [CrossRef]
22. Zhao, Y.; Wang, Z.Q.; Wang, D. Two-Stage Deep Learning for Noisy-Reverberant Speech Enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 53–62. [CrossRef] [PubMed]
23. Li, X.; Girin, L.; Gannot, S.; Horaud, R. Multichannel Speech Separation and Enhancement Using the Convolutional Transfer Function. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 645–659. [CrossRef]
24. Britanak, V.; Rao, K.R. *Cosine-/Sine-Modulated Filter Banks: General Properties, Fast Algorithms and Integer Approximations*; Springer International Publishing: Cham, Switzerland, 2018. [CrossRef]
25. Talmon, R.; Cohen, I.; Gannot, S. Relative Transfer Function Identification Using Convolutional Transfer Function Approximation. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *17*, 546–555. [CrossRef]
26. Jayakumar, E.; Sathidevi, P. An integrated acoustic echo and noise cancellation system using cross-band adaptive filters and wavelet thresholding of multitaper spectrum. *Appl. Acoust.* **2018**, *141*, 9–18. [CrossRef]
27. Lee, C.M.; Shin, J.W.; Jin, Y.G.; Kim, J.H.; Kim, N.S. Crossband filtering for stereophonic acoustic echo suppression. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 1345–1349. [CrossRef]
28. Li, D.; Yin, Q.; Mu, P.; Guo, W. Robust MVDR beamforming using the DOA matrix decomposition. In Proceedings of the 2011 1st International Symposium on Access Spaces (ISAS), Yokohama, Japan, 17–19 June 2011; pp. 105–110. [CrossRef]
29. Habets, E. RIR Generator. Available online: <https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator> (accessed on 11 March 2024).
30. Nielsen, J.K.; Jensen, J.R.; Jensen, S.H.; Christensen, M.G. The Single- and Multichannel Audio Recordings Database (SMARD). In Proceedings of the International Workshop Acoustic Signal Enhancement, Juan les Pins, France, 8–11 September 2014.

31. Habets, E.A.P.; Gannot, S. Generating sensor signals in isotropic noise fields. *J. Acoust. Soc. Am.* **2007**, *122*, 3464–3470. [[CrossRef](#)] [[PubMed](#)]
32. Wada, T.S.; Juang, B.-H. Enhancement of Residual Echo for Robust Acoustic Echo Cancellation. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 175–189. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.