MDPI

Article

[DRAFT] On the Single-Sideband Transform for MVDR Beamformers

Vitor Curtarelli[1,*], Israel Cohen[1]

[1] 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

Abstract: This paper investigates the application of the Single-Sideband Transform (SSBT) for constructing a Minimum-Variance Distortionless-Response (MVDR) beamformer in the context of the convolutive transfer function (CTF) model for short-time transforms. Our study aims to optimize the utilization of SSBT by examining its characteristics and traits. We address a reverberant scenario with multiple sources of contamination, aiming to minimize both undesired sources and reverberation in the output. Through simulations reflecting real-life scenarios, we demonstrate that employing the SSBT – both in a naive and a refined approach – results in superior signal enhancement. The refined approach not only enhances the signal but also ensures the desired distortionless behavior when compared to using only the Short-Time Fourier Transform (STFT).

Keywords: Single-sideband transform; MVDR beamformer; Filter banks; Array signal processing; Signal enhancement.

1. Introduction

Beamformers are an important tool for signal enhancement, having a plethora of applications from hearing aids [1] to source localization [2] to imaging [3,4]. Traditionally, beamforming techniques are used either strictly on the time domain, strictly on the frequency domain, or on the time-frequency domain [5], the later allowing the exploitation of frequency-related information while also dynamically adapting to signal changes over time. While the Short-Time Fourier Transform (STFT) is widely used for time-frequency analysis [6,7], alternative transforms [8–10] can also be employed, offering unique perspectives on signal details.

One such transform is the Single-Sideband Transform (SSBT) [11,12], standing out for its real-valued frequency spectrum. It has been shown that the SSBT works particularly well with short analysis windows [11], lending itself to be useful when working with the convolutive transfer function (CTF) model [13] and filter-banks [**?**] for signal analysis.

Two of the most important goals in beamforming are output noise minimization and the distortionless-ness of the desired signal, both being achieved by the Minimum-Variance Distortionless-Response (MVDR) beamformer [14,15]. As the MVDR beamformer works on the frequency (or time-frequency) domain without any specification on the transform

Citation: Curtarelli, V.; Cohen, I. On the Single-Sideband Transform for MVDR Filter Banks. *Algorithms* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

chosen, it is of interest to explore and compare the performance of this filter, when designing it through different time-frequency transforms.

Motivated by this, our paper explores the SSB transform and its application on the subject of beamforming within the context of the CTF model. We propose an approach for the CTF that allows the separation of desired and undesired speech components for reverberant environments, and employ this approach for designing the MVDR beamformer. We also explore the limitations and traits of the SSBT, and how to properly adapt the MVDR beamformer to this new transform's constraints.

We organized the paper as follows: in Section 2 we introduce the proposed time-frequency transforms, how they're related and which qualities from each are important; Section 3 presents the signal model considered in the time domain, and how it is transferred into the time-frequency domain within the considered framework; and in Section 4 we develop a true-MVDR beamformer with the SSBT, taking into account its features. In Section 5 we present and discuss the results, comparing the studied methods and beamforming techniques obtained; and finally in Section 6 we conclude this paper.

2. STFT and Single-Sideband Transform

When studying signals and systems, often frequency and time-frequency transforms are used in order to change the signal domain [16], allowing the exploitation of different patterns and informations that are inherent to the signal.

Given a time-domain signal $x[n]$, its Short-time Fourier Transform (STFT) [6,7] is given by

$$X_{\mathcal{F}}[l,k] = \sum_{n=0}^{K-1} w[n]x[n-l\cdot O]e^{-\mathrm{j}2\pi k\frac{(n-l\cdot O)}{K}} \tag{1}$$

where $w[n]$ is an analysis window of length K; and O is the overlap between windows of the transform, usually $O = \lfloor K/2 \rfloor$. Even though the STFT is the most traditionally used time-frequency transform, it isn't the only one available. Therefore, exploring different possibilities for such an operation can lead to interesting results.

The Single-Sideband Transform (SSBT) [11] is one such alternative, in which the frequency values are cleverly used such that its spectrum is real-valued. The SSB transform of $x[n]$ is defined as

$$X_{\mathcal{S}}[l,k] = \sqrt{2}\Re\left\{\sum_{n=0}^{L-1} w[n]x[n-l\cdot O]e^{\mathrm{j}2\pi k\frac{(n-l\cdot O)}{K}+\mathrm{j}\frac{3\pi}{4}}\right\} \tag{2}$$

Assuming that $x[n]$ is real-valued, one advantage of using the STFT is that we only need to work with $\left\lfloor \frac{K+1}{2} \right\rfloor + 1$ frequency bins, given its complex-conjugate behavior. Meanwhile, the SSBT needs to use all K possible bins to correctly capture all information of $x[n]$, however it is real-valued.

It is possible to define the SSBT using the STFT (assuming all K bins are available), such that

$$\begin{aligned} X_{\mathcal{S}}[l,k] &= \sqrt{2}\Re\left\{X_{\mathcal{F}}[l,k]e^{\mathrm{j}\frac{3\pi}{4}}\right\} \\ &= -\Re\{X_{\mathcal{F}}[l,k]\} + \Im\{X_{\mathcal{F}}[l,k]\} \end{aligned} \tag{3}$$

which will prove itself to be a very useful formulation.

It is possible to show that, unlike with the STFT, the convolution theorem doesn't hold when using the SSBT. That is, if $y[n] = h[n] * x[n]$, then $Y_{\mathcal{F}}[l,k] = H_{\mathcal{F}}[l,k] * X_{\mathcal{F}}[l,k]$, but $Y_{\mathcal{S}}[l,k] \neq H_{\mathcal{S}}[l,k] * X_{\mathcal{S}}[l,k]$. Nonetheless, as long as any result is first converted into the STFT domain (through Eq. (3)) before being used, it can still be employed to estimate the matrices and signals.

3. Signal and Array Model

Let there be a sensor array without any specific shape, which is comprised of M sensors, within a reverberant environment. In this setting there also are a desired and an interfering sources (name $x[n]$ and $v[n]$), and also uncorrelated noise, $r_m[n]$ (at each sensor m).

We denote $h_m[n]$ as the room impulse response, between the desired signal (at source) and the m-th sensor. We similarly define $g_m[n]$ for the interfering signal at source. From this, we write $y_m[n]$ as the observed signal at the m-th sensor as

$$y_m[n] = h_m[n] * x[n] + g_m[n] * v[n] + r_m[n] \tag{4}$$

We let m' be the reference sensor's index (without compromise, $m' = 1$). We let $x_1[n] = h_1[n] * x[n]$ (and similarly for $v_1[n]$). $b_m[n]$ is the *relative* impulse response between the desired signal (at the reference sensor) and the m-th sensor, such that

$$b_m[n] * x_1[n] = h_m[n] * x[n] \tag{5}$$

and we similarly define $c_m[n]$ such that $c_m[n] * v_1[n] = g_m[n] * v[n]$. Therefore, Eq. (4) becomes

$$y_m[n] = b_m[n] * x_1[n] + c_m[n] * v_1[n] + r_m[n] \tag{6}$$

We can use a time-frequency transform (in here the STFT or the SSBT, as exposed in Section 2) with the CTF model [13] to turn Eq. (6) into

$$Y_m[l,k] = B_m[l,k] * X_1[l,k] + C_m[l,k] * V_1[l,k] + R_m[l,k] \tag{7}$$

where $Y_m[l,k]$ is the transform of $y_m[n]$ (resp. $B_m[l,k]$, $X_1[l,k]$, $C_m[l,k]$, $V_1[l,k]$ and $R_m[l,k]$); l is the window index, and k the bin index, with $0 \leq k \leq K-1$; and the convolution is in the window-index axis.

Assuming that $B_m[l,k]$ is a finite (possibly truncated) response with L_B windows, then

$$B_m[l,k] * X_1[l,k] = \mathbf{b}_m^{\mathsf{T}}[k]\mathbf{x}_1[l,k] \tag{8}$$

in which

$$\mathbf{b}_m[k] = \Big[\, B_m[0\,,\ k]\,,\ B_m[1\,,\ k]\,,\ \cdots\,,\ B_m[L_B-1\,,\ k] \,\Big]^{\mathsf{T}} \tag{9a}$$

$$\mathbf{x}_1[l,k] = \Big[\, B_m[l\,,\ k]\,,\ B_m[l-1\,,\ k]\,,\ \cdots\,,\ B_m[l-L_B+1\,,\ k] \,\Big]^{\mathsf{T}} \tag{9b}$$

and similarly we define $\mathbf{c}_m[k]$ and $\mathbf{v}_1[l,k]$. Note that $\mathbf{b}_m[k]$ doesn't depend on the index l, since the system is time-invariant. With this, Eq. (7) becomes

$$Y_m[l,k] = \mathbf{b}_m^{\mathsf{T}}[k]\mathbf{x}_1[l,k] + \mathbf{c}_m^{\mathsf{T}}[k]\mathbf{v}_1[l,k] + R_m[l,k] \tag{10}$$

Vectorizing the signals sensor-wise, we finally get

$$\mathbf{y}[l,k] = \mathbf{B}^{\mathsf{T}}[k]\mathbf{x}_1[l,k] + \mathbf{C}^{\mathsf{T}}[k]\mathbf{v}_1[l,k] + \mathbf{r}[l,k] \tag{11}$$

where

$$\mathbf{y}[l,k] = \Big[\, y_1[l\,,\ k]\,,\ \cdots,\ y_M[l\,,\ k]\,\Big]^{\mathsf{T}} \tag{12}$$

and similarly for the other variables. In this situation, $\mathbf{B}[k]$ and $\mathbf{C}[k]$ are $L_B \times M$ and $L_C \times M$ matrices respectively; $\mathbf{x}_1[l,k]$ and $\mathbf{v}_1[l,k]$ are $L_B \times 1$ and $L_C \times 1$ vectors respectively; and $\mathbf{y}[l,k]$ and $\mathbf{r}[l,k]$ are $M \times 1$ vectors.

3.1. Reverb-aware formulation

We define Δ as the window-index in which $b_1[n]$ starts, and assume that $b_1[n]$ starts at the start of a window of the transform[1]. We assume that the first window of $\mathbf{B}[k]$ is the desired part of speech, and the rest is an undesired component, which is only reverberation.

With this, we write

$$\mathbf{B}^{\mathsf{T}}[k]\mathbf{x}_1[l,k] = \mathbf{d}_x[k]X_1[l,k] + \sum_{\substack{l'=0 \\ l'\neq\Delta}}^{L_B-1} \mathbf{p}_{B,l'}[k]X_1[l-l',k] \tag{13}$$

where $\mathbf{d}_x[k]$ is the k-th row of $\mathbf{B}[k]$, and $\mathbf{p}_{B,l'}[k]$ is the l'-th row of $\mathbf{B}[k]$. With this, $\mathbf{d}_x[k]X_1[l,k]$ is the desired speech component of $\mathbf{B}^{\mathsf{T}}[k]\mathbf{x}_1[l,k]$, and the summation over l' is the undesired component. We will call $\mathbf{d}_x[k]$ the desired speech frequency response.

We define $\mathbf{p}_{C,l''}$ similarly, such that

$$\mathbf{C}^{\mathsf{T}}[k]\mathbf{v}_1[l,k] = \sum_{l''=0}^{L_C-1} \mathbf{p}_{C,l''}[k]V_1[l-l'',k] \tag{14}$$

From here, we can write

$$\mathbf{y}[l,k] = \mathbf{d}_x[k]X_1[l,k] + \mathbf{w}[l,k] \tag{15}$$

with $\mathbf{w}[l,k]$ being the undesired signal (undesired speech components + interfering source + noise), given by

$$\mathbf{w}[l,k] = \sum_{\substack{l'=0 \\ l'\neq\Delta}}^{L_B-1} \mathbf{p}_{B,l'}[k]X_1[l-l',k] + \sum_{l''=0}^{L_C-1} \mathbf{p}_{C,l''}[k]V_1[l-l'',k] + \mathbf{r}[l,k] \tag{16}$$

It is important to take into account the delay between sensors and the length of the window (in seconds). If the time it takes for the desired signal to go from the reference to the farthest sensor is longer than the length of the window, more than one window will need to be considered desired speech. Assuming that the farthest sensor is a distance δ from the reference, then this isn't a problem as long as $\frac{\delta}{c} < \frac{K}{f_s}$.

[1] This can be easily achieved by left zero-padding all $b_m[n]$ appropriately.

3.2. MVDR beamformer

We use a linear time-invariant filter $\mathbf{f}[l,k]$ to estimate the desired signal at the reference sensor, such that

$$\begin{aligned} Z[l,k] &= \mathbf{f}^{\mathsf{H}}[l,k]\mathbf{y}[l,k] \\ &\approx X_1[l,k] \end{aligned} \tag{17}$$

It is important to note that this filter is time-invariant within the window, but it may change over time depending on the signal.

In order to minimize the undesired signal $\mathbf{w}[l,k]$, we will use an MVDR beamformer [15], whose formulation is

$$\mathbf{f}^{\star}[l,k] = \min_{\mathbf{f}[l,k]} \mathbf{f}[l,k]^{\mathsf{H}} \mathbf{\Phi}_{\mathbf{w}}[l,k]\mathbf{f}[l,k] \text{ s.t. } \mathbf{f}^{\mathsf{H}}[l,k]\mathbf{d}_x[k] = 1 \tag{18}$$

in which $\mathbf{f}^{\mathsf{H}}[l,k]\mathbf{d}_x[k] = 1$ is the distortionless constraint, and $\mathbf{\Phi}_{\mathbf{w}}[l,k]$ is the correlation matrix of the undesired signal, given by

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{w}}[l,k] = &\sum_{\substack{l'=0 \\ l'\neq\Delta}}^{L_B-1} \mathbf{p}_{B,l'}^{\mathsf{H}}[k]\mathbf{p}_{B,l'}[k]\phi_{X_1}[l-l',k] \\ &+ \sum_{l''=0}^{L_C-1} \mathbf{p}_{C,l''}^{\mathsf{H}}[k]\mathbf{p}_{C,l''}[k]\phi_{V_1}[l-l'',k] \\ &+ \mathbf{I}_M\phi_R[l,k] \end{aligned} \tag{19}$$

where $\phi_{X_1}[l,k]$ is the variance of $X_1[l,k]$ (same for $\phi_{V_1}[l,k]$), and $\mathbf{I}_M$ is the $M \times M$ identity matrix, assuming that the distribution of $\mathbf{r}[l,k]$ is the same for all sensors.

The solution to Eq. (18) is given by

$$\mathbf{f}_{\text{mvdr}}[l,k] = \frac{\mathbf{\Phi}_{\mathbf{w}}^{-1}[l,k]\mathbf{d}_x[l,k]}{\mathbf{d}_x^{\mathsf{H}}[l,k]\mathbf{\Phi}_{\mathbf{w}}^{-1}[l,k]\mathbf{d}_x[l,k]} \tag{20}$$

Note that, even though overlapping windows are used when windowing the signal for the time-frequency transform, for simplicity we assume that $X_1[l_1,k]$ is independent of $X_1[l_2,k]$.

4. True-MVDR with the SSB Transform

In the formulation exposed previously, all signals and matrices are within the same domain. That is, in Eq. (18) we have that the distortionless constraint is built such that the beamformer doesn't cause distortion only on the SSBT domain. However, as was explained at the end of Section 2, since the filtering can't be done in the SSBT domain, the beamformer must be converted into the STFT domain (through Eq. (3)) before being applied to the signal. To correctly construct a beamformer that properly fulfills the distortionless constraint, this conversion should be taken into account.

Given the signal $x[n]$, its STFT $X_{\mathcal{F}}[l,k]$ (with $\left\lfloor \frac{K+1}{2} + 1 \right\rfloor$ bins), and its SSBT $X_{\mathcal{S}}[l,k]$ (with K bins), from Eq. (3) it is possible to show that[2]

$$X_{\mathcal{F}}[l,k] = \frac{1}{\sqrt{2}}\left(e^{\mathrm{j}\frac{3\pi}{4}} X_{\mathcal{S}}[l,k] + e^{-\mathrm{j}\frac{3\pi}{4}} X_{\mathcal{S}}[l,K-k]\right) \tag{21}$$

From this, we propose a framework in which we consider both bins k and $K-k$ simultaneously in the SSBT, given that they aren't independent. We thus define $\mathbf{y}'[l,k]$ as

$$\mathbf{y}'[l,k] = \begin{bmatrix} \mathbf{y}[l,k] \\ \mathbf{y}[l,K-k] \end{bmatrix}_{2M\times 1} \tag{22}$$

We similarly define $\mathbf{v}'[l,k]$, from which we define $\mathbf{\Phi}_{\mathbf{v}'}[l,k]$ as its correlation matrix. Under this idea, our filter $\mathbf{f}'[l,k]$ is a $2M \times 1$ vector, from which we can extract the SSBT beamformer $\mathbf{f}[l,k]$ through

$$\mathbf{f}'[l,k] = \begin{bmatrix} \mathbf{f}[l,k] \\ \mathbf{f}[l,K-k] \end{bmatrix}_{2M\times 1} \tag{23}$$

with $\mathbf{f}[l,k]$ being the beamformer for the k-th bin, and $\mathbf{f}[l,K-k]$ for the $[K-k]$-th bin.

Defining $\hat{\mathbf{A}}$ as

$$\hat{\mathbf{A}} = \begin{bmatrix} \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & \cdots & 0 & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & \cdots & 0 \\ 0 & \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & \cdots & 0 & 0 & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & 0 & \cdots & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} \end{bmatrix}_{M\times 2M} \tag{24}$$

then with Eq. (21) it is easy to see that

$$\hat{\mathbf{f}}_{\mathcal{F}}[l,k] = \hat{\mathbf{A}}\mathbf{f}'[l,k] \tag{25}$$

with $\hat{\mathbf{f}}_{\mathcal{F}}[l,k]$ being the obtained beamformer, converted into the STFT domain. From this, the distortionless constraint for the STFT domain can be written for the SSBT domain as

$$\begin{aligned} \hat{\mathbf{f}}_{\mathcal{F}}^{\mathsf{H}}[l,k]\mathbf{d}_{\mathcal{F};x}[k] &= 1 \\ \mathbf{f}'^{\mathsf{H}}[l,k]\hat{\mathbf{A}}^{\mathsf{H}}\mathbf{d}_{\mathcal{F};x}[k] &= 1 \\ \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{D}_x[k] &= 1 \end{aligned} \tag{26}$$

where $\mathbf{d}_{\mathcal{F};x}[l,k]$ is the desired speech frequency response in the STFT domain; and

$$\mathbf{D}_x[k] = \hat{\mathbf{A}}^{\mathsf{H}}\mathbf{d}_{\mathcal{F};x}[k] \tag{27}$$

In this scheme, our minimization problem becomes

$$\mathbf{f}'^{\star}[l,k] = \min_{\mathbf{f}'[l,k]} \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{\Phi}_{\mathbf{v}'}[l,k]\mathbf{f}'[l,k] \text{ s.t. } \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{D}_x[k] = 1 \tag{28}$$

[2] This equation (as well as the derivations going forward) is invalid for $k = 0$, and $k = K/2$ if K is even. However, in those cases $X_{\mathcal{F}}[l,k] = X_{\mathcal{S}}[l,k]$, and the "naïve" SSBT beamformer works.

4.1. Real-valued SSBT true-MVDR beamformer

Given that $\mathbf{D}_x$ is a complex-valued matrix, the solution to Eq. (28) will generally be complex as well, which defeats the purpose of using the SSBT, given that its spectrum is real-valued. Thus, another restriction must be added in order to ensure this desired behavior.

From the distortionless constraint of Eq. (28), we trivially have that

$$\mathbf{f}'^{\mathsf{H}}[l,k]\Re\{\mathbf{D}_x[k]\} = 1 \tag{29a}$$
$$\mathbf{f}'^{\mathsf{H}}[l,k]\Im\{\mathbf{D}_x[k]\} = 0 \tag{29b}$$

which can be put in matricial form,

$$\mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{Q}_x[k] = \mathbf{i}^{\mathsf{T}} \tag{30}$$

with

$$\mathbf{Q}_x[k] = \Big[\, \Re\{\mathbf{D}_x[k]\}\,,\ \Im\{\mathbf{D}_x[k]\} \,\Big]_{2M\times 2} \tag{31a}$$
$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{31b}$$

From this, the minimization problem becomes

$$\mathbf{f}'^{\star}[l,k] = \min_{\mathbf{f}'[l,k]} \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{\Phi}_{\mathbf{v}'}[l,k]\mathbf{f}'[l,k] \text{ s.t. } \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{Q}_x[k] = \mathbf{i}^{\mathsf{T}} \tag{32}$$

whose formulation is the same as the linearly-constrained minimum variance (LCMV) [17] beamformer, and therefore its solution is

$$\mathbf{f}'^{\star}[l,k] = \mathbf{\Phi}_{\mathbf{v}'}^{-1}[l,k]\mathbf{Q}_x[k]\Big(\mathbf{Q}_x^{\mathsf{H}}[k]\mathbf{\Phi}_{\mathbf{v}'}^{-1}[k]\mathbf{Q}_x[k]\Big)^{-1}\mathbf{i} \tag{33}$$

Since all matrices involved in the calculation of this beamformer are now real-valued, then it is trivial that

$$\mathbf{f}'^{\star}_{\text{mvdr}}[l,k] = \mathbf{\Phi}_{\mathbf{v}'}^{-1}[l,k]\mathbf{Q}_x[k]\Big(\mathbf{Q}_x^{\mathsf{T}}[k]\mathbf{\Phi}_{\mathbf{v}'}^{-1}[k]\mathbf{Q}_x[k]\Big)^{-1}\mathbf{i} \tag{34}$$

and, using Eq. (25), we can obtain the desired beamformer $\hat{\mathbf{f}}^{\star}_{\mathcal{F}}[l,k]$, in the STFT domain.

5. Simulations

In the simulations[3], we will use a sampling frequency $f_s = 16\text{kHz}$. The sensor array is assumed an uniform linear array of 10 sensors, with an intersensor distance of 2cm. All RIR's were generated using Habets' room impulse generator [18]. Signals used were from the SMARD [19] and LINSE [20] databases.

The room's dimensions are 4m × 6m × 3m, with a reverberation time of 0.11s. The desired source is located at (2m, 1m, 1m), and its signal is a male voice (SMARD, `50_male_speech_english_ch8_OmniPower4296.flac`).

[3] The code for the simulations can be found in https://github.com/VCurtarelli/py-ssb-ctf-bf.

The interfering source is located (simultaneously) at $(0.5\text{m},\ 5\text{m},\ [0.3 : 0.3 : 2.7]\text{m})$, emulating an open door, with its signal being a babble sound (LINSE database, `babble.mat`). The noise signal is a WGN noise (SMARD database, `wgn_48kHz_ch8_OmniPower4296.flac`). All signals were resampled to the desired sampling frequency. The sensor array is assumed to be an uniform linear array, positioned at $(2\text{m},\ [4.02 : 0.02 : 4.2]\text{m},\ 1\text{m})$, whose sensors are assumed to be omnidirectional and with flat frequency response.

We have that the input SNR between desired and interfering signals is of 5dB, and the SNR between desired and noise signals is of 30dB. The filters were calculated every 25 windows, and consider the previous 25 windows, in order to calculate the correlation matrices.

We will compare the filters obtained through the STFT and SSBT transforms, with T-SSBT denoting the beamformer obtained via the true-distortionless MVDR derived in Section 4, and N-SSBT the naïve approach, in which one would simply use Eq. (20) to calculate the SSBT beamformer. The performance analysis was done only on the STFT domain. In all lineplots, the STFT is presented in red, the N-SSBT in green, and the T-SSBT in blue.

5.1. Results

In this scenario, we assume that the analysis (and synthesis) windows have 32 samples. Figure 2 shows the window-wise averaged gain in SNR, and Fig. 1 the gain in SNR for each window (with the windows being represented by the time started, in seconds), for all methods. In Fig. 3 we have the DSDI for all three methods, window-averaged.

Although it isn't as clear from the per-window results of Fig. 1, Fig. 2 shows us that both beamformers obtained from the SSBT led to a better enhancement of the signal, in terms of the output SNR, with the N-SSBT beamformer having a better performance over (almost) all spectrum, and the T-SSBT beamformer being better for lower frequencies, tying with the STFT for higher ones.

Also, Fig. 3 shows that the T-SSBT filter was able to ensure a distortionless response for the desired signal, a feature that wasn't achieved by the N-SSBT beamformer. This is wholly expected, since the later was naively designed with the MVDR in mind, and wasn't fully planned to achieve a distortionless behavior in the STFT (and therefore the time) domain, while the T-SSBT took this into account on its derivation.

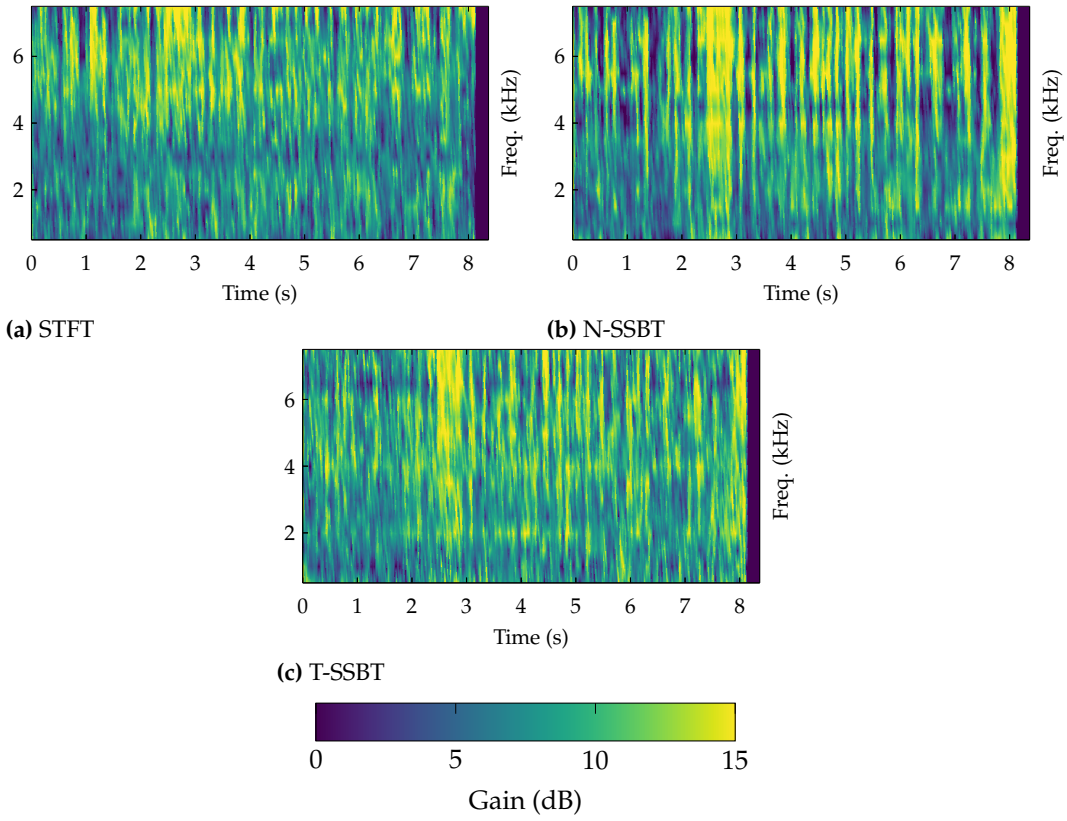

Figure 1. Per-window SNR gain.

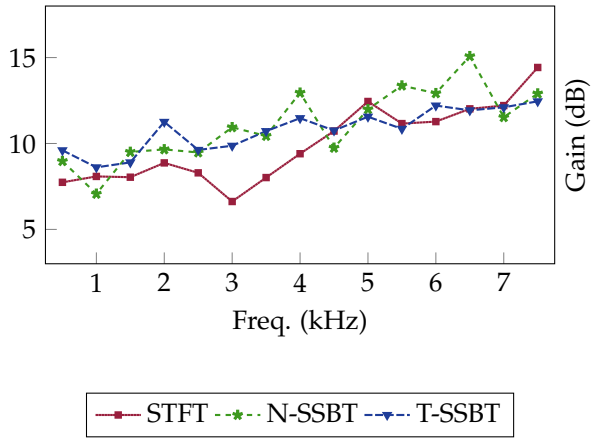

Figure 2. Window-average SNR gain.

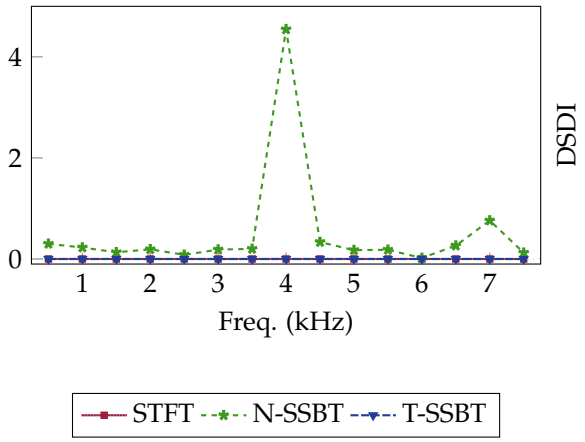

Figure 3. Window-average DSDI.

5.2. Results - 64 samples/window

In this simulation, we changed the number of samples per window from 32 to 64, keeping everything else the same.

In here, from Figs. 4 and 5 we see a similar result to that which was obtained previously, with the N-SSBT beamformer having a better performance overall, but causing some distortion in the desired signal; while the T-SSBT beamformer has a slightly better performance than the one obtained through the STFT, while also having a distortionless behavior, as is seen in Fig. 6.

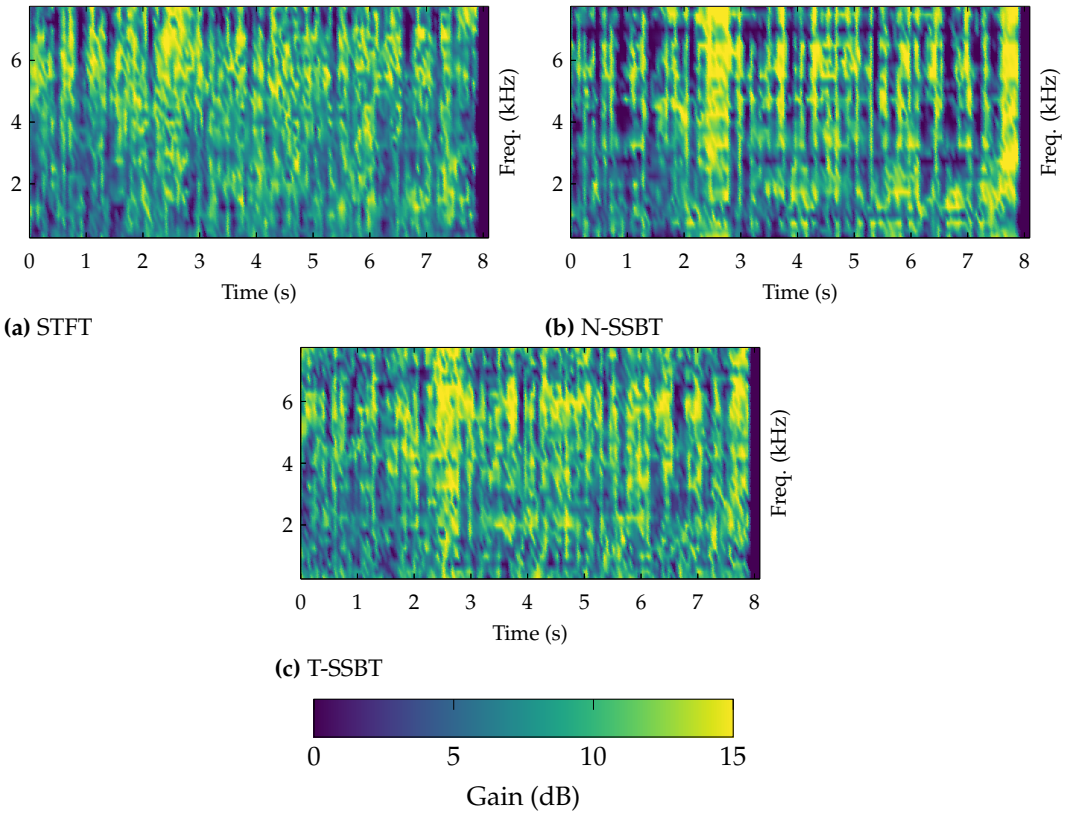

Figure 4. Per-window SNR gain.

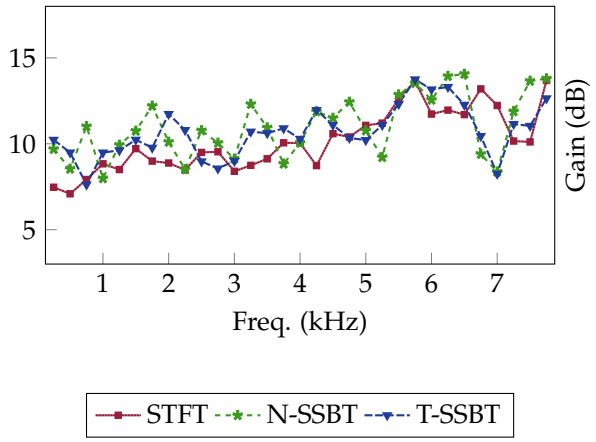

Figure 5. Window-average SNR gain.

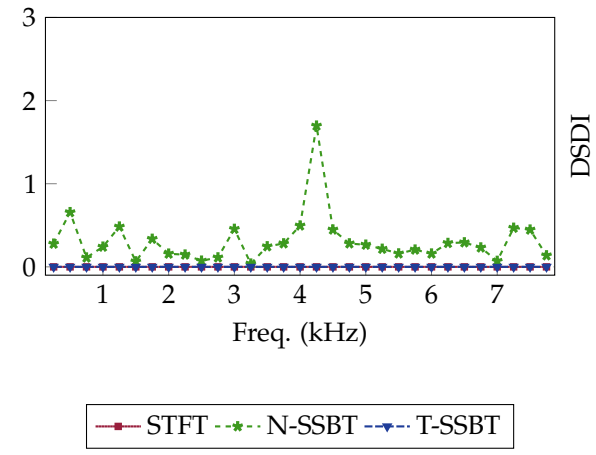

Figure 6. Window-average DSDI.

6. Conclusion

In this study, we investigated the application of the Single-Sideband Transform in beamforming within a reverberant environment, utilizing the convolutive transfer function model for filter bank (i.e., the beamformer) estimation. We implemented a Minimum-Variance Distortionless-Response beamformer to enhance signals in a real-life-like scenario, elucidating the process to achieve a true-distortionless MVDR beamformer when employing the SSB transform. The results demonstrated that both the naive MVDR and the true-MVDR beamformers, designed using the SSBT, outperformed the traditional beamformer obtained via the Short-Time Fourier Transform. The naive MVDR exhibited some distortion on the desired signal, whereas the true-MVDR achieved a distortionless response, as expected.

Future research avenues may explore the integration of this transform into different beamformers, or undertake further comparisons against the established and reliable STFT methodology.

Author Contributions: Conceptualization, I. Cohen and V. Curtarelli; Methodology, V. Curtarelli; Software, V. Curtarelli; Writing—original draft: V. Curtarelli; Writing—review and editing, I. Cohen and V. Curtarelli; Supervision, V. Curtarelli. 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.

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

Abbreviations

The following abbreviations are used in this manuscript:

STFT	Short-Time Fourier Transform
SSBT	Single-Sideband Transform
CTF	Convolutive Transfer Function
MVDR	Minimum-Variance Distortionless-Response
LCMV	Linearly-Constrained Minimum-Variance

References

1. Lobato, W.; Costa, M.H. Worst-Case-Optimization Robust-MVDR Beamformer for Stereo Noise Reduction in Hearing Aids. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **2020**, *28*, 2224–2237. https://doi.org/10.1109/TASLP.2020.3009831.
2. Chen, J.; Kung Yao.; Hudson, R. Source localization and beamforming. *IEEE Signal Processing Magazine* **2002**, *19*, 30–39. https://doi.org/10.1109/79.985676.
3. Lu, Y. BIOMEDICAL ULTRASOUND BEAM FORMING. *Ultrasound in Med. & Biol.* **1994**, *20*, 403–428. https://doi.org/10.1016/0301-5629(94)90097-3.
4. Nguyen, N.Q.; Prager, R.W. Minimum Variance Approaches to Ultrasound Pixel-Based Beamforming. *IEEE Transactions on Medical Imaging* **2017**, *36*, 374–384. https://doi.org/10.1109/TMI.2016.2609889.
5. Benesty, J.; Cohen, I.; Chen, J. *Fundamentals of signal enhancement and array signal processing*; John Wiley & Sons Singapore Pte. Ltd: Hoboken, 2018.
6. Kıymık, M.; Güler, İ.; Dizibüyük, A.; Akın, M. Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. *Computers in Biology and Medicine* **2005**, *35*, 603–616. https://doi.org/10.1016/j.compbiomed.2004.05.001.
7. 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. *The Journal of the Acoustical Society of America* **2021**, *149*, 660–672. https://doi.org/10.1121/10.0003335.
8. Chen, W.; Huang, X. Wavelet-Based Beamforming for High-Speed Rotating Acoustic Source. *IEEE Access* **2018**, *6*, 10231–10239. https://doi.org/10.1109/ACCESS.2018.2795538.
9. Yang, Y.; Peng, Z.K.; Dong, X.J.; Zhang, W.M.; Meng, G. General Parameterized Time-Frequency Transform. *IEEE Transactions on Signal Processing* **2014**, *62*, 2751–2764. https://doi.org/10.1109/TSP.2014.2314061.
10. Almeida, L. The fractional Fourier transform and time-frequency representations. *IEEE Transactions on Signal Processing* **1994**, *42*, 3084–3091. https://doi.org/10.1109/78.330368.
11. Crochiere, R.E.; Rabiner, L.R. *Multirate digital signal processing*; Prentice-Hall signal processing series, Prentice-Hall: Englewood Cliffs, NJ, 1983.
12. Oyzerman, A. Speech Dereverberation in the Time-Frequency Domain. Master's thesis, Technion - Israel Institute of Technology, Haifa, Israel, 2012.

13. Talmon, R.; Cohen, I.; Gannot, S. Relative Transfer Function Identification Using Convolutive Transfer Function Approximation. *IEEE Transactions on Audio, Speech, and Language Processing* **2009**, *17*, 546–555. https://doi.org/10.1109/TASL.2008.2009576.
14. Capon, J. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE* **1969**, *57*, 1408–1418. https://doi.org/10.1109/PROC.1969.7278.
15. Erdogan, H.; Hershey, J.R.; Watanabe, S.; Mandel, M.I.; Roux, J.L. Improved MVDR Beamforming Using Single-Channel Mask Prediction Networks. In Proceedings of the Interspeech 2016. ISCA, September 2016, pp. 1981–1985. https://doi.org/10.21437/Interspeech.2016-552.
16. DeMuth, G. Frequency domain beamforming techniques. In Proceedings of the ICASSP '77. IEEE International Conference on Acoustics, Speech, and Signal Processing, Hartford, CT, USA, 1977; Vol. 2, pp. 713–715. https://doi.org/10.1109/ICASSP.1977.1170316.
17. Habets, E.A.P.; Benesty, J.; Gannot, S.; Naylor, P.A.; Cohen, I. On the application of the LCMV beamformer to speech enhancement. In Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, October 2009; pp. 141–144. https://doi.org/10.1109/ASPAA.2009.5346463.
18. Habets, E. RIR Generator, 2020.
19. Nielsen, J.K.; Jensen, J.R.; Jensen, S.H.; Christensen, M.G. The Single- and Multichannel Audio Recordings Database (SMARD). In Proceedings of the Int. Workshop Acoustic Signal Enhancement, Sep. 2014.
20. Johnson, D.H. Signal Processing Information Database, 2013.