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-windows time-frequency transforms. Our study aims to optimize the utilization of SSBT in this endeavor, 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 show that employing the SSBT – both in a naive and a refined approach – results in superior signal enhancement compared to the Short-Time Fourier Transform (STFT). The refined approach not only enhances the signal but also ensures the desired distortionless behavior, which isn't achieved by the naive one.

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

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:

Copyright: © 2024 by the authors. Submitted to *Algorithms* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

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 analysis.

Among these, the Single-Sideband Transform (SSBT) [11,12], stands out for its real-valued frequency spectrum. It has been shown that the SSBT works particularly well with short analysis windows [11], lending itself useful when working with the convolutive transfer function (CTF) model [13] and filter-banks [14,15] 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 [16,17]. As the MVDR beamformer works on the frequency (or time-frequency) domain without any specification on the transform 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 relevant; Section 3 the signal model considered in the time domain is presented, 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. 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 [18], 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. Thus, exploring different possibilities for such an operation can be useful and lead to interesting results.

The Single-Sideband Transform (SSBT) [11] is one such alternative, in which the frequency values are cleverly calculated such that its spectrum is real-valued, without loss of information. 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.

From Eq. (2) it's easy to see that

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

assuming that all K bins of the STFT are available.

It is possible to show that, unlike with the STFT, the convolution theorem does not hold when employing the SSBT. In other words, 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, by first converting any result into the STFT domain (using Eq. (3)) before utilization, it remains feasible to employ the obtained values for estimating matrices and signals.

3. Signal and Array Model

Let there be a generic sensor array, comprised of M sensors, within a reverberant environment. In this setting there also are a desired and an interfering sources (namely $x[n]$ and $v[n]$), and also uncorrelated noise $r_m[n]$ (at each sensor m). We assume that the sources are spatially stationary, and don't move over time.

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 and, without compromise, set $m' = 1$. 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, define such that

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

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 (here the STFT or the SSBT[1], 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. all other signals); 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

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

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

[1] Although the SSBT doesn't hold the convolution theorem, we will assume it does for the purpose of the formulation.

and in the same way 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 and we assume the sources to be spatially stationary. 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

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

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 therefore the Δ-th window of $\mathbf{B}[k]$ is the desired part of speech, and the rest is an undesired component, comprised only of 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

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

We must consider the sensor delay and window length. If the time for the signal to travel from the reference to the farthest sensor exceeds the window length (in seconds), multiple windows may represent the desired speech. This isn't a problem if $\frac{\delta}{c} < \frac{K}{f_s}$, where

δ is the distance to the farthest sensor, c is the speed of sound, K is the window length, and f_s is the sampling frequency.

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

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

with $\cdot^{\mathsf{H}}$ being the transposed-complex-conjugate operator. 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 [17], whose formulation is

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

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

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

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

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

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 $X_1[l_2,k]$.

3.3. Beamformer metrics

Considering the problem, the metrics of most interest are the gain in signal-to-noise ratio (SNR) and the desired signal distortion index (DSDI), both of these being given by

$$\text{gSNR}[l,k] = \frac{\phi_{V_1}[l,k]\left|\mathbf{f}^{\mathsf{H}}[l,k]\mathbf{d}_x[k]\right|^2}{\mathbf{f}^{\mathsf{H}}[l,k]\mathbf{\Phi_v}[l,k]\mathbf{f}[l,k]} \tag{21}$$

$$v[l,k] = \left|\mathbf{f}^{\mathsf{H}}[l,k]\mathbf{d}_x[k] - 1\right|^2 \tag{22}$$

We can also define the window-averaged gSNR and DSDI as

$$\text{gSNR}[k] = \frac{1}{L_Z} \sum_{l=0}^{L_Y - 1} \text{gSNR}[l,k] \tag{23}$$

$$v[k] = \frac{1}{L_Z} \sum_{l=0}^{L_Y - 1} v[l,k] \tag{24}$$

with L_Z being the number of windows of $Z[l,k]$.

4. True-MVDR with the SSB Transform

In the previous formulation, all signals and matrices operate within the same domain. In Eq. (18), the distortionless constraint is designed to ensure that the SSBT beamformer avoids causing distortion exclusively within the SSBT domain. However, as explained towards the end of Section 2, due to the inherent limitations of filtering in the SSBT domain, the beamformer must undergo conversion into the STFT domain (via Eq. (3)) before filtering. To construct a beamformer that correctly adheres to the distortionless constraint, it is essential to consider this conversion step.

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[2] that

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

From this, we propose a framework in which we consider both bins k and $K-k$ simultaneously in the SSBT beamformer. We thus define $\mathbf{y}'[l,k]$ as

$$\mathbf{y}'[l,k] = \begin{bmatrix} \mathbf{y}[l,k] \\ \mathbf{y}[l,K-k] \end{bmatrix}_{2M \times 1} \tag{26}$$

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, with the first M values being for the k-th bin, and the last M values for the $[K-k]$-th bin. With Eq. (25) it is easy to see that

$$\hat{\mathbf{f}}_{\mathcal{F}}[l,k] = \hat{\mathbf{A}} \mathbf{f}'[l,k] \tag{27}$$

where $\hat{\mathbf{A}}$ is

$$\hat{\mathbf{A}} = \begin{bmatrix} \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & \cdots & 0 & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & \cdots & 0 \\ 0 & \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & \cdots & 0 & 0 & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \frac{e^{\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} & 0 & 0 & \cdots & \frac{e^{-\mathrm{j}\frac{3\pi}{4}}}{\sqrt{2}} \end{bmatrix}_{M \times 2M} \tag{28}$$

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

with $\hat{\mathbf{f}}_{\mathcal{F}}[l,k]$ being the STFT-equivalent beamformer for $\mathbf{f}[l,k]$. From this, it is easy to see that the distortionless constraint for the STFT, within the SSBT domain, is

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

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

$$\mathbf{D}_x[k] = \hat{\mathbf{A}}^{\mathsf{H}}\mathbf{d}_{\mathcal{F};x}[k] \tag{30}$$

In this scheme, our minimization problem becomes

$$\mathbf{f}'^{\star}[l,k] = \min_{\mathbf{f}'[l,k]} \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{\Phi}_{\mathbf{v}'}[l,k]\mathbf{f}'[l,k] \text{ s.t. } \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{D}_x[k] = 1 \tag{31}$$

4.1. Real-valued SSBT true-MVDR beamformer

As $\mathbf{D}_x$ is a complex-valued matrix, the solution to Eq. (31) tends to be complex, contradicting the purpose of utilizing the SSBT. To preserve this desired behavior, an additional constraint is necessary. Forcing $\mathbf{f}'[l,k]$ to be real (which will turn all $\cdot^{\mathsf{H}}$ into $\cdot^{\mathsf{T}}$, pure transpose), from the distortionless constraint of Eq. (29) we trivially have that

$$\mathbf{f}'^{\mathsf{T}}[l,k]\Re\{\mathbf{D}_x[k]\} = 1 \tag{32a}$$

$$\mathbf{f}'^{\mathsf{T}}[l,k]\Im\{\mathbf{D}_x[k]\} = 0 \tag{32b}$$

which can be put in matricial form,

$$\mathbf{f}'^{\mathsf{T}}[l,k]\mathbf{Q}_x[k] = \mathbf{i}^{\mathsf{T}} \tag{33}$$

with

$$\mathbf{Q}_x[k] = \Big[\,\Re\{\mathbf{D}_x[k]\}\,,\ \Im\{\mathbf{D}_x[k]\}\,\Big]_{2M\times 2} \tag{34a}$$

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{34b}$$

From this, the minimization problem becomes

$$\mathbf{f}'^{\star}[l,k] = \min_{\mathbf{f}'[l,k]} \mathbf{f}'^{\mathsf{H}}[l,k]\mathbf{\Phi}_{\mathbf{v}'}[l,k]\mathbf{f}'[l,k] \text{ s.t. } \mathbf{f}'^{\mathsf{T}}[l,k]\mathbf{Q}_x[k] = \mathbf{i}^{\mathsf{T}} \tag{35}$$

whose formulation is the same as the linearly-constrained minimum variance (LCMV) [19] beamformer, and therefore its solution is

$$\mathbf{f}'^{\star}_{\text{mvdr}}[l,k] = \mathbf{\Phi}_{\mathbf{v}'}^{-1}[l,k]\mathbf{Q}_x[k]\left(\mathbf{Q}_x^{\mathsf{T}}[k]\mathbf{\Phi}_{\mathbf{v}'}^{-1}[k]\mathbf{Q}_x[k]\right)^{-1}\mathbf{i} \tag{36}$$

Using Eq. (27), 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 employ a sampling frequency of 16kHz. The sensor array consists of a uniform linear array with 10 sensors spaced at 2cm. Room impulse responses were generated using Habets' RIR generator [20], and signals were selected from the SMARD [21] and LINSE [22] databases.

The room's dimensions are 4m × 6m × 3m (width × length × height), with a reverberation time of 0.11s. The desired source is located at $(2\text{m}, 1\text{m}, 1\text{m})$, it being a male voice (SMARD, `50_male_speech_english_ch8_OmniPower4296.flac`). The interfering source, simulating an open door, is located simultaneously at $(0.5\text{m}, 5\text{m}, [0.3 : 0.3 : 2.7]\text{m})$, with a babble sound signal (LINSE database, `babble.mat`). The noise signal is white Gaussian noise (SMARD database, `wgn_48kHz_ch8_OmniPower4296.flac`). All signals were resampled to the desired frequency.

The uniform linear sensor array is positioned at $(2\text{m}, [4.02 : 0.02 : 4.2]\text{m}, 1\text{m})$, with omnidirectional sensors of flat frequency response. The input SNR between desired and interfering signals is 5dB, and between desired and noise signals is 30dB. Filters are calculated every 25 windows, considering the previous 25 windows to calculate correlation matrices.

We compare filters obtained through the STFT and SSBT transforms. T-SSBT will denote the beamformer obtained via the true-distortionless MVDR from Section 4, and N-SSBT uses Eq. (20) to (naively) calculate the SSBT beamformer. Performance analysis is conducted via the STFT domain, with the SSBT beamformers being converted into it. In line plots, STFT is presented in red, N-SSBT in green, and T-SSBT in blue.

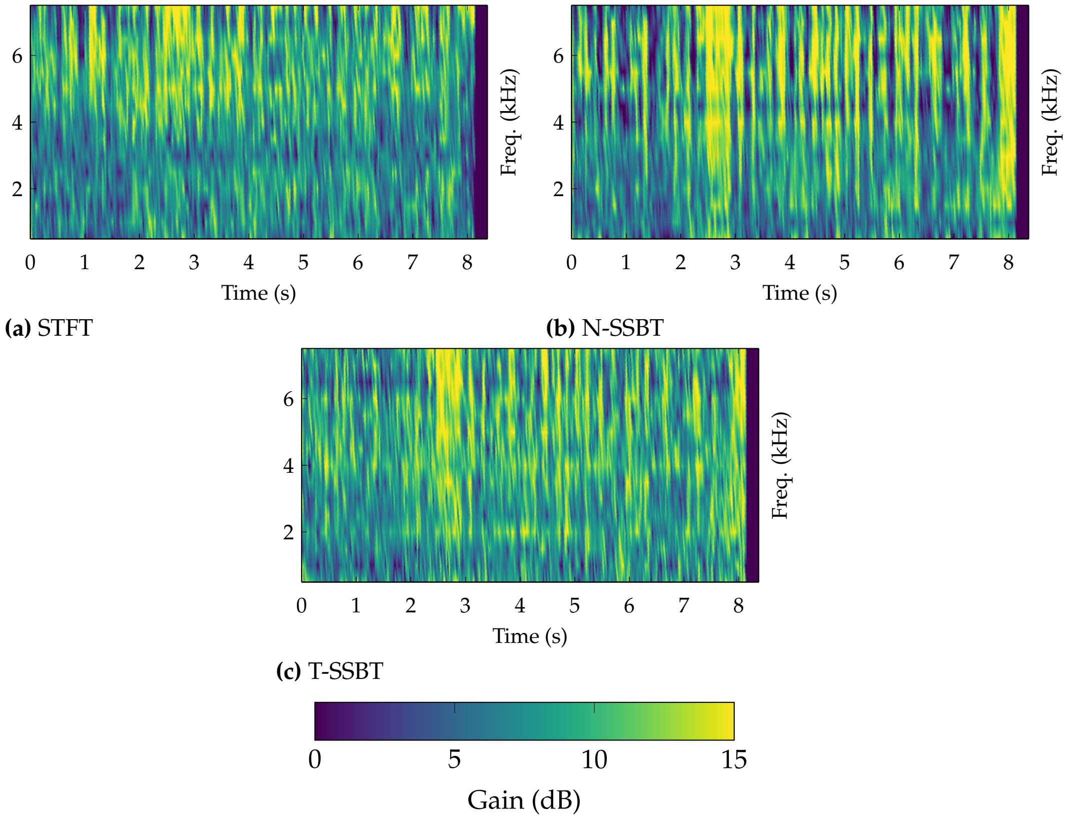
5.1. Results

In this scenario, we assume that the analysis windows have 32 samples. Fig. 1 shows the gain in SNR for each window (with the windows being represented by the time started, in seconds), and Fig. 2 the window-wise averaged gain in SNR, 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 clearly shows that both SSBT beamformers had a better (at most equal) performance than the STFT one, 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 was 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.

[3] The simulation code is available at https://github.com/VCurtarelli/py-ssb-ctf-bf.



(a) STFT

(b) N-SSBT

(c) T-SSBT

Figure 1. Per-window SNR gain.

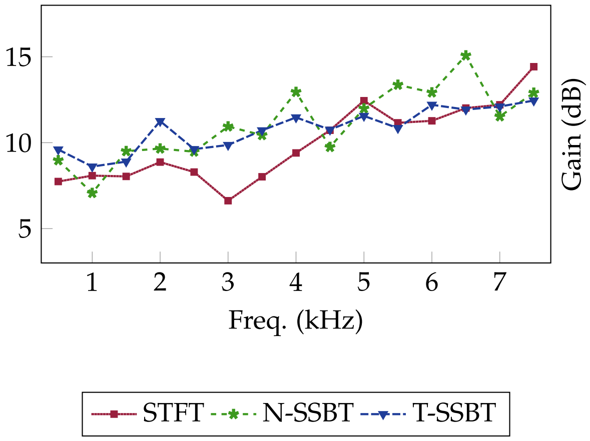


Figure 2. Window-average SNR gain.

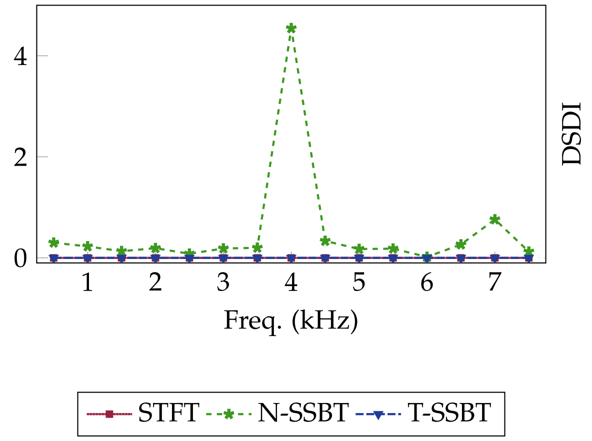
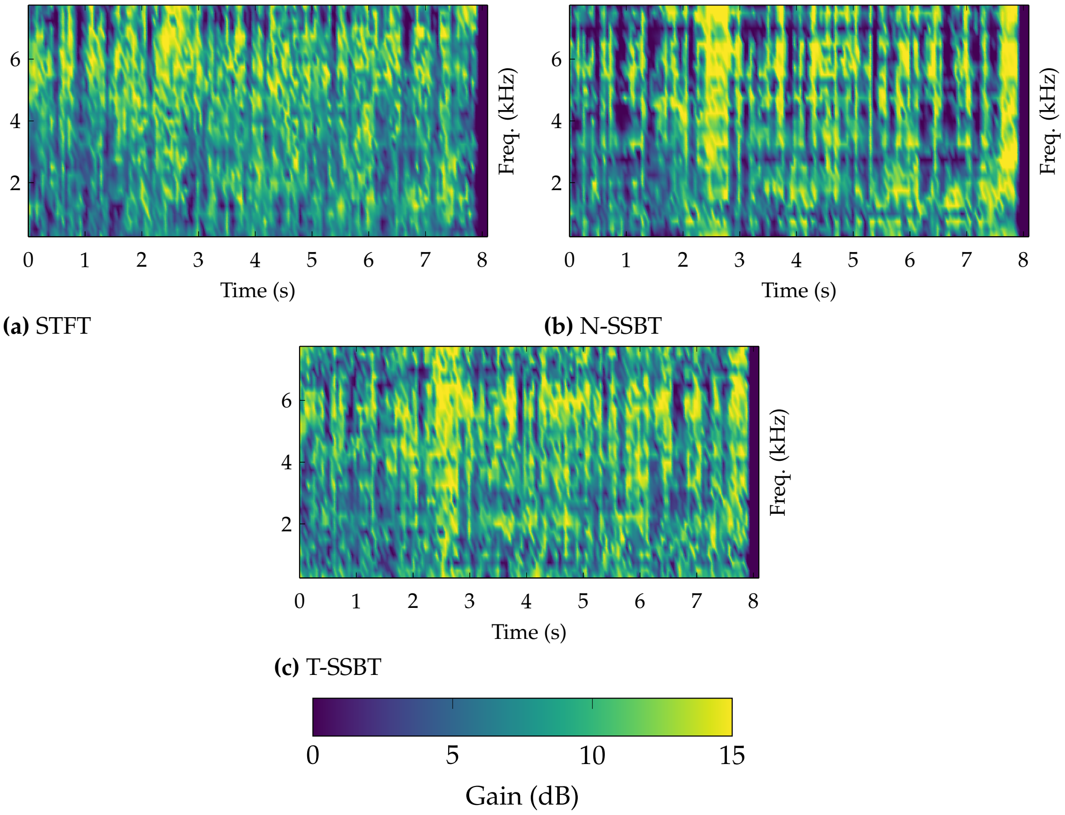


Figure 3. Window-average DSDI.

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; and the T-SSBT beamformer having a slightly better performance STFT one while also having a distortionless behavior, as is seen in Fig. 6.



(a) STFT

(b) N-SSBT

(c) T-SSBT

Figure 4. Per-window SNR gain.

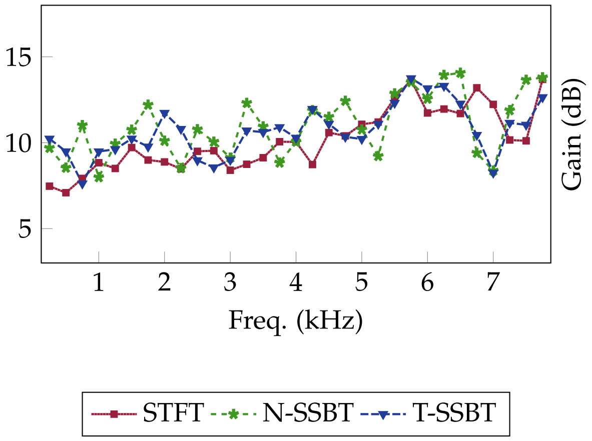


Figure 5. Window-average SNR gain.

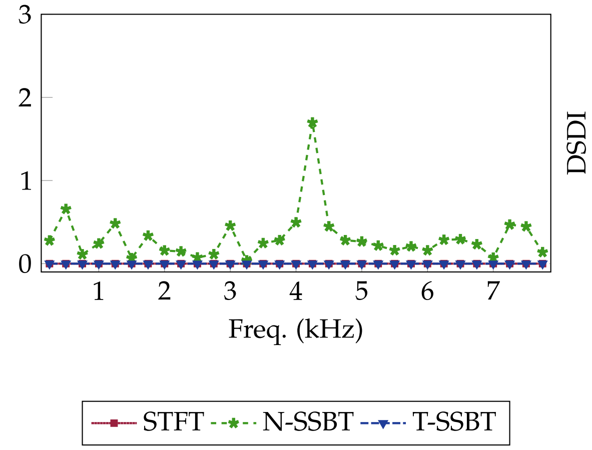


Figure 6. Window-average DSDI.

6. Conclusion 211

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:

CTF	Convolutive Transfer Function
DSDI	Desired Signal Distortion Index
LCMV	Linearly-Constrained Minimum-Variance
MVDR	Minimum-Variance Distortionless-Response
SNR	Signal-to-Noise Ratio
SSBT	Single-Sideband Transform
STFT	Short-Time Fourier Transform

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. Kumatani, K.; McDonough, J.; Schacht, S.; Klakow, D.; Garner, P.N.; Li, W. Filter bank design based on minimization of individual aliasing terms for minimum mutual information subband adaptive beamforming. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, March 2008; pp. 1609–1612. https://doi.org/10.1109/ICASSP.2008.4517933.
15. Gopinath, R.; Burrus, C. A tutorial overview of filter banks, wavelets and interrelations. In Proceedings of the 1993 IEEE International Symposium on Circuits and Systems, Chicago, IL, USA, 1993; pp. 104–107. https://doi.org/10.1109/ISCAS.1993.393668.
16. Capon, J. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE* **1969**, *57*, 1408–1418. https://doi.org/10.1109/PROC.1969.7278.
17. 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.
18. 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.
19. 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.
20. Habets, E. RIR Generator, 2020.
21. 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.
22. Johnson, D.H. Signal Processing Information Database, 2013.

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.