Article

# On the Single-Sideband Transform for MVDR Filter-Banks

**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:** Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

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

## 1. Introduction

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

[1]

**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:

## 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 [? ], 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) [? ] 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) [? ] 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 **[? ]** 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 $\Delta$ 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.

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

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

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

*3.2. MVDR beamformer*

We use an LTI filter $\mathbf{f}[l,k]$ to estimate the desired signal at the reference sensor, such that

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

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

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

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

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

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

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

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

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

[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.

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

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

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

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

### *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

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

whose formulation is the same as the LCMV beamformer, and therefore its solution is

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

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

$$\mathbf{f}'^{\star}[l,k] = \mathbf{\Phi}_{\mathbf{v}'}^{-1}[l,k]\mathbf{Q}_x[k]\left(\mathbf{Q}_x^{\top}[k]\mathbf{\Phi}_{\mathbf{v}'}^{-1}[k]\mathbf{Q}_x[k]\right)^{-1}\mathbf{i} \tag{34}$$

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

**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-cb-lcmv-rect.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| CB | Constant-beamwidth |

## References

1. Loizou, P.C. Speech Enhancement.