Constructing Time-Frequency Dictionaries for Source Separation via Time-Frequency Masking and Source Localisation

Ruairí de Fréin[1], Scott T. Rickard[1], and Barak A. Pearlmutter[2]

[1] Complex & Adaptive Systems Laboratory, University College Dublin, Ireland
[2] Hamilton Institute, National University of Ireland Maynooth, Co. Kildare, Ireland
rdefrein@ee.ucd.ie, scott.rickard@ucd.ie, barak@cs.nuim.ie

Abstract. We describe a new localisation and source separation algorithm which is based upon the accurate construction of time-frequency spatial signatures. We present a technique for constructing time-frequency spatial signatures with the required accuracy. This algorithm for multi-channel source separation and localisation allows arbitrary placement of microphones yet achieves good performance. We demonstrate the efficacy of the technique using source location estimates and compare estimated time-frequency masks with the ideal 0 dB mask.

1 Introduction

Speech is sparse in the time-frequency (T-F) domain, a property which has been exploited for Blind Source Separation (BSS), [4]. Related assumptions, namely the log-max [5] or Windowed Disjoint Orthogonality (WDO) assumption [8] in various transform domains are exploited for decompositions of financial data [6] and images [9]. Localisation can be performed in the time or frequency domain when the technique relies on a sparse representation in a dictionary of pre-computed transfer functions [1,2]. We discuss the challenges involved in constructing a T-F dictionary of spatial signatures for source localisation in T-F in Section 3. A typical office contains recording devices, such as mobile phones, MP3 players, PDAs, hearing aids, and computers all equipped with (largely unused) microphones. Consider a dedicated teleconferencing room, with an arbitrary number of inexpensive microphones. The source location, detected using this sensor array, is used to automatically identify the speaker or indicate the position of the speaker in the room. Our goal is to perform localisation and separation using multiple observations from arbitrarily placed sensors. The T-F domain lends itself to this problem as speech typically has increased WDO [8] and sparsity [7] in the T-F domain than in the time domain or frequency domain.

In an anechoic environment, a continuous time source signal $s_j(t)$ is attenuated and delayed as it propagates the direct path to sensor x_i. The attenuation and delay effect on the jth source received at the ith sensor is (a_{ji}, δ_{ji}), consequently $\hat{s}_{ji}(t) = a_{ji} s_j(t - \delta_{ji})$. I mixture signals are observed, $x_i(t)$, at physical locations x_i, where $h_{ji}(t)$ is the continuous time transfer function from source to sensor.

The source is constrained to lie at one of P arbitrarily placed grid points. We consider a synthetic scenario where the sensors are placed arbitrarily in a 2m × 2m × 2m room in Section 4 and observe the signals,

$$x_i(t) = \sum_{j=1}^{J} \hat{s}_{ji}(t) = h_{1i}(t) \star s_1(t) + h_{2i}(t) \star s_2(t) + \cdots + h_{Ji}(t) \star s_J(t). \quad (1)$$

2   Fractional Delay of Discrete Signals

A continuous time signal $s(t)$ is denoted by $s[n] = s(nT)$ in the discrete time domain where T is the sampling period and $n = 0, 1, 2, \ldots$. A continuous time signal delayed by $\delta \in \mathbb{R}$ seconds is denoted by $s(t - \delta)$. A discrete signal, $s[n]$, can be delayed by an integer, d, number of samples giving $s[n-d]$ or by rounding down, $\lfloor d \rfloor$, if $d = \delta/T$ is non-integer yielding $s[n - \lfloor d \rfloor]$. A source signal $s(t)$ is delayed by δ seconds propagating to a sensor x_i in an ideal anechoic teleconferencing room. Constraining the source physical locations such that the signals can only be delayed by an integer number of samples in the discrete time domain when propagating to each sensor limits the possible source locations. Alternatively, rounding down (denoted by $\lfloor d \rfloor$) introduces error. Although sources are constrained to lie on a grid, this grid can be refined and a space of interest more densely populated to locate arbitrarily placed sources. Non-integer sample delay,

$$s^{\delta}[n] = s(nT - \delta), \quad (2)$$

can be computed using sinc interpolation, given that the signal is bandlimited and sampled at a sufficiently high sampling rate,

$$s^{\delta}[n] = \sum_{n=-\infty}^{\infty} s[n]\mathrm{sinc}(nT - \delta). \quad (3)$$

In practice a finite length approximation of the sinc function leads to error in the estimate of $s^{\delta}[n]$. A non-integer sample delay of a bandlimited signal sampled above the Nyquist rate can also be determined by multiplying the Discrete Fourier Transform of $s[n]$, $\mathrm{DFT}\{s[n]\} = S[k] = \sum_{n=0}^{N-1} s[n]W^{kn}$ where $k = 0, 1, \ldots, N-1$ and $W = e^{-j\frac{2\pi}{N}}$, by a linear phase term W^{kd}. This corresponds to a circular shift of the signal by $d = \delta/T \in \mathbb{R}$ samples. We define the zero-padding function $\mathrm{ZP}\,(b, s[n], e)$ which appends b and e zeros to the beginning and end of the signal respectively. The inverse-pad function $\mathrm{IP}\,(b, s[n], e)$ removes b and e samples from the beginning and end of the signal. Zero-padding by $\lceil d \rceil$, where $\lceil \cdot \rceil$ is the ceiling function, taking the DFT, multiplying by the linear phase term, taking the IDFT and inverse-padding gives the desired result. Defining $\mathrm{IDFT}\{S[k]\} = \frac{1}{N}\sum_{k=0}^{N-1} S[k]W^{-kn}$ to be the inverse DFT, the frequency domain method in (Eqn. 4) is the benchmark method we shall use for the remainder of this work.

$$s^{\delta}[n] = \mathrm{IP}\left(\lceil d \rceil, \mathrm{IDFT}\{\mathrm{DFT}\{\mathrm{ZP}\,(\lceil d \rceil, s[n], 0)\}W^{kd}\}, \lfloor d \rfloor\right) \quad (4)$$

The contributions of this paper are the formulation of the T-F spatial signatures problem, the construction of a practical solution via Algorithm 1 and the synchronized STFT, and a demonstration of the efficacy of the technique by implementing a source localisation algorithm in the discrete T-F domain.

3   Time-Frequency Spatial Signatures

$S[k, m]$ is the Short-Time-Fourier-Transform (STFT) of $s[n]$,

$$\text{STFT}\{s[n]\} = S[k, m] = \sum_{n=mR}^{N-1-mR} s[n]w_a[n - mR]W^{k(n-mR)} \qquad (5)$$

positioned at sample mR where $w_a[n]$ is the analysis window function and R is the number of window hop-size samples. [k,m] are the discrete frequency and time indices respectively. N is the FFT size. The STFT is inverted using the synthesis window $w_s[n]$ and Over-Lap and Add (OLA) re-synthesis. A discrete source signal $s_j[n]$ is delayed by $d \in \mathbb{R}$ samples (neglecting attenuation effects) as it propagates to sensor x_i yielding $s_j^\delta[n]$.

Problem. *Construct a T-F spatial signature $H^\delta[k, m]$ so that,*

$$S_j^\delta[k, m] = S_j^0[k, m]H^\delta[k, m] \quad \text{for } j = 1, \ldots, J \text{ and } \forall |\delta|/T < \Delta. \qquad (6)$$

$S_j^\delta[k, m]$ and $S_j^0[k, m]$ are the synchronized STFT (sSTFT) of $s_j^\delta[n]$ and $s_j[n]$ respectively. We define the sSTFT so that the analysis window is centered on the same portion of the signal for the signal delayed by $|\delta|/T < \Delta$ samples, and the windowed segment of $s^\delta[n]$ is a circular shift of the windowed segment of $s[n]$. Δ is a user defined upper-bound on the range of delays under consideration specified by the maximum expected propagation distance.


The sSTFT is an alternative T-F analysis to the typical STFT method in (Eq. 5). It has the property that if a single source is active in T-F bin $[k, m]$, the T-F representation of the source at each x_i is calculated such that $w_a[n]$ is aligned with the received delayed source $s_j^\delta[n]$ $\forall |\delta|/T < \Delta$.

Definition 1. For a delay $|\delta|/T < N/4$ samples, we define the sSTFT of $s_j^\delta[n]$,

$$S_j^\delta[k, m] = \sum_{n=mR}^{N-1-mR} s_j^\delta[n]w_{az}^\delta[n - mR]W^{k(n-mR)}, \qquad (7)$$

where the window hop-size is $R = N/4$ samples. We define analysis and synthesis windows which are non-zero for $N/2$ samples and zero-padded by $N/2$ zeros,

$$w_{az}[n] = \text{ZP}\left(N/4, w_a[n], N/4\right), \quad w_{sz}[n] = \text{ZP}\left(N/4, w_s[n], N/4\right) \qquad (8)$$
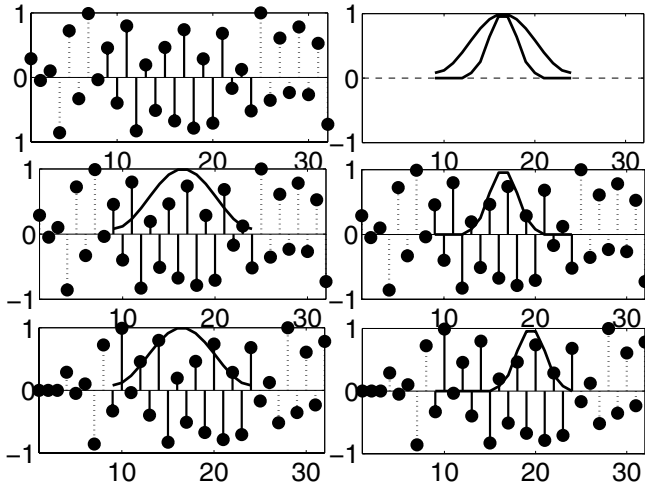
Fig. 1. The STFT (Eq.5) using $w_a[n]$ (which is not zero-padded) does not have the circular shift property in col. 1. The sSTFT (Eq.7) using $w_{az}[n]$ (which is zero-padded) allows the analysis to be synchronized with delays up to $N/4$ samples in col. 2.

to form a pair of windows of length N. Firstly, $w_{az}[n]$ allows the signal $s[n]$ to be shifted by $|\delta|/T < N/4$ samples such that the circular shift property still holds for each local windowed version of $s[n]$. Secondly, $w_{az}[n]$, is aligned with the source signal by delaying it by $d = \delta/T$ samples so that the same samples of $s[n]$ and $s^\delta[n]$ are multiplied by the same samples of $w_{az}^\delta[n]$. We define,

$$w_{az}^\delta[n] = \text{IDFT}\{\text{DFT}\{w_{az}\}[n]W^{kd}\} \text{ and } w_{sz}^\delta[n] = \text{IDFT}\{\text{DFT}\{w_{sz}\}W^{kd}\} \ (9)$$

Fig. 1 illustrates the difference between the STFT and the sSTFT using a signal, $s[n]$, of length 32 samples in column 1 row 1. We index each subplot using the figure number and then row and column index in parenthesis, e.g. Fig. 1(1,1). Samples 9-24 are analyzed using a 16-point FFT and 16-point Hamming window. The Hamming window is superimposed on the signal in Fig. 1(2,1). Samples 9-24 supported by the window, $w_a[n]$, have linear stalks. Samples 1-8 and 25-32 are denoted by dotted stalks. $s[n]$ is delayed by 3 samples in Fig. 1(3,1). Different portions of the signal $s[n]$ and its delayed counter-part are analyzed in Fig. 1(2,1) and Fig. 1(3,1) using the STFT. Samples common to both windowed signals are scaled differently due to the shifted version of the signal in Fig. 1(2,1) and Fig. 1(3,1) relative to the window. The Fourier transform circular shift property does not hold for the local windowed signal Fig. 1(2,1) and its delayed windowed version Fig. 1(3,1). Analyzing $s[n]$ as specified by (Eq.7) using the zero-padded window $w_{az}^0[n]$ (Eq. 8) with $\delta = 0$ in Fig. 1(2,2) and analyzing $s^\delta[n]$ with $w_{az}^\delta[n]$ (Eq. 9) in Fig. 1(3,2) preserves the Fourier transform circular shift property. $w_{az}^0[n]$ is pre-zero-padded and post-zero-padded by $N/4$ samples. $w_{az}^0[n]$ is non-zero for $N/2$ samples. A circular shift of $w_{az}^0[n]$ by $d < N/4$ samples yielding,

$w^\delta[n]$, times $s^\delta[n]$ results in a windowed signal which is a circular shift of the original windowed signal in Fig.1(2,2). This facilitates the construction of T-F spatial signatures using the Fourier transform circular shift property. Due to the analysis window alignment of the sSTFT, the re-synthesis is not dependent on the analyzed signal $s[n]$. The synthesis window $w_{sz}^\delta[n]$ corresponding to $w_{az}^\delta[n]$ is used to re-synthesize any signal accurately using OLA. Synchronized T-F spatial signatures independent of the signal to be delayed and unbiased by wrap-around and windowing can be estimated using Alg. 1 and sSTFT.

Algorithm 1. Synchronized T-F Spatial Signatures

$s[n]$ delayed by $\delta/T < \frac{N}{4}$ samples, resulting in $s^\delta[n]$, is calculated by element-wise multiplication with the matrix $H_\delta[k,m]$ in discrete T-F given the $s[n]$ is analyzed using (Eq. 7) yielding $S^0[k,m]$ using the analysis window w_{az}^0 in (Eq. 8), e.g. $S^\delta[k,m] = S^0[k,m]H_\delta[k,m]$. Using the sample indices q, g

$$H_\delta[k,m] = \frac{\sum_{q=\frac{mN}{4}}^{Q-1-\frac{mN}{4}} \operatorname{sinc}(qT - \frac{mN}{4} - \delta)w_{az}^\delta[q - \frac{mN}{4}]W^{k(q-\frac{mN}{4})}}{\sum_{g=\frac{mN}{4}}^{G-1-\frac{mN}{4}} \operatorname{sinc}(gT - \frac{mN}{4})w_{az}[g - \frac{mN}{4}]W^{k(g-\frac{mN}{4})}}. \quad (10)$$

We form $\boldsymbol{x}[k,m] = [X_1[k,m], \ldots, X_i[k,m], \ldots, X_I[k,m]]^T \in \mathbb{C}^{I\times 1}$ a vector of the observations at each sensor for each T-F point $[k,m]$ using sSTFT. We construct a T-F spatial signatures matrix for each $[k,m]$, $\boldsymbol{D}[k,m] \in \mathbb{C}^{I\times P}$, using Alg. 1 and attenuating each $H_{pi}[k,m]$ using a_{ji}. $\boldsymbol{D}[k,m]$ gives the transfer function, $H_{pi}[k,m]$, for every location, p, in the grid relative to each sensor, x_i, for $[k,m]$.

$$\boldsymbol{D}[k,m] = \begin{matrix} \uparrow \\ \text{Sensor i} \\ \downarrow \end{matrix} \overset{\longleftarrow\text{Location p}\longrightarrow}{\begin{pmatrix} H_{11}[k,m] & \ldots & H_{p1}[k,m] & \ldots & H_{P1}[k,m] \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ H_{1i}[k,m] & \ldots & H_{pi}[k,m] & \ldots & H_{Pi}[k,m] \\ \vdots & \ldots & \vdots & \ldots & \vdots \\ H_{1I}[k,m] & \ldots & H_{pI}[k,m] & \ldots & H_{PI}[k,m] \end{pmatrix}} \quad (11)$$

The J sources $[s_1, \ldots s_j \ldots, s_J]$ are constrained to lie on a subset of the P grid points. We locate a source by estimating the vector $\boldsymbol{c}[k,m] \in \mathbb{C}^{P\times 1}$ which explains the sensor observations in the most parsimonious manner given $\boldsymbol{D}[k,m]$.

$$\boldsymbol{x}[k,m] = \boldsymbol{D}[k,m]\boldsymbol{c}[k,m]. \quad (12)$$

4   Source Localisation and Separation Simulations

We solve each subsystem $[k,m]$ in (Eq. 12) independently. This approach lends itself to real-time parallel implementation on dedicated processors. Assuming WDO in T-F, a single source is active in $[k,m]$. The element of the estimated

$c[k, m]$ with the most energy, for example $p = 25$ in Fig. 2(b), indicates the position of the source. For example in Fig. 2(b) a mixed $L_1 + \lambda L_2^2$ objective function prejudiced in favor of the sparsity constraint, L_1-norm yields a solution with significant energy at one grid location $p = 25$ in comparison with a solution with an emphasis on the L_2^2-norm. The weight λ is a data dependent heuristic. Masking the mixture at x_i, $X_i[k, m]$, using $c_p[k, m]$ as an indicator for each $[k, m]$ for each p of interest, separates the source at p. We consider a mixed objective

$$\min_{\mathbf{c}} \frac{1}{2} \underbrace{\sum_{i=1}^{P} |\mathbf{c}_p|}_{E_1} + \lambda \underbrace{||\mathbf{Dc} - \mathbf{x}||_2^2}_{E_2}, \text{ and define } \tilde{\boldsymbol{D}} = \begin{pmatrix} \text{Re}\{\hat{\boldsymbol{D}}\} & -\text{Im}\{\hat{\boldsymbol{D}}\} \\ \text{Im}\{\hat{\boldsymbol{D}}\} & \text{Re}\{\hat{\boldsymbol{D}}\} \end{pmatrix}. \quad (13)$$

We use an iteratively re-weighted least squares approach in the spirit of [3],

$$\min_{\mathbf{c}} \hat{E}(\mathbf{c}) = \hat{E}_1 + \lambda E_2, \ \hat{E}_1 \equiv \frac{1}{2} \sum_{p=1}^{P} \alpha_p^k |c_p^k|^2, \ \boldsymbol{\alpha} \in \mathbb{R}^{P \times 1} \text{ and } \boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\alpha}). \quad (14)$$

Using a modified objective (Eq. 14) we solve $\frac{\partial \hat{E}(\mathbf{c})}{\partial \mathbf{c}^k} = \mathbf{0}$ for $\mathbf{c}^k$, where k denotes the iteration index and $(\cdot)^H$ denotes the conjugate transpose operation.

$$\left(\boldsymbol{\Lambda} + \lambda \tilde{\boldsymbol{D}}^T \tilde{\boldsymbol{D}}\right) \begin{pmatrix} \text{Re}\{\mathbf{c}\} \\ \text{Im}\{\mathbf{c}\} \end{pmatrix} = \lambda \tilde{\boldsymbol{D}}^T \begin{pmatrix} \text{Re}\{\mathbf{x}\} \\ \text{Im}\{\mathbf{x}\} \end{pmatrix} = (\boldsymbol{\Lambda} + \lambda \boldsymbol{D}^H \boldsymbol{D})\mathbf{c} = \lambda \boldsymbol{D}^H \mathbf{x} (15)$$

Solving (15) iteratively and setting $\alpha_p^{k+1} = 1/|\mathbf{c}_p^k|$ yields a fixed point solution.

In the first experiment we show that construction of accurate T-F spatial signatures, $H_{ji}(k, m)$, is crucial for a sparse solution. Wrap-around and windowing effects inherent in unsynchronized analysis using STFT of (Eq. 5) introduce error and occlude the true solution. We estimate $S^\delta[k, m]$ using (Eq. 7) to analyze $s[n]$. We delay $s[n]$ in T-F using Alg. 1 and re-synthesize using OLA with $w_{sz}^\delta[n]$ and hop-size of $N/4$ (method 1). We compare method 1 with a STFT approach (method 2) e.g., $s[n - d] \approx \text{ISTFT}(\text{STFT}(s[n], w_a, N, N/2)W^{kd}, w_s, N, N/2)$. Method 2 performs analysis of $s[n]$ using the STFT (Eq. 5) with a window $w_a[n]$ of N samples and hop-size $N/2$ (as in Fig. 1 column 1). A linear phase term W^{kd} shifts each frame of the signal in T-F. The resulting signal is re-synthesized using the inverse STFT with a synthesis window $w_s[n]$ of length N and overlap $N/2$ using OLA re-synthesis. SNR is defined as $20 * \log(||s_B^\delta[n]||_2)/||s_B^\delta[n] - s_{est}^\delta[n]||_2)$, where $s_{est}^\delta[n]$ is the estimate of the delayed signal and the benchmark, $s_B^\delta[n]$, is computed using (Eq. 4). We analyze speech from the TIMIT database, sampled at 16kHz, with the parameters $N = 2048$ and $R = N/4$ and $R = N/2$ for method 1 and 2 respectively. Fig. 2(a) illustrates the window wrap-around and windowing effects on the delayed source using the sSTFT (method 1) and STFT (method 2). The sSTFT method exhibits sub-sample dips in SNR due to the numerical instability of the truncated delayed sinc function yet degrades gracefully as a function of delay and achieves an estimate > 40dB when delayed by 511 samples. The SNR of method 2 in Fig. 2(a) decreases rapidly as a function of delay in samples due to windowing effects. These inaccurate T-F spatial signatures are unsuitable for localization via sparse representations.
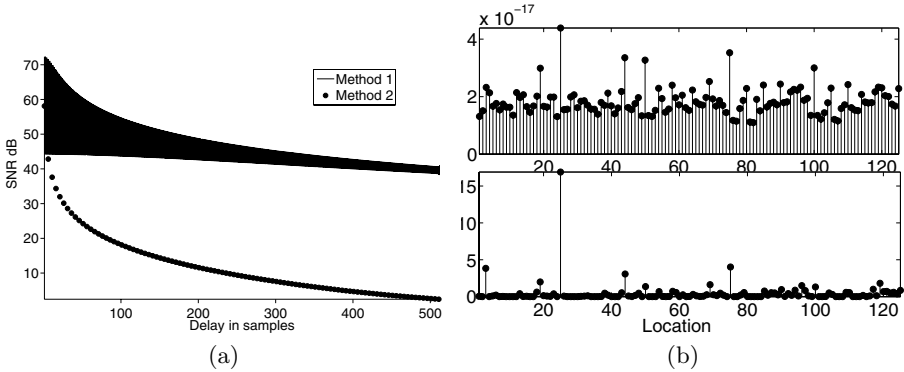
Fig. 2. Fig. 2(a) illustrates the effects of window misalignment and wrap-around using the STFT compared to sSTFT analysis. Row 2 Fig. 2(b) shows that a sparse solution reveals the source location (position 25) compared to a more dense solution (Row 1).

Table 1. Average Localisation and Separation Performance

Source		PSR			WDO			WDO (0dB)			Loc %		
	p	3	5	6	3	5	6	3	5	6	3	5	6
s_1	25	0.89	0.60	0.86	0.88	0.60	0.83	0.92	0.87	0.85	25.0	8.0	6.3
s_2	97	0.90	0.69	0.67	0.89	0.68	0.65	0.92	0.82	0.75	47.0	14.0	11.0
s_3	105	0.80	0.58	0.67	0.79	0.57	0.65	0.87	0.79	0.69	17.0	5.0	5.0
s_4	5	−	0.86	0.70	−	0.85	0.68	−	0.79	0.73	−	24.0	3.0
s_5	62	−	0.94	0.90	−	0.92	0.89	−	0.90	0.86	−	43.0	54.0
s_6	112	−	−	0.77	−	−	0.74	−	−	0.78	−	−	18.0

In the second experiment we perform source localisation, using the sSTFT approach, in a synthetic 2m × 2m × 2m room with grid points every 50cm. We tune (Eq.13) over a range of λ. A sparse solution decreases the Euclidean distance between our signal estimate and the original signal as the trade-off between the reconstruction penalty E_2 and sparsity penalty E_1 is adjusted in (Eq.13). Mixtures of 1 to 6 speakers are generated from the TIMIT database by assigning them randomly to grid-points $\{25, 97, 105, 5, 62, 112\}$. We perform multiple experiments with different initial conditions to test the accuracy of the localisation experiment. We use 10 microphones in each experiment. We choose candidate locations—based on the signal power at each location—by analyzing a subset of the T-F points with the optimal analysis window for each grid point p, e.g. w_{az}^p. Localisation and separation is performed using these optimal windows. The mean results for the $3, 5, 6$ speaker cases are tabulated in Table 1. The T-F mask metrics PSR and WDO introduced in [8] are used for comparison with the 0dB ideal mask defined in [8] to gauge separation performance. Our technique achieves performance comparable with the 0dB mask. A significant percentage of the signal energy is located at the correct source positions.

With regard to existing T-F separation algorithms, relying on relative measurements between multiple observations of the mixture could introduce bias if

the analysis is unsynchronized using STFT (Eq. 5). The DUET algorithm [8] estimates relative attenuations and delays between two closely spaced sensors using the ratios of the STFT of the observations at sensor x_1 and x_2. A bias has been noted in attenuation and delay estimates in [8] where the delay between 2 sensors ranges from -5 to 5 samples and attenuation from -0.15 to 0.15. It is clear from Fig. 1 that using (Eq. 5) to compute $X_1[k, m]$ and $X_2[k, m]$ and taking the ratio of these two channels in T-F, introduces error as the mixture at x_2 is shifted relative to the window and the mixture $x_1[n]$. The circular shift property is not satisfied. The delay estimate is only approximate $\delta[k, m] \approx -\frac{1}{2\pi k} \angle \frac{X_2[k,m]}{X_1[k,m]}$. The samples of the windowed signal $x_1[n]$ common to the windowed $x_2[n]$ are scaled differently due to window misalignment for analysis and so the relative attenuation $\alpha[k, m] \approx |\frac{X_2[k,m]}{X_1[k,m]}|$ is approximate. To conclude, the synchronized STFT method combined with the algorithm for T-F spatial signature construction facilitates the implementation of a global signal shift as a circular shift with in the support of each frame of the analyzed signal. We have presented a sparse source localisation technique using synthetic experiments to motivate this approach. Incorporating sSTFT into a existing techniques, such as DUET and its extensions will lead to gains in accuracy of the mixing parameter estimation. We will extend this approach to the echoic case in future work.

References

1. Malioutov, D.M., Çetin, M., Willsky, A.S.: A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Transactions on Signal Processing 53, 3010–3022 (2005)
2. Model, D., Zibulevsky, M.: Signal Reconstruction in Sensor Arrays using Sparse Representations. Signal Processing 86, 624–638 (2006)
3. Gorodnitsky, I.F., Rao, B.D.: Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. IEEE Transactions on Signal Processing 45, 600–616 (1997)
4. Zibulevsky, M., Pearlmutter, B.A.: Blind source separation by sparse decomposition in a signal dictionary. Neural Computation 13, 863–882 (2001)
5. Reddy, A.M., Raj, B.: Soft Mask Methods for Single-Channel Speaker Separation. IEEE Transactions on Audio, Speech, and Language Processing 15, 1766–1776 (2007)
6. de Frein, R., Drakakis, K., Rickard, S.T.: Portfolio diversification using subspace factorizations. In: Proceedings of the Annual Conference on Information Sciences and Systems, pp. 1075–1080 (2008)
7. Rickard, S.T.: Sparse sources are separated sources. In: Proceedings of the 16th Annual European Signal Processing Conference, Florence, Italy (2006)
8. Yilmaz, O., Rickard, S.T.: Blind separation of speech mixtures via time-frequency masking. IEEE Transactions on Signal Processing 52, 1830–1847 (2004)
9. Hoyer, P.O.: Non-negative sparse coding. In: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, pp. 557–565 (2002)