ROBUST SPATIAL ESTIMATION OF THE SIGNAL-TO-INTERFERENCE RATIO (SIR) FOR NON-STATIONARY MIXTURES

*W. Herbordt, T. Trini, and W. Kellermann**

Telecommunications Laboratory, University Erlangen-Nuremberg
Cauerstraße 7, 91058 Erlangen, Germany
{herbordt,trini,wk}@LNT.de

ABSTRACT

The proposed paper deals with the robust estimation of frequency-dependent signal-to-interference ratios (SIRs) and spatial coherence functions in the DFT domain for non-stationary wideband signals. It can be used to realize, e.g., optimum adaptive beamformers or can be used as a confidence measure in combination with automatic speech recognition in noisy environments.

1. INTRODUCTION

Multi-sensor speech enhancement using, e.g., multichannel noise-reduction or optimum adaptive beamforming, requires separate estimates of cross-power spectral densities (CPSDs) of the desired signal and of the interference, and frequency bin-wise SIRs in order to assure optimum quality of the output signal [1]. A method for estimating the power spectral density (PSD) of -relative to the desired signal- slowly time-varying noise is proposed in, e.g., [2]. In this work, we present a new method for estimating (a) the spatial coherence functions of the desired signal and of the interference between the sensors and (b) the SIR at the sensors for highly non-stationary PSDs. Spatial coherence functions should be slowly time-varying relative to the PSDs. In Section 2, we describe the SIR estimation and analyze its statistical characteristics. Experimental results illustrate the robustness of our approach in Section 3.

2. SIR ESTIMATION USING SPATIAL COHERENCE

The estimation of the SIR is a two-step procedure. A biased estimate of the SIR at the sensors can be determined using the spatial selectivity of a fixed beamformer (Section 2.1). The bias is due to the fact that the beamformer does not separate the desired signal and the interference perfectly. It depends on the spatial coherence of the desired signal and of the interference at the sensors. For correcting the bias, it is thus necessary to estimate the spatial coherence w.r.t. the desired signal and w.r.t. interference separately (Section 2.2), which requires a discrimination between the sources (Section 2.3). The statistical analysis of this estimate of the SIR shows that its variance normalized to the SIR increases with a power of 2 of the true SIR at the sensors with increasing and decreasing SIRs(Section 2.4). This reduces the accuracy of the estimate of the SIR especially at low and high SIRs and requires robustness improvements (Section 2.5).

The estimation is carried out in the discrete Fourier transform (DFT) domain with DFT length N. The discrete frequency index is denoted by μ. The block time index is k. Lower case and upper case bold font represent vector and matrix quantities, respectively. *, T, and H stand for complex conjugation, matrix or vector transposition, and conjugate transposition, respectively.

2.1. Biased spatial estimate of the SIR

For bin-wise spatial estimation of the SIR [1], the temporally windowed DFT-domain sensor signals $\mathbf{x}(k,\mu) = (X_0(k,\mu),\, X_1(k,\mu),\, \ldots,\, X_{M-1}(k,\mu))^T$ of a microphone array with M microphones are used. The array is steered to the position of the desired source.

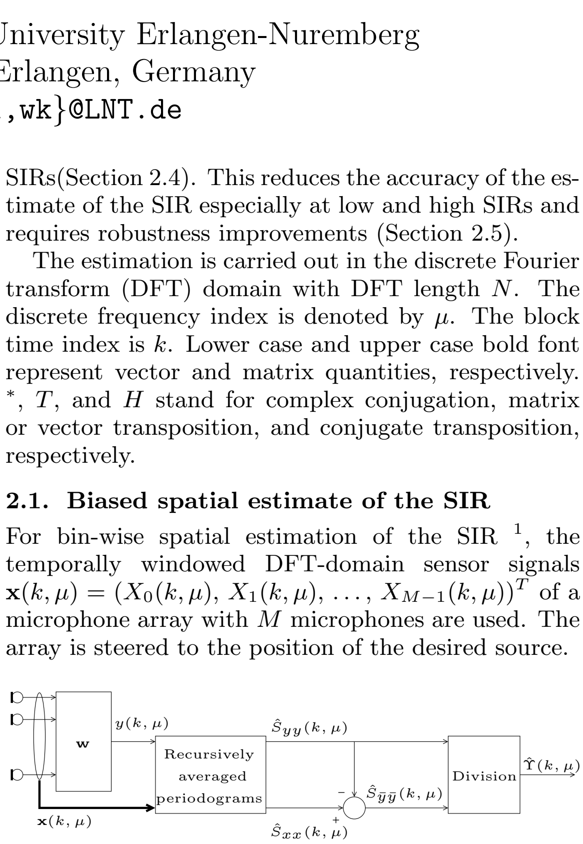

Figure 1: Biased spatial SIR estimate.

According to Figure 1, the sensor signals are weighted and summed up by a fixed uniformly weighted delay&sum beamformer $\mathbf{w}$, yielding the beam-formed signal $Y(k,\mu) = \mathbf{w}^H\mathbf{x}(k,\mu)$. From the sensor signals $\mathbf{x}(k,\mu)$ and from the beamformer output signal $Y(k,\mu)$, recursively averaged windowed periodograms are calculated using an exponential forgetting factor $0 < \beta_S < 1$. From $Y(k,\mu)$, an estimate of the PSD of the beamformer output signal is obtained as $\hat{S}_{yy}(k,\mu) = \mathbf{w}^H\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)\mathbf{w}$, where $\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)$ is the estimated CPSD matrix of the sensor signals. From the sensor signals, we calculate the average PSD $\hat{S}_{xx}(k,\mu) = 1/M \cdot \sum_{m=0}^{M-1} \hat{S}_{x_m x_m}(k,\mu)$. Due to the spatial selectivity of the beamformer, $\hat{S}_{yy}(k,\mu)$ is an estimate of the PSD of the desired signal. Forming the complementary signal by subtracting this estimate of the desired signal from the PSD of the sensor signals $\hat{S}_{xx}(k,\mu)$ thus yields an estimate of the PSD of the interference, i.e.,

$$\hat{S}_{\bar{y}\bar{y}}(k,\mu) = \hat{S}_{xx}(k,\mu) - \hat{S}_{yy}(k,\mu)\,. \qquad (1)$$

[1]We assume that the frequency bins are mutually uncorrelated, which is asymptotically ($N \to \infty$) true. This allows to define the average sensor SIR in the DFT domain as a function of frequency.

*This work was supported by a grant from Intel China Research Center, Beijing, China.

The biased estimate $\hat{\Upsilon}(k,\mu)$ of the SIR at the sensors is finally given by the ratio[2]

$$\hat{\Upsilon}(k,\mu) = \frac{\hat{S}_{yy}(k,\mu)}{\hat{S}_{\bar{y}\bar{y}}(k,\mu)}\,. \qquad (2)$$

The PSDs $\hat{S}_{yy}(k,\mu)$, $\hat{S}_{\bar{y}\bar{y}}(k,\mu)$ in (2) depend on the PSDs of the desired signal and of the interference at the beamformer output, respectively. Assuming spatially homogeneous wave fields, the PSDs w.r.t. the desired signal and w.r.t. the interference at the beamformer output can be written as quadratic forms as $\hat{S}_{dd}(k,\mu)\mathbf{w}^H\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)\mathbf{w}$ and $\hat{S}_{nn}(k,\mu)\mathbf{w}^H\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)\mathbf{w}$. $\hat{S}_{dd}(k,\mu)$ and $\hat{S}_{nn}(k,\mu)$ are the PSDs of the desired signal and of the interference at the sensors, respectively. $\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)$ and $\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)$ are the spatial coherence matrices w.r.t. the desired signal and w.r.t. interference at the sensors, respectively. This allows to write $\hat{\Upsilon}(k,\mu)$ after some rearrangements as follows:

$$\hat{\Upsilon}(k,\mu) = \qquad (3)$$
$$\frac{\widehat{SIR}(k,\mu)F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)) + F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))}{\widehat{SIR}(k,\mu)[1-F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))] + [1-F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))]}\,,$$

where

$$\widehat{SIR}(k,\mu) = \hat{S}_{dd}(k,\mu)/\hat{S}_{nn}(k,\mu)\,, \qquad (4)$$
$$F(\hat{\mathbf{\Gamma}}_*(k,\mu)) = \mathbf{w}^H\hat{\mathbf{\Gamma}}_*(k,\mu)\mathbf{w}\,, \qquad (5)$$

with $\hat{\mathbf{\Gamma}}_*(k,\mu)\,, * \in \{\mathbf{dd},\, \mathbf{nn}\}$. $\hat{\mathbf{\Gamma}}_*(k,\mu)\,, * \in \{\mathbf{dd},\, \mathbf{nn}\}$ corresponds to the PSD of the desired signal and of the interference at the beamformer output normalized by $\hat{S}_{dd}(k,\mu)$ and by $\hat{S}_{nn}(k,\mu)$, respectively. $\widehat{SIR}(k,\mu)$ is the unbiased estimate of the bin-wise SIR, $SIR(k,\mu)$, at the sensors. Equation (3) shows that $\hat{\Upsilon}(k,\mu)$ is an estimate of the bin-wise SIR at the sensors which is biased by functions of the spatial coherence matrices w.r.t. the desired signal and w.r.t. interference. (3) requires $\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu) \neq \hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)$ so that the dependency of $\hat{\Upsilon}(k,\mu)$ on $\widehat{SIR}(k,\mu)$ does not cancel out.

2.2. Bias correction using spatial coherence

Solving (3) for $\widehat{SIR}(k,\mu)$ leads to

$$\widehat{SIR}(k,\mu) = \frac{F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))\cdot[\hat{\Upsilon}(k,\mu)+1] - \hat{\Upsilon}(k,\mu)}{\hat{\Upsilon}(k,\mu) - F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))\cdot[\hat{\Upsilon}(k,\mu)+1]}\,. \qquad (6)$$

We see that an unbiased estimate of the SIR can be calculated with (6) for known *but arbitrary* $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)) \neq F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$. We thus have reduced the problem of the estimation of the SIR to the estimation of spatial coherence matrices. For non-stationary mixtures, the estimation of spatial coherence functions is advantageous compared to a direct estimation of the SIR, since, generally, PSDs of speech and audio signals are strongly time-varying relative to slowly time-varying spatial coherence functions. The PSDs often decay to zero in individual DFT bins, so that $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ can be estimated during inactivity of the interference and of the desired signal in individual DFT bins, respectively. This requires a double-talk detection mechanism with distinction between presence of only the desired signal and only interference.

2.3. Spatial activity detection

For detecting presence of only desired signal and presence of only interference, we use the ratio $\hat{\Upsilon}(k,\mu)$. Although $\hat{\Upsilon}(k,\mu)$ is a biased estimate of the SIR, $\hat{\Upsilon}(k,\mu)$ increases (decreases) with increasing (decreasing) SIR. Thus, if $\hat{\Upsilon}(k,\mu)$ is maximum (minimum), it is very likely that only desired signal (interference) is present. By tracking the maxima (minima) of $\hat{\Upsilon}(k,\mu)$ over time with methods presented in, e.g., [2], lower and upper thresholds $\hat{\Upsilon}_{\mathrm{l}}(k,\mu)$ and $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$ can be determined to distinguish between presence of desired signal alone and interference alone. We introduce the PSD of the beamformer output signal normalized to the average PSD of the sensor signals as

$$\hat{\Psi}(k,\mu) = \frac{\hat{S}_{yy}(k,\mu)}{\hat{S}_{xx}(k,\mu)} = \frac{\mathbf{w}^H\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)\mathbf{w}}{\hat{S}_{xx}(k,\mu)}\,. \qquad (7)$$

$\hat{\Psi}(k,\mu)$ is equal to $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ for presence of only the desired signal and only interference, respectively. We thus can determine $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ according to

$$F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)) = \hat{\Psi}(k,\mu)\Big|_{\hat{\Upsilon}(k,\mu)\geq\hat{\Upsilon}_{\mathrm{d}}(k,\mu)}\,, \qquad (8a)$$
$$F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)) = \hat{\Psi}(k,\mu)\Big|_{\hat{\Upsilon}(k,\mu)\leq\hat{\Upsilon}_{\mathrm{n}}(k,\mu)}\,. \qquad (8b)$$

Usage of (8a), (8b) in (6) finally yields the estimate $\widehat{SIR}(k,\mu)$. This procedure can also be used for estimating the spatial coherence matrices $\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)$ and $\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)$ explicitely.

2.4. Statistical evaluation

For analyzing the variance $\sigma^2_{\widehat{SIR}}(k,\mu)$ of the proposed SIR estimator, we assume wide-sense stationary discrete-time sensor signals $x_m(n)$. Assuming furthermore that $\hat{\Upsilon}(k,\mu) := \Upsilon(\mu)$ is known, then, $\widehat{SIR}(k,\mu)$ is the function of two random variables $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ with variances $\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}})}(k,\mu)$ and $\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}})}(k,\mu)$, respectively. With statistical independence of the random variables $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$, a first-order approximation of the variance $\sigma^2_{\widehat{SIR}}(k,\mu)$ can thus be derived as

$$\sigma^2_{\widehat{SIR}}(k,\mu) \simeq \qquad (9)$$
$$\left(\frac{SIR^2(k,\mu)+SIR(k,\mu)}{F(\mathbf{\Gamma}_{\mathbf{nn}}(\mu)) - F(\mathbf{\Gamma}_{\mathbf{dd}}(\mu))}\right)^2\cdot\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}})}(k,\mu)$$
$$+\left(\frac{SIR(k,\mu)+1}{F(\mathbf{\Gamma}_{\mathbf{nn}}(\mu)) - F(\mathbf{\Gamma}_{\mathbf{dd}}(\mu))}\right)^2\cdot\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}})}(k,\mu)\,.$$

Dividing both sides by $SIR^2(k,\mu)$, it can be seen that the variance of the estimate of the SIR normalized to $SIR^2(k,\mu)$ increases with the power of 2 of the true sensor SIR with increasing and decreasing SIRs and

[2]In [3], the ratio of the variance of the output signal of a fixed beamformer and of a complementary fixed beamformer is introduced for controlling the adaptation of a time-domain adaptive beamformer.

increases linearly with the variance of the coherence estimates.

Next, we have to compute the variances $\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}})}(k,\mu)$ and $\sigma^2_{F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}})}(k,\mu)$. For that, we exclude double-talk of the desired signal and interference, which allows to neglect the activity detector. $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ can then be estimated continuously over time and are equivalent to $\hat{\Psi}(k,\mu)$ (with different PSDs and different spatial coherence matrices). The statistical properties of $\hat{\mathbf{\Gamma}}_*(k,\mu)\,, * \in \{\mathbf{dd},\, \mathbf{nn}\}$ and $\hat{\Psi}(k,\mu)$ are thus equivalent. Since a distinction between the desired signal and interference is not necessary for the following derivations, we give the equations as a function of the discrete-time sensor signals $x_m(n)$ and use the statistical characteristics for the desired signal and for the interference if necessary.

We assume a deterministic weight vector $\mathbf{w}$, and we model the elements $\hat{S}_{x_{m_1}x_{m_2}}(k,\mu)$ of the CPSD matrix $\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)$ as random variables. For specifying the covariance between two CPSDs $\hat{S}_{x_{m_1}x_{m_2}}(k,\mu)$ and $\hat{S}_{x_{m_3}x_{m_4}}(k,\mu)$, the discrete-time sensor signals $x_m(n)$ should be stationary random processes with asymptotic $(n_1 - n_2 \to \pm\infty)$ mutual statistical independence. For $N \to \infty$ and $\mu \neq \{0, \pm N/2\}$, the covariance between the recursively averaged second-order windowed periodograms $\hat{S}_{x_{m_1}x_{m_2}}(k,\mu)$ and $\hat{S}_{x_{m_3}x_{m_4}}(k,\mu)$ can then be derived with $B(k) = (\beta_S^k - 1)/(\beta_S - 1)$ as

$$C_{\hat{S}_{x_{m_1}x_{m_2}}\hat{S}_{x_{m_3}x_{m_4}}}(k,\mu) = \frac{1}{B(k)} S_{x_{m_1}x_{m_3}}(\mu) S_{x_{m_2}x_{m_4}}(\mu)\,. \tag{10}$$

$C_{\hat{S}_{x_{m_1}x_{m_2}}\hat{S}_{x_{m_3}x_{m_4}}}(k,\mu)$ corresponds to the covariance of the second-order windowed periodograms normalized by $B(k)$. The normalization factor $B(k)$ can be interpreted as the effective memory for averaging the second-order windowed periodograms [4].

A first-order approximation of the variance $\sigma^2_F(k,\mu)$ of $\hat{\mathbf{\Gamma}}_*(k,\mu)\,, * \in \{\mathbf{dd},\, \mathbf{nn}\}$ is found by interpreting $\hat{\Psi}(k,\mu)$ after (7) as a function of the complex multivariate random variable $\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)$. By stacking the columns of the $M \times M$ matrix $\hat{\mathbf{S}}_{\mathbf{xx}}(k,\mu)$ into an $M^2 \times 1$ vector $\hat{\mathbf{s}}_{\mathbf{xx}}(k,\mu)$, we find

$$\sigma^2_F(k,\mu) \simeq \left(\frac{\partial \hat{\Psi}(k,\mu)}{\partial \hat{\mathbf{s}}_{\mathbf{xx}}(k,\mu)}\right)^T \mathbf{C}_{\hat{\mathbf{s}}_{\mathbf{xx}}\hat{\mathbf{s}}_{\mathbf{xx}}}(k,\mu) \left(\frac{\partial \hat{\Psi}(k,\mu)}{\partial \hat{\mathbf{s}}_{\mathbf{xx}}(k,\mu)}\right)^*, \tag{11}$$

where the $M^2 \times M^2$ matrix $\mathbf{C}_{\hat{\mathbf{s}}_{\mathbf{xx}}\hat{\mathbf{s}}_{\mathbf{xx}}}(k,\mu)$ is the covariance matrix w.r.t. the complex multivariate random variable $\hat{\mathbf{s}}_{\mathbf{xx}}(k,\mu)$ with elements (10).

Finally, the variance $\sigma^2_{\widehat{SIR}}(k,\mu)$ is obtained by using (11) in (9) for the given wave-field characteristics of the desired signal and of the interference. An example of the normalized variance $\sigma^2_{\widehat{SIR}}(k,\mu)/SIR^2(k,\mu)$ is given in Figure 2 for an office room with reverberation time $T_{60} = 300$ ms and for the setup described in Section 3 with $\beta_S \in \{0.3, 0.7\}$ and with the desired signal arriving from $\theta_\mathrm{d} \in \{\pi/2, 17\pi/36\}$ with broadside steering direction $\pi/2$. In Figure 2 (a) and (b), the normalized variance is shown over the true sensor SIR for $f \simeq 1.5$ kHz and over frequency for different sensor SIRs ($\theta_\mathrm{d} = 17\pi/36$, $\beta_S = 0.7$, $k \to \infty$), respectively. Most importantly, the variance has a minimum around $SIR(\mu) = 0$ dB, where the SIR estimation is most crucial for many applications [1]. $\sigma^2_{\widehat{SIR}}(k,\mu)$ is only little influenced by the exponential forgetting factor and by steering mismatch, as expected from (9), (10), and (11). This is especially important for non-stationary signals, where β_S is small for tracking changes of the PSD, and where mismatch of the steering direction cannot be prevented.

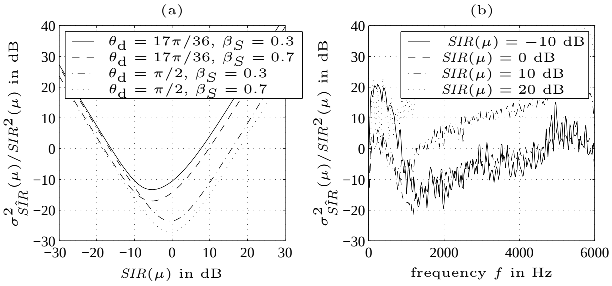

Figure 2: Statistical evaluation.

2.5. Robustness improvement

Section 2.4 showed that the variance of the SIR estimate increases with decreasing and increasing SIR. For improving the accuracy of the SIR estimation, we provide several measures. First, we have to prevent $\widehat{SIR}(k,\mu) < 0$. From (6), we see that $\widehat{SIR}(k,\mu) < 0$, if one of the following two conditions is fulfilled:

$$\hat{\Upsilon}(k,\mu) \;>\; \frac{F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))}{1 - F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))}\,, \tag{12}$$

$$\hat{\Upsilon}(k,\mu) \;<\; \frac{F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))}{1 - F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))}\,, \tag{13}$$

where we assume that the suppression of the interference of the beamformer is greater than the cancellation of the desired signal. For forcing $\widehat{SIR}(k,\mu) > 0$, we solve (12) and (13) for $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$, respectively, and we set the equality sign. We obtain

$$F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)) \;=\; \frac{\hat{\Upsilon}(k,\mu)}{\hat{\Upsilon}(k,\mu)+1}\,, \tag{14}$$

$$F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)) \;=\; \frac{\hat{\Upsilon}(k,\mu)}{\hat{\Upsilon}(k,\mu)+1}\,, \tag{15}$$

respectively. Equation (6) is then modified by replacing the present values of $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ with the right sides of (14) and (15) whenever the conditions (12) and (13) are fulfilled, respectively.

Second, the determination of the spatial coherence functions while $\hat{\Upsilon}(k,\mu)$ is maximum or minimum may lead to over- or underestimation of $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$, respectively. This effect can be reduced by applying a median filter of low order $(3 \ldots 5)$ to $\hat{\Upsilon}(k,\mu)$, which gives $\bar{\Upsilon}(k,\mu)$, and by determining the thresholds $\hat{\Upsilon}_\mathrm{d}(k,\mu)$ and $\hat{\Upsilon}_\mathrm{n}(k,\mu)$ from $\bar{\Upsilon}(k,\mu)$.

Third, if the conditions

$$|F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k-1,\mu)) - \hat{\Psi}(k,\mu)| \;\le\; \Delta_F\,, \tag{16}$$

$$|F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k-1,\mu)) - \hat{\Psi}(k,\mu)| \;\le\; \Delta_F\,, \tag{17}$$

with typically $0.01 < \Delta_F < 0.1$, are fulfilled, it is very likely that only desired signal or only interference is active although the activity detector detects double-talk. In order to allow tracking of such small variations of $\hat{\Psi}(k,\mu)$, $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$ and $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$ are always updated if condition (16) or (17) are met, respectively.

Table 1 summarizes the spatial estimation of the average SIR at the sensors for non-stationary signals for DFT bin μ. In addition to the robustness improvements described previously, the instantaneous estimates of the spatial coherence terms are recursively averaged with a forgetting factor $0 < \beta_F < 1$. The parameters *slope_max* and *rise_min* allow an immediate update of $\hat{\Upsilon}_{\mathrm{d}}(k,\mu)$ and $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$ with the present value of $\bar{\Upsilon}(k,\mu)$ within a certain interval, respectively. The parameter D defines the length of the delay lines $\bar{\Upsilon}_{\mathrm{d}}(k,\mu)$ and $\bar{\Upsilon}_{\mathrm{n}}(k,\mu)$ in which the maxima and minima of $\bar{\Upsilon}(k,\mu)$ are tracked.

1	Compute $\hat{S}_{yy}(k,\mu)$, $\hat{S}_{\bar{y}\bar{y}}(k,\mu)$, and $\hat{S}_{xx}(k,\mu)$
2	Compute $\hat{\Upsilon}(k,\mu) = \hat{S}_{yy}(k,\mu)/\hat{S}_{\bar{y}\bar{y}}(k,\mu)$
3	Compute $\bar{\Upsilon}(k,\mu) = \mathrm{median}_3\{\hat{\Upsilon}(k,\mu)\}$
4	Compute $\hat{\Psi}(k,\mu)$
5	IF $\bar{\Upsilon}(k,\mu) > \hat{\Upsilon}_{\mathrm{d}}(k,\mu)/$*slope_max*
	$\hat{\Upsilon}_{\mathrm{d}}(k,\mu) = \bar{\Upsilon}(k,\mu)$
	Replace all stored values of $\bar{\Upsilon}_{\mathrm{d}}(k-i,\mu)$,
	$i = 1, 2, \ldots, D$, by $\hat{\Upsilon}_{\mathrm{d}}(k,\mu)$
	ELSE
	Find $\hat{\Upsilon}_{\mathrm{d}}(k,\mu)$, the maximum of $\bar{\Upsilon}_{\mathrm{d}}(k-i,\mu)$, $i = 1, 2, \ldots, D$
	$\bar{\Upsilon}_{\mathrm{d}}(k,\mu) = \bar{\Upsilon}(k,\mu)$
6	IF $(\hat{\Upsilon}(k,\mu) \geq \Upsilon_{0,\mathrm{d}}\hat{\Upsilon}_{\mathrm{d}}(k-1,\mu)$
	OR $\lvert\hat{\Psi}(k,\mu) - F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k-1,\mu))\rvert \leq \Delta_F)$
	update $= \hat{\Psi}(k,\mu)$
	ELSE *update* $= F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k-1,\mu))$
	IF $\hat{\Upsilon}(k,\mu) >$ *update*$/(1-$ *update*$)$
	update $= \hat{\Upsilon}(k,\mu)/(\hat{\Upsilon}(k,\mu)+1)$
7	$F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu)) = \beta_F F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k-1,\mu)) + (1-\beta_F)\cdot$ *update*
8	IF $\bar{\Upsilon}(k,\mu) < \hat{\Upsilon}_{\mathrm{n}}(k,\mu)\cdot$ *rise_max*
	$\hat{\Upsilon}_{\mathrm{n}}(k,\mu) = \bar{\Upsilon}(k,\mu)$
	Replace all stored values of $\bar{\Upsilon}_{\mathrm{n}}(k-i,\mu)$,
	$i = 1, 2, \ldots, D$, by $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$
	ELSE
	Find $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$, the minimum of $\bar{\Upsilon}_{\mathrm{n}}(k-i,\mu)$, $i = 1, 2, \ldots, D$
	$\bar{\Upsilon}_{\mathrm{n}}(k,\mu) = \bar{\Upsilon}(k,\mu)$
9	IF $(\hat{\Upsilon}(k,\mu) \leq \Upsilon_{0,\mathrm{n}}\hat{\Upsilon}_{\mathrm{n}}(k-1,\mu)$
	OR $\lvert\hat{\Psi}(k,\mu) - F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k-1,\mu))\rvert \leq \Delta_F)$
	update $= \hat{\Psi}(k,\mu)$
	ELSE *update* $= F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k-1,\mu))$
10	IF $\hat{\Upsilon}(k,\mu) <$ *update*$/(1-$ *update*$)$
	update $= \hat{\Upsilon}(k,\mu)/(\hat{\Upsilon}(k,\mu)+1)$
11	$F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu)) = \beta_F F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k-1,\mu)) + (1-\beta_F)\cdot$ *update*
12	Compute $\widehat{SIR}(k,\mu)$
13	Store $F(\hat{\mathbf{\Gamma}}_{\mathbf{dd}}(k,\mu))$, $F(\hat{\mathbf{\Gamma}}_{\mathbf{nn}}(k,\mu))$, $\bar{\Upsilon}_{\mathrm{d}}(k,\mu)$,
	$\hat{\Upsilon}_{\mathrm{d}}(k,\mu)$, $\bar{\Upsilon}_{\mathrm{n}}(k,\mu)$, and $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$
14	Increment k

Table 1: Summary of the algorithm.

3. EXPERIMENTAL EVALUATION

This algorithm was applied to estimate the SIR of competing male speech signals at an array with $M = 8$ sensors with sensor spacing $d = 4$ cm in an office room with $T_{60} = 300$ ms reverberation time. The desired source and the interfering source are located in broadside direction and in endfire direction, respectively, at a distance of 60 cm to the array center. Figure 1 (a) shows the signal of the desired source $d(t)$ and the interference $n(t)$ over time t. In Figure 1 (b), the decision variable $\hat{\Upsilon}(k,\mu)$ (solid) and the lower and upper thresholds $\hat{\Upsilon}_{\mathrm{d}}(k,\mu)$ (dashed) and $\hat{\Upsilon}_{\mathrm{n}}(k,\mu)$ (dash-dotted) are depicted for frequency $f \simeq 1$ kHz. The estimated SIR, $\widehat{SIR}(k,\mu)$ (bold line) is compared to the true SIR (narrow line), in Figure 1 (d) for the same frequency. Finally, Figure 1 (e) illustrates $\widehat{SIR}(k,\mu)$ (bold line), $SIR(k,\mu)$ (narrow line), and $\hat{\Upsilon}(k,\mu)$ (dashed line) averaged over the frequency range 300 Hz - 5 kHz. We see that the SIR can be accurately estimated, while $\hat{\Upsilon}(k,\mu)$ only provides a rough estimate of the SIR.

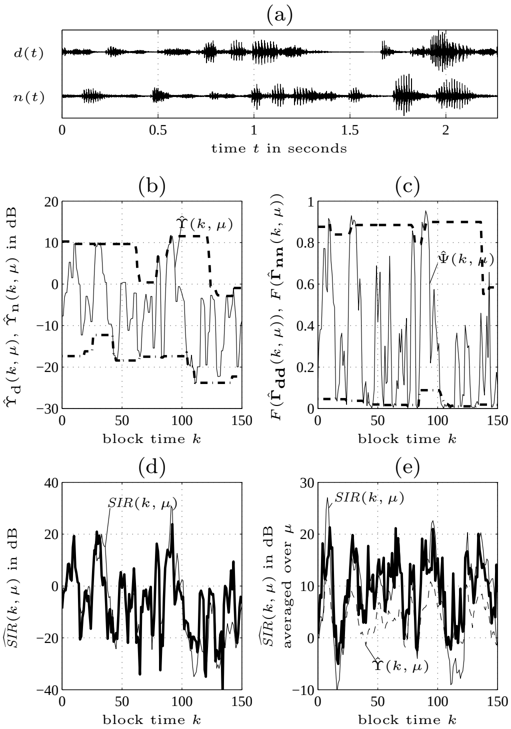

Figure 3: Procedure of the SIR estimation (Sampling rate $f_{\mathrm{s}} = 12$ kHz, $D = 96$, $\beta_F = 0.3$, $\beta_S = 0.34$, *slope_max* $=$ *rise_min* $= 1.25$, $\Upsilon_{0,\mathrm{d}} = 0.89$, $\Upsilon_{0,\mathrm{n}} = 1.12$, $\Delta_F = 0.045$).

4. CONCLUSIONS

In this paper, we presented a new method for estimating the DFT-bin-wise SIR of non-stationary wideband signals. Compared to traditional approaches, our algorithm allows strongly time-varying PSDs. The spatial coherence matrices w.r.t. the desired signal and w.r.t. interference must be slowly time-varying relative to the PSDs. The robustness of our approach was illustrated by experiments in a real reverberant environment.

5. REFERENCES

[1] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds., pp. 155–194. Springer, Berlin, 2003.

[2] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, July 2001.

[3] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2677–2684, October 1999.

[4] D.R. Brillinger, *Time Series-Data Analysis and Theory*, Holt, Rinehart, and Winston, Inc., New York, 1975.