Robust DOA Estimation of Harmonic Signals Using Constrained Filters on Phase Estimates

Karimian-Azari, Sam; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO 2014)

Publication date:
2014

Document Version
Early version, also known as pre-print

ROBUST DOA ESTIMATION OF HARMONIC SIGNALS USING CONSTRAINED FILTERS ON PHASE ESTIMATES

Sam Karimian-Azari ⋆*, Jesper Rindom Jensen* †*, and Mads Græsbøll Christensen*

Audio Analysis Lab, AD:MT, Aalborg University, email: {ska, jrj, mgc}@create.aau.dk

ABSTRACT

In array signal processing, distances between receivers, e.g., microphones, cause time delays depending on the direction of arrival (DOA) of a signal source. We can then estimate the DOA from the time-difference of arrival (TDOA) estimates. However, many conventional DOA estimators based on TDOA estimates are not optimal in colored noise. In this paper, we estimate the DOA of a harmonic signal source from multi-channel phase estimates, which relate to narrowband TDOA estimates. More specifically, we design filters to apply on phase estimates to obtain a DOA estimate with minimum variance. Using a linear array and harmonic constraints, we design optimal filters based on estimated noise statistics. Therefore, the proposed method is robust against different noise scenarios. In colored noise, simulation results confirm that the proposed method outperforms an optimal state-of-the-art weighted least-squares (WLS) DOA estimator.

Index Terms— Audio signal, harmonic model, direction of arrival (DOA), time-difference of arrival (TDOA).

1. INTRODUCTION

Microphone array technology has emerged as a tool to improve audio communication systems such as teleconferencing and hearing aids. Acoustic source localization can be used in audio signal source separation, enhancement, and speech analysis and recognition [1, 2]. It also appears in some microphone array filter designs, e.g., the minimum variance distortionless response (MVDR) beamformer [3] is designed to pass a signal of interest from the known direction of arrival (DOA). Ambient noise in the received signals often causes ambiguities in source localization. Thus, estimating the DOA of an audio source is a challenging problem. In general, DOA estimation methods can be categorized in three groups [4]:

• The time-difference of arrival (TDOA) based estimators that scale the TDOA of successive microphones, e.g., the phase shift [5] and the cross-correlation [6, 7] methods.

• The beamforming based methods that steer the array in a range of possible directions at each frequency bin, and maximize the output power versus the DOA [8], e.g., the steered response power (SRP) method [9].

• The high-resolution DOA estimators based upon spatiospectral correlation matrix estimates, e.g., the MUSIC [10] and ESPRIT [11] methods utilizing the eigenvalue decomposition, and the NLS based method [12].

In terms of computational complexity, the TDOA based estimators possess an advantage over the two other methods in real-time systems [4]. Therefore, we here focus on the TDOA based estimators for audio signals with emphasis on the harmonic structure.

The broadband TDOA estimation method has been proposed in [5] based on the least squares (LS) estimator from the phase of the cross power spectrum. Conventional TDOA estimators are designed assuming a single-source [4,5,13]. However, the harmonic characteristic of audio signals facilitates a remarkable ability to estimate TDOAs of multiple sources which have different fundamental frequencies. The weighted least-squares (WLS) method has been proposed to estimate the DOA from the TDOA estimates of a harmonic signal [14] using the mutual coupling of the phase estimates of the harmonics. Although the WLS DOA estimator [14] attains the Cramér-Rao bound (CRB) in white noise, and outperforms SRP [9] and position-pitch plane based (POPI) [15] methods, the weighting matrices of the WLS method are given by the corresponding Fisher information matrices (FIMs) with the assumption of white Gaussian noise [14].

In this paper, we consider a novel TDOA based approach robust against different noise scenarios. We apply the idea of using the linear relationship between the unwrapped phase estimates across microphones, i.e., the phase shift, and then estimate the DOA from these phase shifts of the harmonics. Since an additive Gaussian noise should be equivalent to an additive phase noise for a high signal-to-noise ratio (SNR) [16], we can filter the phase estimates as the solution to minimize the variance of the phase shifts. We design the minimum variance distortionless response (MVDR) filter using the noise variance estimates from the statistics of the multi-channel phase estimates. We also apply this filtering method on the phase shift estimates of harmonics to estimate the DOA. As a result, the proposed method performs well for different noise scenarios, e.g., in colored noise, since we estimate the noise variance across microphones and harmonics.

The remainder of the present paper is organized as fol-

This research was funded by: ⋆the Villum Foundation, and †the Danish Council for Independent Research, grant ID: DFF 1337-00084.

lows. In the next section, i.e., Section 2, the multi-channel signal model is formulated. Then, we present the DOA estimation method from multi-channel phase estimates in Section 3. Later on, in Section 4, experimental results are reported. In closing, the work is concluded in Section 5.

2. SIGNAL MODEL

2.1. Spatial sampling of a harmonic signal

We model a harmonic signal $s(n)$ at the time instance n as the sum of L sinusoids with a fundamental frequency $\omega_0 \in (0, \pi]$, harmonics with frequencies being integer multiples of the fundamental frequency, the real magnitudes $\alpha_l > 0$, and phases $\varphi_l \in (-\pi, \pi]$ for $l = 1, 2, \ldots, L$:

$$s(n) = \sum_{l=1}^{L} \alpha_l \, e^{j\,(l\omega_0\, n + \varphi_l)}, \tag{1}$$

where $j = \sqrt{-1}$. The complex signal model can be applied to real signals through the Hilbert transform. We assume a microphone array with a set of M omnidirectional microphones that receives a plane wave from the far field, with a DOA $\theta \in [-\pi/2, \pi/2]$. We write the clean multi-channel received signal $\mathbf{x}(n) = [\, x_1(n), x_2(n), \ldots, x_M(n)\,]^T$ as a function of the steering vector $\mathbf{d}_\theta(l\omega_0) = \mathbf{d}(\theta, l\omega_0)$ and the magnitude attenuation of $\boldsymbol{\beta} = \text{diag}\{[\,\beta_1, \beta_2, \ldots, \beta_M]\}$ at each microphone in an anechoic environment, i.e.,

$$\mathbf{x}(n) = \sum_{l=1}^{L} \alpha_l \, e^{j\,(l\omega_0\, n + \varphi_l)} \, \boldsymbol{\beta}\, \mathbf{d}_\theta(l\omega_0), \tag{2}$$

where the superscript ^{T} is the transpose operator, and diag$\{\cdot\}$ denotes a diagonal matrix formed from its vector argument. While different array structures can be considered, we assume a uniform linear array (ULA) structure herein for the proof of our concept with the particular steering vector

$$\mathbf{d}_\theta(\omega) = [\,1,\, e^{-j\omega f_s \tau_0 \sin(\theta)},\, \cdots,\, e^{-j(M-1)\omega f_s \tau_0 \sin(\theta)}]^T, \tag{3}$$

where $\omega \in [0, \pi]$, f_s is the sampling frequency, $\tau_0 = \delta/c$ is the delay between two successive sensors with the distance of δ, and c is the speed of sound.

2.2. Phase noise

We assume that the observed signals are contaminated by Gaussian noise $\mathbf{v}(n) = [\, v_1(n), v_2(n), \ldots, v_M(n)\,]^T \in \mathbb{C}^M$ in the complex form with zero-mean values, i.e.,

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n). \tag{4}$$

The independent noise across the M microphones, i.e., $v_m(n)$ for $m = 1, 2, \ldots, M$, have the real and imaginary uncorrelated parts with the variance of $\sigma_m^2/2$. At a high narrowband SNR, for the lth harmonic, i.e., $\text{SNR}_m^l = (\beta_m \alpha_l)^2/\sigma_m^2 \gg 1$, the additive noise can be converted into an equivalent phase noise $\Delta\varphi_m(l\omega_0)$, and this normally distributed random noise has zero-mean and a variance of $\text{E}\{[\Delta\varphi_m(l\omega_0)]^2\} = 1/(2\,\text{SNR}_m^l)$ [16], where E$\{\cdot\}$ denotes the statistical expectation. Therefore, besides the phases of the harmonics and phase shifts due to the TDOAs, we define a vector containing phase noise as

$$\Delta\boldsymbol{\Phi}_l = [\,\Delta\varphi_1(l\omega_0),\, \Delta\varphi_2(l\omega_0),\, \ldots,\, \Delta\varphi_M(l\omega_0)\,]^T, \tag{5}$$

and we can approximate the noisy signal model in (4) like:

$$\mathbf{y}(n) \approx \sum_{l=1}^{L} \alpha_l \, e^{j\,(l\omega_0\, n + \varphi_l)} \, \mathbf{D}_v(l\omega_0)\, \boldsymbol{\beta}\, \mathbf{d}_\theta(l\omega_0), \tag{6}$$

where $\mathbf{D}_v(l\omega_0) = \text{diag}\{\exp(j\Delta\boldsymbol{\Phi}_l)\}$. For the spatially uncorrelated phase noise, i.e., $\text{E}\{\Delta\varphi_i(l\omega_0)\Delta\varphi_k(l\omega_0)\} = 0$ for $i \neq k$, the covariance matrix of the phase noise vector at the lth harmonic, i.e., $\mathbf{R}_{\Delta\boldsymbol{\Phi}_l} = \text{E}\{\Delta\boldsymbol{\Phi}_l\, \Delta\boldsymbol{\Phi}_l^T\}$, can be shown to be

$$\mathbf{R}_{\Delta\boldsymbol{\Phi}_l} = \text{diag}\Big\{\Big[\, \frac{1}{2\,\text{SNR}_1^l}, \frac{1}{2\,\text{SNR}_2^l}, \ldots, \frac{1}{2\,\text{SNR}_M^l}\,\Big]\Big\}. \tag{7}$$

3. PROPOSED METHOD

Using the model of a harmonic signal and the phase properties of the spatial samples $\mathbf{y}(n)$, we can estimate the source DOA from spatiotemporal signals $\mathbf{Y}(n) = [\,\mathbf{y}(n), \mathbf{y}(n+1), \ldots, \mathbf{y}(n+N-1)\,] \in \mathbb{C}^{M\times N}$. First, we estimate the phase shifts of each harmonic across the microphones, which are related to the TDOAs of narrowband signals, and then estimate the DOA from the collection of phase shift estimates, inspired by [14]. To achieve an optimal solution from the noisy estimates, we estimate both phase shifts and DOA sequentially with minimum variance, based on the assumption that the fundamental frequency and also the number of harmonic components are given, e.g., by using a multi-channel pitch estimation [12, 17, 18] along with multi-channel model order estimation [19].

3.1. Phase shift estimate

Following the harmonic signal model (2), the phase of the lth harmonic in the mth channel of a ULA is

$$\Phi_{l,m} = \varphi_l - (m-1) l\omega_0 f_s \tau_0 \sin(\theta). \tag{8}$$

We write the collection of M phases for the lth harmonic, i.e., $\boldsymbol{\Phi}_l = [\,\Phi_{l,1}, \Phi_{l,2}, \ldots, \Phi_{l,M}\,]^T$, as a linear combination of the signal phase φ_l and the phase shift ψ_l as

$$\begin{aligned}\boldsymbol{\Phi}_l &= \big[\,\mathbf{1}_M\,,\, (\mathbf{1}_M - \boldsymbol{\Gamma}_M)\,\big] \begin{bmatrix} \varphi_l \\ l\omega_0 f_s \tau_0 \sin(\theta) \end{bmatrix} \\ &= \boldsymbol{\Pi}_M \begin{bmatrix} \varphi_l \\ \psi_l \end{bmatrix},\end{aligned} \tag{9}$$

where $\mathbf{1}_M$ is the all-ones column vector of length M, and $\mathbf{\Gamma}_M = [\,1, 2, \ldots, M\,]^T$. The matrix $\mathbf{\Pi}_M = [\,\mathbf{1}_M\,, (\mathbf{1}_M - \mathbf{\Gamma}_M)\,] \in \mathbb{R}^{M\times 2}$ is defined based upon the number of microphones and the linear relationship between the TDOAs of a ULA. This can be generalized to other arrays.

Following the approximated signal model in (6), the phase estimates $\hat{\mathbf{\Phi}}_l$, which are estimated from $\mathbf{y}_m(n) = [\,y_m(n), y_m(n+1), \ldots, y_m(n+N-1)\,]$ for $m = 1, \ldots, M$, can be written like

$$\hat{\mathbf{\Phi}}_l = \mathbf{\Phi}_l + \Delta\mathbf{\Phi}_l, \tag{10}$$

where $\Delta\mathbf{\Phi}_l$ is the phase noise vector. The covariance matrix of $\Delta\mathbf{\Phi}_l$, which describes the spatial properties of noise statistics, is defined like

$$\mathbf{R}_{\Delta\mathbf{\Phi}_l} = \mathrm{E}\{(\hat{\mathbf{\Phi}}_l - \mathrm{E}\{\hat{\mathbf{\Phi}}_l\})\,(\hat{\mathbf{\Phi}}_l - \mathrm{E}\{\hat{\mathbf{\Phi}}_l\})^T\}. \tag{11}$$

Although the consecutive M phases lie on a continuous line, which are originated from the first microphone with a zero phase shift, the phase estimates are wrapped in $(-\pi, \pi]$. Therefore, we have to unwrap these phase estimates, e.g., using the unwrapping algorithm in [20].

We apply a filter $\mathbf{W} \in \mathbb{R}^{M\times 2}$ to estimate the parameter vector $[\,\varphi_l,\ \psi_l\,]^T$ like

$$\begin{bmatrix} \hat{\varphi}_l \\ \hat{\psi}_l \end{bmatrix} = \mathbf{W}^T\hat{\mathbf{\Phi}}_l = \mathbf{W}^T\mathbf{\Pi}_M \begin{bmatrix} \varphi_l \\ \psi_l \end{bmatrix} + \mathbf{W}^T\Delta\mathbf{\Phi}_l, \tag{12}$$

with the constraint that $\mathbf{W}^T\mathbf{\Pi}_M = \mathbf{I}_{2\times 2}$, where $\mathbf{I}_{2\times 2}$ is an identity matrix of size 2. The mean square error (MSE) of an unbiased estimator is given by

$$\mathrm{MSE}\left\{ \begin{bmatrix} \hat{\varphi}_l \\ \hat{\psi}_l \end{bmatrix} \right\} = \mathrm{tr}\{\mathbf{W}^T\mathbf{R}_{\Delta\mathbf{\Phi}_l}\mathbf{W}\}, \tag{13}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a square matrix. Later on, to approach a minimum variance of the parameter vector estimator in (12), the minimum variance distortionless response (MVDR) method is derived by

$$\begin{aligned} \min_{\mathbf{W}} \quad & \mathrm{tr}\{\mathbf{W}^T\mathbf{R}_{\Delta\mathbf{\Phi}_l}\mathbf{W}\} \\ \text{subject to} \quad & \mathbf{W}^T\mathbf{\Pi}_M = \mathbf{I}_{2\times 2}. \end{aligned} \tag{14}$$

Using the method of Lagrange multipliers, the optimal filter $\mathbf{W}_{\mathrm{MVDR}}$ is given by

$$\mathbf{W}_{\mathrm{MVDR}} = \mathbf{R}^{-1}_{\Delta\mathbf{\Phi}_l}\mathbf{\Pi}_M(\mathbf{\Pi}_M^T\,\mathbf{R}^{-1}_{\Delta\mathbf{\Phi}_l}\mathbf{\Pi}_M)^{-1}. \tag{15}$$

3.2. DOA estimate

While the DOA of the lth harmonic can be obtained from the phase shift estimate $\hat{\psi}_l$, i.e., $\hat{\theta}_l = \sin^{-1}(\hat{\psi}_l\,/\,l\omega_0 f_s\tau_0)$, the DOA of the harmonic source can be estimated from the L estimates with minimum variance, where all the phase shifts $\mathbf{\Psi} = [\,\psi_1, \psi_2, \ldots, \psi_L\,]^T$ are modeled like

$$\mathbf{\Psi} = \omega_0 f_s \tau_0 \sin(\theta)\,\mathbf{\Gamma}_L. \tag{16}$$

We can show that the phase shift estimates are also approximately distorted by Gaussian noise $\Delta\mathbf{\Psi} = [\,\Delta\psi_1,\ \Delta\psi_2,\ \ldots,\ \Delta\psi_L\,]^T$ like

$$\hat{\mathbf{\Psi}} = \mathbf{\Psi} + \Delta\mathbf{\Psi}. \tag{17}$$

The covariance matrix of $\Delta\mathbf{\Psi}$ which describes spectral properties of noise is defined like

$$\begin{aligned} \mathbf{R}_{\Delta\mathbf{\Psi}} &= \mathrm{E}\{\Delta\mathbf{\Psi}\Delta\mathbf{\Psi}^T\} \\ &= \mathrm{E}\{(\hat{\mathbf{\Psi}} - \mathrm{E}\{\hat{\mathbf{\Psi}}\})\,(\hat{\mathbf{\Psi}} - \mathrm{E}\{\hat{\mathbf{\Psi}}\})^T\}. \end{aligned} \tag{18}$$

Herein, we apply a filter $\mathbf{h} \in \mathbb{R}^L$ to estimate the sinusoidal function of the DOA from the phase shift estimates like

$$\sin(\hat{\theta}) = \mathbf{h}^T\hat{\mathbf{\Psi}} = \omega_0 f_s\tau_0 \sin(\theta)\,\mathbf{h}^T\mathbf{\Gamma}_L + \mathbf{h}^T\Delta\mathbf{\Psi}. \tag{19}$$

With the distortionless constraint that $\mathbf{h}^T\mathbf{\Gamma}_L = 1/\omega_0 f_s\tau_0$, the MSE of an unbiased estimator is given by $\mathrm{MSE}\{\sin(\hat{\theta})\} = \mathbf{h}^T\mathbf{R}_{\Delta\mathbf{\Psi}}\mathbf{h}$. To approach a minimum MSE estimator, the optimal filter is given as the solution to the following problem:

$$\begin{aligned} \min_{\mathbf{h}} \quad & \mathbf{h}^T\mathbf{R}_{\Delta\mathbf{\Psi}}\mathbf{h} \\ \text{subject to} \quad & \mathbf{h}^T\mathbf{\Gamma}_L = 1/\omega_0 f_s\tau_0. \end{aligned} \tag{20}$$

Subsequently, the optimal filter $\mathbf{h}_{\mathrm{MVDR}}$ is found like

$$\mathbf{h}_{\mathrm{MVDR}} = \frac{1}{\omega f_s\tau_0}\,\mathbf{R}^{-1}_{\Delta\mathbf{\Psi}}\,\mathbf{\Gamma}_L\,(\mathbf{\Gamma}_L^T\,\mathbf{R}^{-1}_{\Delta\mathbf{\Psi}}\,\mathbf{\Gamma}_L)^{-1}, \tag{21}$$

and the optimal DOA can be estimated with a minimum variance like

$$\hat{\theta} = \sin^{-1}(\mathbf{h}^T_{\mathrm{MVDR}}\hat{\mathbf{\Psi}}). \tag{22}$$

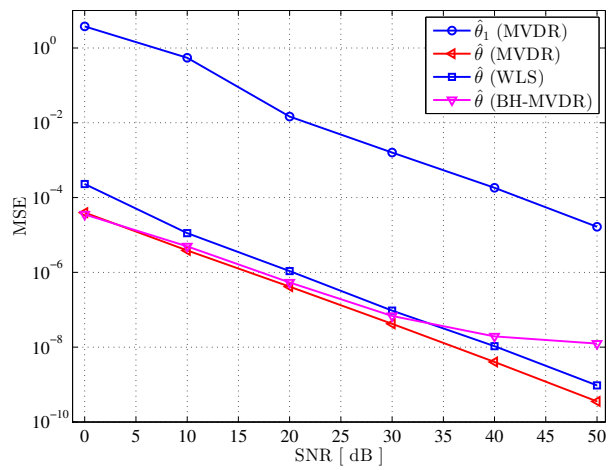

Fig. 1. MSE of DOA estimates of a harmonic signal versus SNR levels of colored noise.

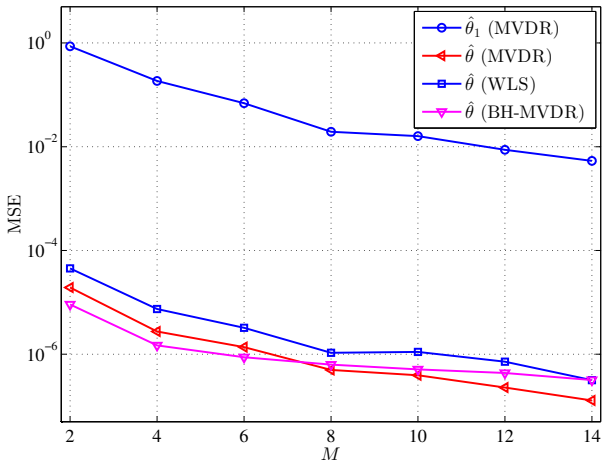

Fig. 2. MSE of DOA estimates of a harmonic signal, contaminated by colored noise in SNR = 20 dB, versus number of microphones.

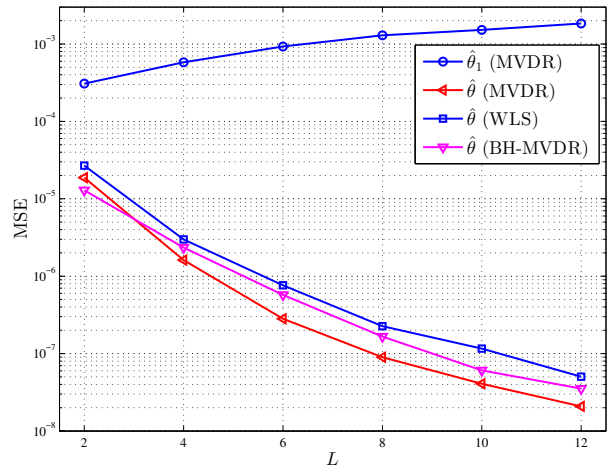

Fig. 3. MSE of DOA estimates of a harmonic signal, contaminated by colored noise in SNR = 30 dB, versus number of harmonics.

4. SIMULATION RESULTS

We evaluate the performance of the proposed filtering method for estimating the DOA of a complex harmonic signal in an anechoic environment. We compare the MSE of the proposed method in 200 Monte-Carlo simulations with the broadband MVDR beamforming method with harmonic emphasis (BH-MVDR) [18] and the WLS DOA estimator in [14]. We investigate the performance of the proposed method in different situations, i.e., SNR levels, and the number of microphones and harmonics are varied.

In the experiments, we assume that we know the fundamental frequency and number of harmonics, and then estimate the phases of multi-channel signals individually using the least-squares (LS) method, because the results of the LS method are asymptotically efficient in colored noise with a large number of samples [21]. With a high enough number B of phase estimates, which are assumed stationary, the time averaging should converge to the statistical expectation, i.e.,

$$\mathrm{E}\{\hat{\mathbf{\Phi}}_l\} = \frac{1}{B}\sum_{b=0}^{B-1}\hat{\mathbf{\Phi}}_l(b) - bl\omega_0\mathbf{1}_M, \tag{23}$$

$$\mathrm{E}\{\hat{\mathbf{\Psi}}\} = \frac{1}{B}\sum_{b=0}^{B-1}\hat{\mathbf{\Psi}}(b), \tag{24}$$

where $\hat{\mathbf{\Phi}}_l(b)$ are estimated from $\mathbf{Y}(b)$, and $\hat{\mathbf{\Psi}}(b)$ is the second vector element of $\mathbf{W}_{\mathrm{MVDR}}^T[\,\hat{\mathbf{\Phi}}_l(b) - bl\omega_0\mathbf{1}_M\,]$ for $l = 1, 2, \ldots, L$. Then, we also estimate the asymptotic covariance matrices in (11) and (18) by time averaging. These matrices can be guaranteed to be full rank in the two filter designs (15) and (21) by choosing $B \geq \max(M, L)$.

We use a ULA with $M = 10$ omnidirectional microphones that $\delta = 0.04$ m, $f_s = 8.0$ kHz, $c = 343.2$ m/s at 20°C, and add colored Gaussian noise which is generated by passing a complex white Gaussian noise with zero-mean through an autoregressive (AR) filter given by $1/(1 - 1.3z^{-1} + 0.4z^{-2})$ in Z-transform. In all simulations, we place a synthetic signal at $\theta = \pi/6$, where $\omega_0 = 0.06\pi$ and $L = 10$ with uniform random distributed phases φ_l, identical magnitudes $\alpha_l = 1$, and the magnitude attenuation $\boldsymbol{\beta} = \mathrm{diag}\{[\,0.25, 0.5, 1.0, 0.67, 0.33, 0.5, 0.67, 1.0, 0.34, 0.4, 0.5, 0.67, 1.0, 2.0\,]\}$ for a maximum number of 14 microphones. Then, we estimate $B = 100$ numbers of the phases of each channel individually from $N = 128$ samples.

At different SNR levels of the colored noise, Figure 1 shows that the proposed MVDR filtering method outperforms both the WLS and BH-MVDR methods for SNR > 5 dB. In 20 dB of SNR, Figure 2 shows that the MVDR filtering method outperforms the WLS method and the BH-MVDR method for $M \geq 8$. In Figure 3, we also see that the DOA estimates of a signal with a high number of harmonics will be more robust than lower numbers. Moreover, the proposed method outperforms the WLS DOA estimator, using $M = 10$ in SNR = 30 dB. In general, the results of all experiments confirm that the DOA estimates between all harmonics has significantly lower uncertainty than the DOA estimate of the first harmonic, i.e., $\hat{\theta}_1$.

5. CONCLUSION

We have shown that an additive Gaussian noise can cause a Gaussian phase noise on multi-channel signals, and proposed a DOA estimation method emphasizing the harmonic signal model. We applied a filtering method on multi-channel phase estimates to approach optimal results by estimating spatial and spectral noise statistics. We have shown that the proposed method outperforms the WLS DOA estimator in colored noise, which is an optimal solution in white Gaussian noise outperforming some state-of-the-art methods [14].

6. REFERENCES

[1] M. Brandstein and D. Ward, eds., *Microphone Arrays - Signal Processing Techniques and Applications*. Springer-Verlag, 2001.

[2] J. Benesty, Y. Huang, and J. Chen, *Microphone Array Signal Processing*, vol. 1. Springer-Verlag, 2008.

[3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[4] M. S. Brandstein and H. Silverman, "A practical methodology for speech source localization with microphone arrays," *Comput. Speech Language*, 1997.

[5] Y. Chan, R. Hattin, and J. B. Plant, "The least squares estimation of time delay and its use in signal detection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, pp. 217–222, Jun 1978.

[6] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, pp. 320–327, Aug. 1976.

[7] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, 2003.

[8] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Education, Inc., 2005.

[9] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays - Signal Processing Techniques and Applications* (M. S. Brandstein and D. B. Ward, eds.), ch. 8, pp. 157–180, Springer-Verlag, 2001.

[10] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, Mar. 1986.

[11] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, Jul. 1989.

[12] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 21, pp. 923 –933, May 2013.

[13] J.-M. Yang and H.-G. Kang, "Two-stage source tracking method using a multiple linear regression model in the expanded phase domain," *EURASIP J. on Applied Signal Process.*, vol. 2012, no. 1, pp. 1–19, 2012.

[14] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Statistically efficient methods for pitch and DOA estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 3900–3904, 2013.

[15] M. Wohlmayr and M. Képesi, "Joint position-pitch extraction from multichannel audio," in *Proc. Interspeech*, pp. 1629–1632, Aug. 2007.

[16] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear regression (corresp.)," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 832–835, 1985.

[17] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 409–412, Mar. 2012.

[18] S. Karimian-Azari, J. R. Jensen, and M. G. Christensen, "Fast joint DOA and pitch estimation using a broadband MVDR beamformer," in *Proc. European Signal Processing Conf.*, Sept. 2013.

[19] S. Karimian-Azari, J. R. Jensen, and M. G. Christensen, "Fundamental frequency and model order estimation using spatial filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2014.

[20] K. Itoh, "Analysis of the phase unwrapping algorithm," *Applied Optics*, vol. 21, no. 14, pp. 2470–2470, 1982.

[21] P. Stoica, A. Jakobsson, and J. Li, "Cisoid parameter estimation in the colored noise case: asymptotic Cramér-Rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Signal Process.*, vol. 45, pp. 2048–2059, Aug. 1997.