DIRECTION-OF-ARRIVAL ESTIMATION OF SPEECH SOURCES UNDER ALIASING CONDITIONS

Vinod V. Reddy and Andy W. H. Khong, Members, IEEE.

Nanyang Technological University, Singapore
Email: vreddy@ntu.edu.sg, Andykhong@ntu.edu.sg

ABSTRACT

Due to practical considerations the microphone spacing is increased to achieve improved resolution by violating the spatial Nyquist criterion. Accompanied aliasing components adversely affect the identifiability of the source direction peaks. We investigate the effect of aliasing on the spatial spectrum of the steered minimum variance distortionless response (STMV) method and propose a novel multi-stage scheme assisted by subband decomposition for suppressing aliasing components. The performance of the proposed technique, evaluated with simulations and recorded room responses, reflects the improvement in the identifiability of accurate source directions under aliasing conditions.

I. INTRODUCTION

In applications such as teleconferencing and human-computer interface, estimation of speech source directions from microphone recordings is an essential task. With practical advantages under consideration, the number of microphones and their spacing are important design constraints. Violating the spatial Nyquist criterion to serve this objective such that the microphone spacing $d > 0.5\lambda_{\min}$, where $\lambda_{\min}$ is the wavelength corresponding to the highest signal frequency f_{u}, introduces ambiguities in the direction-of-arrival (DOA) estimates. This phenomenon is referred to as the spatial aliasing condition.

DOAs for wideband speech signals can be estimated from the spatial spectrum of a steered beamformer [1]. The data acuired from a microphone array across the speech bandwidth is appropriately transformed to estimate a steered covariance matrix (STCM) for each scan direction. The steered minimum variance distortionless response (STMV) is then used for coherent wideband DOA estimation [1]. A common framework has been presented in [2] which parameterizes the STCM with path delays corresponding to the scan directions and show the use of STCM in the steered response power-phase transform (SRP-PHAT) and the multi-channel cross correlation (MCCC) [3].

Estimation of the time-difference-of-arrival (TDOA) for speech source localization has been presented in [4]–[6]. While the technique presented in [4] is a special case of SRP-PHAT presented for a microphone pair, the techniques proposed in [6], [7] estimate source TDOA from the unmixing filters obtained via convolutive blind source separation (BSS). By averaging the angular spectrum across the entire frequency range these techniques estimate TDOA from the raised angular spectrum floor due to aliasing. However, they do not explicitly attempt to suppress aliasing components.

DOA estimation under aliasing conditions has been recently explored using sparse signal representation [8], [9]. A multi-dictionary, multiple measurement signal model is employed for identifying true source directions across various frequency bins. While these techniques resolve for true source directions using frequency components within a small band, the DOA and aliasing information present across the entire signal bandwidth may not be fully utilized. The use of wider bandwidth is essential particularly for quasi-stationary speech signals where the signal statistics change with time.

II. RELATION TO PRIOR WORK

Under aliasing conditions, it is shown in [10] that the average spectrum of a DOA estimator across the entire speech spectrum results in a raised spectrum floor. The TDOA-based algorithms [6], [7], [11] also average their angular spectrum across the entire frequency spectrum in order to minimize the aliasing components while maintaining the source direction peaks. As shown in [12], the averaged spectrum obtained with these algorithms exhibit ambiguous aliasing peaks over the raised spectrum floor. A multi-stage DOA estimation scheme has been presented in [12] to suppress the aliasing peaks from the identified alias-free speech subband. In this work, we first present the effect of aliasing on the STMV algorithm. We then regulate the influence of the spatial spectrum obtained from the alias-free subband on successive stages of DOA estimation performed on various subbands. DOA estimation performance is studied with simulations and recorded room responses.

III. EFFECT OF SPATIAL ALIASING AND THE STMV ALGORITHM

III-A. Signal Model

Within a reverberant environment, the output of the mth microphone of an $M-$microphone uniform linear array

(ULA) can be modeled as

$$x_m(n) = \sum_{p=1}^{P} \mathbf{h}_{m,p}^T \mathbf{s}_p(n) + v_m(n),\ m = 1, 2, ..., M, \quad (1)$$

where $\mathbf{h}_{m,p} = [h_{m,p}(0)\ \ldots\ h_{m,p}(L_h - 1)]^T$ is the L_h-length room impulse response between the mth microphone and the pth source out of the total P speech sources present, $\mathbf{s}_p(n) = [s_p(n)\ s_p(n-1)\ \ldots\ s_p(n - L_h + 1)]^T$ is the pth source signal vector and $v_m(n)$ is the uncorrelated additive noise of the mth microphone.

Transforming the microphone array output into time-frequency domain using $K-$point short-time Fourier transform (STFT), we can express the array output vector for each time-frequency bin as

$$\begin{aligned}\underline{\mathbf{x}}(k,l) &= \sum_{p=1}^{P} \underline{\mathbf{h}}_p(k) \underline{s}_p(k,l) + \underline{\mathbf{v}}(k,l) \\ &= \sum_{p=1}^{P} \mathbf{a}(k,\theta_p) \underline{s}_p(k,l) + \underline{\mathbf{b}}(k,l) + \underline{\mathbf{v}}(k,l),\end{aligned} \quad (2)$$

where $\underline{\mathbf{v}}(k,l) \in \mathbb{C}^{M\times 1}$ is the STFT of the additive noise vector across the microphone array, $\underline{s}_p(k,l)$ is the STFT of the pth source signal and $\underline{\mathbf{h}}_p(k) = [\underline{h}_{1,p}(k)\ \ldots\ \underline{h}_{M,p}(k)]^T$ denotes the transfer function between the pth source and all the microphones. Assuming a diffused sound field $\underline{\mathbf{h}}_p(k)$ is decomposed into the direct-path component and a reverberant component [13] where the diffused sound propagation component for all the sources can be cumulatively represented by $\underline{\mathbf{b}}(k,l)$. For a ULA with microphone spacing d, the steering vector along θ_p at frequency f_k is given by

$$\mathbf{a}(k,\theta_p) = \left[1\ e^{-\jmath\frac{2\pi f_k}{c} d\cos\theta_p}\ \ldots\ e^{-\jmath\frac{2\pi f_k}{c}(M-1)d\cos\theta_p}\right]^T, \quad (3)$$

where $\jmath = \sqrt{-1}$, c is the speed of sound and k is the bin index corresponding to f_k. When $d > \lambda_{\min}/2$, where $\lambda_{\min} = c/f_{\mathrm{u}}$, the above steering vector is not unique, thus introducing spatial aliases. Irrespective of the algorithm employed for DOA estimation, aliasing components for a source along θ are positioned at [10]

$$\cos\phi = \cos\theta \mp \frac{\gamma c}{f_k d},\ \ \gamma \in \mathbb{Z}. \quad (4)$$

The look directions $\phi \in [0,\pi]$ satisfying (4) include the source direction θ for $\gamma = 0$ and aliasing angles θ^{a} when $\gamma \neq 0$.

III-B. Review of the STMV algorithm

For each look direction ϕ, the array snapshots $\underline{\mathbf{x}}(k,l)$ across the speech bandwidth are steered to the direction ϕ with the transformation matrix $\mathbf{T}(k,\phi) = \mathrm{diag}(\mathbf{a}(k,\phi))$. The STCM is then estimated as

$$\mathbf{R}(\phi) = \sum_{k=k_{\mathrm{l}}}^{k_{\mathrm{u}}} \mathbf{T}^H(k,\phi)\mathbf{R}(k)\mathbf{T}(k,\phi), \quad (5)$$

where the variables k_{l} and k_{u} denote, respectively, the frequency-bin indices corresponding to the lower and higher frequencies f_{l} and f_{u} for sources spanning a bandwidth of $[f_{\mathrm{l}},\ f_{\mathrm{u}}]$. One can alternatively employ PHAT-compensated STCM in order to provide moderate robustness to reverberation [4] which can be obtained as

$$\widetilde{\mathbf{R}}(\phi) = \sum_{k=k_{\mathrm{l}}}^{k_{\mathrm{c}}} \mathbf{T}^H(k,\phi)\left(\underline{\mathbf{x}}(k,l)\underline{\mathbf{x}}^H(k,l) \oslash \left|\underline{\mathbf{x}}(k,l)\underline{\mathbf{x}}^H(k,l)\right|\right)\mathbf{T}(k,\phi), \quad (6)$$

where $\oslash$ denotes componentwise division of two matrices, $|.|$ denotes a matrix with the absolute value of its elements. The STMV estimator can then be formulated as an optimization problem which solves for $\mathbf{w}_\phi$ for each ϕ, given by [2]

$$\begin{aligned}&\underset{\mathbf{w}_\phi}{\text{minimize}} \quad \mathbf{w}_\phi^H \widetilde{\mathbf{R}}(\phi)\mathbf{w}_\phi \\ &\text{subject to} \quad \mathbf{w}_\phi^H \mathbf{1}_M = 1,\end{aligned} \quad (7)$$

where $\mathbf{1}_M$ is an $M \times 1$ vector of ones. The spatial spectrum of STMV is then

$$\Psi_{\mathrm{STMV}}(\phi) = \frac{1}{\mathbf{1}_M^T \widetilde{\mathbf{R}}^{-1}(\phi)\mathbf{1}_M} \quad (8)$$

and the DOA of the P sources is given by

$$\widehat{\theta} = \underset{\phi\in\Theta_P}{\mathrm{argmax}}\ \Psi_{\mathrm{STMV}}(\phi), \quad (9)$$

where Θ_P is the $P-$dimensional parameter space.

III-C. Effect of spatial aliasing

In order to study the effect of aliasing for a wideband source scenario, we first assume the source has a constant power σ_s^2 across the bandwidth $[f_{\mathrm{l}}, f_{\mathrm{u}}]$ for mathematical tractability. The STCM without PHAT compensation is then given by

$$\mathbf{R}(\phi) = \sigma_s^2 \sum_{k=k_{\mathrm{l}}}^{k_{\mathrm{u}}} \mathbf{a}(k,\theta-\phi)\mathbf{a}^H(k,\theta-\phi) + (f_{\mathrm{u}} - f_{\mathrm{l}})\sigma_v^2 \mathbf{I}_M, \quad (10)$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, σ_v^2 is the additive noise variance and

$$\begin{aligned}\mathbf{a}(k,\theta-\phi) &\triangleq \mathbf{T}^H(k,\phi)\mathbf{a}(k,\theta) \\ &= \left[1\ e^{-\jmath 2\pi(d/c)f_k\eta}\ \ldots\ e^{-\jmath 2\pi(M-1)(d/c)f_k\eta}\right]^T,\end{aligned} \quad (11)$$

with $\eta = \cos\theta - \cos\phi$. Since $\Psi_{\mathrm{STMV}}(\phi)$ cannot be simplified to a closed-form analytic solution, we study the spectrum for $\phi = \theta$ and $\phi \neq \theta$ independently. For $\phi = \theta$, (10) simplifies to

$$\mathbf{R}(\theta) = (f_{\mathrm{u}} - f_{\mathrm{l}})\left(\sigma_s^2 \mathbf{1}_M \mathbf{1}_M^T + \sigma_v^2 \mathbf{I}_M\right) \quad (12)$$

since $\mathbf{a}(k,\theta-\phi) = \mathbf{1}_M, \forall k$. The spatial spectrum is then obtained using the matrix inversion lemma as

$$\Psi_{\mathrm{STMV}}(\theta) = (f_{\mathrm{u}} - f_{\mathrm{l}})\left(\sigma_s^2 + \frac{\sigma_v^2}{M}\right). \quad (13)$$

For the case where $\phi \neq \theta$, we employ the following identity for any positive definite $M \times M$ matrix $\mathbf{C} > 0$ [14, pp. 452]

$$\left(\mathbf{1}_M^T \mathbf{C}^{-1}\mathbf{1}_M\right)^{-1} \mathbf{1}_M \mathbf{1}_M^T \leq \mathbf{C}. \quad (14)$$

Adopting this identity by pre- and post-multiplying (14) with $\mathbf{1}_M^T$ and $\mathbf{1}_M$ and $\mathbf{C} = \mathbf{R}(\phi)$, we note that

$$\begin{aligned}\Psi_{\mathrm{STMV}}(\phi) &\leq \frac{1}{M^2}\mathbf{1}_M^T\mathbf{R}(\phi)\mathbf{1}_M \\ &\leq \frac{1}{M^2}\mathbf{1}_M^T\left[\sigma_s^2\sum_{k=k_1}^{k_\mathrm{u}}\mathbf{a}(k,\theta-\phi)\mathbf{a}^H(k,\theta-\phi)+\sigma_v^2\mathbf{I}_M\right]\mathbf{1}_M \\ &\leq \frac{\sigma_s^2}{M^2}\sum_{f_k=f_1}^{f_\mathrm{u}}\sum_{i,j=0}^{M-1}e^{-\jmath 2\pi(i-j)(d/c)f_k\eta}+(f_\mathrm{u}-f_1)\frac{\sigma_v^2}{M}.\end{aligned} \tag{15}$$

Simplifying the summation term yields

$$\Psi_{\mathrm{STMV}}(\phi) \leq \sigma_s^2\sum_{f_k=f_1}^{f_\mathrm{u}}\frac{\mathrm{sinc}^2\left[M(d/c)f_k\eta\right]}{\mathrm{sinc}^2\left[(d/c)f_k\eta\right]}+(f_\mathrm{u}-f_1)\frac{\sigma_v^2}{M}. \tag{16}$$

Adding and subtracting $\sigma_s^2(f_\mathrm{u}-f_1)$ to the right-hand terms of (16) and using (13), we have

$$\Psi_{\mathrm{STMV}}(\phi) \leq \Psi_{\mathrm{STMV}}(\theta)-\sigma_s^2\sum_{f_k=f_1}^{f_\mathrm{u}}\left\{1-\frac{\mathrm{sinc}^2\left[M(d/c)f_k\eta\right]}{\mathrm{sinc}^2\left[(d/c)f_k\eta\right]}\right\}. \tag{17}$$

It is important to note that $\left|\frac{\mathrm{sinc}[M(d/c)f_k\eta]}{\mathrm{sinc}[(d/c)f_k\eta]}\right| = 1, \forall f_k$ when $(d/c)f_k\eta = 0, \forall f_k$ which implies that along the source direction, $\Psi_{\mathrm{STMV}}(\phi) = \Psi_{\mathrm{STMV}}(\theta)$ given by (13). Since aliasing position corresponding to the source direction varies with frequency as observed in (4), we have $\eta > 0$ for all other directions. Therefore, $\left|\frac{\mathrm{sinc}(M(d/c)f_k\eta)}{\mathrm{sinc}((d/c)f_k\eta)}\right| < 1$ results in $\Psi_{\mathrm{STMV}}(\phi) < \Psi_{\mathrm{STMV}}(\theta)$.

IV. PROPOSED MULTI-STAGE STMV ALGORITHM (STMV-MS)

From (4), we observe that the aliasing effect due to each frequency component occurs at integral multiples of $c/(f_k d)$. While aliasing effect can be avoided within $[0, 180^\circ]$ completely by choosing $d \leq \lambda_{\min}/2 = c/(2f_\mathrm{u})$, having $d > \lambda_{\min}/2$ allows aliasing effect due to some frequencies within the subband. However, it is important to note that a small subband within the speech bandwidth is alias-free although Nyquist criterion is violated. We identify this subband $[f_1, f_\mathrm{c}]$ with the critical frequency f_c defined as

$$f_\mathrm{c} = \frac{c}{2d}, \tag{18}$$

and $f_1 \leq f_\mathrm{c} \leq f_\mathrm{u}$. This frequency is derived from (4) by ensuring that the first aliasing component is situated at maximum distance from the source direction, i.e., $|\cos\phi - \cos\theta| = 2$. We estimate the PHAT-compensated STCM, denoted by $\widetilde{\mathbf{R}}_1(\phi)$, for the first subband $[f_1, f_\mathrm{c}]$ as

$$\widetilde{\mathbf{R}}_1(\phi) = \sum_{k=k_1}^{k_\mathrm{c}}\mathbf{T}^H(k,\phi)\left(\underline{\mathbf{x}}(k,l)\underline{\mathbf{x}}^H(k,l)\oslash\left|\underline{\mathbf{x}}(k,l)\underline{\mathbf{x}}^H(k,l)\right|\right)\mathbf{T}(k,\phi). \tag{19}$$

Solving the optimization problem in (7) employing $\widetilde{\mathbf{R}}_1(\phi)$ provides alias-free spatial response given by

$$\widetilde{\Psi}_{\mathrm{stmv},1}(\phi) = \frac{1}{\mathbf{1}_M^T\widetilde{\mathbf{R}}_1^{-1}(\phi)\mathbf{1}_M}, \tag{20}$$

where the lower-case subscript denotes multi-stage processing and the second subscript denotes the subband index. This low-resolution spatial spectrum provides a coarse estimate of the source directions that can be employed in the optimization criterion for the second subband $[f_\mathrm{c}, f_1]$ given by

$$\begin{aligned}&\underset{\mathbf{w}_\phi}{\text{minimize}} \quad \mathbf{w}_\phi^H\widetilde{\mathbf{R}}_2(\phi)\mathbf{w}_\phi \\ &\text{subject to} \quad \mathbf{w}_\phi^H\mathbf{1}_M = \left\{\widetilde{\Psi}_{\mathrm{ms},1}(\phi)\right\}^\alpha,\end{aligned} \tag{21}$$

where $\widetilde{\mathbf{R}}_2(\phi)$ denotes the PHAT-compensated STMV estimated for the second subband and α is a scaling factor which regulates the direction-dependent weight provided to the constraint. The spatial spectrum for the second stage STMV is then given by

$$\widetilde{\Psi}_{\mathrm{stmv},2}(\phi) = \frac{\left\{\widetilde{\Psi}_{\mathrm{stmv},1}(\phi)\right\}^{2\alpha}}{\mathbf{1}_M^T\widetilde{\mathbf{R}}_2^{-1}(\phi)\mathbf{1}_M}. \tag{22}$$

When d is substantially large giving $f_\mathrm{u} > 2f_\mathrm{c}$, a careful observation reveals that the subband $[f_\mathrm{c}, 2f_\mathrm{c}]$ is affected by only the first aliasing component corresponding to $\gamma = \pm 1$ in (4) where as the subband $[2f_\mathrm{c}, f_\mathrm{u}]$ is affected by components corresponding to $\gamma = \pm 2$. Generalizing this subband decomposition for an arbitrary value of d, we can decompose the speech bandwidth into $Q = \lceil f_\mathrm{u}/f_\mathrm{c}\rceil$. This decomposition constrains the maximum microphone spacing such that

$$d < \frac{c}{2f_1} = \frac{\lambda_{\max}}{2}. \tag{23}$$

This condition is essential for the first subband to have sufficient data for estimating an alias-free spectrum. Violating this condition will result in aliasing components intruding into the first subband $[f_1, 2f_\mathrm{u}]$.

Generalizing the STMV approach in (22) for Q-stage STMV, we obtain

$$\widetilde{\Psi}_{\mathrm{stmv},Q}(\phi) = \frac{\left\{\widetilde{\Psi}_{\mathrm{stmv},Q-1}(\phi)\right\}^{2\alpha}}{\mathbf{1}_M^T\widetilde{\mathbf{R}}_Q^{-1}(\phi)\mathbf{1}_M} = \prod_{q=1}^{Q}\frac{1}{\left\{\mathbf{1}_M^T\widetilde{\mathbf{R}}_q^{-1}(\phi)\mathbf{1}_M\right\}^{(2\alpha)^{Q-q}}}. \tag{24}$$

While the higher subband spatial spectra improve the resolution of $\widetilde{\Psi}_{\mathrm{stmv},Q}(\phi)$, aliasing components are suppressed by the lower subband spatial spectra which are less affected by aliasing. The spectrum $\widetilde{\Psi}_{\mathrm{stmv},Q}(\phi)$ therefore is expected to provide unambiguous DOA estimation.

V. SIMULATION RESULTS

The performance of STMV-MS is evaluated using both synthetic and recorded impulse responses. Microphone recordings of 1 s duration are sampled at 16 kHz and bandlimited to $[400, 4000]$ Hz and divided into time frames each of 2048 samples over which DOA estimation is performed independently and averaged subsequently.

Recorded room impulse responses acquired using a five-microphone ULA in a room of size 6 m×6 m×3 m with $d = 0.12$ m ($\approx 1.5\lambda_{\min}$). With this microphone spacing, $Q = 3$ subbands are constructed for STMV-MS. The two

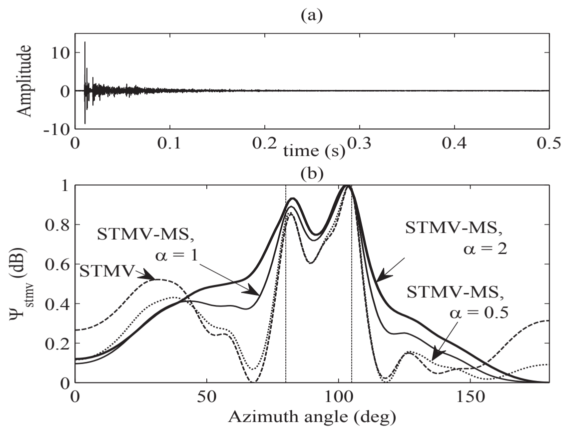

Fig. 1. (a) Recorded room impulse response acquired between the first source and first microphone and spatial spectra for (b) Spatial spectra of STMV and STMV-MS for $\alpha = 0.5, 1$ and 2.

sources are positioned at $80°$ and $105°$ w.r.t. the array axis, 3 m away from the array centroid. The reverberation time of the room impulse response is estimated to be 0.5 s. Two speech signals are convolved with the acquired impulse responses and additive white Gaussian noise is added to the received signal to achieve an SNR of 10 dB. The impulse response from the first source to the first microphone is shown in Fig. 1(a). The spatial spectra obtained using STMV and STMV-MS with $\alpha = 0.5, 1$ and 2 are normalized to unity and plotted in Fig. 1(b).

We note that STMV and STMV-MS resolve the two sources directions due to the large microphone spacing. However, an additional peak is observed in the spatial spectrum of STMV at $32°$ because of aliasing. This false peak is partially suppressed by STMV-MS with $\alpha = 0.5$ while the proposed algorithm successfully suppresses the undesired peak when $\alpha = 1$ and 2. We therefore infer that the direction-dependent weights provided by the previous subband spatial spectrum successfully suppresses the aliasing component when the scaling factor satisfies $\alpha \geq 1$.

Since STMV and STMV-MS have the same underlying working principle, the accuracy of DOA estimates obtained by STMV and STMV-MS are similar. We therefore evaluate only the performance of STMV and STMV-MS to identify the source direction peaks from the spectrum floor of their normalized spatial spectra. To quantify the distinction of source direction peaks from the spectrum floor, we employ an identifiability metric defined as [12]

$$\mathrm{I} = 10 \log_{10} \frac{\widehat{\Psi}(\theta)}{\mathrm{mean}\left\{ \widehat{\Psi}(\phi)\big|_{\phi \neq \theta} \right\}}, \tag{25}$$

where $\widehat{\Psi}(\phi)$ denotes the spatial spectrum normalized to unity and $\mathrm{mean}\left\{ \widehat{\Psi}(\phi)\big|_{\phi \neq \theta} \right\}$ is the spectrum floor mean of the normalized spatial spectrum over the azimuth range excluding the source directions. To assess the identifiability of STMV and STMV-MS, we perform a simulation using two speech sources randomly selected from a database of eighteen sources for each trial. Room impulse response is synthesized using the method of images [15] for two sources positioned at $70°$ and $100°$ w.r.t. a ULA of five microphones. DOA estimation is performed on the array outputs obtained at 20 dB using the STMV and STMV-MS. We evaluate the identifiability metric from the normalized spatial spectra averaged over 100 trials for various microphone spacings. The identifiability metric is plotted against d in Fig. 2 for $\alpha = 0.5, 1, 2$.

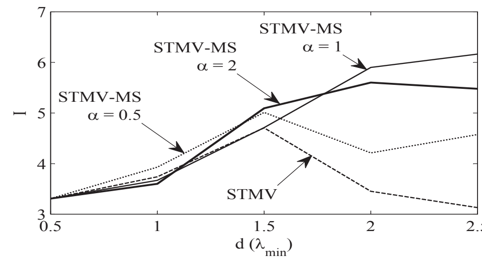

Fig. 2. Variation of identifiability metric with d for STMV and STMV-MS.

For STMV-MS with $\alpha = 0.5$, we note that I increases until $d = 1.5\lambda_{\min}$ and then decreases with further increase in d. This is because, the contribution of aliasing on the spatial spectrum is not significant for the source directions chosen. This is supported by the fact that I increases until $d = 1.5\lambda_{\min}$ even for the baseline STMV algorithm. For $d > 1.5\lambda_{\min}$, the aliasing components adversely affect the identifiability of the source direction peaks. We observe from Fig. 2 that the identifiability under such scenario is improved by STMV-MS for $\alpha \geq 1$. We therefore conclude that the proposed multi-stage STMV algorithm provides unambiguous DOA estimation when $\alpha \geq 1$.

VI. CONCLUSION

For wideband speech signals we showed that the aliasing effect results in a raised specturm floor due to its dependence on frequency. Noting that different frequency bands within the speech spectrum is influenced by different amount of alaising, we proposed a subband-based mulit-stage steered minimum variance method for unambiguous DOA estimation. Simulation results with recorded room impulse responses and the method of images exhibit an improvement in the identifiability of the source directions from the spatial spectrum of the new scheme by the suppression of aliasing components for scaling factors satisfying $\alpha \geq 1$.

VII. REFERENCES

[1] J. Krolik and D. Swingler, “Multiple broad-band source location using steered covariance matrices,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1481–1494, 1989.

[2] J. Dmochowski, J. Benesty, and S. Affes, “Direction of arrival estimation using the parameterized spatial correlation matrix,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1327–1339, 2007.

[3] J. Benesty, J. Chen, and Y. Huang, “Time-delay estimation via linear interpolation and cross correlation,” *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 5, pp. 509–519, 2004.

[4] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[5] W. Zhang and B. D. Rao, “A two microphone-based approach for source localization of multiple speech sources,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, pp. 1913–1928, Nov. 2010.

[6] F. Nesta and M. Omologo, “Generalized state coherence transform for multidimensional TDOA estimation of multiple sources,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 246 –260, Jan. 2012.

[7] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, “TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1490–1503, Aug. 2011.

[8] M. M. Hyder and K. Mahata, “Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation,” *IEEE Trans. Signal Process.*, vol. 58, pp. 4646–4655, Sept. 2010.

[9] Z. Tang, G. Blacquiere, and G. Lues, “Aliasing-free wideband beamforming using sparse signal representation,” *IEEE Trans. Signal Process.*, vol. 59, pp. 3464–3469, July 2011.

[10] J. Dmochowski, J. Benesty, and S. Affes, “On spatial aliasing in microphone arrays,” *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1383 –1395, Apr. 2009.

[11] C. Blandin, A. Ozerov, and E. Vincent, “Multi-source TDOA estimation in reverberant audio using angular spectra and clustering,” *Signal Process.*, vol. 92, no. 8, pp. 1950 – 1960, 2012.

[12] V. Reddy, A. Khong, and B. Ng, “Unambiguous speech DOA estimation under spatial aliasing conditions,” *IEEE/ACM Trans. Audio, Speech, and Language. Process.*, 2014.

[13] T. Gustafsson, B. Rao, and M. Trivedi, “Source localization in reverberant environments: Modeling and statistical analysis,” *IEEE Trans. Speech, Audio Process*, vol. 11, no. 6, pp. 791–803, 2003.

[14] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton University Press, 2009.

[15] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.