ROBUST AUDIO LOCALIZATION WITH PHASE UNWRAPPING

Kainan Chen[1], *Jürgen T. Geiger*[1], *Walter Kellermann*[2]

[1]Huawei German Research Center
Munich, Germany

[2]Chair of Multimedia Communications and Signal Processing
Friedrich-Alexander-Universität Erlangen-Nürnberg
Erlangen, Germany

ABSTRACT

Most of multichannel sound source Direction Of Arrival (DOA) estimation algorithms suffer from spatial aliasing problems. The phase differences between a pair of microphones are wrapped beyond the spatial aliasing frequency. A common solution is to adjust the distance between the microphones to obtain a suitable aliasing frequency, and take only the frequency band below the aliasing frequency for localization. With correct phase unwrapping, a broader frequency band can be utilized for localization. In this paper, we investigate a method for phase unwrapping solving the spatial aliasing problem for scenarios with a single source and high-level diffuse background noise (around 0dB SNR). The aliasing frequency is estimated from the signal, and is used to unwrap a phase difference vector. Pre- and post-processing steps are applied to increase the robustness. Our experiments with a large number of simulated and real signals demonstrate the robustness of our method in noise.

Index Terms— Sound Source Localization, DOA, Spatial Aliasing, Phase Unwrapping

1. INTRODUCTION

Multichannel sound source localization is an active research topic since several decades. Localization algorithms for sources in the far-field can be categorized into DOA and Time Difference Of Arrival (TDOA) estimation. Although state-of-the-art algorithms in both categories can localize very accurately in low-noise free-field conditions, it is still challenging to localize sound sources with interference, reverberation, and high-level noise background noise [1] [2] [3].

The spatial aliasing frequency is related to the distance between the microphones and the DOAs of the incident sources [4]. The simplest solution is to choose the distance between the microphones to cover the full frequency band within the aliasing frequency. This method is widely used in localization and also beamforming [4] [5] [6] [7] [8]. To increase the aliasing frequency, these methods reduce the distance between the microphones at the cost of accuracy.

Most of the narrow-band sound source localization algorithms that estimate the delay or phase difference between two microphones face the same problem of spatial aliasing when the energy of the source concentrates at high frequencies, e. g. MUltiple SIgnal Classification (MUSIC) [9] and Estimation of Signal Parameter via Rotational Invariance Technique (ESPRIT) [10]. For these algorithms, spatial aliasing results in phase wrapping. One method to overcome spatial aliasing is to unwrap the phase. Once the phase unwrapping problem is solved, the full frequency band (up to the Nyquist sampling rate) can be used.

As a basic phase unwrapping method, Itoh's algorithm [11] simply goes through all phase samples along the frequency axis, one by one, and, if the phase difference between the current and the next sample is larger than π (smaller than $-\pi$), subtracts 2π from or adds 2π to all following samples, respectively. This method works only for signals with sufficiently low levels of additive noise or other impairments. A more robust phase unwrapping algorithm uses a Kalman filter [12] and transforms the phase unwrapping problem into a state estimation problem. The estimated phase is unwrapped through a state space model for the phase function which ensures phase continuity. The main feature of this method is to combine phase unwrapping with simultaneous noise reduction, overcoming the drawback that other general methods have to implement the elimination of phase noise before phase unwrapping. The authors of [5] [13] propose a multi-stage DOA estimation where the received signal is first decomposed into subbands of equal width. An unambiguous low-accuracy DOA is estimated from the first subband. Then aliasing components are suppressed in the second subband, and this process is repeated for the next bands.

In general, all prior art can be categorized into three different categories. First, 2D phase unwrapping algorithms (e. g. in image processing), which try to solve a similar but different problem, for example [14] [12]. There, the original unwrapped phase map is not linear and the corresponding algorithms do not suffer from the lack of excitation energy at low frequencies. Second, sequential unwrapping algorithms, such as the one by Itoh [11] or Kalman filter [12]. They process the phase samples sequentially along the frequency axis. Third, algorithms that evaluate specific frequencies and not all the frequencies together, e. g. [5] [13].

While there are numerous concepts for localization in reverberant environments involving blind system identification [15] [16] [17] [18] [19], we only point out here that in [20], it is reported that the energy of reverberation is more distributed at low frequencies in normal acoustic environments. So extending the usable frequency range with our algorithm can also reduce the effect of reverberation, and it is confirmed by our experiments.

Our algorithm is designed for a scenario with a single point sound source and high-level non-stationary acoustical background noise. The goal of our method is to improve the localization robustness by utilizing a wider frequency range. The individual steps of our algorithm can be summarized as source subspace estimation, aliasing frequency estimation, wrapping direction estimation, post processing, and denoising. Each of these steps is performed directly for the entire frequency range, as opposed to, for example, in algorithms based on an IIR filter [21], where the unwrapping is done sequentially.

The rest of the paper is organized as follows. In Section 2, the

The research leading to these results has received funding from the European Commision Union Seventh Framework Programme (FP7/2007/2013) under grant agreement 607480 LASIE.

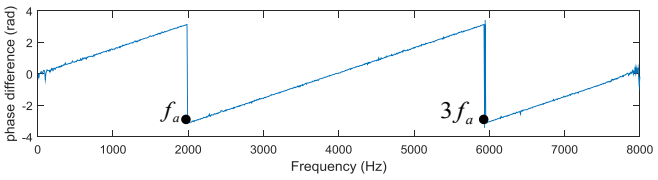

Fig. 1. Observed spatial aliasing in phase difference vector Φ. Target signal is white noise, DOA 60°; pink background noise, spatially white; SNR=20 dB; $\Delta d = 0.1$m

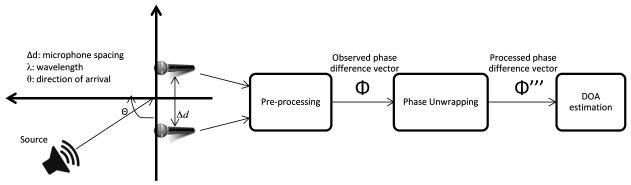

Fig. 2. System overview

problem is formulated. Our algorithm is described in Section 3. This is followed by Section 4, where simulated and real tests are shown and some result analysis is given. Finally in Section 5, we conclude our paper.

2. PROBLEM FORMULATION

The spatial aliasing frequency f_a is related to the distance between two microphones Δd and θ, where θ is the incident direction of the sound source [22],

$$f_a = \frac{|\sin\theta|\Delta d}{\pi}. \quad (1)$$

We take n as the n_{th} frequency bin at frequency f_n, $n \in [1, \ldots, N]$, and $f_N = \frac{f_s}{2}$, where f_s denotes the sampling rate. The observed phase differences Φ_n,

$$\Phi_n = \frac{2\pi f_n \sin\theta\ \Delta d}{c} \mod 2\pi \quad (2)$$

are mapped onto $(-\pi, \pi]$ for the algorithms that are mentioned in [4]. The estimated phase difference of the subband n is denoted as Φ_n, from which we form the phase difference vector $\mathbf{\Phi} = [\Phi_1, \ldots, \Phi_n, \ldots, \Phi_N]$. An example of observed $\mathbf{\Phi}$ with the spatial aliasing phenomenon is shown in Fig. 1. It can be seen that $\mathbf{\Phi}$ is wrapped at f_a and $3f_a$.

3. PROPOSED APPROACH

The system overview is shown in Fig. 2. A pair of microphones receives the source signal from DOA θ, which is the final result we want to estimate. Pre-processing can be any algorithm (e.g. ESPRIT [10]) that estimates the phase difference vector $\mathbf{\Phi}$. After our algorithm's reconstruction (unwrapping) and denoising, we obtain the clean phase difference vector $\mathbf{\Phi}'''$ for the final DOA estimation.

3.1. System Details

The details of our method are shown in Fig. 3. In the part of pre-processing of phase unwrapping, we first estimate the target source strength of the observed phase difference vector $\mathbf{\Phi}$, so we can roughly remove some frequency bands for better analysis. Then we estimate the aliasing frequency f_a and wrapping direction (the wrapping direction defines whether π has to be added or subtracted for unwrapping). Based on these two estimates, we can carry out the phase unwrapping.

Theoretically, after phase unwrapping, $\mathbf{\Phi}''$ should be a straight line assuming the source signal propagates in a free sound field. In

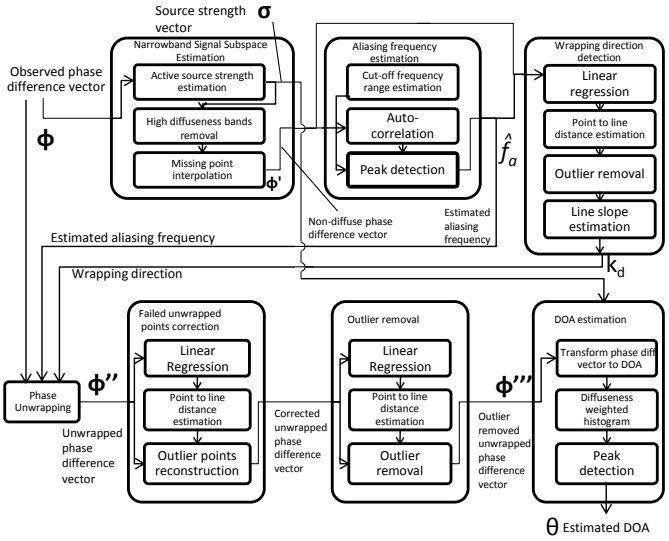

Fig. 3. System details

order to mitigate the effects of noise, we apply the two steps of failed unwrapped points correction and outlier removal. At the final step, we do not apply the conventional conversion of phase difference to DOA [10], but add some statistical processing, based on the estimated source strength vector, for further accuracy improvements.

3.2. Narrow-band Signal Subspace Estimation

The goal of this step is to remove those subbands that do not contain enough energy of the source subspace and are thus unusable for localization. Considering our targeted use case of a single source in diffuse background noise, a simple method is used here to find relevant frequency bins. The method is very similar to source subspace estimation methods introduced in [9] [10]. First, the eigenvalues of the covariance matrix of each narrow-band signal spectrum are obtained and normalized. We denote σ_i as the normalized largest eigenvalue for the i_{th} frequency band, as an element of the normalized target source strength vector $\boldsymbol{\sigma} = [\sigma_1, \sigma_i, \ldots, \sigma_N]$. We set a threshold τ_d, and decide that, if $\sigma_i < \tau_d$, the signal subspace component at the i_{th} frequency band is considered to be too weak and thus not relevant for localization (Instead of a simple fixed threshold, more sophisticated methods for noise threshold estimation could be used, such as minimum statistics [23]). After removing the detected noise bands, the removed points are interpolated using linear interpolation.

3.3. Aliasing Frequency Estimation

As a first step, the possible range of the aliasing frequency is determined, to preclude large errors when estimating the aliasing frequency $\hat{f}_a$. From Eq. (2) we know that when Δd is fixed, we can find the minimum aliasing frequency by $\min f_a = \frac{c}{\Delta d}$. Our ultimate goal is not to estimate perfectly the aliasing frequency, but to have all the wrapped phase observations unwrapped. Thus, given our post-processing (failed unwrapped points reconstruction), we do not always have to unwrap the phase difference. It was empirically found that the post-processing algorithm performs well as long as the number of wrapped phase observations remains below a certain threshold. Thus, we limit the estimation result in the range of $\frac{c}{\Delta d} < \hat{f}_a < \tau_a f_N, \tau_a \in [0.5, 1]$.

We apply auto-correlation to $\mathbf{\Phi}'$, and denote the resulting vector as $\mathbf{R}_\Phi$. $\mathbf{R}_\Phi$ is symmetrical, so we only consider half of it. Theoretically, if the vector $\mathbf{\Phi}'$ is wrapped, the peak of vector $\mathbf{R}_\Phi$ is at

0 in x-axis, the second peak is at the position of $2f_a$, and the minimum is at f_a. The positions of these points are periodical, as shown in Fig. 1. When f_a is in the range of $[\frac{f_N}{2}, f_N]$, $\mathbf{R}_\Phi$ does not show a full period, so we detect the minimum (rather than the peak) and use it as an estimate for the aliasing frequency $\hat{f}_a$. When $f_a > \frac{f_s}{2}$ (no phase wrapping occurs), the minimum of the autocorrelation of $\mathbf{\Phi}^{'}$ and $\hat{f}_a$ are not meaningful. Whether $\hat{f}_a$ is useful is determined as described in Section 3.6. If $\hat{f}_a$ is not meaningful, the steps of phase unwrapping are not performed, but only the step of removing outliers for denoising (Section 3.7).

3.4. Wrapping Direction Estimation

Before $\mathbf{\Phi}$ is unwrapped, it is determined in which direction it should be unwrapped, i.e., whether π should be added or subtracted('Wrapping Direction estimation'). Assuming the phase is wrapped ($f_a < \frac{f_s}{2}$) and noisy, we can not use linear regression or another simple method to estimate the wrapping direction. We extract two segments from $\mathbf{\Phi}^{'}$; the first one is from 0Hz to the estimated aliasing frequency $\hat{f}_a$ (at frequency bin $N_a = \frac{N\hat{f}_a}{f_s}$), and the second one is from N_a to $3N_a$. In order to improve the robustness to errors in the estimation of $\hat{f}_a$, we limit the frequency bin ranges of the segments to $[0, 0.75N_a]$ and $[1.4N_a, 2.6N_a]$, respectively. Then we further denoise these two segments again, with the same process on each segment. An example of the process on the first segment is described below.

First, we find the best fitting line for the segment using linear regression. The fitting line is expressed as

$$\hat{\Phi}_i^{'} = ki + b \ \ (b = 0 \ \text{ for } \ i < N_a). \tag{3}$$

Then we compute the vertical distance d_i of the i_{th} point to the corresponding fitting line,

$$d_i = \Phi_i^{'} - \hat{\Phi}_i^{'} \tag{4}$$

We consider a point to be noise and remove it if

$$\begin{cases} \frac{\text{card}\{d_i | d_i > 0\}}{N_a d_i} & < \tau_o \ \text{ if } \ d_i > 0 \\ \frac{\text{card}\{d_i | d_i \leq 0\}}{N_a d_i} & < \tau_o \ \text{ if } \ d_i \leq 0 \end{cases} \quad \tau_o \in (0, 1), \tag{5}$$

where card$\{\}$ denotes cardinality measure.

We apply linear regression again to each denoised segment; the resulting slopes are denoted as k_1 and k_2. Then we compute the averaged distance deviation for both segments, denoted as v_1 and v_2. Denoting the length of the segments as l_1 and l_2, we estimate the wrapping direction k_d by the sign of $k_1 f(v_1, l_1) + k_2 f(v_2, l_2)$. Here we define the weighting function as $f(v, l)$.

3.5. Phase Unwrapping

Given the frequency bin of the estimated aliasing frequency N_a and the wrapping direction k_d, we obtain the unwrapped phase difference vector as

$$\Phi_i^{''} = \Phi_i + 2k_d \lfloor \frac{N_a + i}{2N_a} \rfloor \pi. \tag{6}$$

Here $\lfloor x \rfloor$ denotes the floor function of x and $\mathbf{\Phi}^{''} = [\Phi_1^{''}, \ldots, \Phi_i^{''}, \ldots, \Phi_N^{''}]$ is the unwrapped phase difference vector.

3.6. Failed Unwrapped Points Correction

The method described in this section is simple, but very efficient. An example with failed unwrapping points is given in Fig. 4, and we want to additionally unwrap them. The failed unwrapped points detection is similar to the denoising part in Section 3.4 – first find the best fitting line for $\mathbf{\Phi}^{''}$, and denote k as the slope of the fitting line. Then we correct the failed unwrapped points by,

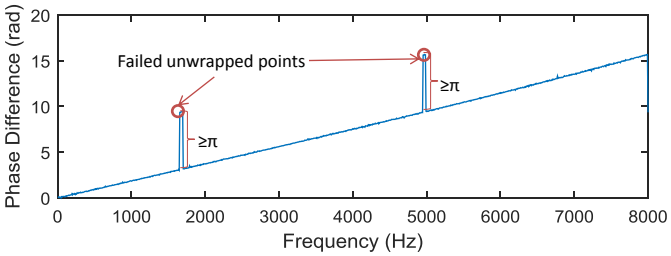

Fig. 4. Failed unwrapping points

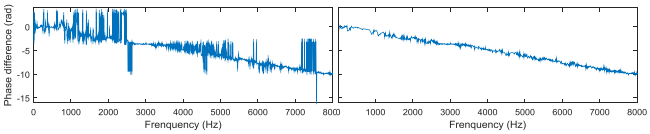

Fig. 5. Before and after failed unwrapping points reconstruction and outlier removal denoising

$$\hat{\Phi}_i^{''} = \Phi_i^{''} + \text{sign}(\Phi_i^{''} - ki)2\pi, \text{ if } |\Phi_i^{''} - ki| > 2\pi \tag{7}$$

We also check whether it is necessary to do phase unwrapping by

$$C = \sum_{i=1}^{N} |\Phi_i| - \tau_c \sum_{i=1}^{N} |\hat{\Phi}_i^{''} - ki|, \tag{8}$$

here τ_c denotes a constant parameter. If $C > 0$, we assume what we did was correct and continue with the next steps with $\hat{\mathbf{\Phi}}^{''}$, otherwise we use the observed phase difference vector $\mathbf{\Phi}$ for the following steps. This helps to avoid large errors in the case of small DOA (signal arriving from the front), when f_a is close to $f_s/2$.

3.7. Outlier Removal for Denoising

The steps to find outliers are the same as the part in Section 3.4, and all of the points for which the distance is higher than the threshold τ_k are removed from $\hat{\mathbf{\Phi}}^{''}$. Then we estimate the error by the sum of all distances from the points to the fitting line divided by the number of points. The process of detecting and removing outliers is iteratively repeated until the error is sufficiently small (averaged distance is smaller than τ_o, ($\tau_o > 0$) or up to 50% of the points are removed. Fig. 5 shows an example before and after failed unwrapping points reconstruction and outlier removal denoising. The result of this step is denoted as $\mathbf{\Phi}^{'''}$.

3.8. From Processed Phase Difference Vector to DOA

In the final step, we transform $\mathbf{\Phi}^{'''}$ to DOA by the conventional method as described in [10]. We compute the histogram of the DOAs from all of the narrow-bands using the method described in [24], and additionally the narrow-band results are weighted by the target source strength vector $\boldsymbol{\sigma}$. After applying a first order moving average filter to the histogram, the location of the highest peak shows the final estimated DOA.

4. EVALUATION

The evaluation is done by both simulated data and real recordings with many different types and energy levels of sources from different DOAs, as well as different noise types. We use a uniform linear array (ULA) with 5 microphones for both simulations and recordings. The distance between the closest microphones Δd is 0.1m.

4.1. Parametrisation

We implemented our method in the short-time Fourier transform (STFT) domain with window length 0.2s, hop size 0.1s, sampling

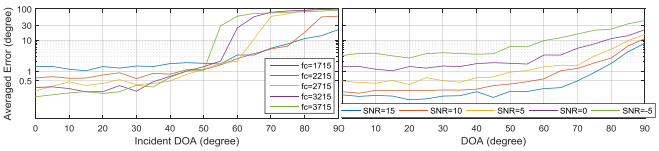

Fig. 6. Simulation results based on ESPRIT, with different f_c, SNR=0dB (left) and different SNR, $f_c = 1715$Hz (right).

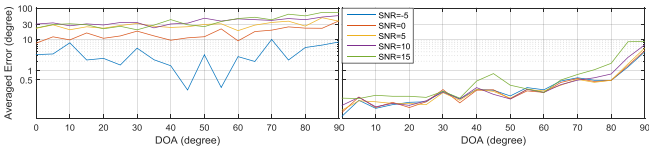

Fig. 7. Simulation results based on ESPRIT, for ESPRIT with Itoh's method (left) and our method (right).

frequency $48k$Hz and FFT size 16384. For most kinds of energy distributions of sound sources, we take only the frequency bins below $8k$Hz for localization.

We have many functions combined, so the thresholds and parameters settings are very important for the system's robustness. We used only a small set of signals to optimize the parameters $\tau = [\tau_d\ \tau_a\ \tau_k\ \tau_c\ \tau_o]$ and the function $f(l, v)$. All of the following experiments in real-recordings and simulations use the parameters $\tau = [0.3\ 0.37\ 0.75\ 1\ 0.5]$, $f(l, v) = \frac{l}{v^2}$.

4.2. Simulation Results

The signals were generated with different SNR at 15,10,5,0,-5dB, and different DOAs of the incident plane wave source from the side (90°) to the front (0°) in free field scenario. The target source signal is white noise, and the background noise is incoherent (spatially white) pink. Pink noise was chosen because it better resembles the background noise in a realistic scenario, such as a crowded public space. Furthermore, reverberation resulting from background noise tends to have a similar shape as pink noise [20]. For each DOA and each SNR, a 3s test data set is generated.

An experimental analysis on how the cutoff frequency f_c affects the DOA estimation is shown in Fig. 6 on the left side, averaged over all blocks and expressed in terms of absolute DOA errors in degrees. In this experiment, $f_a = 1715Hz$, we can see that if we take $f_c > f_a$ the errors from off-broadside angles increase. The right side of Fig. 6 shows how SNR affects the localization. We can see that the accuracy degrades with lower SNR.

The comparison between the Itoh algorithm [11] and our method is shown in Fig. 7. We can see that our method is much better than [11] in all of the tests. Comparing to the ESPRIT with fixed $f_c = 1715$Hz, generally, without unwrapping, the accuracy degrades when the SNR is low and the source is closer to the side (more than 40°). Our phase unwrapping method reconstructs the phase difference vector. Although the error increases towards the sides, its average is within reasonable limits (1°). In [18] [25], for a single pair of microphones or ULA, the accuracy for localizing the sources from off-broadside is lower, because the spatial aliasing problem limits the usable frequency range when DOAs go from 0° to $\pm 90^\circ$. With our algorithm, the results show that the errors are low for all the directions.

4.3. Real Recording Results

The recording setup is the same as for the experiments in the simulation (now using 5 microphones of type AKG C562CM in the center of a lab room with measured reverberation time T60 of 0.28s). The

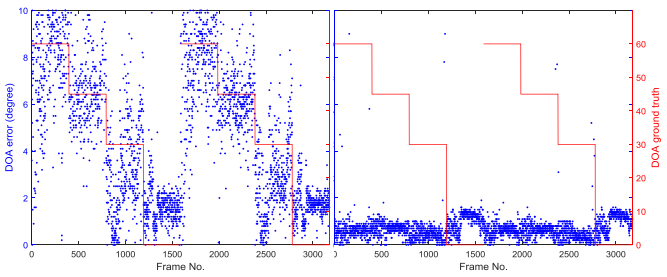

Fig. 8. DOA error for the real recordings, for the baseline ESPRIT with fixed $f_c = 1715Hz$ (left) and our method (right).

Algorithm	Ave Err, ($Err < 10^\circ$)	% of Err Frames ($Err > 10^\circ$)
ESPRIT $f_c 1.7k$	4.5°	11.41%
ESPRIT $w.PU$	0.86°	3.61%

Table 1. Evaluation results

background signals rendered by 22 surround speakers are recorded in Marienplatz in Munich, Germany and on a bridge above Isar river using an EigenMike [26]. The target signals include 15 speech recordings from the GRID corpus [27] and 37 events of the types breaking glass, gunshot, screaming, dog barking, and all kinds of alarms. Target sources are played from single speakrs with 3 to 4 meters distance to the microphone array, so we can assume the waves were plane when they arrived to the microphones. The SNR is estimated using free-field conditions and the plane wave assumption, neglecting reverberation of the target sources. The resulting SNR estimates are in the range of $[-5, 10]$dB.

We analyzed 3176 windows of size 16384. Fig. 8 shows the ESPRIT DOA estimation result and ESPRIT with our phase unwrapping algorithm, (the errors in 10°). The first half of each figure (windows $1 - 1588$) is with the background of Marienplatz, and the second half (windows $1589 - 3176$) for the scenario of the bridge. The ground truths of the source DOA are shown in the red axis. The overall error for the original ESPRIT is higher than that in the simulation, but is significantly reduced with our algorithm. We observe that the effect of noise is reduced when a wider frequency band is used with our phase unwrapping. We can also see for the original ESPRIT that the errors from the sides are much higher than from the front, but with our phase unwrapping, they are consistently low for all the DOAs and both types of background noise.

Evaluation results are summarized in Table 1. $f_c 1.7k$ denotes cut-off frequency at 1715Hz and $w.PU$ denotes with phase proposed phase unwrapping. We analyze the error rate when the localization works, and how often the localization does not work (error $> 10^\circ$). From the result we can find that the average estimation error is reduced by a factor of 5 with our phase unwrapping, and our algorithm failed three times less frequently than the original ESPRIT. The experiments also showed that the chosen values for τ were robust.

5. CONCLUSIONS

We presented an approach to solve the spatial aliasing problem by a robust phase unwrapping algorithm for a single point source scenarios. The results obtained with simulated and real signals show that, with our algorithm, localization algorithms such as ESPRIT, work with higher accuracy for various kinds of sound sources under diffuse noise. Future work will be directed towards applying our method to multiple point sources.

6. REFERENCES

[1] J. Valin, F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," *Robotics and Autonomous Systems*, vol. 55(3), pp. 216 – 228, 2007.

[2] A. Ihlefeld and B. Shin-Cunningham, "Effect of source spectrum on sound localization in an everyday reverberant room," *The Journal of the Acoustical Society of America*, vol. 130(1), pp. 324 – 333, 2011.

[3] J. H. DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*, Ph.D. thesis, Brown University, 2000.

[4] J. Dmochowski, J. Benesty, and S. Affès, "On spatial aliasing in microphone arrays," *IEEE Transactions on Signal Processing*, vol. 57(4), pp. 1383 – 1395, 2009.

[5] V. V. Reddy, A. W. Khong, and B. P. Ng, "Unambiguous speech DOA estimation under spatial aliasing conditions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22(12), pp. 2133 – 2145, 2014.

[6] M. Amin, K. Ahmed, and Z. Chowdhury, "Estimation of direction of arrival (DOA) using real-time array signal processing," in *International Conference on Electrical and Computer Engineering*, 2008.

[7] J. Dmochowski, J. Benesty, and S. Affès, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15(4), pp. 1327–1339, 2007.

[8] J. Nikunen and T. Virtanen, "Direction of arrival based spatial covariance model for blind sound source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22(3), pp. 727 – 739, 2014.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions*, vol. 34(3), pp. 276 – 280, 1986.

[10] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions*, vol. 37(7), pp. 984–995, 1989.

[11] K. Itoh, "Analysis of the phase unwrapping algorithm," *Applied Optics*, vol. 21(14), pp. 2470, 1982.

[12] R. Krämer and O. Loffeld, "Presentation of an improved phase unwrapping algorithm based on kalman filters combined with local slope estimation," *ERS SAR Interferometry*, vol. 406, pp. 253, 1997.

[13] V. V. Reddy and A. W. Khong, "Direction-of-arrival estimation of speech sources under aliasing conditions," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[14] G. Nico, G. Palubinskas, and M. Datcu, "Bayesian approaches to phase unwrapping: theoretical study," *IEEE Transactions on Signal processing*, vol. 48(9), pp. 2545 – 2556, 2000.

[15] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *the Journal of the Acoustical Society of America*, vol. 107(1), pp. 384–391, 1999.

[16] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann, "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2005.

[17] A. J. V. Lombard, *Localization of Multiple Independent Sound Sources in Adverse Environments*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, 2012.

[18] 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 Transactions on Audio, Speech, and Language Processing*, vol. 19(6), pp. 1490 – 1503, 2011.

[19] F. Nesta and M. Omologo, "Cooperative Wiener-ICA for source localization and separation by distributed microphone arrays," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.

[20] J. Backus and R. Baskerville, *The Acoustical Foundations of Music*, W W Norton, 2nd edition, 1977.

[21] M. Navarro, J. Estrada, M. Servin, J. Quiroga, and J. Vargas, "Fast two-dimensional simultaneous phase unwrapping and low-pass filtering," *Optics express*, vol. 20(3), pp. 2556 – 2561, 2012.

[22] H. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*, John Wiley & Sons, 2004.

[23] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on speech and audio processing*, vol. 9(5), pp. 504–512, 2001.

[24] B. Ottersten and T. K, "Direction-of-arrival estimation for wide-band signals using the esprit algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 38(2), pp. 317 – 327, 1990.

[25] H. Teutsch, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, 2005.

[26] mh acoustics LLC, "www.mhacoustics.com," 2016.

[27] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *The Journal of the Acoustical Society of America*, vol. 120(5), pp. 2421 – 2424, 2006.