Improved Direct-path Dominance Test for Speaker Localization in Reverberant Environments

Lior Madmoni and Boaz Rafaely
The Department of Electrical and Computer Engineering,
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Emails: liormad@post.bgu.ac.il; br@bgu.ac.il.

***Abstract*—Speaker localization in real environments is a fundamental task for many audio signal processing applications. Many localization methods fail when the environment imposes challenging conditions, such as reverberation. Recently, a method for direction of arrival (DOA) estimation of speakers in reverberant environments was developed, which utilizes spherical arrays. This method uses the direct-path dominance (DPD) test to select time-frequency bins that contain spatial information on the direct sound. In this work, it is shown that when the threshold of the DPD test is lowered to select more bins for the estimation process, it falsely identifies bins dominated by reverberant sound, reducing DOA estimation accuracy. In this paper, a new DPD test is developed, which evaluates the extent to which the measured plane-wave density can be represented by a single plane-wave. While being more computationally expensive than the original test, it is more robust to reverberation, and leads to an improved DOA estimation. The latter is demonstrated by simulations of a speaker in a reverberant room.**

***Index Terms*—Speaker localization, reverberation, spherical arrays.**

I. Introduction

Direction-of-arrival (DOA) estimation of speakers is of great interest in the audio signal processing community, and is employed in tasks such as noise reduction, interference suppression, source tracking, and natural human-robot interaction. Approaches for DOA estimation include subspace algorithms [1], beamforming [2], and time-delay estimation [3]. When introducing reverberation, which is present in many real life scenarios, the estimation becomes more challenging, because the recorded speech signal is comprised of both the direct sound from the source and reflections from other directions.

Recently, approaches for DOA estimation under reverberation were developed, that extract information arriving directly from the source, by learning the relative transfer function between microphones [4], [5], or by applying the direct-path dominance (DPD) test to identify time-frequency (TF) bins in the short-time Fourier transform (STFT) domain that are dominated by the direct sound [6]–[9]. The latter studies did not require learning of transfer functions, and were based on measurements from a spherical microphone array, utilizing the spherical harmonics (SH) domain. This enables frequency smoothing without the use of focusing matrices, in order to decorrelate coherent signals. Recent extensions of this approach were presented that are robust to variations in the acoustic environment [10], and that are more computationally efficient [11].

In this paper, it is shown that the DPD test, originally developed in [6], can falsely select TF bins dominated by reverberant sound. This may degrade the performance of DOA estimation by the MUltiple SIgnal Classification (MUSIC) algorithm [1], which is then employed on TF bins that are contaminated by reflections. After demonstrating the shortcomings of the original DPD test, a new DPD test is developed with the aim of overcoming the limitations of the original test. While the original DPD test examines the ratio between the first two singular values of the cross-spectrum matrix, the new DPD test measures more explicitly the extent to which the measured data can be modeled by a single plane-wave. It is shown that TF bins that pass the new test are composed mainly of the direct sound. Then, a computer simulation of DOA estimation demonstrates the superiority of the new test.

II. Overview of the DPD test

This section overviews the DPD test [6], used for speaker localization in reverberant environments.

Consider a spherical array comprised of Q omni-directional microphones with the sphere's center located at the origin of a spherical coordinate system denoted by (r, θ, ϕ). The microphones are located at $\{\Omega_q \equiv (\theta_q, \phi_q)\}_{q=1}^{Q}$. Assuming the sound field is comprised of L far field sources at wavenumber k, and with DOAs of $\{\Psi_l = (\theta_l, \phi_l)\}_{l=1}^{L}$, the measured pressure can be described as [12]

$$\mathbf{p}(k) = \mathbf{V}(k, \mathbf{\Psi})\mathbf{s}(k) + \mathbf{n}(k), \tag{1}$$

where $\mathbf{p}(k) = \left[p(k, r, \Omega_1), p(k, r, \Omega_2), \ldots, p(k, r, \Omega_Q)\right]^T$ is a $Q \times 1$ vector containing the pressure at the microphones, $\mathbf{s}(k) = \left[s_1(k), s_2(k), \ldots, s_L(k)\right]^T$ is an $L \times 1$ vector holding the amplitudes of the signals, $\mathbf{V}(k, \mathbf{\Psi})$ is a $Q \times L$ matrix with columns containing the steering vectors between each source and microphone [12], $\mathbf{n}(k) = \left[n_1(k), n_2(k), \ldots, n_Q(k)\right]^T$ is a $Q \times 1$ vector of noise components, and $(\cdot)^T$ is the transpose operator. Since the array is of spherical geometry, the measurements can be transformed to the SH domain [13]–[15], and decomposed into plane-wave density (PWD) coefficients [6]:

$$\mathbf{a_{nm}}(k) = \mathbf{Y}^H(\mathbf{\Psi})\mathbf{s}(k) + \tilde{\mathbf{n}}(k), \tag{2}$$

where the $(N+1)^2 \times 1$ vector $\mathbf{a_{nm}}(k) = \left[a_{00}(k), a_{1(-1)}(k), a_{10}(k), \ldots, a_{NN}(k)\right]^T$ holds the

ISBN 978-90-827970-1-5 © EURASIP 2018

noisy PWD coefficients in the SH domain, $\mathbf{Y}^H(\mathbf{\Psi}) = \left[\mathbf{y}^*(\Psi_1), \mathbf{y}^*(\Psi_2), \ldots, \mathbf{y}^*(\Psi_L)\right]$ is an $(N+1)^2 \times L$ matrix with columns $\mathbf{y}(\Psi_l) = \left[Y_0^0(\Psi_l), Y_1^{-1}(\Psi_l), \ldots, Y_N^N(\Psi_l)\right]^T$, which hold the SH functions $Y_n^m(\cdot)$ of order n and degree m, and with N denoting the maximal SH order, usually set to $N = \lceil kr \rceil$ and satisfying $(N+1)^2 \leq Q$ [15]. $\tilde{\mathbf{n}}(k)$ is an $(N+1)^2 \times 1$ vector holding the measurement noise components in the SH domain, $(\cdot)^*$ is the complex conjugate operator, and $(\cdot)^H$ is the Hermitian operator.

Next, the model in (2) is transformed to the STFT domain with (τ, ω) as the time and frequency indices, respectively:

$$\mathbf{a_{nm}}(\tau, \omega) = \mathbf{Y}^H(\mathbf{\Psi})\mathbf{s}(\tau, \omega) + \tilde{\mathbf{n}}(\tau, \omega). \tag{3}$$

As discussed in [6], in order to overcome the effects of coherent room reflections, local TF correlation matrices are computed as the sample correlation matrices averaged over frequency:

$$\begin{aligned}\tilde{\mathbf{R}}_a(\tau, \omega) =& \frac{1}{J_\tau J_\omega} \sum_{j_\omega=0}^{J_\omega-1} \sum_{j_\tau=0}^{J_\tau-1} \mathbf{a_{nm}}(\tau - j_\tau, \omega - j_\omega) \\ & \times \mathbf{a_{nm}}^H(\tau - j_\tau, \omega - j_\omega),\end{aligned} \tag{4}$$

where J_τ and J_ω are the number of time and frequency bins for the averaging, respectively.

Next, the singular value decomposition (SVD) of the correlation matrix in (4) is computed. If $\tilde{\mathbf{R}}_a(\tau, \omega)$ is dominated by a single source (assumed to be the direct sound), its first singular value should be dominant as well. This lead to the DPD test:

$$\mathcal{A}_{\text{DPD}} = \left\{(\tau, \omega) : \frac{\sigma_1(\tau, \omega)}{\sigma_2(\tau, \omega)} > \mathcal{TH}_{\text{DPD}}\right\}, \tag{5}$$

where $\mathcal{A}_{\text{DPD}}$ is a set of TF bins that passed the DPD test, $\sigma_1(\tau, \omega)$ and $\sigma_2(\tau, \omega)$ are the largest and second largest singular values of $\tilde{\mathbf{R}}_a(\tau, \omega)$, respectively, and $\mathcal{TH}_{\text{DPD}}$ is a threshold to be chosen sufficiently larger than 1.

In the final stage, the MUSIC algorithm is applied to the bins that passed the DPD test. First, the MUSIC spatial spectrum is computed:

$$S_{\text{MUSIC}}(\Omega, \tau, \omega) = \frac{1}{||\mathbf{U}_n^H(\tau, \omega)\mathbf{y}^*(\Omega)||^2}, \quad \forall(\tau, \omega) \in \mathcal{A}_{\text{DPD}} \tag{6}$$

where Ω are all possible DOAs, and $\mathbf{U}_n(\tau, \omega)$ is the noise subspace of $\tilde{\mathbf{R}}_a(\tau, \omega)$ [1], assuming a single source. DOA estimation is performed next by finding the direction Ω corresponding to the peak of the spatial spectrum

$$\Omega_{\text{DPD}} = \left\{\Omega : \arg\max_{\Omega} S_{\text{MUSIC}}(\Omega, \tau, \omega) \quad \forall(\tau, \omega) \in \mathcal{A}_{\text{DPD}}\right\}. \tag{7}$$

Finally, a single DOA estimate can be produced by taking the mean value of the set Ω_{DPD}.

In [6], it was suggested to perform a further step of clustering, to facilitate DOA estimation of multiple speakers. Clustering was also applied in [16] to improve the accuracy of estimation with a single speaker. Gaussian mixture model (GMM) with five Gaussians was employed, such that the final DOA estimate is given by the mean of the dominant cluster. This clustering method was shown to significantly improve the estimation accuracy under challenging reverberation conditions.

III. Failure of the DPD test

In this section, some scenarios for which the DPD test may fail are presented, highlighting the limitations of the original test. While the DPD test has been shown to be robust to reverberation [6], matrix $\tilde{\mathbf{R}}_a(\tau, \omega)$ that is dominated by room reflections rather than the direct sound may still yield a relatively high ratio of $\frac{\sigma_1(\tau,\omega)}{\sigma_2(\tau,\omega)}$, allowing TF bins dominated by reverberant sound to pass the test. In other words, although a single plane-wave sound field leads to a high singular-value ratio, this may be a sufficient condition but not a necessary one, as will be demonstrated in the following example. Consider the case where the TF bins used for the averaging of matrix $\tilde{\mathbf{R}}_a(\tau, \omega)$ are dominated by a few bins that have a high power, but which represent a reverberant sound field. In this case, the SVD of matrix $\tilde{\mathbf{R}}_a(\tau, \omega)$ may yield a single high power singular vector representing the reverberant sound field, with the remaining singular vectors having smaller singular values. The corresponding MUSIC spectrum will therefore point to a DOA that may not correspond to the source, which will in turn add bias to the DOA estimation. To overcome this phenomenon, $\mathcal{TH}_{\text{DPD}}$ could be set to a sufficiently high value. However, raising $\mathcal{TH}_{\text{DPD}}$ may lead to only a small number of bins passing the test, which may require a long signal duration to accumulate a set of bins that is reliable for DOA estimation. A method to reduce the bias due to a relatively low $\mathcal{TH}_{\text{DPD}}$ is to use GMM clustering, as was described in the previous section.

An example illustrating the failure of the DPD test is presented next. A room of dimensions $8 \times 5 \times 3\,\text{m}$ was simulated using the image method, with reverberation time of $T_{60} = 1\,\text{s}$, and a critical distance of $0.6\,\text{m}$. A spherical array of $Q = 32$ nearly-uniformly distributed microphones [15] was positioned $2\,\text{m}$ away from a speaker that was modeled by a point source. Speech signals taken from the TIMIT database [17] were played at the source position, and an STFT was applied to the microphone signals with a Hann window of 512 samples and with 50% overlap. In addition, sensor noise with signal-to-noise ratio of $40\,\text{dB}$ was added. Matrix $\tilde{\mathbf{R}}_a(\tau, \omega)$, as in (4), was calculated for each TF bin with $J_\tau = 2,\ J_\omega = 15$, and the DPD test was performed as described in (5) with $\mathcal{TH}_{\text{DPD}} = 4$, such that 3,000 bins passed the test out of approximately 45,000 bins for the 5 seconds long speech signal. The clean and reverberant speech spectrograms are presented in Fig. 1(a) and Fig. 1(b), respectively. The set $\mathcal{A}_{\text{DPD}}$ is marked on Fig. 1(b) by black color. Notice that some bins, such as those circled by the purple-color ellipse, are located at relatively reverberant regions on the spectrogram, indicated by the smeared speech signal. On the other hand, some bins, such as those circled by the pink-color ellipse, are located at the onsets of the speech signal, i.e., regions with a relatively high

ISBN 978-90-827970-1-5 © EURASIP 2018

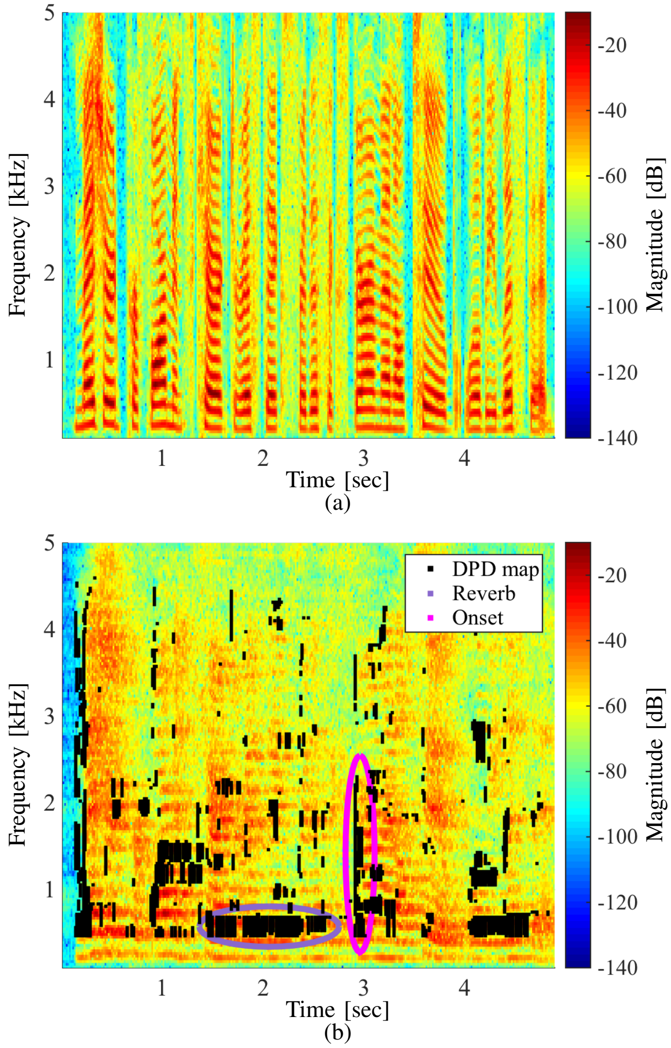

Fig. 1. Spectrogram of (a) clean speech and (b) reverberant speech. The set $\mathcal{A}_{\text{DPD}}$, as in (5) with $\mathcal{TH}_{\text{DPD}} = 4$ is colored in black. Bins that are circled by the pink ellipse correspond to the onset of speech, while bins that are circled by the purple ellipse correspond to reverberant speech.

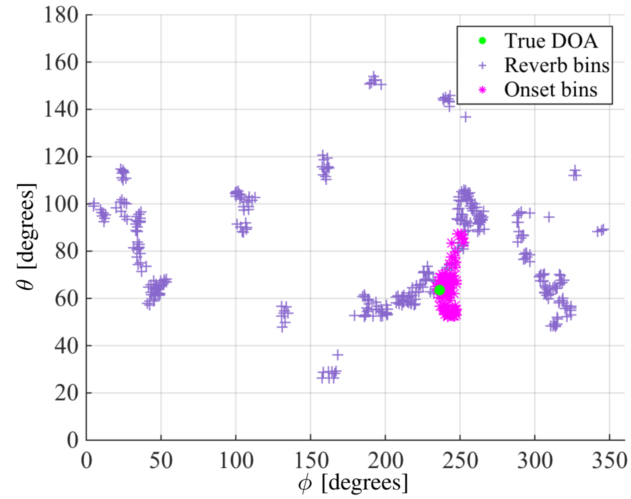

Fig. 2. DOA estimates from bins dominated by reverberant sound (purple), and bins close to the onsets of speech (pink), corresponding to the marked bins in Fig.1(b). The true DOA is marked by a green dot.

energy that appear after a period of silence. The DOAs from these two sets were estimated using (7), and are presented on a (θ, ϕ) map in Fig. 2. They are marked with the corresponding colors, i.e., purple and pink. The true DOA of the source is also marked on the figure, with a green dot. It is clear that the bins dominated by reverberant sound are less likely to carry source direction information, since the corresponding DOAs are more sparsely scattered, compared to the estimates from the onset bins, which are clustered more tightly round the true DOA. This example motivates the development of an improved DPD test that is more robust to reverberation, as presented next.

IV. A NEW DPD TEST BASED ON PROXIMITY TO A SINGLE PLANE-WAVE

In this section, a new DPD test is presented. The same system model as described in Section II is assumed for this test as well. Matrix $\tilde{\mathbf{R}}_a(\tau, \omega)$ is then calculated similarly to (4), and its SVD is computed. Let $\mathbf{u}_1(\tau, \omega)$ denote the singular vector of $\tilde{\mathbf{R}}_a(\tau, \omega)$ corresponding to the highest singular value. Next, the proximity of $\mathbf{u}_1(\tau, \omega)$ to a single plane-wave representation, i.e., $\mathbf{y}^*(\Omega)$, is computed for all possible directions, and the value corresponding to the direction with the closest proximity is selected as the proximity measure. The following MUSIC-based measure is used to compute the proximity to a single plane-wave (SPW):

$$SPW(\tau, \omega) = \max_{\Omega} \frac{1}{\left\| \mathbf{P}^{\perp}_{\mathbf{u}_1(\tau,\omega)} \mathbf{y}^*(\Omega) \right\|^2}, \tag{8}$$

where

$$\mathbf{P}^{\perp}_{\mathbf{u}_1(\tau,\omega)} = \mathbf{I} - \frac{\mathbf{u}_1(\tau, \omega)\mathbf{u}_1^H(\tau, \omega)}{\|\mathbf{u}_1(\tau, \omega)\|^2} \tag{9}$$

is the projection to the subspace which is orthogonal to $\mathbf{u}_1(\tau, \omega)$, and $\mathbf{I}$ is the identity matrix of size $(N+1)^2$. The measure proposed in (8) can be replaced by other scale-invariant measures for the proximity of two vectors, but this is left for future research. Motivated by (8), the new test is called the DPD-SPW (direct path dominance - single plane wave), and is now formulated as

$$\mathcal{A}_{\text{DPD-SPW}} = \left\{ (\tau, \omega) : SPW(\tau, \omega) > \mathcal{TH}_{\text{DPD-SPW}} \right\}, \tag{10}$$

where $\mathcal{A}_{\text{DPD-SPW}}$ is the set of TF bins that passed the DPD-SPW test, and $\mathcal{TH}_{\text{DPD-SPW}}$ is a threshold to be chosen. Since the denominator of (8) is in the range of $[0, 1]$, with zero suggesting that $\mathbf{u}_1(\tau, \omega)$ represents a SPW, the threshold should be chosen to satisfy $\mathcal{TH}_{\text{DPD-SPW}} \gg 1$. Finally, the DOA estimation from each bin is given by

$$\Omega_{\text{DPD-SPW}} = \Big\{ \Omega : \arg\max_{\Omega} \frac{1}{\left\| \mathbf{P}^{\perp}_{\mathbf{u}_1(\tau,\omega)} \mathbf{y}^*(\Omega) \right\|^2}, \; \forall (\tau, \omega) \in \mathcal{A}_{\text{DPD-SPW}} \Big\}. \tag{11}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

Note that the DOA estimation from each bin can be calculated already in (8), as the direction Ω that corresponds to the selected maximum value, and so (11) does not require new computations.

The proposed DPD-SPW test is different from the original DPD test, as discussed and demonstrated next. Assuming the sound field is comprised of only the direct sound, $\tilde{\mathbf{R}}_a(\tau,\omega)$ will be dominated by a single singular vector representing a single plane-wave, and both the DPD and DPD-SPW tests will yield a high score. However, if $\tilde{\mathbf{R}}_a(\tau,\omega)$ is dominated by a single singular vector, which does not represent the direct sound, but rather reverberant sound composed of a superposition of plane-waves, the DPD-SPW test will yield a low score, due to the dissimilarity to a single plane-wave (demonstrated in Section V). On the other hand, as was demonstrated in Section III, the original DPD test may still score high for this case. This suggests that a single plane-wave sound field is a necessary condition to pass the DPD-SPW test, compared to a sufficient condition to pass the original DPD test, although a more comprehensive theoretical analysis is proposed for future study.

It is important to note that the new algorithm for DOA estimation based on (11) is more computationally expensive than the original algorithm based on (7). This is because it requires a search over Ω for all TF bins, whereas the algorithm in (7) requires a search over Ω only for TF bins that passed the DPD test in (5). However, the cost of computation may be balanced by an improvement in performance, as presented next.

V. Simulation Study

In this section, a simulation study comparing the new DPD-SPW test, described in Section IV, and the original DPD test, described in Secion II, is presented. The aim is to evaluate the robustness of the DPD-SPW test to reverberation, as well as to compare the performance of DOA estimation based on the two tests. The simulation setup is the same as described in Section III, with a few additions that are described next.

A. Analysis of the DPD-SPW test

In this subsection, bins that passed the DPD-SPW test are analyzed and compared to bins that passed the original DPD test. The DPD-SPW test was computed as in (10) with $\mathcal{TH}_{\text{DPD-SPW}} = 4.2$, such that 3,000 bins passed the test, which is the same number as for the DPD test with $\mathcal{TH}_{\text{DPD}} = 4$, as in Fig. 1(b). In this case, the set $\mathcal{A}_{\text{DPD-SPW}}$ is marked on the reverberant speech spectrogram by black color and is presented in Fig. 3. As can be seen, the majority of bins that passed the test are associated with the onsets of the speech signal. This demonstrates the behavior of the test as discussed in Section IV; bins dominated by reverberant speech are less likely to pass the DPD-SPW test (compared to Fig. 1(b) and the DPD test), because it explicitly tests the proximity to a single plane wave representation.

Next, $\Omega_{\text{DPD-SPW}}$ was estimated using (11) and is presented in Fig. 4 on a (θ,ϕ) map. The figure also shows the DOAs

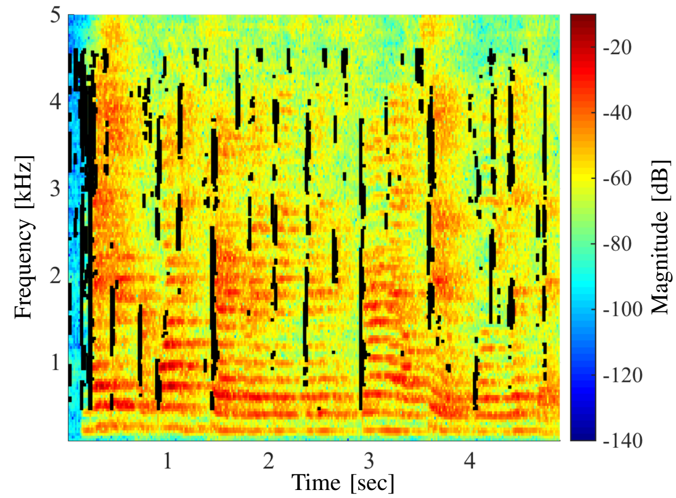

Fig. 3. Spectrogram of reverberant speech with the map of bins that passed the DPD-SPW test in (10) with $\mathcal{TH}_{\text{DPD-SPW}} = 4.2$, colored in black.

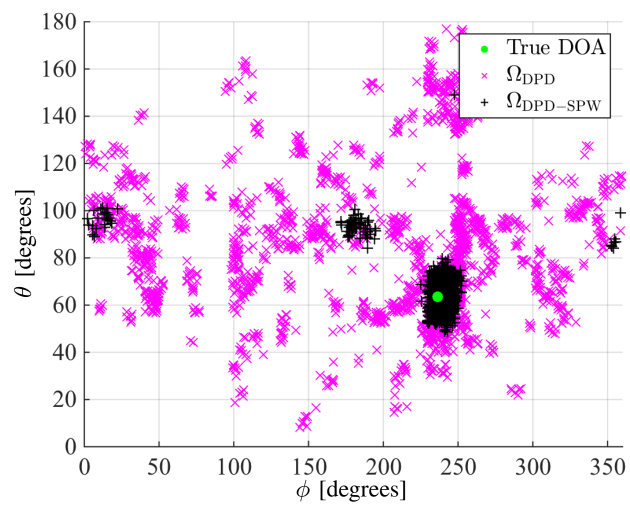

Fig. 4. DOA estimates from the algorithm in (11) based on the DPD-SPW test with $\mathcal{TH}_{\text{DPD-SPW}} = 4.2$, marked black and the DOA estimates from the algorithm in (7) based on the DPD test with $\mathcal{TH}_{\text{DPD}} = 4$, marked pink. The true DOA is marked by a green dot.

set Ω_{DPD} from the bins that passed the original DPD test, marked with a black color in Fig. 1(b). Fig. 4 shows that the vast majority of the estimates in $\Omega_{\text{DPD-SPW}}$ are located near the true DOA, marked by a green dot. This is in contrast to the estimates in Ω_{DPD}, which are scattered in a wide range of directions away from the true source direction. This demonstrates that the TF bins that passed the DPD-SPW test carry spatial information that is dominated by the direct sound.

B. DOA estimation analysis

In this subsection, performance analysis of four DOA estimation algorithms based on the DPD and the DPD-SPW tests is presented. Performance is compared with respect to the number of bins that passed each test, by adjusting the test threshold. In this study, five different speakers were simulated

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE I

MEAN VALUE OF DOA ESTIMATION ERROR FOR DIFFERENT NUMBER OF BINS THAT PASSED EACH TEST, AND FOR REVERBERATION TIME OF $T_{60} = 1$ S. THE ALGORITHM COMPARED ARE: $\mathcal{D}_{\text{DPD}}$, $\mathcal{D}_{\text{DPD-SPW}}$, $\mathcal{D}_{\text{DPD-GMM}}$, AND $\mathcal{D}_{\text{DPD-SPW-GMM}}$.

No. of bins	$\mathcal{D}_{\text{DPD}}$	$\mathcal{D}_{\text{DPD-SPW}}$	$\mathcal{D}_{\text{DPD-GMM}}$	$\mathcal{D}_{\text{DPD-SPW-GMM}}$
100	$13.4°$	$1.1°$	$4.4°$	$0.9°$
500	$17.3°$	$1.2°$	$2.0°$	$0.9°$
1,000	$18.4°$	$1.4°$	$2.0°$	$1.3°$
2,000	$22.0°$	$1.1°$	$1.7°$	$1.6°$
5,000	$23.6°$	$1.3°$	$1.8°$	$1.6°$
7,000	$24.8°$	$2.8°$	$1.8°$	$1.8°$

with speech signals taken from the TIMIT database. The four DOA estimation algorithms are detailed below:

1) $\mathcal{D}_{\text{DPD}}$: The set of DOAs was estimated by (7). The final DOA estimate was computed as the mean of Ω_{DPD}.
2) $\mathcal{D}_{\text{DPD-GMM}}$: The same as $\mathcal{D}_{\text{DPD}}$, with GMM clustering composed of three Gaussians applied to Ω_{DPD}, such that the final DOA estimate is the mean of the dominant Gaussian, as described in [16].
3) $\mathcal{D}_{\text{DPD-SPW}}$: The set of DOAs was estimated by (11). The final DOA estimate was computed as the mean of $\Omega_{\text{DPD-SPW}}$.
4) $\mathcal{D}_{\text{DPD-SPW-GMM}}$: The same as $\mathcal{D}_{\text{DPD-SPW}}$, with GMM clustering composed of three Gaussians applied to $\Omega_{\text{DPD-SPW}}$, such that the final DOA estimate is the mean of the dominant Gaussian.

The estimation errors for each algorithm averaged over all five speakers, are presented in Table I. Notice that the $\mathcal{D}_{\text{DPD}}$ algorithm does not perform very well, for all cases. However, GMM clustering significantly improves the performance of this algorithm. The $\mathcal{D}_{\text{DPD-SPW}}$ algorithm outperforms the $\mathcal{D}_{\text{DPD}}$ algorithm, and has performance that is comparable to the $\mathcal{D}_{\text{DPD-GMM}}$ algorithm. Also note that GMM clustering does not necessarily enhance the performance of $\mathcal{D}_{\text{DPD-SPW}}$. GMM may not fit the distribution of the data in this case. Future research may include the study of the distribution of these DOA estimates.

In summary, this simulation study demonstrated that the DPD-SPW test is more robust to reverberation than the original DPD test. The DOA estimation algorithm based on this test outperforms the method described in (7) for the case studied here.

VI. CONCLUSIONS

In this paper, a new direct-path dominance test was developed, denoted the DPD-SPW test, which is more robust to reverberation than the original DPD test. The test measures the proximity of the first singular vector of the measurement correlation matrix to a single plane-wave. The new test is more computationally complex than the DPD test, but demonstrated superior performance in DOA estimation under high reverberation. Future work could include a comprehensive theoretical analysis of the new test to complement the simulation results, as well as a comparison to the DPD test based on the directivity of the sound field [11].

REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar 1986.

[2] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, April 1988.

[3] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug 1976.

[4] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of the direct-path relative transfer function for supervised sound-source localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2171–2186, 2016.

[5] X. Li, L. Girin, F. Badeig, and R. Horaud, "Reverberant sound localization with a robot head based on direct-path relative transfer function," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 2819–2826.

[6] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 10, pp. 1494–1505, 2014.

[7] A. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 2296–2300.

[8] D. Pavlidi, S. Delikaris-Manias, V. Pulkki, and A. Mouchtaris, "3D localization of multiple sound sources with intensity vector estimates in single source zones," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1556–1560.

[9] S. Delikaris-Manias, D. Pavlidi, V. Pulkki, and A. Mouchtaris, "3D localization of multiple audio sources utilizing 2D DOA histograms," in *Signal Processing Conference (EUSIPCO), 2016 24th European*. IEEE, 2016, pp. 1473–1477.

[10] B. Rafaely, D. Kolossa, and Y. Maymon, "Towards acoustically robust localization of speakers in a reverberant environment," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, March 2017, pp. 96–100.

[11] B. Rafaely and K. Alhaiany, "Speaker localization using direct path dominance test based on sound field directivity," *Signal Processing*, vol. 143, pp. 42–47, 2018.

[12] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation and modulation theory*. Wiley Online Library, 2002, vol. 1.

[13] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–1781.

[14] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 2. IEEE, 2002, pp. II–1949.

[15] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[16] B. Rafaely and D. Kolossa, "Speaker localization in reverberant rooms based on direct path dominance test statistics," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6120–6124.

[17] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguistic Data Consortium*, vol. 10, no. 5, p. 0, 1993.

ISBN 978-90-827970-1-5 © EURASIP 2018