AN INTEGRATED FRAMEWORK FOR MULTI-CHANNEL MULTI-SOURCE LOCALIZATION AND VOICE ACTIVITY DETECTION

Mohammad J. Taghizadeh[1,2], *Philip N. Garner*[1], *Hervé Bourlard*[1,2], Hamid R. Abutalebi[3,1] and Afsaneh Asaei[1,2]

[1]Idiap Research Institute, Martigny, Switzerland
[2]Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[3]Yazd University, Yazd, Iran

{mtaghizadeh,pgarner,hbourlard,habutalebi,aasaei}@idiap.ch

ABSTRACT

Two of the major challenges in microphone array based adaptive beamforming, speech enhancement and distant speech recognition, are robust and accurate source localization and voice activity detection. This paper introduces a spatial gradient steered response power using the phase transform (SRP-PHAT) method which is capable of localization of competing speakers in overlapping conditions. We further investigate the behavior of the SRP function and characterize theoretically a fixed point in its search space for the diffuse noise field. We call this fixed point the *null* position in the SRP search space. Building on this evidence, we propose a technique for multi-channel voice activity detection (MVAD) based on detection of a maximum power corresponding to the *null* position. The gradient SRP-PHAT in tandem with the MVAD form an integrated framework of multi-source localization and voice activity detection. The experiments carried out on real data recordings show that this framework is very effective in practical applications of hands-free communication.

Index Terms: Multi-source localization, Multi-channel voice activity detection, Steered Response Power (SRP) localization, Diffuse noise field

1. INTRODUCTION

Speaker localization is a demanding area of research in hands-free speech communication using microphone arrays. In such applications, accurate knowledge of the speaker location is essential for an effective beampattern steering and interference suppression. This task gets even more challenging in meeting acquisition and conference recordings due to the presence of competing speakers [1]. We will briefly review the main approches to address this issue as follows:

I. High Resolution Spectral Estimation: Several algorithms have been proposed based on high resolution spectral estimation, such as minimum variance spectral estimation, auto-regressive modeling and various techniques based on eigen-analysis such as Multiple Signal Classification (MUSIC). These approaches are based on analysis of the received signals' covariance matrix, hence need an accurate estimation of the source signals, and impose a stationarity assumption. The underlying hypotheses are hardly realistic in case of speech signals as well as the room acoustics and the results are not very promising [2].

II. Time Difference Of Arrival (TDOA) Estimation: A common localization approach is based on TDOA estimation of the sources with respect to a pair of sensors. This approach is very practical if the placement of the microphones provides an accurate 3D estimation of the delays. Some commercial products such as automatic steering of cameras for video-conferences have been developed based on this idea [3]. In such applications, an updating rate of 300ms for location information is possible even in unfavorable acoustic conditions. However, in the scenario of multiple-target tracking and adaptive beam-steering, higher update rate is usually beneficial [4]. The generalized cross correlation (GCC) is the most celebrated technique for TDOA estimation. The basic idea is to find the peak of the cross-correlation function of the signal of two microphones. A weighting scheme is usually applied to increase the robustness of this approach to noise and multi-path effects. The maximum likelihood (ML) weighting is theoretically optimal when there is an uncorrelated noise source and there is no reverberation effect. In practice however, the performance of GCC-ML is highly degraded due to reverberation, and the Phase Transform (PHAT) yields better results [5].

Alternative TDOA estimation approaches are based on room impulse response identification. The basic idea behind this approach is that the acoustic channel defined for each speaker-microphone pair is a function of the speaker location. Hence, identifying the room impulse response enables us to compute TDOAs and localize the speakers. When there is no prior knowledge about the microphone array geometry, this scenario could be formulated as a blind Multiple-Input Multiple-Output (MIMO) channel identification problem.

The solution usually incorporates blind source separation at the pre-processing step and resolves the ambiguity of the acoustic mixing process by localization along with the separation of the individual sources [6, 7].
Some other alternatives for TDOA estimation are based on singular value decomposition for estimation of the room impulse response which is very practical for the speech signal but requires at least 250ms of data to converge [8].

III. Beamformer Steered Response Power (SRP): Finally, it is possible to localize the speaker directly based on the beamformer output power. In this approach, the space is scanned by steering the beam-pattern and finding the maximum power. The delay-and-sum beamformer, minimum variance beamformer and generalized side-lobe canceler have been the most effective methods for speaker localization [9]. Unlike TDOA-based approaches, SRP-based localization approaches have a higher effective update rate, i.e., they can work with much shorter frames even in adverse acoustic conditions; hence, they are practically appropriate for realistic applications, especially in multi-party scenarios [10]. Different filtering proposals have been used in SRP techniques, among which the phase-transform filter (PHAT) has been shown to provide a robust localization framework [11].

This paper is organized as follows: The general concepts of SRP localization approaches are introduced in 2.2. We then provide theoretical as well as empirical evidence that the SRP output power for the silent frames exhibits a peak corresponding to a fixed point in its search space. Relying on this observation, we formulate a multi-channel voice activity detection (MVAD) in Section 2.3. In Section 2.4 a multi-speaker modification of SRP-PHAT for localization of competing sources is proposed by applying a spatial gradient function on the beamformer output. We further carry out some experiments on the real data recordings to evaluate the proposed framework in Section 3. Conclusions are drawn in Section 4.

2. MULTI-SOURCE LOCALIZATION AND VOICE ACTIVITY DETECTION

2.1. Signal model

We consider a scenario in which M microphones record the signal of L sources; the single-channel received signal, $x_m(t)$, is composed of two components: (1) a filtered version of the original signal, s_l, which has been convolved with the source-microphone room impulse response, $h_{m,l}$ and (2) an uncorrelated independent additive noise $n_m(t)$

$$x_m(t) = \sum_{l=1}^{L} s_l(t) * h_{m,l} + n_m(t) \tag{1}$$

2.2. SRP-PHAT source localization

The general procedure of the beamforming applies filter-and-sum on the input microphone-channels. The filters are usually adapted in order to enhance the source signal whilst suppressing the interference; hence the beamformer output is maximized when the beampattern is focused accurately towards the speaker. In the SRP localization, the output power is used for a 3D scanning of the space where the maximum power corresponds to the location of the active speaker. To state it concisely, the Generalized Cross Correlation (GCC) is defined as

$$R_{m,n}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(G_m(\omega)X_m(\omega))(G_n(\omega)X_n(\omega))^* e^{j\omega\tau}d\omega \tag{2}$$

where X and G are the Fourier transform of the signal and filter, respectively. Defining the weighting function $\Psi_{m,n}(\omega) = G_m(\omega)G_n^*(\omega)$, the GCC function would be

$$R_{m,n}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Psi_{m,n}(\omega)X_m(\omega)X_n^*(\omega)e^{j\omega\tau}d\omega \tag{3}$$

The PHAT weighting function is defined as

$$\Psi_{m,n}(\omega) = |X_m(\omega)X_n^*(\omega)|^{-1} \tag{4}$$

Substituting 4 into 3 and taking the summation of all possible microphone pairs, the SRP-PHAT is obtained

$$P(\rho,\theta,\varphi) = 2\pi\sum_{m,n} R_{m,n}(\tau_{m,n}) \quad m,n \in \{1,2,...,M\} \tag{5}$$

where $\tau_{m,n}$ is the time difference of arrival of the source signal located at $\kappa(\rho,\theta,\varphi)$ to the two microphones m and n. Note that the source location is represented in spherical coordinates where ρ denotes the range and θ and φ correspond to the azimuth and elevation, respectively.

The largest peak corresponds to the dominant speaker located at

$$\kappa(\hat{\rho},\hat{\theta},\hat{\varphi}) = \arg\max_{\rho,\theta,\varphi} P(\rho,\theta,\varphi) \tag{6}$$

2.3. Multi-Channel VAD

In this section, we will investigate the SRP-PHAT formulation when the input is a diffuse noise, which is often the case in realistic environments without presence of any active speaker. We characterize theoretically the existence of a predefined point for the SRP function for the diffuse noise; hence, there is no speech activity. Suppose that n_i and n_k represent the noises at microphones i and k respectively. The cross spectral density between n_i and n_k is

$$\Phi_{ik}(\omega) = N_i(\omega)N_k^*(\omega), \tag{7}$$

where $N_i(\omega)$ and $N_k(\omega)$ are Fourier transform noises i and k respectively. and the coherence between noises n_i and n_k is

$$\Gamma_{ik}(\omega) = \frac{\Phi_{ik}(\omega)}{\sqrt{\Phi_{ii}(\omega)\Phi_{kk}(\omega)}}. \tag{8}$$

For the diffuse noise, we have [12]

$$\Gamma_{ik}(\omega) = \mathrm{sinc}\left(\frac{\omega d_{ik}}{c}\right), \quad (9)$$

where d_{ik} is the distance between the two microphones and c is the speed of sound. Substituting equation 9 into equation 3, we obtain

$$\begin{aligned} R_{i,k}(\tau) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\Gamma_{ik}(\omega)}{|\Gamma_{ik}(\omega)|}e^{j\omega\tau}d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{sinc}(\frac{\omega d_{ik}}{c})}{|\mathrm{sinc}(\frac{\omega d_{ik}}{c})|}e^{j\omega\tau}d\omega. \end{aligned} \quad (10)$$

In order to find the maximum of the SRP-PHAT function, we compute the derivative w.r.t. τ; hence

$$\frac{\partial R_{i,k}(\tau)}{\partial\tau} = \frac{j}{2\pi}\int_{-\infty}^{\infty}\frac{\omega\mathrm{sinc}(\frac{\omega d_{ik}}{c})}{|\mathrm{sinc}(\frac{\omega d_{ik}}{c})|}e^{j\omega\tau}d\omega = 0. \quad (11)$$

The above equality holds for $\tau = 0$; hence the maximum of $R_{i,k}(\tau)$ is obtained for a point with equal distance to the two microphones. The same argument is true for all microphone pairs; therefore on the direction perpendicular to the microphone array the closest point to the center of the microphone array is where the output power of the SRP is maximized. Obviously, this has a strong dependence on the elevation and less sensitivity to the azimuth (θ). Since this is obtained only when there is no speech activity, we call it the *null* point of the SRP search space. We can exploit this fact to detect voice activity in the acquired multi-channel speech frames.

The integrated framework of SRP localization and MVAD reduces the complexity of speech analysis in microphone-array applications such as hands-free speech recognition. The previous proposals on MVAD which takes advantage of the extra information provided by additional sensors [13, 14] increases the computational load. A few others have been also published recently based on Gaussianity assumption of the frequency components [15] or non-uniform phase assumption [16]. In practice however, these hypotheses are not realistic. On the other hand, the technique that we propose here is based on a realistic model of the acoustic conditions as a diffuse noise field and imposes no computational load on the source localization and beamforming designed for data acquisition. Moreover we don't need any training or threshold optimization which is a common computational load in any VAD structure.

2.4. Spatial Gradient SRP-PHAT

In this section, we further exploit the integrated frame work of SRP localization and MVAD, and extend it for multi-party scenarios. The PHAT transform whitens the microphone signals; hence yields sharper peaks at the output power corresponding to the actual location of the L sources. In multi-speaker scenarios, the localization of L competing sources amounts to the detection of the largest L peaks of the beamformer output power. In practice however, the SRP-PHAT output has many local maxima due to the multi-path effect which make the extraction of the largest L peaks very difficult. Considering the fact that the SRP has a discrete search space, we first apply a three dimentional box filtering (averaging) defined as follows:

$$\bar{P}(\rho_i, \theta_i, \varphi_i) = \frac{\sum_{c=-1}^{1}\sum_{b=-1}^{1}\sum_{a=-1}^{1} P(\rho_{i-c}, \theta_{i-b}, \varphi_{i-a})}{27} \quad (12)$$

To find the second source location, we have to remove the data corresponding to the dominant speaker from the search space. Therefore, the data points from all directions of ρ, θ, φ which correspond to the negative spatial gradient of the SRP output power ($\bar{P}$) are discarded. The directional derivative of $\bar{P}$ at point κ in direction $\boldsymbol{u}$ is obtained by

$$\nabla_{\boldsymbol{u}}\bar{P}(\kappa) = \lim_{h\to 0^+}\frac{\bar{P}(\kappa + h\boldsymbol{u}) - \bar{P}(\kappa)}{h} = \nabla\bar{P}(\kappa)\boldsymbol{u}, \quad (13)$$

where $\boldsymbol{u}$ is the unit vector and ∇ on the right denotes the gradient and

$$\nabla\bar{P}(\rho,\theta,\varphi) = \frac{\partial\bar{P}}{\partial\rho}e_\rho + \frac{1}{\rho}\frac{\partial\bar{P}}{\partial\varphi}e_\varphi + \frac{1}{\rho\sin\varphi}\frac{\partial\bar{P}}{\partial\theta}e_\theta, \quad (14)$$

where $e_\rho, e_\theta, e_\varphi$ are the canonical basis vectors of the coordinate system. Then, the directional derivative defined in equation 13 is computed at the location of the largest peak denoted by κ in 26 $\boldsymbol{u}$ directions. Hence,

$$\boldsymbol{u} \in \left\{\frac{ie_\rho + je_\theta + ke_\varphi}{\sqrt{i^2+j^2+k^2}}; i,j,k \in \{-1,1,0\}, i^2+j^2+k^2 \neq 0\right\} \quad (15)$$

Then in all directions as long as the gradient function has a negative value, we take a small step $\Delta d = \rho\,\varepsilon\sqrt{i^2+j^2+k^2}$ with $0 < \varepsilon \ll 1$ to the next data point and this procedure is continued until all the data points with negative gradient are discarded from the search space. The residual is then searched to find the maximum power corresponding to the second dominant speaker. This procedure is continued until the SRP maximum corresponds to the *null* point in the search space. The number of active speakers at each frame is determined by detecting this *null* point in the SRP residual.

3. EXPERIMENTS

In this section, we present experimental results on the proposed integrated framework of SRP-PHAT localization and MVAD based on (1) simulated data with the diffuse noise field and (2) real recording using the MONC as well as RT09 databases.

3.1. Diffuse Noise Field Simulation and Results

We consider a scenario in which three white noise sources are located at random positions in the room. The room impulse responses are generated with the image model technique

[17] using intra-sample interpolation, up to 15^{th} order reflections and omni-directional microphones. The corresponding reflection ratio, β used by the image model was calculated via Eyring's formula:

$$\beta = \exp(-13.82/[\mathrm{c} \times (\mathrm{L_x}^{-1} + \mathrm{L_y}^{-1} + \mathrm{L_z}^{-1}) \times \mathrm{T}]) \quad (16)$$

where $\mathrm{L_x}, \mathrm{L_y}$ and $\mathrm{L_z}$ are the room dimensions, c is the sound velocity in the air ($\approx$342m/s) and T is the room reverberation time. In our experiments T=300ms and the room direct-paths are discarded from the impulse responses for generation of the semi-diffuse noise signals [18]. Three noise sources are randomly positioned in the room and a circular microphone array with 8-channels and diameter of 20cm located at the center of the room records the diffuse noise.

The SRP-PHAT run on the multi-channel diffuse noise signal exhibits a consistent peak corresponding to the nearest point to the maximum elevation to the center of the array. The experiments are carried out for 128ms frames with 50% overlap. The maximum elevation in search space of our simulation as well as real data tests are 85° and 75° respectively. As illustrated in Fig.1 the *null* point exists at the ρ=5cm and $\varphi = 85^\circ$ in the search space. As expected, the azimuth value is almost random. These results provide empirical evidence of the formulation derived in section 2.3. This joint framework of gradient SRP-PHAT localization and voice activity detection is shown to work well under the assumption of the diffuse noise field for the real environments. In the following section, we conduct some experiment on the real data recordings.

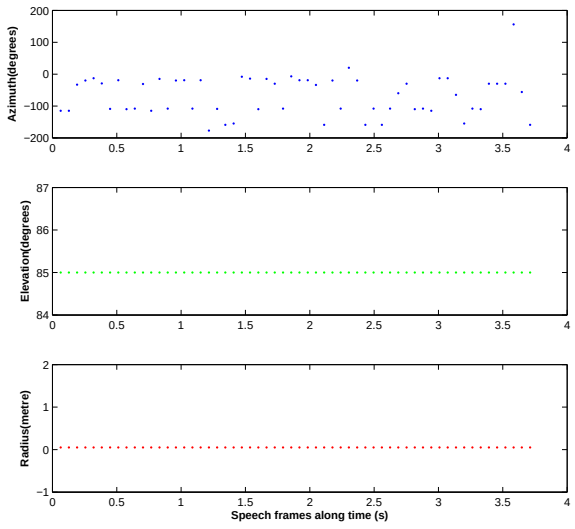

Fig. 1. SRP-PHAT localization in diffuse noise field

3.2. Speech Database

We have evaluated our framework using two databases recorded in real environments: (1) The Multichannel Overlapping Numbers Corpus (MONC) [19]. We have used the following two recording scenarios; S1: one speaker located at L1 (78cm,135°,23°) without overlapping, S12: one competing speaker located at L2 (78cm,45°,23°), (2) Rich transcription (RT09) is structured for mixing metadata extraction and speech-to-text (STT) technologies. We use this database of a precise evaluation of MVAD using 8-channel microphone recordings. The details are explained in [20]. We use file `EDI_20071128_1000_ci01_NONE.sph` that was recorded in the IMR meeting room by array1 at Edinburgh. We use the ICSI ground truth, that is, a hand transcription automatically aligned with the data. In this sense, the ground truth can contain errors; however, the results are still informative. File `EDI_20071128-1000.rttm` was used as the ground truth.

3.3. Single Speaker Localization and MVAD

In the first scenario, we run our algorithm on S1. The Signal-to-noise Ratio (SNR) is estimated about 9dB. The results are depicted in Fig. 2. The 3D search space of SRP-PHAT consists of 150,000 points. The nearest vertical point to the center of the array (SRP *null*-point) is N (ρ=0.45m,φ=75°). The *null*-point is detected when there is no speech activity in the frame, e.g. the first two frames in the Fig. 2. The high accuracy of the proposed MVAD can be seen in Fig. 2(e). For instance, there exists a high energy noisy region between 3.62 and 3.94 seconds which has been correctly identified as a non-speech part of the signal. By removing the silent frames using MVAD, Fig. 3 is obtained. Note that the joint localization-VAD framework exhibits highly accurate results as the standard deviation (SD) of azimuth estimation is 0.5° and the SD of elevation estimation is 2°. Accurate estimation of the range however, is not possible.

3.4. Multi-Speakers Localization and MVAD

The second scenario considers overlapping speech segments. In Fig. 4, generic SRP-PHAT has been used along with MVAD. As we can observe, only the dominant speaker is detected at each frame. The silent frames detected by MVAD are shown at azimuth = 0. As the figure illustrates, the dominant speaker is localized very accurately, the SD of azimuth estimation is 1°. If we use the spatial gradient modification of SRP-PHAT, we can localize both the dominant as well as the inferior speakers precisely at each frame. The results of this experiment are depicted in Fig. 5. The results are shown for the azimuth estimation. Upon the detection of a peak at elevation = 75°, the MVAD has detected a noisy region where no speech activity is present. The dominant speaker is indicated by circles; the inferior speaker is extracted by the

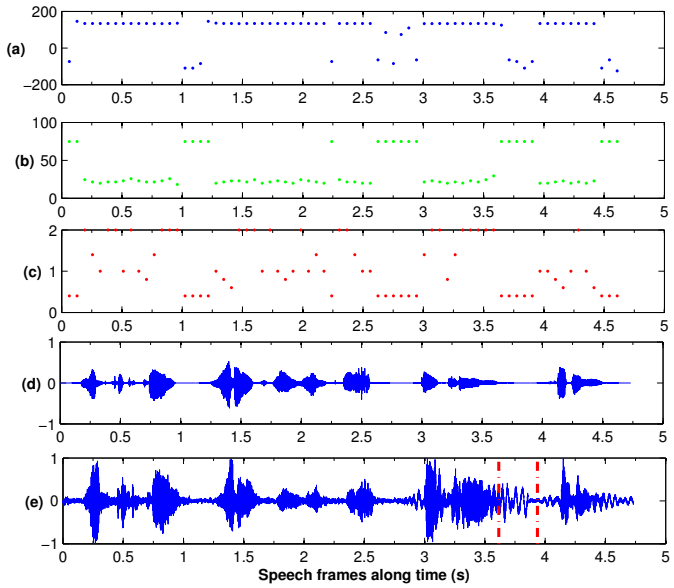

Fig. 2. Speaker localization using SRP-PHAT in non-overlapping conditions. (a) estimated azimuth (degrees), (b) estimated elevation (degrees), (c) estimated range (metre), (d) clean speech waveform, (e) distant speech recorded by microphone array

gradient method and it is denoted by dots. The number of active speakers is determined when the *null*-point is detected in the gradient SRP-PHAT residual.

In our final experiment, we evaluate the proposed MVAD on part of the RT09 database. A total 315s of speech signal is processed in frames of length 256ms with 50% overlap. The speech material is taken from an 83s and another 232s segment of a file. This is in order to avoid physical noise such as door slams in the background and such that more than 18% of frames are silent. This enables us to have a sound evaluation of MVAD. It is not an exhaustive test; we only aim to have an evaluation on a modern corpus.

The total error rate for MVAD is 6.4%, which consists of 2.7% missed speech and 3.7% false alarms. The proposed MVAD is practical in meeting recordings, which are usually moderate SNR speech but highly reverberant situations. The sample spectrogram and the speech waveform are illustrated in Fig. 6. The recognized silent parts (output of MVAD) are indexed with boxes in a yellow strip. As the figure shows, the signal is highly noisy. The majority of the errors in MVAD happen at the transitions of silent and speech.

4. CONCLUSIONS

We proposed an integrated framework for multi-channel multi-source localization and voice activity detection which is very effective in real acoustic conditions and practical

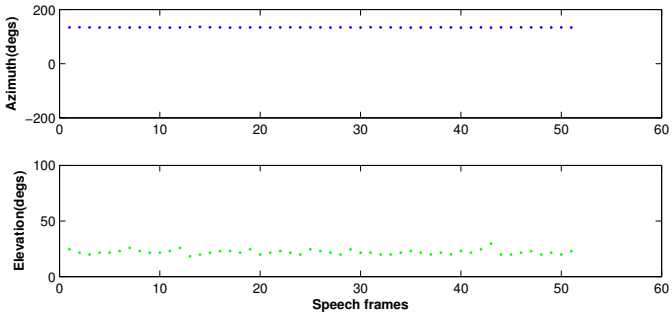

Fig. 3. Improvement of joint source localization and voice activity detection framework in non-overlapping condition

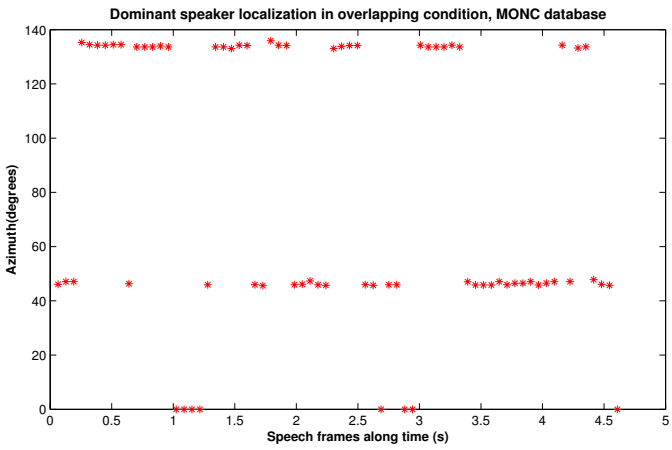

Fig. 4. Dominant speaker localization in overlapping condition using SRP-PHAT on MONC.

hands-free speech scenarios. Our method exploits the SRP localization technique. We introduced a spatial gradient modification to SRP-PHAT for localization of competing sources. We further worked out the SRP search space for the diffuse noise field and characterized a fixed point corresponding to the SRP peak for non-speech frames. This formulation led to introducing another application of the gradient SRP-PHAT as an MVAD. Experiments conducted on real data recordings showed that the framework could exhibit highly accurate results for multi-source localization and voice activity detection in microphone array applications, in particular in highly reverberant environments, such as aircraft cockpits and automobile interiors, where the noise fields are usually diffuse.

5. ACKNOWLEDGEMENTS

This work was supported by the Swiss National Science Foundation under the National Center of Competence in Research (NCCR) on "Interactive Multi-modal Information Management" (IM2) and the European Union 7^{th} Framework

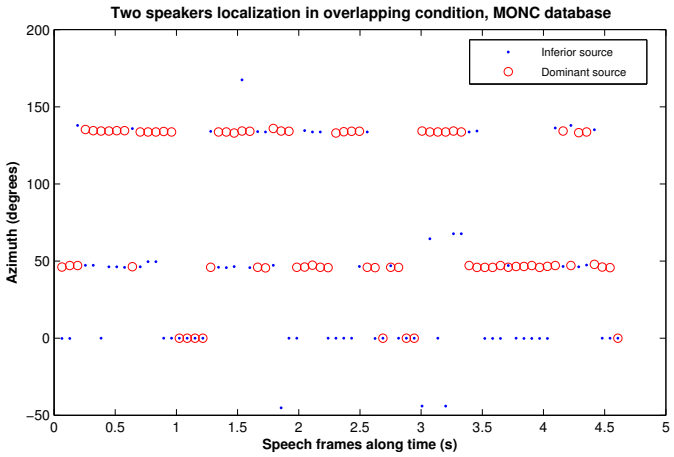

Fig. 5. Two competing speakers localization using spatial gradient extension of SRP-PHAT on MONC.

Programme IST Integrating Project "Together Anywhere Together Anytime" (TA2, FP7-214793).

6. REFERENCES

[1] A. Waibel, M. Bett, F. Metze, K. Ries, T. Schaaf, T. Schultz, H. Soltau, H. Yu, and K. Zechner, "Advances in automatic meeting record creation and access," in *Proceedings of ICASSP*, 2001.

[2] J. Dmochowski, S. Benesty, and S. Affes, "Broadband music: opportunities and challenges for multiple source localization," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007.

[3] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proceedings of ICASSP*, 1997.

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

[5] M. Omologo and P. Svaizer, "Acoustic source localization in noisy and reverberant environments using csp analysis," in *Proceedings of ICASSP*, 1996.

[6] J. Chen, J. Benesty, and A. Huang, "Mimo acoustic signal processing," in *HSCMA Workshop, Invited Talk*, 2005.

[7] H. Buchner, R. Aichner, and W. Kellerman, "Trinicon: A versatile framework for multichannel blind signal processing," in *Proceedings of ICASSP*, 2004.

[8] Y. Huang, J. Benesty, and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for real-time acoustic source localization," in *Proceedings of ICASSP*, 1999.

[9] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," in *IEEE Trans. Ant. Prop.*, 1982.

[10] A. Asaei, M. J. Taghizadeh, M. Bahrololum, and M. Ghanbari, "Verified speaker localization utilizing voicing level in split-bands," *Signal Processing*, vol. 89, 2009.

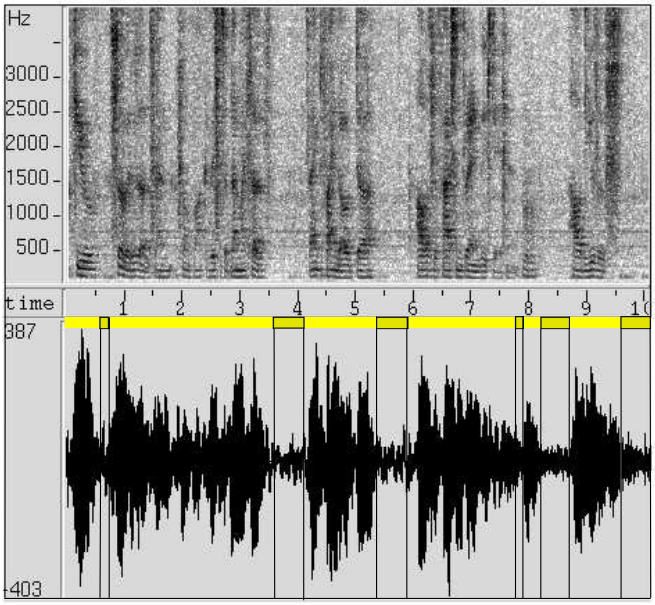

Fig. 6. Top: spectrogram and Bottom: speech waveform. Silent parts that recognized by MVAD have been showed with boxes in yellow strip.

[11] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," in *PhD Thesis, Brown University*, 1993.

[12] J. Bitzer, K. Kammeyer, and K. U. Simmer, "An alternative implementation of the superdirective beamformer," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999.

[13] N. Doukas, P. Naylor, and T. Stathaki, "Voice activity detection using source separation techniques," in *Eurospeech*, 1997.

[14] Q. Zou, X. Zou, M. Zhang, and Z. Lin, "A robust speech detection algorithm in a microphone array teleconferencing system," in *Proceedings of ICASSP*, 2001.

[15] I. Potamitis, "Estimation of speech presence probability in the field of microphone array," in *IEEE Signal Processing Letters*, 2004.

[16] G. Kim and N. I. Cho, "Voice activity detection using phase vector in microphone array," in *Electronics Letters*, 2007.

[17] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of Acoustic Society of America*, vol. 65, 1979.

[18] E. Habets, "Generating sensor signals in isotropic noise fields," in *J. Acoust. Soc. Am. Volume 122, Issue 6,*, 2007.

[19] "The multichannel overlapping numbers corpus," Idiap resources available online:, http://www.cslu.ogi.edu/corpora/monc.pdf.

[20] "The 2009 (rt-09) rich transcription meeting recognition evaluation plan," Rich Transcription Evaluation Project:, http://www.itl.nist.gov/iad/mig/tests/rt/2009/docs/rt09-meeting-eval-plan-v2.pdf.