Array Steered Response Time-Alignment for Propagation Delay Compensation for Acoustic Localization

Pasi Pertilä

Department of Signal Processing

Tampere University of Technology, P.O.Box 553, FIN-33101, Tampere, Finland

Tel: +358-4-0849-0786, Fax +358-3-3115-4989

e-mail: pasi.pertila@tut.fi

Abstract - Locating a source using direction estimates from spatially separated microphone arrays has been widely researched. The directions are traditionally obtained by steering the array to a direction that maximizes the response. A more robust approach is to directly combine the array steered responses and maximize the resulting spatial likelihood function. However, in the atmosphere the propagation delay from source to array is typically significant, causing true source position and the estimated direction to differ. Localization methods using directional estimates that compensate this delay exist. This work proposes delay compensation for steered array responses to increase localization performance. The method is shown to improve robustness against source movement.

I. INTRODUCTION

With a far-field sound source the array response represents source direction information. Time delay estimation (TDE) can be applied for wide-band signal direction of arrival (DOA) estimation. TDE estimation methods include the widely known generalized cross-correlation (GCC) method [1]. The steered response power using phase transform (SRP-PHAT) uses GCC values to calculate the array steered response for the broad-band source signal [2]. Beamforming is typically applied for narrow-band signal DOA estimation [3].

Source localization using DOA estimates (bearings only localization) has been extensively studied. The DOA vector is typically obtained by maximizing the

array steered response. If the maximum results from a spurious peak the location estimate will be corrupted. To improve localization robustness the steered responses of the arrays are combined to build a spatial likelihood function (SLF) of source location. The method results in a non-Gaussian and non-linear state estimation problem. This approach has no obvious closed form solutions, however the SLF can be evaluated in any point of interest. The source can be located by directly maximizing the SLF or by using sequential Bayesian methods such as particle filters [4, 5].

Steered responses from multiple arrays have been directly used to locate a narrow-band source [6]. Here, for wide-band signals the GCC is used to calculate the steered response algorithm for the arrays.

A moving source introduces Doppler effect, i.e., samples emitted at constant interval from source are temporally clustered for an approaching source and stretched when a source is moving away. Here, the observed signal is assumed quasi-stationary between closely spaced microphones during a short processing frame (20 ms). For long processing frames (300 – 500 ms) the Doppler effect becomes more significant and could be compensated [7]. In [8] modeling the Doppler effect is studied.

Acoustical DOA observations from arrays appear to follow the true source position due to propagation delay. This biases localization. Methods to compensate the delay have been proposed for DOA based localization. In [9] an iterative method for source cross-fixing is presented using time-alignment of array samples. In [10] DOA observations from arrays are time-aligned for each candidate source position and the source location is solved iteratively. In [11] an online method using recursive Bayesian techniques to locate the source from DOA observations is proposed. However, the concept

Thanks to M.Sc. Teemu Korhonen for assistance.

of time-alignment has not been reported for steered response arrays. This paper proposes a steered response time alignment algorithm for steered array response based localization.

The outline of this paper is as follows. Section II motivates the use of propagation delay in source localization. Section III discusses the steered array response based localization and presents the proposed delay compensation method for localization in a sequential Bayesian framework. The method is tested with simulations in Section IV, and Section V presents the results. Section VI concludes the discussion.

II. PROPAGATION DELAY

The limited speed of sound causes significant signal propagation delay between source (located at $\mathbf{r}$) and i th receiver (located at $\mathbf{m}_i$). This delay is sometimes referred as the time of flight (TOF). The locations are defined in 2D Cartesian coordinate system without loss of generalization¹. The i th microphone signal is written as delayed and scaled version of the source signal

$$x_i(r_i, t) = A \cdot s(t - r_i/c) + v_i(t), \quad (1)$$

where r_i is Euclidean distance from microphone i to source, i.e., $r_i = \|\mathbf{m}_i - \mathbf{r}\|$, A is amplitude, $s(t)$ is source signal, c is speed of sound, and v_i is white noise that is independent between channels.

The received signal (1) can be written in the form

$$x_i(r_i, t + r_i/c) = A \cdot s(t) + v_i(t + r_i/c). \quad (2)$$

In (2) the direct path signal is time-aligned at the source emission location (at time t).

The signal is processed in frames of length T_f samples from data gathered at sampling rate F_s . It is assumed that the microphones are grouped in arrays and the inter-array microphone distance is small compared to array to source distance.

The quantized propagation delay from source to array l (located at $\mathbf{p}_l$) is

$$\tau(\mathbf{p}_l, \mathbf{r}) = \lfloor F_s \cdot \|\mathbf{p}_l - \mathbf{r}\| / (c \cdot T_f) \rfloor, \quad (3)$$

where $\lfloor \cdot \rfloor$ notes rounding to nearest frame. For an example, a 200 m source distance with parameter values from Table 1 corresponds to a propagation delay (3) of 27 processing frames.

III. STEERED RESPONSE BASED LOCALIZATION

For closely spaced array microphones the wavefront is approximately planar. A wavefront impinging from di-

rection $\mathbf{k}$ has an arrival time difference between microphone pair $p = \{i, j\}$:

$$\Delta\tau_p(\mathbf{k}) = (\mathbf{m}_i - \mathbf{m}_j)^T \mathbf{k} / (\|\mathbf{k}\| \cdot c). \quad (4)$$

The GCC is here estimated between signals (1) at time t as [1]

$$R_p^t(\tau) = \mathcal{F}^{-1}\{W_p(k)X_i(k)X_j(k)^*\}, \quad (5)$$

where $X_j(k)^*$ is a complex conjugate transpose of the DFT of the j th microphone signal, k is discrete frequency, $\mathcal{F}^{-1}\{\cdot\}$ denotes inverse DFT, $W_p(k) = |X_i(k)X_j(k)^*|^{-1}$ is phase transform (PHAT) weighting [1], and $|\cdot|$ is absolute value. The pairwise GCCs from array l are denoted $\mathbf{R}_l$, where $l = 1, \dots, L$, and L is the number of arrays.

A. Steered Array Response Combination

The directional response from array l is the product of GCCs steered to direction $\mathbf{k}_l$ with far-field SRP-PHAT [2]

$$P(\mathbf{R}_l^t | \mathbf{k}_l) = \sum_p R_p^t(\Delta\tau_p(\mathbf{k}_l)), \quad (6)$$

where $P(\cdot)$ denotes conditional likelihood, and $\mathbf{k}_l = \mathbf{r} - \mathbf{p}_l$. The responses from L separate arrays are here combined by multiplication to build the SLF:

$$P(\mathbf{R}_{[1:L]}^t | \mathbf{r}) = \prod_{l=1}^L P(\mathbf{R}_l^t | \mathbf{k}_l). \quad (7)$$

The likelihood of source located at $\mathbf{r}$ is obtained by (7). A straightforward localization solution is to maximize (7) as a function of source position $\mathbf{r}$.

B. Applying the propagation model

Using the time-aligning concept (2) the proper temporal time frame from each array is selected for each source location according to the TOF value (3). The temporally aligned data frame is then used to calculate the array steered response (6) and the responses from all arrays is then combined as

$$P(\mathbf{R}_{[1:L]}^{\mathbf{t}} | \mathbf{r}) = \prod_{l=1}^L P(\mathbf{R}_l^{t+\tau(\mathbf{p}_l, \mathbf{r})} | \mathbf{k}_l), \quad (8)$$

where $\mathbf{t} = [t, t+1, \dots, t + \max_l \{\tau(\mathbf{p}_l, \mathbf{r})\}]$ and indicates that additional time frames of GCC values are required for the time alignment, compared to (7).

As in the case of no time-alignment the source position could be estimated by maximizing (8) as a function of source position. However, sequential processing methods have been shown to be more robust in

¹The used vector presentation is extendable to 3D.

indoor sound source localization [12] than the direct maximization, and are therefore considered here.

C. Sequential Source State Estimation

The sound source position is estimated using a Monte Carlo -based sequential Bayesian state estimation scheme, known as particle filtering (PF). For a more complete discussion on PF methods refer to [5, 4]. The specific PF method applied here is known as sampling importance resampling (SIR). The source motion is modeled as a Langevin process as proposed in [13]. The Langevin process is described for source x-coordinate position and velocity ($x_t, \dot{x}_t$) at discrete time t [14]:

$$\dot{x}_t = a_x \cdot \dot{x}_{t-1} + b_x \cdot F, \quad (9)$$

$$x_t = x_{t-1} + \Delta T \cdot \dot{x}_t, \quad (10)$$

$$a_x = \exp(-\beta \cdot \Delta T), \quad (0 \leq a_x \leq 1), \quad (11)$$

$$b_x = \bar{v} \sqrt{1 - a_x^2}, \quad (0 \leq b_x \leq \bar{v}), \quad (12)$$

where ΔT is time period, β is rate-constant, a_x and b_x are Langevin coefficients, $\bar{v}$ is steady-state RMS velocity and $F \sim \mathcal{N}(0, 1)$ is an excitation parameter. The parameter values are detailed in Table 1. The y-coordinate values are identical.

The PF approximates the posterior density with a set of N_j weighted random samples $\mathcal{X}_t = \{\mathbf{x}_t^j, w_t^j\}_{j=1}^{N_j}$ for each frame t , where the state is

$$\mathbf{x}_t = \{x_t, y_t, \dot{x}_t, \dot{y}_t\}. \quad (13)$$

The approximate posterior density is written as

$$P(\mathbf{x}_{0:t} | \mathbf{R}_{[1:L]}^{1:t}) \approx \sum_{j=1}^{N_j} w_t^j \delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^j), \quad (14)$$

where the scalar weights $w_t^1, \dots, w_t^{N_j}$ sum to unity, and $\delta(\cdot)$ is the Dirac's delta function. The estimate of source position is here obtained with the median operator of the position variables of the source state as

$$\hat{\mathbf{r}}_t^{\text{ME}} = \text{median}\{\mathbf{r}_t^1, \mathbf{r}_t^2, \dots, \mathbf{r}_t^{N_j}\}, \quad (15)$$

where $\mathbf{r}_t^i$ are the x,y position values of the i th particle $\mathbf{x}_t^i$.

IV. SIMULATIONS

The proposed approach is tested with simulations. A source emits pseudo-random noise with a flat spectrum at frequencies ω in the range [20, 10k] Hz. The i th microphone signal at time t is modeled similar to the

Fig. 1. The received signal spectrum of the passing source shows the effect of Doppler shift and geometrical and atmospheric attenuation. The line displays the source distance from microphone. Microphone is located at $\langle 49.5, 200 \rangle$, see Fig. 2.

solution of the spherical wave equation [3]

$$x_i(r_i, t) = \sum_{\omega} A(r_i, \omega, T, H) \cdot \exp\{j(\omega(t - r_i/c) + \phi(\omega))\} + w_i, \quad (16)$$

where $A(r_i, \omega, T, H)$ models the geometrical and atmospheric attenuation, $\phi(\omega)$ is the pseudo-random phase, and w_i is i.i.d. white Gaussian noise. The attenuation $A(\cdot)$ is implemented according to [15] using 20 °C temperature (T) and 70 % humidity (H) as parameters.

The source is moving from x,y coordinates $\langle -80, 150 \rangle$ to $\langle 460, 150 \rangle$ (m) in 8.1 seconds, i.e., at the speed of 66.7 m/s. The spectrogram and the distance of the passing sound source observed from a microphone at $\langle 49.5, 200 \rangle$ is displayed in Fig. 1. The frequency shift caused by the Doppler effect of is visible from the received signal spectrum. The shift is strongest at 2 s time which represents the closest point of approach (CPA) at 50 m distance where the 2nd derivative of radial distance is at its highest. The atmospheric attenuation is evident in the end of recording since it is most significant at high frequencies for distant sources. In total, three arrays each consisting of four microphones are simulated with sampling rate specified in Table 1.

TABLE 1. Signal and Langevin process parameters.

Parameter	Symbol	Value
Sampling frequency	F_s	48 000 Hz
Frame length	T_f	1024 samples
Time period	ΔT	0.0213 s
Rate-constant	β	5 Hz
RMS-velocity	$\bar{v}$	66 m/s
Number of particles	N_j	5000

The 2D arrays are located at $\langle 200, 0 \rangle$, $\langle 50, 200 \rangle$, and $\langle 250, 250 \rangle$. The arrays are similar in shape. Three array microphones are placed in a triangle with 1 m edge length. The fourth microphone located at center of the triangle.

To illustrate the spatial likelihood function shape Fig. 2 shows the output for the traditional method (7) in panel 2(a) and the proposed method (8) is displayed in panel 2(b). The output is calculated for a fixed grid of cell size 1 m². In panel (a) the maximum response direction from each array points behind the source due to propagation delay. In panel (b) data from adjacent time frames is used for hypothetical each source distance. Therefore, the observations seem to curve behind the source position. Note that in panel (b) the source is located at the “time of emission” not at “time of most recent sample.” The estimation of the on-line source position requires a prediction operation since no acoustic information from the real-time source position is available.

V. RESULTS

Two localization approaches are tested. The traditional approach of not utilizing a propagation delay model is analyzed for reference using (7). The proposed approach (8) is the second method. The results are given for the mean RMS error as a function of microphone signal average SNR in Fig. 3. The x-axis indicates the SNR for each microphone signal for the whole recording, and y-axis displays the RMS error estimated by

$$\text{RMSE}\{\hat{\mathbf{r}}^{\text{ME}}\} = \sqrt{\frac{1}{N} \sum_t (\mathbf{r}(t) - \hat{\mathbf{r}}_t^{\text{ME}})^2}, \quad (17)$$

where N time frames bounded between $t \in [1.24, 5.74]$ seconds are used. The data from the beginning and end of simulation is omitted to lessen errors due to filter convergence and end of data. Additionally, the particles were centered near the true source position in the beginning of calculation. Since the PF is a stochastic method the results of 100 trial runs were averaged.

The results show a significant decrease in the RMS error when the propagation delay is included in the estimation. In every SNR value the proposed method outperforms the traditional approach. When the propagation delay is not included a constant bias exists in the RMS error. This is removed by the proposed approach.

Fig. 3. The RMS error of localization with simulations is displayed. The x-axis indicates signal SNR and y-axis is RMS error. The method with propagation delay (∇) increases localization accuracy, compared to not using propagation delay to time-align observations ($\square$).

VI. CONCLUSIONS

This paper discusses sound source localization using propagation delay between the unknown source position and the receiving microphone arrays. The steered responses of the array are proposed to be time-aligned using the propagation delay and then combined. The resulting source location distribution is estimated with a particle filtering approach. The effect of utilizing the propagation delay is tested with simulations. In the simulations a broadband signal of a moving source is subjected to geometrical and atmospheric attenuation and noise. The results indicate that utilizing the proposed approach reduces the localization RMS error.

VII. REFERENCES

- [1] C. Knapp and G. Carter, “The Generalized Correlation Method for Estimation of Time Delay,” *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. 24, no. 4, pp. 320 – 327, Aug 1976.
- [2] A. Johansson, N. Grbic, and S. Nordholm, “Speaker localisation using the far-field SRP-PHAT in conference telephony,” in *International Symposium on Intelligent Signal Processing and Communication Systems, Kaohsiung, Taiwan ROC*, 2002.
- [3] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Signal Processing Series. Prentice Hall, Upper Saddle River, NJ 07458, 1993.
- [4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174 – 188, 2002.
- [5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Statistics

(a) Traditional far-field SRP-PHAT based SLF.

(b) Far-field SRP-PHAT based SLF with time-alignment.

Fig. 2. An illustration of the 2D spatial likelihood function (SLF) using steered array responses directly (left) and with proposed time-alignment (right). The SNR is set to -5 dB in each channel. The endpoints of source path (“□”), source location (“○”), and microphone array locations (“▽”) are illustrated. The source velocity is 66.7 m/s.

for Engineering and Information Science. New York, NY: Springer-Verlag, 2001.

- [6] A. Ali, T. Collier, L. Girod, K. Yao, C.E. Taylor, and D.T. Blumstein, “An Empirical Study of Collaborative Acoustic Source Localization,” in *Proceedings of the 6th international conference on Information processing in sensor networks*, 2007, pp. 41–50.
- [7] B. Ferguson and K. Lo, “Passive wideband cross correlation with differential Doppler compensation using the continuous wavelet transform,” *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3434 – 3444, 1999.
- [8] P. Teal, “Tracking wide-band targets having significant Doppler shift,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 489 – 497, 2007.
- [9] F. M. Dommermuth, “A Simple Procedure for Tracking Fast Maneuvering Aircraft Using Spatially Distributed Acoustic Sensors,” *J. Acoust. Soc. Am.*, vol. 82, pp. 1418 – 1424, 1987.
- [10] P. Pertilä, M. Parviainen, T. Korhonen, and A. Visa, “Moving Sound Source Localization in Large Areas,” in *2005 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2005)*, 2005, pp. 745 – 748.
- [11] Y. Guo, A. Xuea, and D. Penga, “A recursive algorithm for bearings-only tracking with signal time delay,” *Signal Processing*, vol. 88, no. 6, pp. 1539 – 1552, June 2008.
- [12] Eric A. Lehmann, *Particle Filtering Methods for Acoustic Source Localisation and Tracking*, Ph.D. thesis, Australian National University (ANU), Canberra, Australia, July 2004.
- [13] J. Vermaak and A. Blake, “Nonlinear filtering for speaker tracking in noisy and reverberant environments,” in *Int. Conf. Acoust., Speech, and Signal Process., (ICASSP’01)*, 2001, vol. 5, pp. 3021 – 3024.
- [14] C. Voges, P. Bauer, and T. Fingscheidt, “A particle filtering algorithm for audiovisual speaker localisation,” in *Workshop on Positioning, Navigation and Tracking (WPNC)*, 2007.
- [15] ISO:9613-1:1993(E), “International Standard 9613-1: Acoustics – Attenuation of sound during propagation outdoors – Part 1: Calculation of the absorption of sound by the atmosphere,” 1993.