Sound Source Tracking Using Microphone Arrays

WANG PENG and WEE SER
Center for Signal Processing
School of Electrical & Electronic Engineering
Nanayang Technological Univerisy
SINGAPORE, 639798

Abstract: - The capability of being able to track the moving sound sources is one of the most important characteristics of a microphone array system. In this paper, a simple source tracking scheme that employs the statistical property of the measured signal is introduced. Performance analysis and simulation results are also presented.

Key-Words: - **Microphone array, Source tracking, DOA estimation**

1 Introduction

Microphone arrays have attracted much attention in literature in recognition of their good potential to replace the traditional sound capture systems (microphones) [1-6]. The most desirable characteristic of microphone arrays is their good noise reduction and interference suppression ability, which makes them apparently superior solution for badly-conditioned environments. Another appreciated feature is their capability in sound source tracking, which brings in the benefit of self-adaptable directivity. The source tracking ability is widely considered as a byproduct of the direction of arrival (DOA) estimation, and thus has seldom been studied as an independent topic in literature.

Although the tracking of a moving sound source can be achieved by consecutively estimating its DOA, it is sometimes advantageous to employ a dedicated tracking mechanism. In this paper, such a mechanism is described. By exploiting an obvious fact that the spatial samplings on the same wavefront have equal amplitude and thus minimum variance, a simple source tracking scheme has been developed.

2 Problem Formulation

Leaving out the distortion introduced by the sampling process, the measurement of the j^{th} sensor can be expressed as follows:

$$x_j(t) = \sum_{i=0}^{M-1} s_i(t-\tau_{ij}) + n_j(t) \quad (j = 1, \ldots, N) \quad (1)$$

where M is the number of the sound sources, N is the number of the sensors, $s_i(t)$ is the i^{th} source signal, $n_j(t)$ is the additive noise detected by the j^{th} sensor, and τ_{ij} is the corresponding propagation delay. In this paper, our study will be confined to far-field sources and uniform linear arrays (ULAs). Accordingly,

$$\tau_{ij} = \tau_{i1} + (j-1)d\sin\theta_i / c \quad (2)$$

where c is the sound propagation velocity, d is the sensor spacing, and θ_i is the DOA of the i^{th} source. Without loss of generality assume s_0 to be the desired source. Given the value of its DOA θ_0 at a time instant, the objective of source tracking is to trace the deviation of θ_0 thereafter.

3 Sound Source Tracking

Under the supposition that the measured signal is void of interference, (1) can be simplified to

$$x_j(t) = s_0(t-\tau_{0j}) + n_j(t) \quad (j = 1, \ldots, N) \quad (3)$$

Use the term, measurement slice, to describe a collection of samples recorded by different sensors, i.e.,

$$\mathbf{x} = \{x_j(t_j),\ j=1, \ldots, N\} \quad (4)$$

where j does not necessarily to be consecutive. The variance of a measurement slice is defined to be

$$\sigma_x{}^2 = E\left(\sum_{j=1}^{N}\left(x_j(t_j) - \frac{1}{N}\sum_{j=1}^{N} x_j(t_j)\right)^2\right) \quad (5)$$

Further suppose that the additive noises are pairwise independent Gaussian processes with zero mean and variance σ_n^2. Substituting (3) into (5), the following equation is obtained

$$\sigma_x^{\ 2} = \sigma_s^{\ 2} + (N-1)\sigma_n^{\ 2} \tag{6}$$

where σ_s^2 is the variance of the signal components observed by the array, i.e.,

$$\sigma_s^{\ 2} = E\left(\sum_{j=1}^{N}\left(s_0(t_j-\tau_{0j}) - \frac{1}{N}\sum_{j=1}^{N}s_0(t_j-\tau_{0j})\right)^2\right) \tag{7}$$

Obviously, σ_x^2 is minimized when σ_s^2 takes the value of zero. If the source signal is nonperiodic within the time interval of concern, $\sigma_s^2=0$ implies

$$t_i - \tau_{0i} = t_j - \tau_{0j} \ \ \forall \ i,j \in \{1,\ldots,N\} \tag{8}$$

i.e., the signal components of this measurement slice are from the same wavefront. Substituting (2) into (8) and exploiting the far-field assumption, we have

$$\theta_0 = \arcsin(c \times t_\Delta / d) \tag{9}$$

where t_Δ, called the unit time span, is the time difference between two adjacent elements, i.e.,

$$t_\Delta = \frac{t_i - t_j}{i-j} \tag{10}$$

As is demonstrated above, once a measurement slice with minimum variance is detected, the DOA of the source is also learnt. Therefore, the problem of source tracking can be solved by searching over the received data for the qualified measurement slices.

The relationship between the movement of the source and the location of a qualified measurement slice is further revealed by differentiating both sides of (9) with regard to time, and the result is

$$\cos\theta_0 \frac{d\theta_0}{dt} = \frac{c}{d}\frac{dt_\Delta}{dt} \tag{11}$$

Let D be the distance from the source to the array, and $v_{0\mathrm{r}}$ be the rotation velocity of the source that accounts for the deviation of θ_0. $v_{0\mathrm{r}}$ can be calculated by projecting the source velocity along the direction

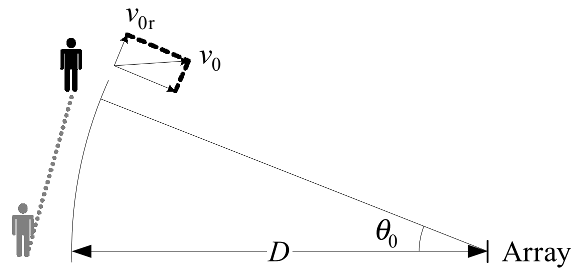

Fig. 1 Projection of the source velocity

perpendicular to the line connecting the array and the source (see Fig. 1).

Noting that $d\theta_0/dt$ is associated with $v_{0\mathrm{r}}$ as

$$\frac{d\theta_0}{dt} = \frac{v_{0\mathrm{r}}}{D} \tag{12}$$

therefore (11) can be rewritten as follows

$$\frac{dt_\Delta}{dt} = \frac{v_{0\mathrm{r}}d}{cD}\cos\theta_0 \tag{13}$$

Assume that T_{E} is the time interval between two consecutive estimations and $v_{0\mathrm{r}}$ is invariant within this time period. It can be derived that

$$\Delta t_\Delta \approx \frac{v_{0\mathrm{r}}d}{cD}\cos\theta_0 \times T_{\mathrm{E}} \tag{14}$$

Forcing T_{E} to be m multiple of the sampling interval T_{S}, we have

$$\Delta t_\Delta < \frac{v_{0\mathrm{r}}d}{cD} m T_{\mathrm{S}} \tag{15}$$

Using this expression we can approximately estimate the change of the unit time span Δt_Δ if the value of m is specified, and thus to outline the suitable searching area; or, calculate the value of m if the searching area is predefined, and thus to decide the favorable estimation interval T_{E}.

Let us use a 4-senor ULA to clarify the above statement. Assume that the distance-aperture-ratio (D/Nd) is 10, $v_{0\mathrm{r}} = 5$ m/s, and $c = 342$ m/s. Substituting these values into (15), we obtain $\Delta t_\Delta < mT_s/2736$, that is to say, the change of the unit time span Δt_Δ within the estimation interval T_{E} is less than one sampling interval

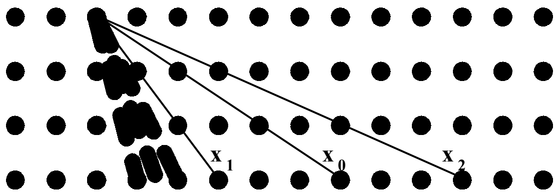

Fig. 2 Illustration of the searching area using a 4-sensor ULA

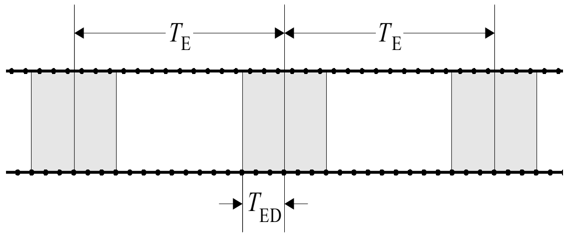

Fig. 3 Equivalent distance

T_S as long as m is smaller than 2736. In this case the searching area is as shown in Fig. 2, where, the dot matrix represents the received data sequences, the row and the column spacing corresponds to the sensor spacing d and the sampling interval T_S, the line $\mathbf{x}_0$ indicates the forecasted position of a qualified slice, in accordance with the previous DOA estimate, and $\mathbf{x}_1$ and $\mathbf{x}_2$ illustrate the upper- and the lower- bounds for the slope of the qualified slice, or to say, define the range of the unit time span.

The qualified measurement slice can then be determined using the following unit decision rule:

$$\sigma_x^2(\mathbf{x}_i) = \min\left\{\sigma_x^2(\mathbf{x}_0),\ \sigma_x^2(\mathbf{x}_1),\ \sigma_x^2(\mathbf{x}_2)\right\} \Rightarrow \mathbf{x}_i \qquad (16)$$

In previous discussion it has been implicitly assumed that at the estimation instant there can be found a measurement slice whose elements are from the same wavefront. However, in practice this is hardly possible because of the incontinuity of the data sequences. When the position of the wavefront lies between two measurement slices, e.g. $\mathbf{x}_1$ and $\mathbf{x}_0$ shown in Fig. 2, the tracking scheme is expected to be able to correctly decide which one of the two slices is the better estimate, i.e., provides more faithful DOA information. Due to the random nature of the source signal and the sampling positions, the decision drawn out from one set of measurement slices, e.g. $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2\}$, is very likely to be incorrect even for clean signal. Fortunately, we do not have to count on the reliability of every decision.

Recall that the DOA information is conveyed by the slopes of the measurement slices (the unit time spans). All parallel slices around the estimation instant are therefore equivalent for our study of source tracking. The maximum time distance from the estimation instant (see Fig. 3), within which all the parallel slices carry the same copy of the DOA information, has thus been called equivalent distance, denoted as T_{ED}. The value of T_{ED} varies between 0 and $0.5T_E$. Since the equivalent distance has been introduced to provide a statistically average effect, cautious calculation of T_{ED} is usually not necessary and a somewhat arbitrary choice as $0.3T_{ED}$ is often adequate. The final decision can now be based on all the equivalent measurement slices, by finding out which equivalent slice group contains the most unit decisions.

The batch processing method described above, by itself, can not guarantee the consistency of the attained decisions. It must be combined with either the interpolation or the upsampling technique, to sufficiently counteract the negative effect introduced by the uncertainty of the sampling positions and the source signal. The effect of the interpolation (or the upsampling) is to secure the general monotonicity of the signal variance σ_s^2 within T_{Span}, the time span of a measurement slice set, and hereby to legitimize the batch decision result. A direct derivative is obtained as follows

$$T_{Span} \ll T_{Sig} \qquad (17)$$

where T_{Sig} is the local time period of the source signal and T_{Span} is the time span of a measurement slice set observed by the source, i.e., the length of the source signal covered by the measurement slices involved in one unit decision. For that shown in Fig. 2, when the corresponding position of a wavefront falls into the coverage of $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2\}$, T_{Span} equals to $6T_s$. Obviously, a sampling rate much higher than Nyquist's is here suggested.

With sufficient signal resolution offered by the interpolation or the upsampling, the batch processing method will be able to give out highly reliable decisions, at least, when there exists no interference and noise.

4 Performance Analysis

4.1 Resolution

The resolution of the proposed sound source tracker, that is, the minimum deviation of the DOA that can be detected, is determined by the difference of the slopes of neighboring measurement slices, e.g. $\mathbf{x}_1$ and $\mathbf{x}_0$ shown in Fig. 2. It can be derived from (9) that

$$\Delta\theta_{\min} \approx \frac{cT_{\mathrm{S}}}{d}\sec\theta_0 \tag{18}$$

Obviously, the source resolving ability of the proposed tracker can be enhanced by increasing the sampling rate T_{S}^{-1} or the sensor spacing d. The latter solution may incur spatial ambiguity [7] and therefore must be applied very discreetly. Another feasible method is to revise the measurement slice formation. Some alternative formations have been illustrated in Fig. 2 with dot line, where the measurement slices are constructed by aligning samples inconsecutive in space, which in effect is equivalent to increase the sensor spacing d. Using these alternative slice formats may also help to release the high demand on the sampling rate since the time spans of the slices may be significantly reduced. However, because fewer samples can be exploited in the calculation of the received signal variance σ_x^2, this method suffers the loss of the robustness against the noise and the interference.

Note that very high tracking resolution is not usually necessary since the spatial resolving capability of the array is limited (the beamwidth). However, some amount of 'redundant' resolution is valuable for the tracking: it increases the tolerance of the array to error decisions and hereby protects the captured sound from being severely spoiled by mistracking.

4.2 Noise and interference

It is not easy to accurately predict the behavior of the noise and the interference. Here we only give several manifest remarks.

The proposed sound source tracking is based on the searching for local minima of the variance of the received signal σ_x^2. If the size of the measurement slices is sufficiently large, the additive noise equally shifts the variance and has no apparent influence on the decision making. The maximal possible size of the measurement slices is restricted by the number of the sensors in the array; therefore, in pursuit of good tracking, relatively larger array should be considered in highly noisy environment. Increasing the sensor spacing d is another method to deal with the additive noise. It can be effectively enhance σ_s^2, the signal component of σ_x^2, and consequently improves the robustness against the noise. However, as has been mentioned earlier, this method may deteriorate the spatial unambiguity.

The presence of interference is possibly a positive factor against the noise when the variances of the interference and the desired signal are approximately in phase. However, it is more general to be destructive and sometimes may even seriously mislead the tracking. Fortunately, when the following condition holds, the interfering effect may be randomized and the interference can then be simply treated as a source of noise

$$T'_{\mathrm{Span}} > 0.5T_{\mathrm{I}} \tag{19}$$

where T_{I} is the local time period of the interference signal and T'_{Span} is the time span of a member of a measurement slice set observed by the interference. Recall the relationship given in (17) and note that

$$T'_{\mathrm{Span}} > (N-1)\frac{d}{c}\left|\sin\theta_{\mathrm{I}} - \sin\theta_0\right| - T_{\mathrm{Span}} \tag{20}$$

where θ_{I} is the DOA of the interference signal. It follows from (19) that

$$(N-1)\frac{d}{c}\left|\sin\theta_{\mathrm{I}} - \sin\theta_0\right| > T_{\mathrm{Span}} + 0.5T_{\mathrm{I}} \tag{21}$$

Obviously, increasing the array aperture, i.e., $(N\text{-}1)d$, is helpful in suppressing the misleading effect of the interference, by converting it into 'noise'.

This expression can be used to estimate the minimum DOA difference between the desired source and the interference that is required by safe tracking when T_{Span} has been fixed. It also places a constraint on the sampling rate, which should be taken into consideration in the design, together with (19).

It should be reminded here the importance of the formation of the measurement slices. A good formation may bring much better performance against the noise and the interference. This should be given particular attention for large arrays and slow moving sources, where, obviously, more formations are available for exploitation.

5 Simulation Results

An 8-sensor array has been studied in the conducted simulations. Suppose that the sensor spacing is 5cm, and the equivalent distance is 100 sampling periods. For simplicity, the source signal is assumed to be a sine wave of 1 kHz and it is sampled at 40 kHz. The rates of error decision at different DOA, derived from 20,000 trials, are shown in Fig. 4. Note that maximum error decision rates appear at around 5°, which in this case corresponds to the midpoint of two neighboring measurement slices.

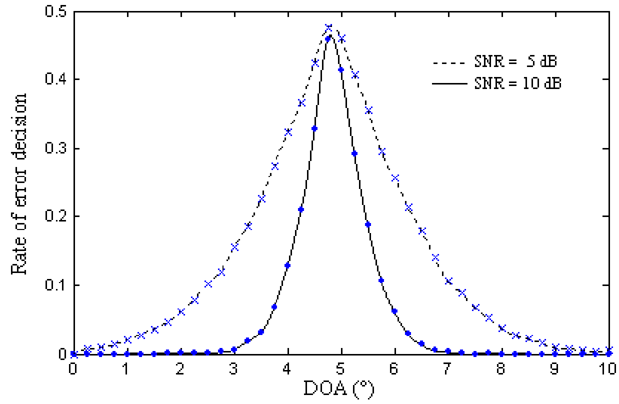

Fig. 4 Rates of error decisions
(N=8, d=5cm, SNR=5, 10dB)

The effect of increasing the array size and the sensor spacing is illustrated in Fig. 5 and 6 respectively. In Fig. 5 a 16-sensor array with sensor spacing 5 cm is studied, while in Fig. 6 an 8-sensor array with sensor spacing 10 cm is investigated. Note that in deriving Fig. 6 the source signal has been sampled at 20 kHz so as to maintain the same tracking resolution. As can be obviously observed, both the methods reduce the rates of error decision apparently, and in comparison, increasing the array size is more effectual.

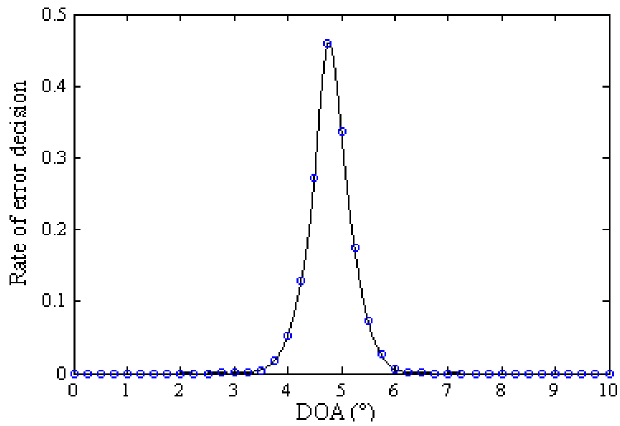

Fig.5 Rates of error decisions
(N=16, d=5cm, SNR=5dB)

It has also been certified in the simulations, that the behavior of the interference is quite like noise when (19) is satisfied.

For wideband sources as speech, bandpass filtering can be employed to obtain a narrowband version of the measured signal, and thus to avoid the unaffordable interpolation or upsampling rate associated with direct exploitation of the original signal. Besides, several narrowband versions can be considered together to further reduce the rate of error decisions.

6 Concluding Remarks

In this paper a sound source tracking scheme using microphone arrays has been introduced. Based on the obvious fact that the measured data from the same signal wavefront have minimum variance, the proposed tracking is easy to understand and implement. Furthermore, the existence of multiple ways to control its performance gives designers more freedom in planning the tracking according to the specific requirements. However, due to the high upsampling or interpolation rate that is associated with sufficient spatial resolution and reliable tracking, large amount of memory is necessary.

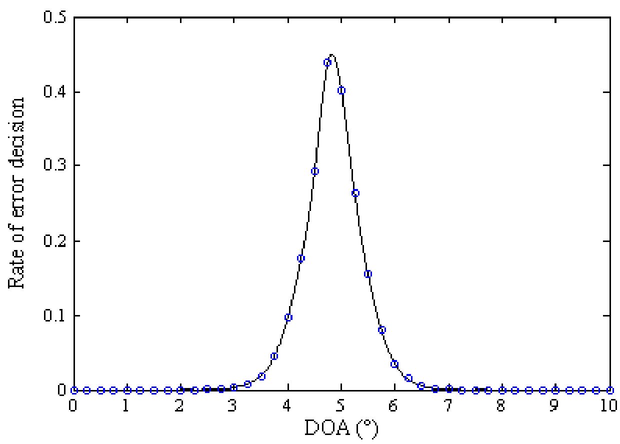

Fig. 6 Rates of error decisions
(N=8, d=10cm, SNR=5dB)

The application of the method presented in this paper is not necessarily to be constrained to source tracking. It can also be adopted in DOA estimation to refine the results, or even be used to estimate DOA directly.

References:

[1] J. L. Flanagan and H. F. Silverman, Material for international workshop on microphone array systems: theory and practice, *Technical Report LEMS-113*, Division of Engineering, Brown University, Providence, RI 02912.

[2] H. Silverman, Some analysis of microphone arrays for speech data acquisition, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1669-1712, Dec. 1987.

[3] Y. Grenier, A microphone array for car environments, *Speech Communications*, vol. 12, pp. 25-39, Jan. 1993.

[4] O. Hoshuyama, A. Sugiyama, and A. Hirano, A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 47, pp. 2677-2684, Oct. 1999.

[5] C. Marro, Y. Mahieux, and K. U. Simmer, Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering, *IEEE Trans. Speech, Audio Processing*, vol. 6, pp. 240-259, May 1998.

[6] Q. Lin, E. Jan, and J. Flanagan, Microphone arrays and speaker identification, *IEEE Trans. Speech, Audio Processing*, vol. 2, pp. 622-629, Oct. 1994.

[7] B. D. Van Veen and K. M. Buckley, Beamforming: a versatile approach to spatial filtering, *IEEE ASSP Magazine*, pp. 4-24, Apr. 1988.