MULTIPLE ACOUSTIC SOURCE LOCALIZATION BASED ON MULTIPLE HYPOTHESES TESTING USING PARTICLE APPROACH

Yeongseon Lee, Ted S. Wada, Biing-Hwang (Fred) Juang

Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA 30332

(yslee, twada, juang)@ece.gatech.edu

ABSTRACT

Localization of multiple acoustic sources in a non-ideal environment has a number of difficulties, among which are accurate acoustic feature estimation for multiple sources and association uncertainty between measurements and their corresponding sources. This paper focuses more on the latter and proposes an algorithm based on a multiple-hypothesis framework for both a measurement model and a measurement association model to localize multiple sources. A conditional data likelihood model based on a measurement hypothesis is proposed and implemented using particles. Simulation results demonstrate that the proposed algorithm is capable of localizing the positions of multiple sources with a small number of microphones without any prior knowledge when the amount of reverberation is moderate.

Index Terms— time difference of arrival estimation, multiple acoustic source localization, data association, multiple hypotheses, data likelihood

1. INTRODUCTION

Acoustic source localization using time difference of arrival (TDOA) measurements from multiple microphone pairs has been an active research area with applications in robots, teleconferences, surveillance, and object tracking. However, in spite of much research for accurately estimating TDOAs in a reverberant and noisy mixing condition, there is no representative algorithm that can be generally applied for multiple source localization in such a non-ideal acoustic environment. Recently, rather than attacking reverberation directly, non-linear Bayesian filtering has been proposed [1, 2] to utilize the fact that the measurements from true sources have temporal consistency while the measurements caused by reverberation or noise may not display such a relationship. Even though this approach has been shown to be quite efficient and robust, the tracking results are strongly dependent on initial source location and associations of TDOA estimates to true sources. For real applications, an effective source location initialization is thus necessary.

This paper proposes an algorithm that can be applied for the initialization of multiple acoustic source tracking. To handle multiple sources, reverberation, and noise all together, TDOA measurements are cast into a multiple-hypothesis framework as in [1, 3]. To associate unlabeled TDOA measurements to multiple sources, what we refer to as measurement association hypotheses are set up, the process of which is similar to target-to-measurement association in [4] but unique since we associate each hypothesis to only one source. A conditional data likelihood for each hypothesis is evaluated using Monte-Carlo technique [5], where particles are generated in the vicinity of possible source locations. The conditional data likelihoods are then compared to each other to select the best hypothesis candidate if the likelihood clearly indicates the existence of a source. The paper also discusses the effect of integer description of measured TDOAs for source localization in two specific microphone array geometries, to which we refer here as a centralized or a distributed array, that have been ignored by most of the past literatures. In particular, the localization performance is analyzed using the two types of array to find a proper microphone array structure.

2. ESTIMATION OF TDOAS FROM MULTIPLE SOURCES

Generally the TDOA of a signal from source k measured at a pair of mic i and mic j is defined by

\[ \tau_{ij,k} = \frac{1}{c} \left( \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2} - \sqrt{(x_j - x_k)^2 + (y_j - y_k)^2 + (z_j - z_k)^2} \right), \]  

(1)

where \( c \) is the speed of sound, \((x_i, y_i, z_i)\) and \((x_j, y_j, z_j)\) are the known positions of mics i and j, respectively, and \((x_k, y_k, z_k)\) is an unknown position of source k.

When there are two signals \( s_1(t) \) and \( s_2(t) \) from two sources, as an example of multiple sources, the recorded signal at mic i, represented by \( x_i(t) \), in an ideal, free-field environment is given by

\[ x_i(t) = \alpha_{i1}s_1(t - t_{i1}) + \beta_{i2}s_2(t - t_{i2}), \]

(2)

where \( \alpha_{i1} \) and \( \beta_{i2} \) are the attenuation factors from source 1 and source 2 to mic i, respectively, and \( t_{i1} \) and \( t_{i2} \) are arbitrary time delays. Then the cross-power spectrum of signals measured at mics i and j is

\[ G_{ij}(\omega) = X_i(\omega)X_j^*(\omega) \]

(3)

\[ = \alpha_{i1}\alpha_{j1}|S_1(\omega)|^2e^{-\omega(t_{i1}-t_{j1})} + \beta_{i2}\beta_{j2}|S_2(\omega)|^2e^{-\omega(t_{i2}-t_{j2})} + \alpha_{i1}\beta_{j1}S_1(\omega)S_2^*(\omega)e^{-\omega(t_{i1}-t_{j2})} \]

(4)

\[ + \alpha_{i2}\beta_{j1}S_1^*(\omega)S_2(\omega)e^{-\omega(t_{i2}-t_{j1})} \]

\[ = \alpha_{i1}\alpha_{j1}|S_1(\omega)|^2e^{-\omega\tau_{i1j}} + \beta_{i2}\beta_{j2}|S_2(\omega)|^2e^{-\omega\tau_{i2j}} \]

(5)

where \( X_i(\omega) \) is the Fourier transform of \( x_i(t) \) and \( \tau_{i1j} = t_{i1} - t_{j1} \) and \( \tau_{i2j} = t_{i2} - t_{j2} \) are the TDOAs. The last two cross-terms in (4) become zeros due to the assumption of uncorrelated acoustic sources. Thus TDOAs corresponding to the two sources may be estimated by \( \tau_{i1j} \) and \( \tau_{i2j} \) that maximize \( G_{ij} \).

Among many methods developed in the past to improve the accuracy of TDOA estimation, the phase transform (PHAT) [6] is still one of the most effective when using a speech signal in a moderately
reverberant environment. PHAT in terms of (5) is given by
\[ R_{ij}(\tau) = \int_{-\infty}^{\infty} \frac{\alpha_1|S_1(\omega)|^2 e^{-j\omega \tau_{ij}} + \beta_2|S_2(\omega)|^2 e^{-j\omega \tau_{ij}^2}}{\alpha_1|S_1(\omega)|^2 + \beta_2|S_2(\omega)|^2} e^{j\omega \tau} d\omega. \] (6)
Since \( R_{ij}(\tau) \) can have spurious peaks as well as true peaks in a noisy environment, TDOAs from the candidate sources are obtained by finding the peaks that are greater than a certain threshold \( \xi_1 \):
\[ \tau = \{ \tau_{ij} : R_{ij} > \xi_1 \}. \] (7)

In practice the TDOA estimate is expressed in terms of number of samples, given by
\[ \hat{\tau}_{ij} = \text{int}(\tau_{ij}F_s) \text{ samples}, \] (8)
where \( \text{int}(\cdot) \) is the nearest integer function and \( F_s \) the sampling frequency, since the sampling is performed in discrete time such that a TDOA measurement is discretized into intervals of one sampling period.

3. LOCALIZATION OF MULTIPLE SOURCES

Localizing an acoustic source can be described as a process of finding the best intersection in the 3D source-microphone space by associating the TDOA measurements from multiple microphone pairs to a source. However, in multiple source environment, the mapping between TDOA measurements and the corresponding sources is ambiguous. Moreover, TDOA estimates may not be reliable due to interference from reverberation or noise. Therefore, one real challenge for acoustic source localization, especially for multiple sources, is correctly associating each unlabeled TDOA peak to a particular source.

3.1. Measurement association hypotheses

Without knowing the associations between TDOA measurements and sources, we need to hypothesize a measurement-to-source association. For example, [4] derives each hypothesis by drawing one TDOA measurement from each microphone pair and associating the TDOAs across all microphone pairs such that the combination is associated to one source.

Let \( \tau^p = \{ \tau^p_j : j = 1, 2, \ldots, M^p \} \) be the TDOA measurements from microphone pair \( p \), where \( M^p \) is the number of peaks detected according to (7), and let \( \mathbf{z} \) be a set of TDOAs at one time instance for the entire microphone array, represented by a vector \( \mathbf{z} = [\tau^1, \ldots, \tau^{N^r}] \).

The total number of measurement association hypotheses is a product of the number of measurements from each microphone pair plus one (for invalid measurement), i.e., \( \prod_{p=1}^{N^r} (M^p + 1) \). For the example in Table 1, the number is \((2 + 1)(3 + 1) = 12\). When the number of microphone pairs is increased to \( N^r = 4 \) and each microphone pair gives three TDOAs measurements, the total number of associations is \((3 + 1)^3 = 256\). Thus the number of hypotheses can quickly get out of hand, and a large amount of computation would then be required for the evaluation of all hypotheses. Therefore, the hypotheses with TDOAs that are bigger than admissible values for each microphone pair are eliminated beforehand. Those with not enough TDOA measurements for localization in a 3D space are also removed.

3.2. Conditional data likelihood

A data likelihood is generally defined as \( p(\mathbf{z}|x) \), where \( \mathbf{z} \) is a measurement set and \( x \) is a state, i.e., the 3D locations of multiple sources. As we cannot evaluate \( p(\mathbf{z}|x) \) directly since the association between the measurement set and the locations of all sources is unknown, by conditioning on a given hypothesis \( r_h \), the data likelihood is simplified as
\[ p(\mathbf{z}|x, r_h) = p(\tau^1, \ldots, \tau^{N^r}|x_k, r_h) \]
\[ = \prod_{p=1}^{N^r} p(\tau^p_j|x_k), \] (10)
\[ = \prod_{p:|p|=0} p_C(p) \prod_{p:|p|=0} p(\tau^p_j|x_k) \] (11)

That is, \( x \) is replaced by \( x_k \) for a particular source \( k \) in (10) since each measurement association hypothesis maps to one source. The factorization in (11) is possible due to statistical independence of measurements across microphone pairs, where \( j(p, h) \) represents the measurement label corresponding to microphone pair \( p \) and measurement association hypothesis \( r_h \), e.g., the entries in Table 1, and \( \tau^p_j|x_k \) is TDOA from microphone pair \( p \) for a given hypothesis \( r_h \). The data likelihood is further factored into two sets in (12): the first term is the probability of clutter, denoted by \( p_C \), which is usually assumed to be a uniform probability over the measurement space of microphone pair \( p \), and the second term is the data likelihood for the measurement from a true source, approximated by Gaussian
\[ p(\tau^p_j|x_k) \propto \exp\left[ -\frac{1}{2} (\tau^p_j-x_k^p)^T \Sigma^{-1} (\tau^p_j-x_k^p)-\frac{1}{2} \right], \] (13)

where \( \Sigma \) is TDOA from microphone pair \( p \) for a given state \( x_k \) and \( \Sigma^{-1} \) is by practice designed to be a diagonal matrix. In our case, \( \tau^p_j|x_k \) is estimated by PHAT, whereas \( \Sigma \) is determined through the particles approach as discussed in the next section.

3.3. Source localization using conditional data likelihoods in particle implementation

The next and final step is to find the right measurement association hypotheses corresponding to true sources using the conditional likelihoods. Here we implement the conditional data likelihoods with
particles through Monte-Carlo approach \cite{5} since it eliminates a triangulation process for source localization.

First, $N$ particles are generated over a possible support area by using a uniform distribution for each particle $x^{(n)}$. Next, the data likelihood in (12) is evaluated for each hypothesis $r_h$ such that the maximum value of $p(z|x^{(n)}, r_h)$ and the corresponding $x^{(n)}$ are determined. These maximum conditional data likelihood values are then compared among all measurement association hypotheses and gated by using a threshold. Choosing a proper threshold is very critical when the measurements are noisy. Therefore, we choose the threshold to be small enough in order to admit more hypotheses than the actual number of sources. Since the chosen measurement association hypotheses include many subsets of each other, the best hypotheses may be selected from each subset. Finally, the source locations are estimated from pre-selected hypotheses when the mean-square error between the estimated and the generated TDOA measurements is less than some threshold $\varsigma_2$:

$$\hat{x}_k = \{ x_k : \sum_{p=1}^{N_x} (\tau^p_{j(p,h)} - \hat{\tau}^p(x_k))^2 < \varsigma_2 \}. \quad (14)$$

4. MICROPHONE POSITIONING

In general, source localization using TDOAs becomes less precise as a source moves farther away from the microphone array. Specifically, the degradation in the localization performance in one direction is observed to be worse when compared to the other directions especially when all microphones are located close to each other.

For example, Figure 1(a) shows TDOA contours calculated over the x-y space at a fixed location along the z-axis by using the so-called centralized array, whose microphone positions used during our experiment are provided in Table 2. As the last two plots in Figure 1(a) show, microphone pairs 5 and 6 contribute to poor localization performance along the y-axis since they give almost the same TDOAs over the upper half of the x-y space when the TDOA is approximated as integer value of samples by practice. Consequently, the intersection of the TDOAs for a source at a far field is represented as a long beam in the direction of the source rather than a point around the actual source location as illustrated in Figure 1(b).

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Mic # & Centralized array & Distributed array \\
\hline
& (x,y,z) cm & (x,y,z) cm & \\
\hline
1 & (300,40,150) & (300,10,120) & \\
2 & (330,40,150) & (380,10,120) & \\
3 & (270,40,150) & (340,10,160) & \\
4 & (300,10,150) & (340,10,80) & \\
5 & (300,70,150) & (590,250,120) & \\
6 & (300,40,180) & (590,330,120) & \\
7 & (300,40,120) & (590,290,80) & \\
8 & & (590,290,160) & \\
\hline
Mic pairs & (1,2),(1,3),(1,4) & (1,2),(1,3),(1,4) & \\
& (1,5),(1,6),(1,7) & (5,6),(5,7),(5,8) & \\
\hline
\end{tabular}
\caption{The positions of microphones.}
\end{table}

The localization precision can be improved by distributing microphones around a tracking area. The positions of microphones in the so-called distributed array that we tested are provided in Table 2. In such an array configuration, microphones can be distributed in many ways. Here we divide microphones into two groups that are distributed over different locations, where each group consists of four microphones. As shown in Figure 1(c), the overlapping TDOA area becomes much more localized when the distributed microphone array with two groups is used.

5. SIMULATIONS

5.1. Simulation settings

A simulation is performed in a room with the dimension of $6m \times 4m \times 3m$. The reference point (0,0,0) is set at the bottom-left corner of the room. Microphones are positioned in either the centralized array or the distributed array as listed in Table 2. Both microphone arrays use 6 microphone pairs. Microphone signals are generated by imaging methods \cite{7} using two sources, a male and a female, whose voices were captured at the sampling rate of 16 KHz and whose locations are according to various scenarios as listed in Table 3. White Gaussian noise of 30 dB signal-to-noise ratio (SNR) is added to the microphone signals generated by the imaging method.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Source 1 & Source 2 & Description of Sources \\
\hline
1 (100,100,120) & (500,100,120) & Far apart at a near field \\
2 (100,360,120) & (150,360,120) & Near each other at a far field \\
3 (100,360,120) & (500,380,150) & Far apart at a far field \\
\hline
\end{tabular}
\caption{Three scenarios for source locations.}
\end{table}

5.2. Simulation results

5.2.1. Estimation of TDOAs

The estimation of TDOAs from multiple sources is difficult even in a moderately reverberant environment. Several reverberation conditions were tested: $T_{60} = 150$, 300, and 600 ms with SNR = 30 dB. However, PHAT was able to provide reliable TDOAs for performing the localization for all three scenarios only when $T_{60} = 150$ ms.

Figure 2 shows the estimated TDOAs for Scenario 2, which is the most difficult scenario of all since the sources are located very close to each other at a far field. PHAT using $\varsigma_1 = 0.2$ in (7) was able...
to detect multiple TDOAs within ±1 sample error for each source with negligible missed detection error. We also observed that the accuracy of the estimated TDOA is strongly related to the distance between a source and microphones. If the distance was short even though the reverberation is high, the TDOA estimation was very accurate. The estimated TDOA was also better if microphones were placed in the center of a room and not along the walls.

5.2.2. Localization of multiple sources

The proposed localization method was successfully tested on the two array configurations. The results are shown in Figure 3, where the plots at the left column are from using the centralized array while those at the right column are from using the distributed array. As expected, using the centralized array exhibits a large error in the y direction, whereas the distributed array allows for much more accurate localization of source positions.

6. CONCLUSIONS

This paper presented an algorithm based on a measurement hypotheses framework for both a measurement model and a measurement association model to localize multiple sources without any prior knowledge. A conditional data likelihood model based on a measurement association hypothesis was proposed and implemented using the particles approach. The simulation results showed that the proposed algorithm is able to localize two sources in a moderately reverberant and noisy environment. However, we still need to improve the performance of TDOA estimation when the reverberation is more severe, e.g., we may have to include other types of sensors such as a camera to supplement the direction information.

7. REFERENCES