A Hybrid Approach for Speaker Tracking using Time of Arrival with Concave-Convex Procedure

Dhanya R, Smitha K S

Abstract—Single source localization problem using Time f Arrival (ToA) technique is described here.Time of Arrival is the travel time of a radio signal from a single transmitter to a remote single receiver. Among various models for localization measurement of ToA is relatively direct since by identifying and locating known samples from transmitted source signal, the signal arrival time can be determined. The corresponding unknown source-measurement associations can be incorporated into optimization. An efficient three step algorithm is used to solve this optimization problem which includes the steps of course location estimation, determination of source-measurement association and source location refinement. This approach simplify the problem with convex relaxation and approximation techniques. Here a popular optimization package like CVX is used. The proposed algorithm has low computational complexity and is feasible for real time applications.

Index Terms—Time of Arrival measurement, Voice Activity Deduction(VAD), optimization, convex relaxation, course location estimation.

I. INTRODUCTION

A central topic in spoken-language-systems research is what's called speaker diarization, and it gives the answer who spokes when. Speaker diarization is an essential function of any program that automatically annotated audio or video recordings. Speaker tracking aims to detect segments corresponding to a known set of target speakers speaker diarization aims to detect speakers without any prior knowledge about them. Telephone conversations, recorded meeting and broadcast news audio are some applications of speaker diarization[1],[2]. The two common approaches of speaker diarization are audio segmentation and speaker detection. The VAD system is used for selecting speech segments from an input signal and ignores non-speech segments.

Signal measurement is the key to tracking and localization. Several measurement models have been proposed in the literature. It includes measurements of Time Difference of Arrival (TDOA)[3], Time of arrival (TOA)[4], Angle of Arrival (AoA)[5], Received Signal Strength (RSS) [6], combinations of above [7] etc. In TDOA, the time difference of arrival of two signals observed by two adjacent microphones is taken. The mathematical relations between the TDoAs and the observed signals are nonlinear and non-injective. Hence the linear estimators cannot be used to obtain the estimations. In [8], source tracking using TDOA with expectation maximization algorithm is described.

TOA (time of arrival) and TDOA (time difference of arrival) methods use geometric relationships based on distances or distance differences between the sensor nodes and source nodes. The TOA method uses the transit time between transmitter and receiver to find distance directly, whereas the TDOA method finds the location from the differences of the arrival times measured on pairs of transmission paths between the target and fixed nodes and EM algorithm is more effective [9]-[10]. Both TDOA and TOA are based on the time-of-flight (TOF) principle of distance measurement, where the parameter, time interval, is converted into distance by multiplied with the speed of propagation. Among these various models, measurement of TOA is relatively direct since by identifying and locating known samples from transmitted source signal, the signal arrival time can be determined. In this work, it is assumed that the cooperation of source and sensor nodes is such that at sensor nodes the propagation time of the signal can be found.

In many papers[9]-[10] multiple source localization with unknown source-measurement associations is described. In this work, single source localization using Time of Arrival technique is addressed. The unknown source-measurement associations can be characterized by binary variables and so can be incorporated into optimization. An optimal search method is one that always finds the best solution or *a* best solution, if there is more than one. Optimization is vital to modern speech and natural language processing systems. An efficient three step algorithm is used to solve this optimization problem which includes the steps of course location estimation, determination of source-measurement association and source location refinement. This approach simplify the problem with convex relaxation and approximation techniques.

The remainder of the paper is organized as follows. In section II the system model and optimization problem are described. Section III describes results before conclusions in section IV.

II. METHODOLOGY

A. System Model

Consider a model of the room with M sensor nodes receiving signals from a source, S who is moving. The locations of the sensor nodes are known. The problem is to track the unknown source node. The conceptual diagram of the problem is shown in Fig.1.

Revised Version Manuscript Received on August 13, 2015.

Dhanya R, Electronics and Communication Engineering Department, LBS Institute of Technology, Thiruvananthapuram, Kerala, India.

Smitha K S, Electronics and Communication Engineering Department, LBS Institute of Technology, Thiruvananthapuram, Kerala, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

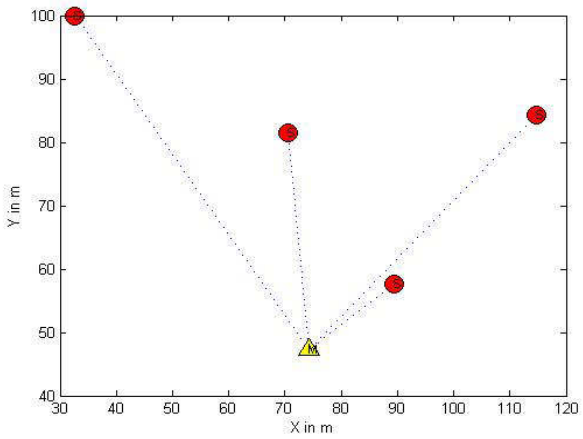

Fig.1. Illustration of sensor nodes and speaker constellation

The above figure shows a room model created which consists of a single speaker and four sensor nodes. The triangle represents the source node and the red circles show the sensor nodes, ie, microphones.

The received signals from each sensor nodes can be written as

$$z_m(t,k) = \sum_s a_m(t,k)v(t,k) + n_m(t,k) \qquad (1)$$

where $z_m(t,k)$ is the signal received at the m-th sensor node

$a_m(t,k)$ denotes the acoustic transfer function

v(t,k) is the speech signal emanating from the speaker

and $n_m(t,k)$ is the Additive white Gaussian noise.

t=0,….T-1 denotes the time index and k=0,…..K-1 denotes the frequency index.

All the signals received by the sensor nodes are sent to a decision fusion center where measured signals are processed to obtain estimations.

Let $z = vec_{m,t,k} z_m(t,k)$ (2)

where vec operation denotes the vector concatenation of all received sensor node signals in frequency mode.

The acoustic transfer function can be written as

$$a_m(t,k) \cong \frac{1}{\left\| p_s - p_m \right\|} \exp(-j \frac{2\Pi k}{K} \frac{t_i}{T_s}) \qquad (3)$$

where T_s denotes the sampling period

p_s and p_m are speaker s and sensor node m locations and

t_i is the time of arrival

t_i is the time of arrival of the source node at the i-th sensor node and has the form

$$t_i = \frac{1}{c}\left\| x_i - y_j \right\| + \tau_j + n_i \qquad (4)$$

where c is the speed of light

τ_j is the unknown initial transmission time instant of the source and

n_i is the TOA measurement of noise.

All sensor nodes send their TOA measurements to a decision fusion center where the measured data is processed to find the estimations of the source node locations. But the initial transmission time of the source node is unknown. The sensor nodes can only sense the arriving signals. The measured TOA values can be written in matrix form.

Let t_i be TOA measurement vector at the i-th sensor node and P_i be the source-measurement association of the i-the sensor node. The permuted version of t_i is obtained by applying source measurement associations to t_i and is denoted by $\bar{t}_i$.

$$\bar{t}_i = P_i t_i = [t_1, ...t_M]^T, \forall i = 1, ...N. \qquad (5)$$

B Optimization Problem

Here a TOA based source localization is described and for that an optimization problem is described. In the literature, previously mentioned it is assumed that no prior knowledge about the number and identities of speaker in the tracking process. . If the speakers are known apriori, for speaker segmentation traditional speaker identification algorithm can be used. However, in many cases, such as continuous speech stream from live news broadcasting or a meeting, the apriori knowledge of speaker identities and the number of speakers are often not available or difficult to obtain. Even in well-structured news broadcasting, we cannot assume that the anchor persons are always the same. Therefore, it is desirable to perform unsupervised speaker change detection and tracking algorithm in audio content analysis. In this work, adaptive filter which is based on Time of Arrival (TOA) technique for tracking accuracy is proposed. First room model is initialized. Then Voice Activity Deduction (VAD) is done on audio files received from each microphone. The VAD system is used for selecting speech segments from an input signal and ignores non-speech segments. The Time of Arrival of each signal in each microphone is taken and the adaptive filter is used to sample the space of possible speaker locations and to fuse the bearing measurements from audio sources. The proposed approach incorporates kinematic information of moving speaker by using an estimator for each speaker in order to constrain the evolution of the location measurements and then fuses the location estimates of the same speaker from multiple microphone arrays for better coverage of the sensed environment and directly accounts for the measurement origin uncertainty. From (5), p_i^m depends on the start transmission time and location of the k_m-th source node which are also unknown variables. Thereby, with TOA measurement vectors from all the sensor nodes, a joint estimate of p_i^m, τ_{k_m} and y_{k_m} are needed. First, we define the resulting estimation error at the i-th sensor for the m-th source as

$$e(i-1)M + m) = p_i^m t_i - \frac{1}{c}\left\| x_i - y_{k_m} \right\| - \tau_{k_m}$$
$$\forall i = 1, ..., N, m = 1, ..., M \qquad (6)$$

Where estimation error vector **e** is denoted by $[e_{(i-1)M+m}]$.

The optimization problem is

$$e_i p_i \in \Pi^{MxM} \; y_{k_m} \tau_{k_m} \qquad \|e\|_q$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

subject to

$$e(i-1)M+m$$

$$= p_i^m t_i - \frac{1}{c}\|x_i - y_{k_m}\| - \tau_{k_m} \qquad (7)$$

The problem involves mixed integer variables, ie, it includes both discrete and continuous optimization variables and non-convex continuous optimization and so is difficult to solve. So the approach to solve this optimization problem consists of three steps.

(a) First course location of the source node is estimated. It is a difficult one because it is a non-convex mixed integer problem. For solving this permutation matrix is converted into convex non-integer constraint. This convex approximation is also called concave-convex procedure. A number of observations we get using this procedure is shown in Fig.2.

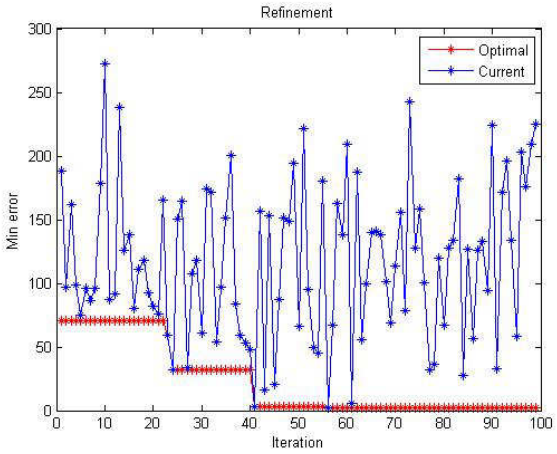

Fig.2. approximations obtained using concave-convex procedure

(b)Then in the second step source-measurement association is found and then from all these obtained values we get the source locations. Now the problem is convex and a software tool such as CVX can be used to solve this.

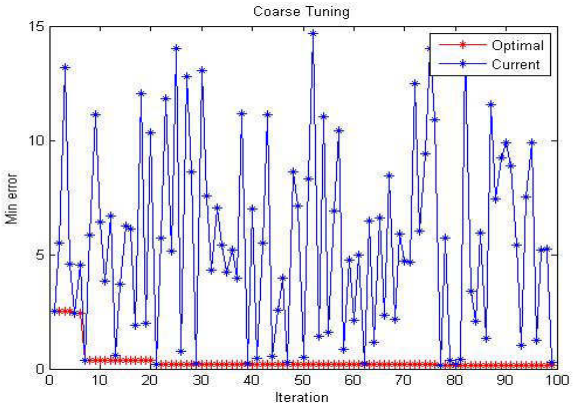

Fig.3. Estimated values after optimization

From the first step we get a set of initial estimates of source locations. And the in the second step we can find the source-measurement association from the permutation matrix, P_i, so it is possible to improve the accuracy of the location. Now all the estimations obtained from the permutation matrix can be refined to get closer values.

III. RESULTS

This work is simulated using MATLAB with the help of software packages CVX and ISM. For the simulation a room model is created using ISM with three sensor nodes as shown in Fig.4.

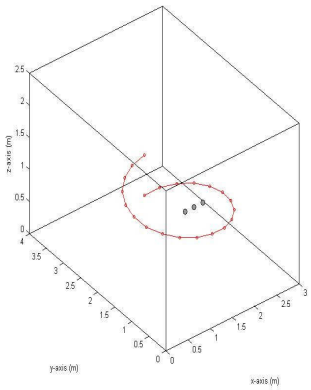

Fig.4. A room model with moving speaker

As in the previous work if we are considering the TDOA of different signals at different sensor nodes, we get the tracking positions as shown in Fig.4. Instead of TDOA as in this work, if we are considering TOA with optimization, the tracking position is obtained as in Fig.5.

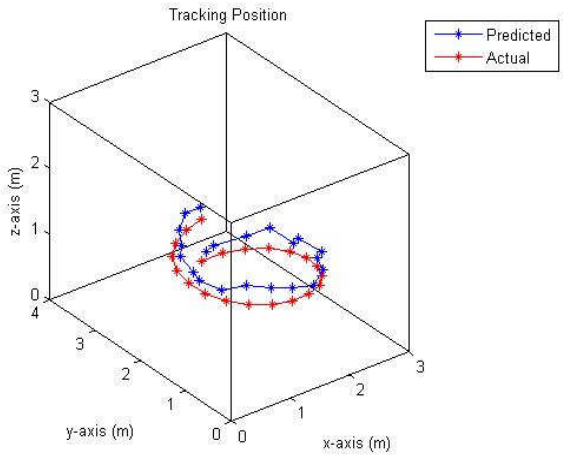

Fig 5. Speaker tracking using TDOA method

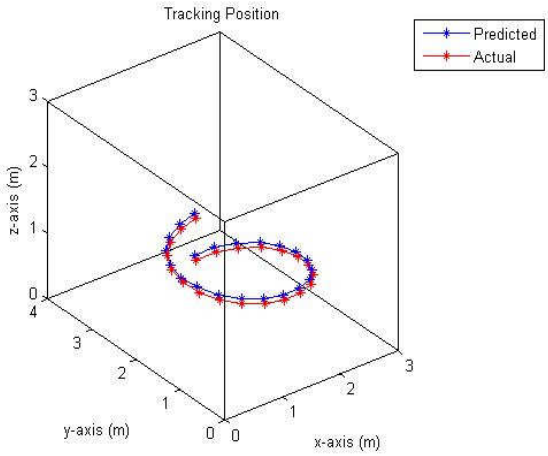

Fig 5. Speaker tracking using TOA method

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

The mean value of the error in the tracking using TDOA method is obtained as 0.1825 and in this work the error can be reduced and is 0.0617. A speaker with circular as well as noncircular motion can be tracked. Fig.6 and 7 shows the tracking of a speaker with noncircular motion.

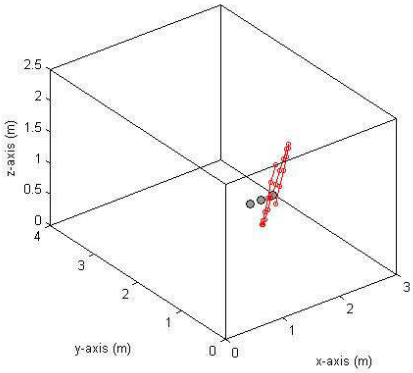

Fig.6 Speaker with noncircular motion

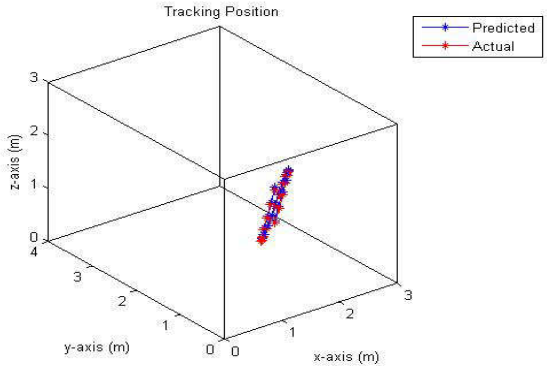

Fig.7. Tracking of a speaker with noncircular motion

IV. CONCLUSION

Tracking the locations of a speaker in the presence of background noise and reverberation is of great interest in a number of applications. Speaker recognition includes speaker identification and verification. In the existing tracking methods, it is assumed that the input speech belongs to a speaker whose is known apriori. In many applications, such as a real-time conversation or news broadcasting, the speech signal is continuous and no information is there about the beginning and ending of the speech segment. Therefore, for indexing speech streams based on speaker analysis based on audio track, first it is necessary to find speaker change points before the speaker can be identified. Speaker tracking is essential in many applications, such as meeting, audio/video browsing etc. and speaker change detection is a preliminary processing for speaker tracking.

In previous studies, it is assumed that there is no prior knowledge about the number and the identities of speaker in tracking process. If the speakers are registered apriori, traditional speaker identification algorithm can be used for speaker segmentation. However, in many cases, such as continuous speech stream from live news broadcasting or a meeting, the apriori knowledge of speaker identities and the number of speakers are often not available or difficult to obtain. Even in well-structured news broadcasting, we cannot assume that the anchor persons are always the same. Therefore, it is desirable to perform unsupervised speaker change detection and tracking algorithm in audio content analysis.

In this work, adaptive filter based on Time of Arrival (TOA) technique for tracking accuracy is proposed. The Time of Arrival of each signal in each microphone is taken and the adaptive filter is used to sample the space of possible speaker locations and to fuse the bearing measurements from audio sources. The proposed approach incorporates kinematic information of moving speaker by using an estimator for each speaker in order to constrain the evolution of the location measurements and then fuses the location estimates of the same speaker from multiple microphone arrays for better coverage of the sensed environment and directly accounts for the measurement origin uncertainty. The method requires low computational complexity and is feasible for real-time applications. The effectiveness of the approach can be illustrated by extensive simulation study on tracking a single moving speaker.

REFERENCES

1. N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, III, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.
2. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Process. Mag.*, vol. 22, no.4, pp. 24–40, Jul. 2005.
3. N. Patwari, A. O.Hero, III, M. Perkins,N. S.Correal, andR. J.O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.
4. K. Yang, G. Wang, and Z.-Q. Luo, "Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2775–2784, Jul. 2009.
5. X. Li, "Collaborative localizationwith received-signal strength inwireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3807–3817, Nov. 2007.
6. D. Niculescu and B. Nath, "Ad hoc positioning system (APS) using AOA," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM'03)*, San Francisco, CA, Mar./Apr. 2003, vol. 3, pp. 1734–1743.
7. L. Cong andW. Zhuang, "Hybrid TDOA/AOAmobile user location for wideband CDMA cellular systems," *IEEE Trans. Wireless Commun.*, vol. 1, no. 3, pp. 439–447, Jul. 2002.
8. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pp. 1–38, 1977.
9. Ofer Schwartz, Sharon Gannot, "Speaker tracking using Recursive EM Algorithms", IEEE transactions on audio, Speech And Language Processing, Vol. 22, No.2, Month 2014.
10. H Shen, Zhi Ding, Soura Dasgupta, Chunming Zhao, "Multiple Source Localization in Wireless Sensor Networks Based on Time of Arrival Measurement", IEEE Transactions on Signal Processing, Vol.62, No.8, April 15 2015

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.