Multiple Gunshots Source Algorithm based on Integration of Cluster Analysis

Shu LI
Hunan City University, Yiyang, Hunan, 413000, China

Abstract: Gunshots positioning system can quickly determine the position of the gunshots source. It will help to protect the safety of soldiers on the battlefield, assist police to crack down on violent fear of molecules. On how to improve the performance of multiple gunshots source localization algorithm gunshots proposed multiple gunshots source localization algorithm based on integration of sub frame estimation and cluster analysis. First, a signal is divided into 8 sub frames. Each sub frame signal using the phase variation is calculated in response to the power control weighting function, which searches the sound source position to obtain the maximum value of the sub frame estimation. Because the signal in the time domain gunshot is sparse, these estimates correspond to a plurality of sound source position. Then the estimated values of the sub frame is divided into several categories using the convergence of clustering algorithm. Finally controlled in response to the sub frame by the average estimated value of the power function was evaluated to give the final gunshot source location estimation. Experimental results show that in the case of 2-3 gunshots, the positioning performance of the algorithm is much higher than traditional algorithms.

Keywords: Gunshots Positioning System; Cluster Analysis; Sound Source Signal; Multiple Gunshots Source; Subframe; Anti-terror Attacks

1. Introduction

Microphone array system can be widely used in car hands-free communication [1], speech separation [2], the sound source monitoring [3] and other fields. Sound source localization based on microphone array technology has played a key role in these systems. Such as hands-free communication systems, it may use the beam source array according to the estimated orientation algorithm to suppress the formation of interference and noise. Since the 1970s, researchers at home and abroad to gunshot positioning method conducted in-depth research, the various positioning algorithm and used the actual sound source localization system [4-6]. At present, on the basis of a single sound source localization based on a growing number of researchers have begun to focus on multiple sound source localization techniques.

In the real environment, such as ordinary buildings environment, in addition to an array of received signals direct sound but also have reflected sound and background noise. Reflected in a number of obstacles repeatedly claimed to reverb, which is an important factor affecting the positioning performance. In the case of multiple sources of gunshots, the sound source localization algorithm must also have the ability to distinguish between multiple sources. Literature [7] proposed multiple sound source localization algorithm to simulate the human auditory, in the real world can achieve 4 position of the sound source, but the algorithm is computationally expensive and still difficult practical. Literature proposed the formation of pyramid-based multiple sound source localization algorithm. The proposed algorithm

1.1. Traditional SRP-PHAT Multiple Sound Source Localization Algorithm

Based on the received signal sub frame sound source localization techniques divided the array microphone array into sub frame first. Then estimate the location of the sound source signal according to one or more frames. As mentioned above, the time-frequency analysis algorithm requires tens to hundreds of frame signal in order to estimate the position of the plurality of sound sources. The length of frame signal is several tens of milliseconds, so this type of algorithm is used only for a stationary sound source positioning. SRP-PHAT algorithm only need one data to estimate the location of the sound source, with a wider range of practical value.

1.2. Improved SRP-PHAT Sound Source Localization Algorithm

The basic idea of the improved algorithm is based on speech signal sparsity in the time domain. The signal frame is divided into several sub frame to process. Suppose that frame length is T, which is divided into N_{SF} sub frames. The sub frames do not overlap, the sub frame size is $T_{SF} = T / N_{SF}$. For each sub frame using the algorithm described in Section 2 to give the SRP functions

$\hat{P}_n^{\text{PHAT}}(\boldsymbol{q})$ of each sub frame, the sound source position to obtain the sub frame estimation.

$$\hat{\boldsymbol{q}}_n = \arg\max_{\boldsymbol{q}\in Q} \hat{P}_n^{\text{PHAT}}(\boldsymbol{q}),\quad n = 1,2,\ldots,N_{SF} \qquad (1)$$

According to the sparsity of assumptions, these estimates include multiple sound source location.

AC clustering will be divided into N_c classes, but not every class corresponds to a sound source. Similar to a single sound source, due to the influence of reverberation and noise, some of the sub frame may get the wrong estimate. These estimates are far away from the true value, which will become a separate category clustering. The number of class $C(k)$ elements is $|C(k)|$, which means that sub frames have been estimated that there are similar. The greater the $|C(k)|$ value, $C(k)$ contained in the real sound source location estimates that the greater the possibility. We set the threshold γ_{th}, leaving only the classes $|C(k)| > \gamma_{\text{th}}$ that can be expressed as $\{C(k),\ k = 1:\tilde{N}_c\}$, apparently $\tilde{N}_c \le N_c$. The larger γ_{th} helps to reduce the estimated error, but may discard some of the correct estimate. Its value should be appropriately selected according to the number of gunshots. A class has a number of estimates, they correspond to the same source, from these estimates also need to select a best estimate. In order to evaluate these estimates, we define the function of the average sub-frame SRP

$$\hat{P}_{\text{avg}}^{\text{PHAT}}(\boldsymbol{q}) = \frac{1}{N_{SF}}\sum_{n=1}^{N_{SF}} \hat{P}_n^{\text{PHAT}}(\boldsymbol{q}) \qquad (2)$$

The k_{th} class in the best estimate for

$$\hat{\boldsymbol{q}}_k^* = \arg\max_{\boldsymbol{q}\in C(k)} \hat{P}_{\text{avg}}^{\text{PHAT}}(\boldsymbol{q}),\ k = 1,\ldots,\tilde{N}_c \qquad (3)$$

Due to the sudden of gunshots signal, multiple gunshots in a frame signal is not necessarily completely overlap. For example two gunshots, one frame of a signal may contain multiple gunshots or may contain only one repeated gunshot. We use N_s represent the number of the gunshots source, assuming its value is known, N_a represents a number of gunshots echo. The value is unknown, $\hat{N}_a$ represents the estimated values, said there should be $\hat{N}_a \le N_s$. The target of multiple sound source localization algorithm is based on a frame data to estimate the number and location of gunshot source. According to formula (9), we obtain the sound source position estimated at $\{\hat{\boldsymbol{q}}_k^*,\ k = 1:\tilde{N}_c\}$. If $\tilde{N}_c \le N_s$, the above estimate is the final output, and $\hat{N}_a = \tilde{N}_c$. If $\tilde{N}_c > N_s$, but also choose an estimated value of $\hat{N}_a$ from $\{\hat{\boldsymbol{q}}_k^*,\ k = 1:\tilde{N}_c\}$, then $\hat{N}_a = N_s$. According to the value of $|C(k)|$ and $\hat{P}_{\text{avg}}^{\text{PHAT}}(\hat{\boldsymbol{q}}_k^*)$ to evaluate the pros and cons of $\hat{\boldsymbol{q}}_k^*$. For example, two estimation values $\hat{\boldsymbol{q}}_k^* \in C(k)$ and $\hat{\boldsymbol{q}}_l^* \in C(l)$, if $|C(k)| > |C(l)|$, $\hat{\boldsymbol{q}}_k^*$ is considered superior to $\hat{\boldsymbol{q}}_l^*$, and vice versa. If $|C(k)| = |C(l)|$, then compare the value of $\hat{P}_{\text{avg}}^{\text{PHAT}}(\hat{\boldsymbol{q}}_k^*)$ and $\hat{P}_{\text{avg}}^{\text{PHAT}}(\hat{\boldsymbol{q}}_l^*)$, where the larger the value corresponding to the estimated better. The proposed modified SRP-PHAT multiple sound source localization algorithm using pseudo-code description of the process, as shown in Table 2.

Table 1. Improved SRP-PHAT Sound Source Localization Algorithm Processes

1. one input signal was divided into N_{SF} subframes, to calculate SRP functions $\hat{P}_n^{\text{PHAT}}(\boldsymbol{q})$
2. According to the formula (7) to give N_{SF} subframes estimate value $\hat{\boldsymbol{q}}_n$ $\hat{\boldsymbol{q}}_n$
3. Select L = 'simple', use AC clustering estimated marge N_{SF} sub frames together into N_c class
4. retain the classes only meet the conditions $\lvert C(k)\rvert \ge \gamma_{\text{th}}$, which is expressed as $\{C(k),\ k = 1:\tilde{N}_c\}$
5. According to the formula (8), calculated $\hat{P}_{\text{avg}}^{\text{PHAT}}(\boldsymbol{q})$ .In accordance with formula (9) to give $\{\hat{\boldsymbol{q}}_k^*,\ k = 1:\tilde{N}_c\}$
6. Make $\hat{N}_a = \min(\tilde{N}_c, N_s)$, depending on the value of $\lvert C(k)\rvert$ and $\hat{P}_{\text{avg}}^{\text{PHAT}}(\hat{\boldsymbol{q}}_k^*)$ to evaluate $\hat{\boldsymbol{q}}_k^*$, whichever is optimal points $\hat{N}_a$, denoted as $\hat{\boldsymbol{q}}_{sk},\ k = 1,\ldots,\hat{N}_a$
7. Output ultimate source position estimation gunshot $\hat{\boldsymbol{q}}_{sk},\ k = 1,\ldots,\hat{N}_a$

2. Experiment and Analysis

In order to contrast the performance of improved algorithm with the original algorithm, this research recorded a 7.62 mm assault rifle firing at the time of the blast wave signal as the sound source. The maximum position measurement error is taken 2cm in the simulation, and the maximum delay estimation error is 5 microseconds, for pure sound source localization experiment. The microphone array for collecting gunshots signals is a radius of 1.5 m of uniform circular array, which is placed in a level. The number of microphones is 8, shown in Figure 1. DOA vector pointing from the origin point to the sound source, and the vector is defined as ζ by formula (3). We take collection database Numbers "recorder-2014-04-0001" as a group of data, which has three shots. The microphone array and gunshots location are shown in Figure 2.

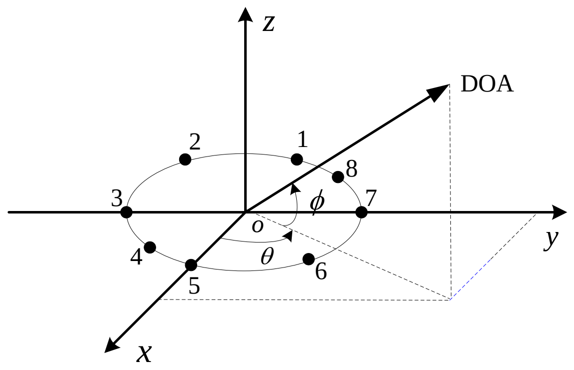

Figure 1. Microphone Array and Sound Source DOA Vector

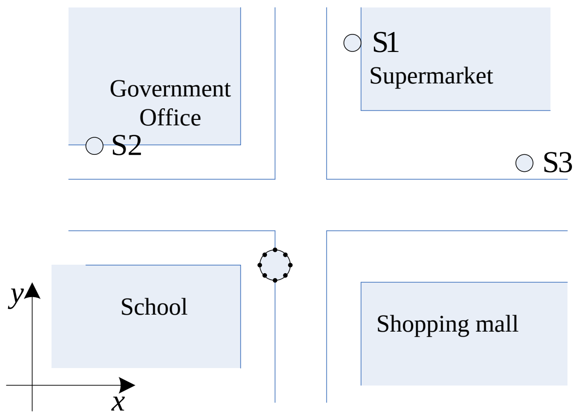

Figure 2. Microphone Array and Gunshots Source Location

Three gunshots orientation are: $\boldsymbol{q}_{s1}=(74^{\circ},21^{\circ})$, $\boldsymbol{q}_{s2}=(-7^{\circ},24^{\circ})$ and $\boldsymbol{q}_{s3}=(-50^{\circ},14^{\circ})$. In the process of shooting incident, shot first single occurrence, that is s1 and s2, s1 and s3, s2 and s3, then three people shot together. In each case, we take 12.5s data, a total of 50s data. Signal sampling frequency is 16kHz. The frame length is $T=1024$ point(64ms). The number of sub frames per frame is $N_{SF}=8$. Each sub frame is added hanning window, the frame overlap rate is zero. After removing the silent gap, a total of 625 frames data used in the experiment, including 439 of two gunshots, and 186 frames of three gunshots.

The main purpose of the algorithm is to calculate the minimum area containing the target. First, freely choose two nodes detected the target information to calculate the intersection boundary of two circular detection areas. Then find the intersecting circular detection area between the line and $[T_1,T_2]$ through an iterative approach. After these two steps, can find the node satisfying the condition, and may calculate the straight line d_2 passing through the two points $\{R_1,R_2\}$. Then using the same method to calculate a straight line d_2 through the circular area of $[R_1,R_2]$. Finally, the area obtained is the cross area of c_i, c_j and c_k, as shown in Figure 3. "+" represents the target position.

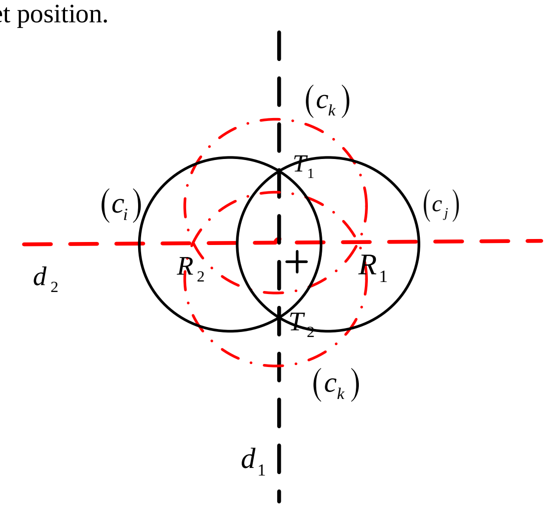

Figure 3. Several Aimple Sound Source Estimates

Whether traditional SRP-PHAT algorithm, or the proposed algorithm, must be in source space search to get the sound source position estimation. According to the positional relationship between the array and the sound source, we set the search range of the horizontal angle θ is $-180^{\circ}\sim180^{\circ}$, and the range of the elevation ϕ is $0^{\circ}\sim45^{\circ}$, steps are 1°. Uniform circular array was used in the experiment plane array, and has smaller aperture. The array for elevation resolution is very low, so only in accordance with horizontal positioning accuracy to evaluate the performance of the algorithm. Each frame data input can get $\hat{N}_a$ point estimate results. If the horizontal angle of the k_{th} point estimate with a mobile shooting angle true value is less than 5°, this point is the correct estimate, otherwise is the error estimate. The correct estimation of the cumulative number of points obtained in all the frames, and then divided by the total number of frames to get the estimated activity gunfire correct rate:

$$R_c=\frac{\sum_{i=1}^{I}\alpha_c(i)}{\sum_{i=1}^{I}N_a(i)} \tag{4}$$

Similarly the estimated error rate can be obtained

$$R_e=\frac{\sum_{i=1}^{I}\alpha_e(i)}{\sum_{i=1}^{I}N_a(i)} \tag{5}$$

In the above two formulas, I is used for signal frame of the experiment, $\alpha_c(i)$ and $\alpha_e(i)$ represent the number of correct estimate and the number of error estimate obtained by the i_{th} frame data. $N_a(i)$ is the true value of the number of gunshots moving for i_{th} frame data. It is noteworthy that the number of moving gunshots of each frame is unknown, so that the sum of correct rate and error rate is not to 1, which reflects the performance of positioned algorithm.

In order to facilitate researchers conducted sound source localization experiments, AV16.3 database also provides a single pure gunshot, these recordings synchronized with the array recordings, so we can determine in the i_{th} frame time. Each sound source is not active, the i_{th} frame data $N_a(i)$ can be obtained.

We first consider the case of two gunshots. Because each frame of data is divided into eight sub frames, the probability that each sound source frame occurs in two or more sub frames. Therefore we set $\gamma_{\text{th}} = 2$ in Table 2. Consider the extent and source close to the estimate of discrete sub frame is large, set $d_{\text{th}} = 16^{\circ}$ in Table 1. For convenience of description, the literature [8] algorithm referred to as SRP-PHAT, the proposed improved algorithm referred to as SRP-PHAT-SF. Two algorithms were used to make two gunshots sound source localization experiments, the results is shown in table 3.

Table 2. Two Gunshots Cases Positioning Performance Comparison of Two Kinds of Algorithm

Algorithm and evaluation index	SRP-PHAT		SRP-PHAT-SF	
	R_c (%)	R_e (%)	R_c (%)	R_e (%)
s1 and s2	70.4	12	76	8
s1 and s3	59.93	16.85	62.92	15.36
s2 and s3	68.05	10.9	68.8	6.39
average	66.03	13.28	69.09	9.96

From Table 3 can be seen, the proposed algorithm has significantly improved positioning performance than the literature [8]. Where the average correct rate increased by about 3%, the average error rate reduced by 3.32%.

We then consider the case of three gunshots. Due to the number of sound sources are more, in order to raise the correct rate, set $\gamma_{\text{th}} = 1$, the same as the rest of the parameters with two gunshots. The experimental results are shown in Table 4. Seen from Table 4, the case of three gunshots, the proposed algorithm can greatly improve the correct rate, and error rate increased slightly.

Table 3. Three Gunshots Cases Positioning Performance Comparison of Two Kinds of Algorithm

Algorithm	R_c (%)	R_e (%)
SRP-PHAT	50.3	18.05
SRP-PHAT-SF	64.71	18.46

Randomly placed three tetrahedral arrays, according to the geometric relationship to calculate blast wave arrival times of the three microphone array. Then the coordinates of the three microphone array, the arrival time as a known quantity, respectively SRP-PHAT-SF location algorithm with this algorithm simulation 50 times, the distribution of gunfire positioning error as shown in Figure 4. According to calculations, SRP-PHAT-SF location algorithm gunfire positioning error is 1.704m, the average location error gunfire algorithm is 1.576m, which can achieve more accurate than the SRP-PHAT-SF position location algorithm.

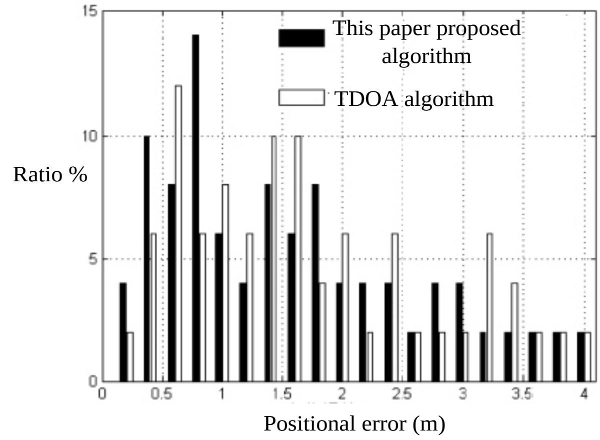

Figure 4. Gunshot Positioning Error Distribution

SRP-PHAT-SF location algorithm cannot eliminate the practical application of multipath interference caused by the delay estimation errors, in order to show the superiority of the proposed method, add 1-4 8ms in TDE 8 array delay estimation errors positioning the two methods is shown in Figure 5.

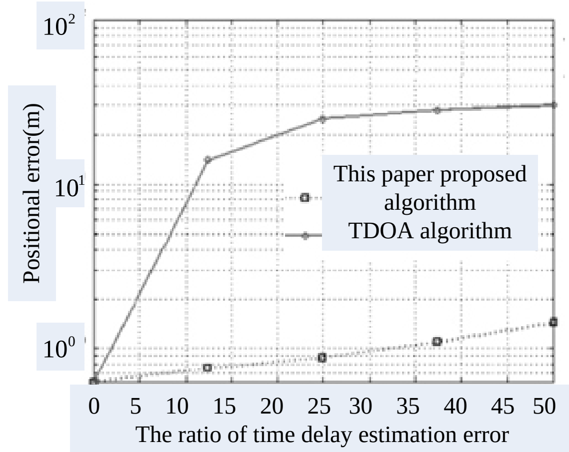

Figure 5. Delay Estimation Errors Affect the Proportion of Positioning Error

As can be seen in Figure 5, when the proportion of time delay estimation error is 0%, the two methods positioning error is almost the same. When the proportion of time delay estimation error increases, SRP-PHAT-SF positioning error increases linearly. When the number of the microphone array increases, the position error increases and decreases the number of the array shown in Figure 6. But with the increase in the number of arrays, the hardware requirements are also increased by the simulation results, so it should be between the array and the number of positioning accuracy trade-off.

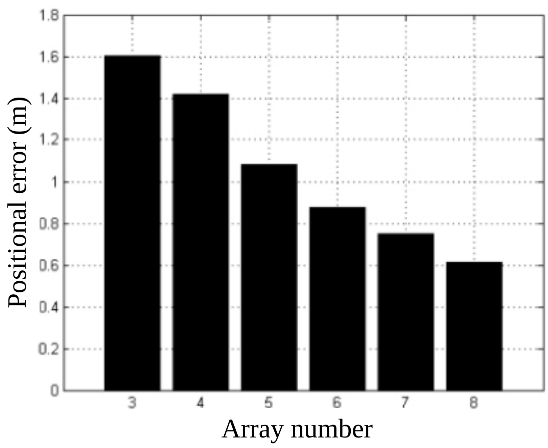

Figure 6. The Influence of Array Number of Positioning Error

3. Conclusion

In the real environment, noise around is an important factor affecting the performance of sound source localization algorithm. SRP-PHAT algorithm of reverberation has stronger robustness and has been widely applied. But the spatial resolution of SRP-PHAT is low. In the case of multiple gunshots, the space for each sound source peaks often overlap, resulting in this algorithm has poor performance of the multiple sound source localization. This paper proposed fusion sub frame estimate and cluster analysis of multiple sound source localization algorithm. One frame of shots signal is divided into eight sub frames. The sub frame is calculated SRP function to search the strongest sound source position sub frame in each sub frame. Since the speech signal has time domain sparsity, most sound source localization estimates obtained for each sub frame can not correspond to the same source, include multiple sound source location. Using AC cluster algorithm to these estimates is divided into several classes, and using the average to evaluate the estimate through function SRP, thereby obtain the final sound source position estimation. Use real environment 2 ~ 3 shots microphone array data recording experiment. The results show that the improved algorithm can effectively improve the positioning performance.

References

[1] Andrew R,Peters G P, Lennox J. Approximation and regional aggregation in multi-regional input-output analysis for national carbon footprint accounting . Economic Systems Research, 2009, 21 (3): 311-335.

[2] Boyd J P,Fitzgerald W J,Mahutga M C,Smith D A. Computing continuous core/periphery structures for social relations data with MINRES/SVD, Social Networks, 2010, 32 (2): 125-137.

[3] Dietzenbacher E. More on Multipliers. Journal of Regional Science, 2005, 45 (2): 421-426,

[4] Holub H W,Schnabl H. Qualitative input-output analysis and structural information. Economic Modelling, 2012, 2(1): 67-73.

[5] Holub H W? Tappeiner G. A general qualitative technique for the comparison of economic structures. Quality & Quantity, 2010, 22 (3): 293-310.

[6] Kranich J. Agglomeration, vertical specialization, and the strength of industrial linkages. Papers in Regional Science, 2011, 90 (1): 159-178.

[7] Lee S,Yooka S H,Kimb Y. Centrality measure of complex networks using biased random walks. The European Physical Journal, 2010, 68 (2): 277-281.

[8] Midmore P,Munday M,Roberts A. Assessing industry linkages using regional input-output tables. Regional Studies, 2006, 40 (3): 329-343