

## Nâng cao chất lượng phát hiện sự kiện âm thanh trong bài toán định vị nguồn âm theo nguyên lý TDOA

Trần Công Thìn<sup>1</sup>, Nguyễn Trung Kiên<sup>1\*</sup>, Bùi Ngọc Mỹ<sup>1</sup>, Nguyễn Huy Hoàng<sup>2</sup>

<sup>1</sup>Viện Khoa học và Công nghệ quân sự;

<sup>2</sup>Học viện Kỹ thuật quân sự.

\*Email: kiennt67@gmail.com

Nhận bài: 28/3/2022; Hoàn thiện: 19/4/2022; Chấp nhận đăng: 01/6/2022; Xuất bản: 28/6/2022.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.80.2022.60-70>

### TÓM TẮT

Bài báo trình bày một số kết quả nghiên cứu giải pháp nâng cao chất lượng định vị nguồn âm sử dụng nguyên lý TDOA, trong trường hợp loại nguồn âm cần định vị được xác định trước. Trong đó đề xuất giải pháp cải thiện chất lượng phát hiện sự kiện âm thanh dựa trên bộ lọc tương quan kết hợp với tiền xử lý tín hiệu bằng kỹ thuật phân tích thành phần độc lập ICA. Trên cơ sở phân tích lý thuyết và tiến hành mô phỏng Monte-Carlo trên Matlab, kết hợp với dữ liệu được thu thập trong điều kiện thực tế bài báo sẽ đánh giá hiệu quả của giải pháp được đề xuất, qua đó cho phép nâng cao chất lượng định vị nguồn âm theo nguyên lý TDOA.

**Từ khoá:** Định vị nguồn âm; TDOA; ICA.

### 1. MỞ ĐẦU

Trong các hệ thống định vị nguồn âm thanh nói chung và định vị nguồn âm sử dụng nguyên lý TDOA (Time Differences of Arrival) nói riêng, phát hiện sự kiện âm thanh (PHSKAT) là bước đầu tiên, ảnh hưởng lớn tới khả năng định vị chính xác nguồn âm, tránh trường hợp định vị sai nguồn âm, hoặc bỏ sót nguồn âm cần định vị. Đặc biệt, với các hệ thống định vị nguồn âm ngoài trời, khoảng cách định vị xa, nhiễu và tạp âm tác động lớn tới tín hiệu, qua đó làm giảm khả năng phát hiện chính xác sự kiện âm thanh [3, 8]. Hiện nay, có nhiều kỹ thuật khác nhau được sử dụng nhằm PHSKAT, chủ yếu dựa trên đặc trưng biên độ, tần số tín hiệu, hoặc áp dụng kỹ thuật học sâu (Deep Learning), tuy nhiên, trong bài toán định vị nguồn âm, việc PHSKAT cần được thực hiện trong thời gian ngắn, sẵn sàng cho các bước xử lý tiếp theo nhằm đảm bảo tính thời gian thực của hệ thống. Các công trình nghiên cứu kỹ thuật PHSKAT dùng trong hệ thống định vị nguồn âm đã được công bố phần lớn dựa trên các đặc trưng về phổ và năng lượng của tín hiệu, thích hợp với việc phát hiện các nguồn âm có cường độ mạnh, phổ năng lượng tập trung vào một khoảng tần số hẹp [5, 6]. Với các nguồn âm có khoảng cách định vị xa, tín hiệu thu được trên các cảm biến có biên độ thấp, phổ năng lượng tín hiệu trải dài trên một khoảng rộng, chưa có nhiều nghiên cứu đưa ra phương pháp PHSKAT phù hợp [3, 6]. Bài báo sẽ trình bày một số kết quả nghiên cứu đánh giá khả năng PHSKAT của các kỹ thuật dựa trên biên độ tín hiệu, bộ lọc tương quan. Trong đó, đề xuất giải pháp cải thiện hiệu quả PHSKAT bằng cách sử dụng bộ lọc tương quan kết hợp với giải pháp tiền xử lý tín hiệu sử dụng kỹ thuật phân tích thành phần độc lập ICA (Independent Component Analysis). Các kỹ thuật này được sử dụng tương đối phổ biến trong xử lý tín hiệu âm thanh nói chung, tuy nhiên khi được kết hợp áp dụng phù hợp vào việc PHSKAT sẽ nâng cao khả năng phát hiện, qua đó góp phần nâng cao chất lượng định vị nguồn âm sử dụng nguyên lý TDOA.

### 2. PHÁT HIỆN SỰ KIỆN ÂM THANH TRONG BÀI TOÁN ĐỊNH VỊ NGUỒN ÂM THEO NGUYÊN LÝ TDOA

#### 2.1. Định vị nguồn âm theo nguyên lý TDOA

Phương pháp định vị nguồn âm theo nguyên lý TDOA được xây dựng trên cơ sở ước tính hiệu thời gian đến  $\tau_{ij}$  của tín hiệu âm thanh thu được trên các cảm biến [1].

$$\tau_{ij} = \tau_i - \tau_j \tag{1}$$

Trong đó:  $\tau_i$  và  $\tau_j$  lần lượt là thời gian tín hiệu truyền từ nguồn âm tới cảm biến thứ  $i$  và  $j$ . Triển khai công thức hiệu thời gian đến theo khoảng cách  $r_i, r_j$  từ nguồn âm tới các cảm biến và vận tốc lan truyền của âm thanh trong không khí  $v$ , thu được phương trình biểu diễn một hyperboloid trong không gian 3 chiều [1] :

$$\tau_{ij} = \frac{r_i}{v} - \frac{r_j}{v} = \frac{\| \mathbf{x}_s - \mathbf{m}_i \| - \| \mathbf{x}_s - \mathbf{m}_j \|}{v} \tag{2}$$

Với  $\mathbf{x}_s = [x_s, y_s, z_s]^T$ ,  $\mathbf{m}_i = [x_i, y_i, z_i]^T$ ,  $\mathbf{m}_j = [x_j, y_j, z_j]^T$  lần lượt là tọa độ của nguồn âm và tọa độ của hai cảm biến thứ  $i$ , và thứ  $j$  trong không gian.

Để xác định tọa độ nguồn âm  $\mathbf{x}_s$  trong không gian 3 chiều, cần giao hội của ít nhất 3 hyperboloid, tương ứng cần ít nhất 4 cảm biến âm thanh (từ 1 - 4) để tạo thành 3 cặp cảm biến độc lập. Như vậy, để định vị nguồn âm theo nguyên lý TDOA, cần thực hiện hai bước, ước tính hiệu thời gian đến  $\tau_{ij}$  và giải hệ phương trình để xác định tọa độ của nguồn âm.

Tuy nhiên, trước khi thực hiện hai bước trên, cần phải PHSKAT cần định vị, từ đó trích xuất chính xác cửa sổ chứa sự kiện âm thanh, là cơ sở cho việc ước tính  $\tau_{ij}$ . Đặc biệt với các hệ thống định vị nguồn âm trong đó âm thanh định vị được biết trước như định vị điểm nổ, phát hiện bắn tỉa,... thì việc PHSKAT càng đóng vai trò quan trọng, đảm bảo loại bỏ trường hợp định vị sai nguồn âm hoặc bỏ sót âm thanh cần định vị [3].

Trong bài toán định vị nguồn âm theo nguyên lý TDOA, các kỹ thuật PHSKAT yêu cầu phải được thực hiện trong thời gian ngắn nhằm trích xuất chính xác cửa sổ chứa sự kiện âm, đảm bảo tính thời gian thực của hệ thống. Do đó, các kỹ thuật sử dụng thường dựa trên đặc trưng biên độ hoặc bộ lọc tương quan [3]. Một số nghiên cứu PHSKAT dựa trên mạng Neural tích chập (Convolutional Neural Networks-CNN) [5, 6] hoặc mạng mạng Neural hồi tiếp (Recurrent Neural Networks-RNN) [6, 7], tuy nhiên, hiệu quả của các kỹ thuật này trong bài toán định vị nguồn âm chưa được đánh giá cụ thể, đặc biệt ở tính thời gian thực của hệ thống. Bên cạnh đó, để phát hiện sự kiện theo các kỹ thuật trên đòi hỏi phải xây dựng được các bộ thư viện hoàn thiện cho nhiều loại âm thanh cần định vị.

**2.2. Phát hiện sự kiện âm thanh theo biên độ**

Phương thức đơn giản để phát hiện một sự kiện âm thanh là dựa vào biên độ của tín hiệu, khi đó bộ đề Neyman-Pearson về xác suất phát hiện tín hiệu [9] được sử dụng để tính ngưỡng phát hiện tín hiệu theo công thức :

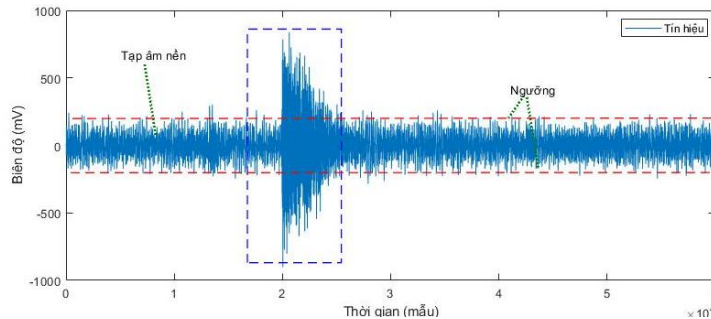
$$L(z) = \frac{p(z | H_{p_1})}{p(z | H_{p_0})} > \gamma \tag{3}$$

Trong đó:

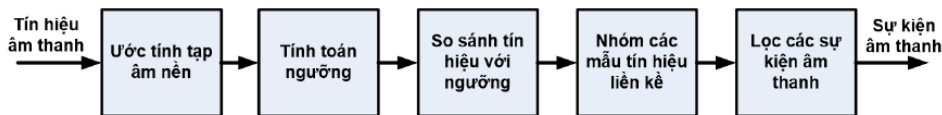
- $L(z)$  là tỉ lệ khả năng xảy ra sự kiện âm thanh;
- $H_{p_0}$  là giả thuyết không có sự kiện âm thanh;
- $H_{p_1}$  là giả thuyết có sự kiện âm thanh;
- $z$  là tập quan sát;
- $p(z)$  là hàm mật độ xác suất;
- $\gamma$  là ngưỡng phát hiện.

Theo công thức 3, để một tín hiệu được xác định là tồn tại sự kiện âm thanh thì tỉ lệ xảy ra sự kiện  $L(z)$  phải lớn hơn ngưỡng phát hiện  $\gamma$ , trong đó,  $H_{p_0}$  và  $H_{p_1}$  lần lượt là giả thuyết về tập âm và sự kiện âm thanh. Hình 1 minh họa một sự kiện âm thanh là tiếng nổ đầu nòng của súng

AK47, tín hiệu này là một sự kiện âm thanh điển hình với các thông số đặc trưng chính bao gồm: biên độ, thời gian tồn tại, tần số. Tín hiệu này có thể được phân tách khỏi tạp âm nền bởi các ngưỡng phát hiện.



Hình 1. Phát hiện sự kiện âm thanh theo biên độ.



Hình 2. Các bước phát hiện sự kiện âm thanh theo biên độ.

Thuật toán phát hiện sự kiện dựa trên công thức 3 được thể hiện trong hình 2. Toàn bộ quá trình bao gồm năm giai đoạn, cụ thể như sau: Bước đầu tiên ước tính nền nhiễu của tín hiệu âm thanh đầu vào để cung cấp cho bước thứ hai là tính toán ngưỡng âm và dương. Các ngưỡng phát hiện này là không cố định mà thay đổi theo cường độ tạp âm. Bước thứ ba so sánh các mẫu tín hiệu với các ngưỡng. Các mẫu trên ngưỡng dương hoặc dưới ngưỡng âm được coi là thành phần của một tín hiệu chứa sự kiện âm thanh. Bằng cách nhóm liên tục các tín hiệu này theo một cửa sổ trượt có thể thu được tín hiệu của sự kiện âm thanh hoàn chỉnh. Tuy nhiên, có khả năng các sự kiện âm thanh thu được không phải là sự kiện âm thanh cần phát hiện, do đó bước cuối cùng lọc sự kiện âm thanh là cần thiết để loại bỏ các sự kiện sai. Một sự kiện tiếng nổ đầu nòng thực sự phải tuân theo điều kiện sau:

$$(A \geq A_m) \cap (N \geq N_m) \quad (4)$$

Trong đó,  $A$  và  $N$  lần lượt là biên độ và thời gian trung bình của sự kiện,  $A_m$  và  $N_m$  là giá trị nhỏ nhất của chúng.

### 2.3. Phát hiện sự kiện âm thanh sử dụng bộ lọc tương quan

Hệ số tương quan, ký hiệu là  $r_{XY}$  được sử dụng để đo lường mức độ phụ thuộc tuyến tính của hai biến ngẫu nhiên  $X$  và  $Y$  tính theo công thức [2, 9]:

$$r_{XY} = \frac{\text{cov}(X, Y)}{\delta_X \cdot \delta_Y} = \frac{E((X - EX) \cdot (Y - EY))}{\delta_X \cdot \delta_Y} \quad (5)$$

trong đó:  $\text{cov}(X, Y) = E((X - EX) \cdot (Y - EY))$  là hiệp phương sai của  $X$  và  $Y$ ;

$\delta_X$  và  $\delta_Y$  lần lượt là độ lệch chuẩn của hai biến  $X, Y$ .

Hệ số tương quan  $r_{XY}$  thể hiện mức độ phụ thuộc tuyến tính của hai biến ngẫu nhiên  $X$  và  $Y$ , khi  $|r_{XY}|$  càng gần 1 thì tính chất phụ thuộc tuyến tính giữa hai biến  $X$  và  $Y$  càng chặt, khi  $|r_{XY}|$  càng gần 0 thì sự phụ thuộc tuyến tính này càng lỏng lẻo, nếu  $|r_{XY}| = 0$  thì  $X$  và  $Y$  không tương quan.

Tiến hành so sánh tín hiệu chuẩn  $X_{S_n}$  và tín hiệu thu được trên các cảm biến  $X_{i_n}$  sử dụng

hàm tương quan có dạng [2, 6, 9]:

$$r(m) = \frac{\sum_{n=1}^N (X_s(n) - \bar{X}_s)(X_i(n) - \bar{X}_i)}{\sqrt{\sum_{n=1}^N (X_s(n) - \bar{X}_s)^2} \sqrt{\sum_{n=1}^N (X_i(n) - \bar{X}_i)^2}} \quad (6)$$

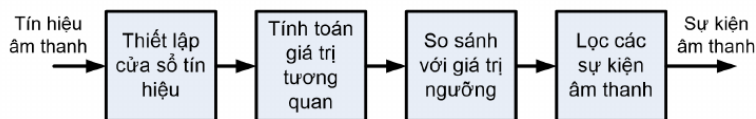
trong đó:  $\bar{X}_s$  và  $\bar{X}_i$  tương ứng là giá trị trung bình các phần tử của vector tín hiệu mẫu  $X_s$  và tín hiệu  $X_i$  thu được ở cảm biến thứ  $i$ .

Khi đó, giá trị tương quan cực trị được xác định theo biểu thức :

$$r_{MAX} = \arg \max_m \{r(m)\} \quad (7)$$

Giá trị  $r_{MAX}$  thể hiện mức độ tương quan giữa tín hiệu mẫu và tín hiệu thu được, do đó, có thể sử dụng  $r_{MAX}$  để đánh giá tỉ lệ khả năng xảy ra sự kiện âm thanh thông qua việc so sánh với giá trị ngưỡng  $r_{ng}$ . Nếu  $r_{MAX} \geq r_{ng}$  có thể đánh giá rằng, tín hiệu thu được trên cảm biến có tồn tại thành phần tín hiệu truyền tới từ nguồn âm, qua đó, khẳng định vừa có sự kiện âm thanh xảy ra, ngược lại nếu  $r_{MAX} < r_{ng}$  thì sự kiện âm thanh không tồn tại. Giá trị ngưỡng  $r_{ng} \in [0,1]$ , việc chọn giá trị  $r_{ng}$  có vai trò quan trọng phụ thuộc vào môi trường và khoảng cách định vị cũng như dạng tín hiệu cần định vị.

Để áp dụng bộ lọc tương quan vào việc PHSKAT, một mô hình các bước thực hiện được xây dựng như trên hình 3 bao gồm 4 bước cụ thể như sau.

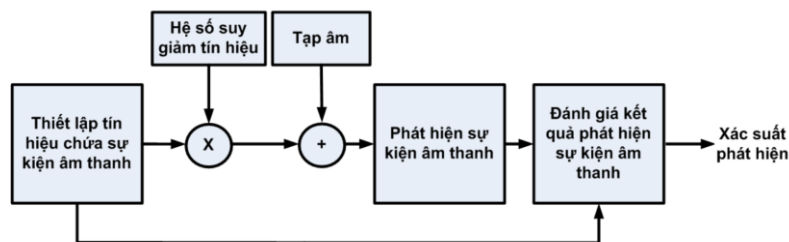


**Hình 3.** Các bước phát hiện sự kiện âm thanh dùng bộ lọc tương quan.

Bước 1, thiết lập cửa sổ tín hiệu thu được từ các cảm biến sao cho độ dài của cửa sổ bằng đúng độ dài của tín hiệu chuẩn  $X_{s,n}$ . Bước tiếp theo tính toán giá trị tương quan của tín hiệu thu được và tín hiệu mẫu theo công thức 6 và 7, sau đó tiến hành so sánh giá trị tương quan  $r_{MAX}$  với giá trị ngưỡng  $r_{ng}$ . Trong quá trình tính toán giá trị tương quan, khi có sự kiện âm thanh sẽ tồn tại một nhóm các giá trị lớn hơn mức ngưỡng, do đó, bước tiếp theo cần thiết phải lọc các sự kiện âm thanh được xác định liền kề để đồng nhất thành một sự kiện âm thanh chính xác duy nhất.

**2.4. Mô phỏng đánh giá hiệu quả giải pháp**

Để đánh giá khả năng PHSKAT của các giải pháp đã nêu, một chương trình mô phỏng Monte-Carlo được xây dựng trên Matlab có cấu trúc như trên hình 4.



**Hình 4.** Mô phỏng phát hiện sự kiện âm thanh.

Trong đó, một tín hiệu âm thanh chứa tiếng nổ đầu nòng của súng AK47 được nạp vào hệ thống, mức độ suy giảm tín hiệu và tạp âm được thiết lập. Trên cơ sở tín hiệu đó, các thuật toán PHSKAT theo biên độ và bộ lọc tương quan được sử dụng để PHSKAT.

Kết quả tương ứng với các mức SNR (Signal to Noise Ratio) khác nhau được đưa vào đánh giá xác suất PHSKAT được xác định theo biểu thức:

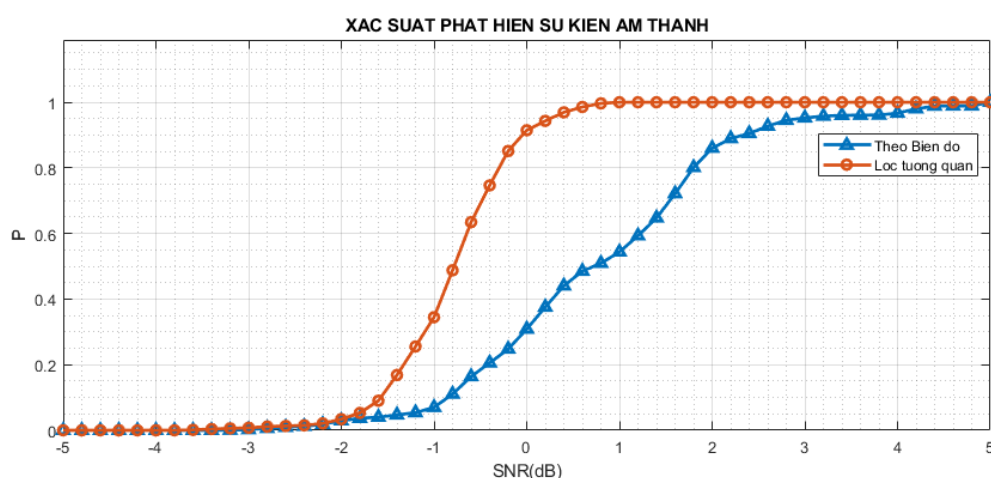
$$p = \frac{1}{N} A \quad (8)$$

Trong đó:

$p$  là xác suất phát hiện đúng sự kiện âm thanh;

$A$  là các sự kiện âm thanh được phát hiện đúng;

$N$  là số lượng các sự kiện âm thanh được đưa vào mô phỏng.



Hình 5. Xác suất phát hiện sự kiện âm thanh

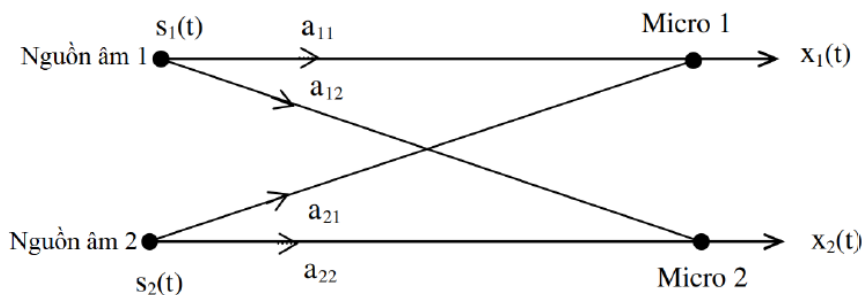
Kết quả mô phỏng được thể hiện trên hình 5, có thể nhận thấy, đối với phương pháp PHSKAT theo biên độ, khi tỉ số  $SNR > 3 \text{ dB}$ , xác suất phát hiện sự kiện âm thanh  $p \approx 1$ . Tuy nhiên, khi SNR giảm, xác suất PHSKAT  $p$  giảm, khi  $SNR < -1 \text{ dB}$  các sự kiện âm thanh không được phát hiện  $p \approx 0$ . Với phương pháp PHSKAT sử dụng bộ lọc tương quan, khi tỉ số  $SNR > 0.5$  xác suất PHSKAT  $p \approx 1$  xác suất này giảm khi  $SNR < 0$ , tuy nhiên, độ dốc lớn hơn khi so sánh với phương pháp sử dụng theo biên độ.

Từ kết quả mô phỏng có thể nhận thấy, phương pháp PHSKAT sử dụng bộ lọc tương quan có hiệu quả cao hơn so với phương pháp phát hiện theo biên độ. Tuy nhiên, khi tỉ số SNR thấp hiệu quả của bộ lọc tương quan suy giảm tương đối nhanh, điều này khiến giải pháp hoạt động không hiệu quả khi cự ly định vị nguồn âm xa, đòi hỏi phải có những bước tiên xử lý tín hiệu thích hợp nhằm nâng cao chất lượng PHSKAT.

### 3. NÂNG CAO CHẤT LƯỢNG PHÁT HIỆN SỰ KIỆN ÂM THANH

#### 3.1. Phân tích thành phần độc lập

Phân tích thành phần độc lập ICA là một phương pháp xử lý tín hiệu được xây dựng nhằm tách tín hiệu nhiều chiều trộn lẫn vào nhau thành các tín hiệu độc lập. Thuật toán ICA có nhiều ứng dụng rộng rãi trong nhiều bài toán khác nhau như xử lý tín hiệu, phân tích dữ liệu,... Với bài toán định vị nguồn âm sử dụng nguyên lý TDOA cần nhiều cảm biến âm thanh tại các vị trí khác nhau trong không gian, việc sử dụng thuật toán ICA để tách âm thanh cần định vị ra khỏi các nguồn âm khác nhằm xác định chính xác sự kiện âm thanh là một ý tưởng khả thi.



**Hình 6.** Mô hình hai nguồn âm trộn lẫn.

Trên hình 5 biểu diễn hai nguồn âm được thu lại bởi hai cảm biến khác nhau, khi đó, tín hiệu thu được trên các cảm biến có dạng [9, 10]:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{aligned} \quad (9)$$

Trong mô hình hệ thống ICA, các tín hiệu âm thanh như  $s_i(t)$ ,  $x_i(t)$  được xem là các biến ngẫu nhiên, do đó, các tín hiệu này thường được biểu diễn dưới dạng các vector. Tổng quát hóa mô hình hệ thống ICA ta được công thức [9]:

$$\begin{aligned} \mathbf{x} &= \mathbf{A}\mathbf{s} \\ \mathbf{x} &= \sum_{i=1}^n \mathbf{a}_i s_i \end{aligned} \quad (10)$$

Mục đích của kỹ thuật phân tích thành phần độc lập nhằm giải bài toán như công thức 10, về bản chất là tìm các nghiệm  $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$ , hay chính xác hơn là xác định ma trận  $\mathbf{W} = \mathbf{A}^{-1}$ , trong đó, thỏa mãn các điều kiện ràng buộc bao gồm: Các nguồn tín hiệu  $s_i$  được xem là độc lập thống kê với nhau; ma trận trộn  $\mathbf{A}$  là ma trận vuông khả nghịch; tối đa chỉ có một nguồn tín hiệu gốc có phân bố Gauss. Trên thực tế không thể tìm được nghiệm  $\mathbf{s}$  một cách trực tiếp mà phải sử dụng các phương pháp thống kê, trong đó, nguồn âm  $\mathbf{s}$  được ước lượng thông qua phép đo tính phi Gauss. Theo định lý giới hạn trung tâm, tổng của nhiều biến ngẫu nhiên có phân bố gần với phân bố Gauss hơn các biến ngẫu nhiên độc lập ban đầu [10]. Ở mô hình ICA vector ngẫu nhiên  $\mathbf{x}$  là tổng của các vector nguồn  $\mathbf{s}$  độc lập, khi đó, vector  $\mathbf{x}$  sẽ có phân bố gần với phân bố Gauss. Nếu việc trộn được đảo ngược lại theo cách nào đó thì các tín hiệu nhận được sẽ ít Gauss hơn. Do đó, ước lượng ICA nhằm đến cực tiểu hóa tính Gauss tức cực đại hóa tính phi Gauss bởi vì điều này sẽ cho đưa tới các thành phần độc lập. Để xác định tính phi Gauss, phương pháp phương pháp xấp xỉ Negentropy được sử dụng [10]. Đây là phương pháp đo lường khoảng cách thông tin tới tín hiệu có phân bố chuẩn có cùng kỳ vọng và phương sai, Negentropy được tính theo công thức [11]:

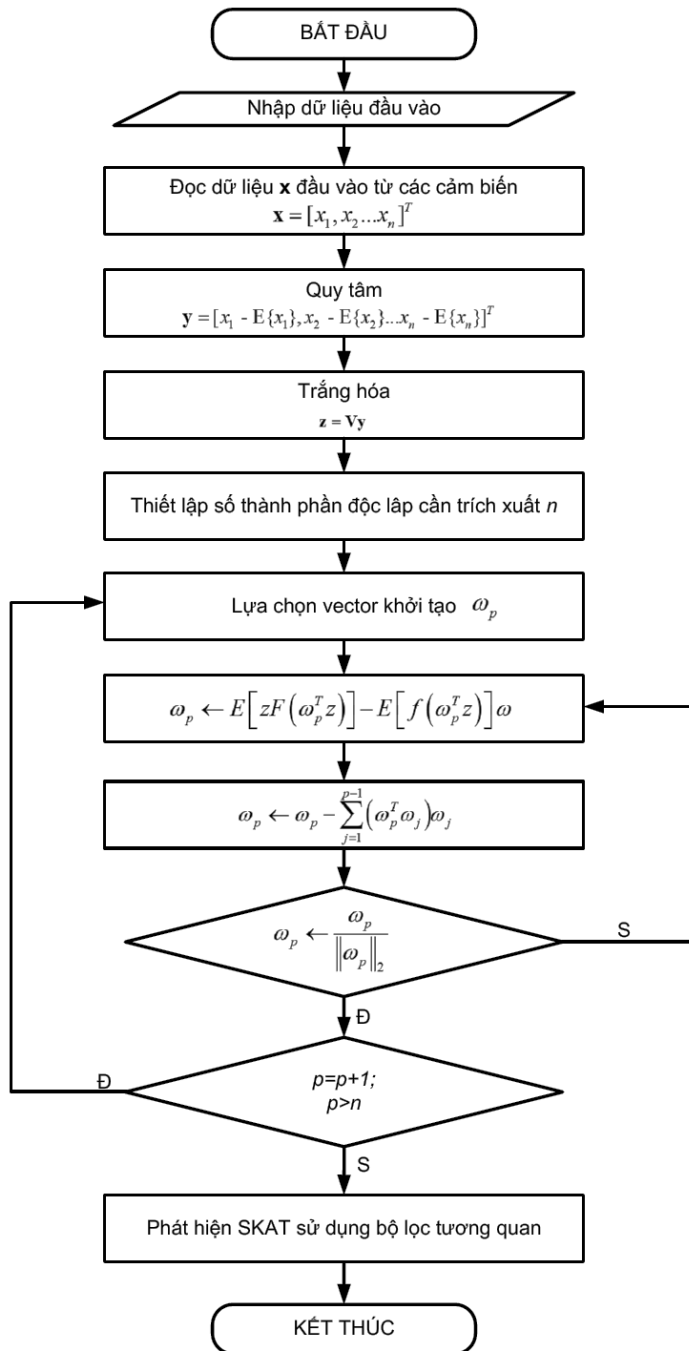
$$J(\mathbf{x}) = H(\mathbf{x}_g) - H(\mathbf{x}) \quad (11)$$

Trong đó,  $\mathbf{x}_g$  là một biến ngẫu nhiên Gauss của cùng một ma trận tương quan như  $\mathbf{x}$ . Tuy nhiên, việc ước lượng Negentropy rất khó thực hiện, thực tế Negentropy được xấp xỉ dựa trên hàm đối tượng  $G_i$  [11].

$$J(\mathbf{x}) \approx [\mathbb{E}(G(\mathbf{x}_g)) - \mathbb{E}(G(\mathbf{x}))]^2 \quad (12)$$

Hàm đối tượng  $G_i$  có nhiều sự lựa chọn khác nhau, tuy nhiên, hàm logcosh có dạng như công thức 13 được chứng minh là tương đối hiệu quả và dễ thực hiện, cũng là hàm được sử dụng trong giải pháp của bài báo.

$$G_1 = \frac{1}{a_1} \text{logcosh}(a_1 u) \text{ với } 1 \geq a_1 \leq 2 \quad (13)$$



**Hình 7.** Lưu đồ thuật toán PHSKAT sử dụng tiền xử lý ICA.

Thuật toán ICA được thực hiện tuần tự theo các bước như trên hình 7, trong đó, tín hiệu thu được trên các cảm biến trước hết cần được loại bỏ nhiễu trắng hay còn gọi là quá trình quy tâm. Quá trình quy tâm được thực hiện theo công thức:

$$\mathbf{y} = \mathbf{x} - E\{\mathbf{x}\} \quad (14)$$

Trong đó:  $\mathbf{y}$  là tín hiệu sau quy tâm;

$E\{x\}$  là trị trung bình của vector dữ liệu  $\mathbf{x}$ .

Tiếp theo là quá trình trắng hóa, đây thực chất là phép biến đổi ma trận trộn  $A$  trở nên trực giao dựa trên thực hiện phép nhân ma trận  $V$  với vector dữ liệu  $y$ .

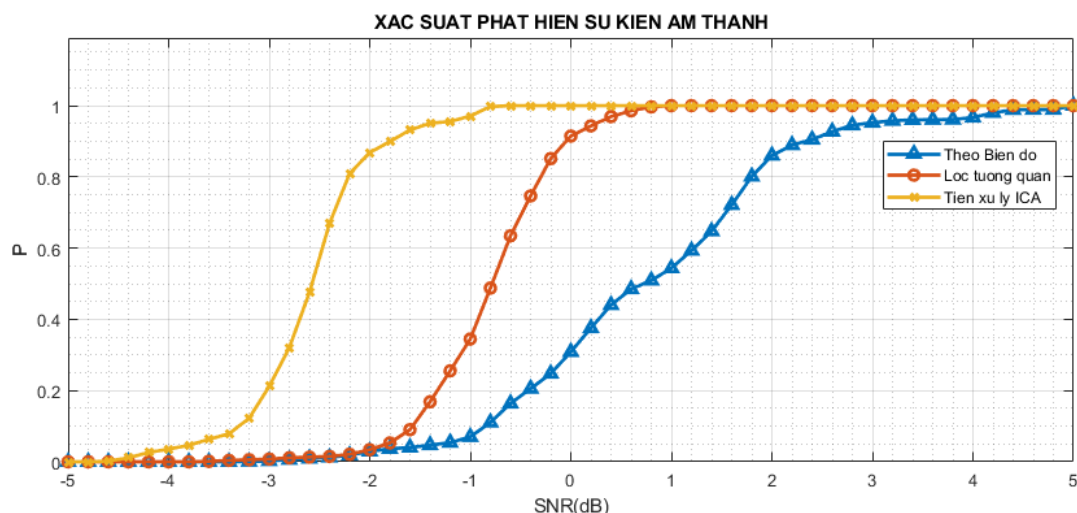
$$z = Vy \tag{15}$$

Với  $V$  là ma trận làm trắng được tính thông qua triển khai trị riêng EVD (Eigenvalue Decomposition) của ma trận hiệp phương sai  $E\{yy^T\} = EDE^T$ .

Tín hiệu sau khi được trắng hóa sẽ được đưa vào tính toán xác định thành phần độc lập, quá trình thực hiện thông qua các bước sau:

1. Lựa chọn ngẫu nhiên ma trận khởi tạo  $\omega_p$
2. Xác định hàm Negentropy:  $\omega_p \leftarrow E[zF(\omega_p^T z)] - E[f(\omega_p^T z)]\omega$
3. Kiểm tra tính hội tụ của hàm:  $\omega_p \leftarrow \omega_p - \sum_{j=1}^{p-1} (\omega_p^T \omega_j) \omega_j$

Quá trình trên sẽ được thực hiện tuần tự đến khi hàm  $\omega_p$  hội tụ, qua đó xác định được thành phần độc lập của tín hiệu s thông qua công thức 10.



Hình 8. Xác suất phát hiện sự kiện âm thanh sau tiền xử lý ICA.

Tiến hành mô phỏng khả năng PHSKAT theo mô hình như trên hình 4, trong đó, bộ tiền xử lý tín hiệu được đưa vào đánh giá, kết quả xác suất PHSKAT được thể hiện như trên hình 8.

Có thể nhận thấy, bộ phân tích âm thành phần độc mang lại hiệu quả đáng kể. Với  $SNR > -2$  dB xác suất PHSKAT của giải pháp  $p \approx 1$ , khả năng của giải pháp chỉ suy yếu khi  $SNR < -3$  dB. Điều đó cho thấy, trong điều kiện môi trường có tạp âm mạnh, giải pháp kết hợp giữa bộ phân tích thành phần độc lập và lọc tương quan mang lại hiệu quả PHSKAT hiệu quả, phù hợp áp dụng cho bài toán định vị nguồn âm sử dụng nguyên lý TDOA.

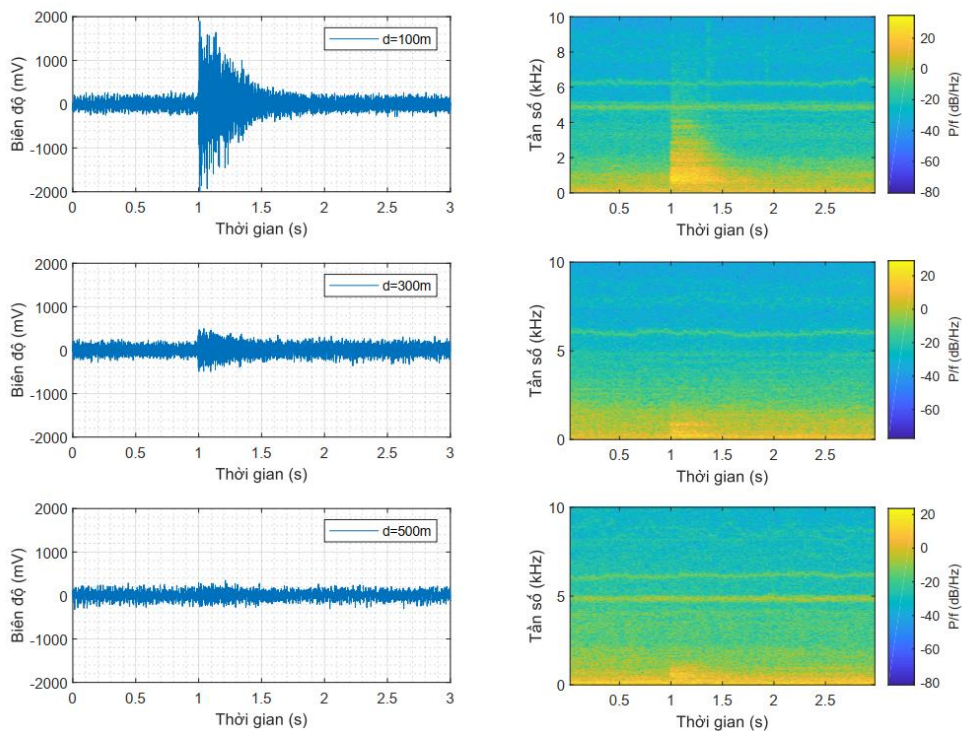
### 3.2. Một số kết quả thực nghiệm của giải pháp

Để đánh giá hiệu quả của của thuật toán giải pháp PHSKAT, một hệ 08 cảm biến âm thanh INMP401 của hãng InvenSense được sử dụng để thu nhận tín hiệu âm thanh, đây là loại cảm biến vi cơ tiên tiến có độ nhạy cao, kích thước nhỏ gọn. Những cảm biến này được bố trí bố trí trên hai hình hình tứ diện đều lồng nhau có độ dài cạnh 1 m như trên hình 9, việc bố trí các cảm biến nằm trên tứ diện đều giúp tín hiệu thu được đồng đều trên các hướng tới [3].



**Hình 9.** Cụm cảm biến thu tín hiệu âm thanh.

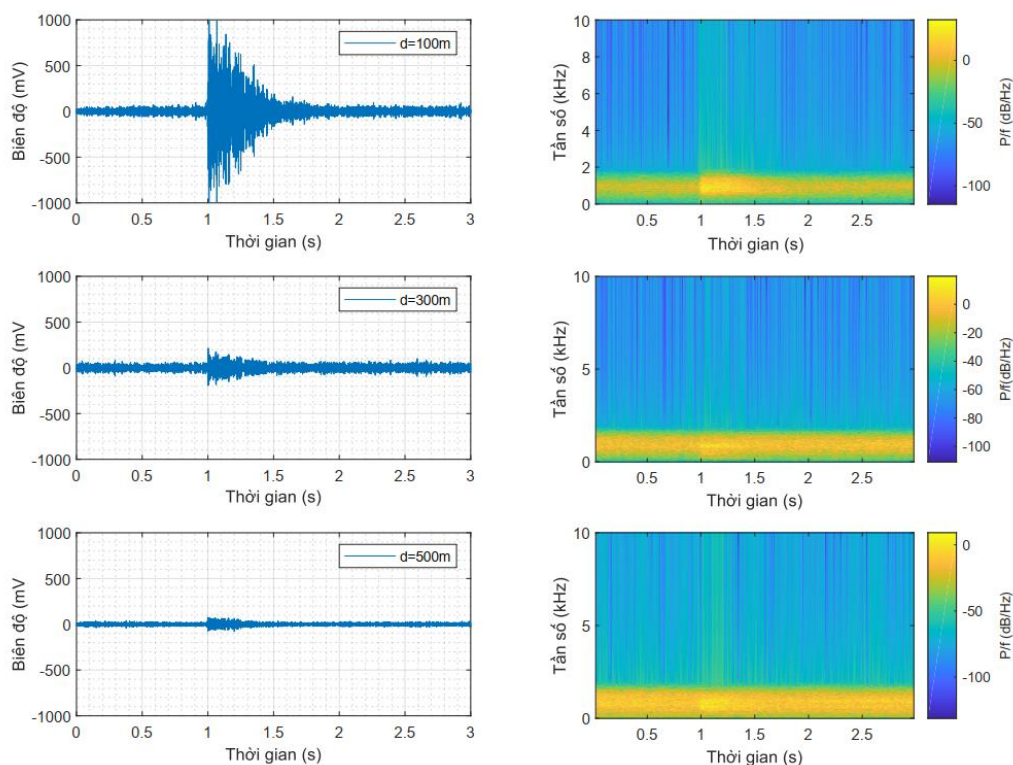
Tiến hành phát tín hiệu âm thanh tạo giả tiếng nổ đầu nòng của súng AK47 qua một loa nén công suất 30 W tại các cự ly khác nhau trong điều kiện môi trường thực tế. Tín hiệu sẽ được thu đồng thời trên 08 cảm biến, những tín hiệu này được số hóa và lưu trữ với tần số lấy mẫu  $f_s=20\text{ kHz}$  làm cơ sở cho các bước xử lý tín hiệu và định vị nguồn âm thanh.



**Hình 10.** Tín hiệu âm thanh thu được tại các cự ly khác nhau.

Hình 10 thể hiện dạng và giản đồ phổ của tín hiệu trên các cảm biến ở các cự ly khác nhau. Có thể nhận thấy ảnh hưởng của tạp âm lên tín hiệu tại các cự ly khác nhau là rất rõ ràng, tại cự ly nhỏ hơn 300 m hình ảnh tín hiệu là tương đối rõ ràng và phân biệt với nền, mức độ năng lượng tương ứng với tần số âm thanh có thể nhận thấy tương đối rõ. Tuy nhiên, khi cự ly lớn hơn 300 m việc phân biệt tín hiệu và tạp âm tương đối khó khăn, phổ năng lượng tín hiệu cũng bị hòa lẫn vào năng lượng của tạp âm.

Trên hình 11 là tín hiệu thu được sau bộ phân tích thành phần độc lập, có thể nhận thấy tín hiệu sau bộ lọc nổi bật trên nền nhiễu, tương ứng trên giản đồ histogram khoảng tần số chứa tín hiệu cũng được thể hiện rõ ràng hơn.



**Hình 11.** Tín hiệu âm thanh sau bộ phân tích ICA tại các cự ly khác nhau.

Kết quả sau khi tiến hành lọc tương quan xác định sự kiện âm thanh cho tín hiệu sau bộ lọc được thể hiện trên bảng 1, hệ số tương quan trung bình của tín hiệu sau bộ lọc ICA tại mọi cự ly đều vượt trội so với tín hiệu không qua bộ lọc.

**Bảng 1.** Hệ số tương quan trung bình tại các cự ly.

$r_{MAX}$ \ Cự ly	100m	200m	300m	400m	500m
Tín hiệu nguyên bản	0.727	0.643	0.332	0.218	0.154
Tín hiệu sau ICA	0.854	0.775	0.702	0.617	0.529

Việc cải thiện giá trị hệ số tương quan cho phép lựa chọn  $r_{ng}$  có giá trị lớn hơn, đảm bảo được độ tin cậy và khả năng phát hiện chính xác sự kiện âm thanh, đồng thời loại bỏ các trường hợp phát hiện nhầm hay bỏ sót sự kiện cần định vị. Kết quả này là tiền đề cho phép nâng cao chất lượng định vị nguồn âm.

#### 4. KẾT LUẬN

Trên cơ sở phân tích lý thuyết và tiến hành mô phỏng trên Matlab, cho thấy giải pháp PHSKAT sử dụng bộ lọc tương quan kết hợp bộ phân tích thành phần độc lập mang lại hiệu quả phát hiện cao, đặc biệt trong điều kiện cự ly định vị xa, tín hiệu âm thanh chịu ảnh hưởng của

nhieu và tạp âm. Giải pháp được chứng minh thông qua chương trình mô phỏng, cũng như trên dữ liệu tín hiệu âm thanh thu được trên các cảm biến trong môi trường thực tế. Hiệu quả của giải pháp cho phép nâng cao chất lượng định vị nguồn âm nói chung và định vị nguồn âm sử dụng nguyên lý TDOA nói riêng, giúp xác định chính xác sự kiện âm thanh cần định vị, tránh các trường hợp định vị nhầm hoặc bỏ sót sự kiện âm thanh.

### TÀI LIỆU THAM KHẢO

- [1]. Trần Công Thìn, Bùi Ngọc Mỹ, Nguyễn Huy Hoàng, Phạm Văn Hòa, "Xây dựng giải pháp định vị nguồn âm theo nguyên lý TDOA trong điều kiện vận tốc âm thanh biến đổi", Hội thảo Ứng dụng Công nghệ cao vào thực tiễn - 60 năm phát triển Viện KH-CN quân sự, (2020).
- [2]. Lê Bá Long, *Sách hướng dẫn học tập Xác suất thống kê*. Hà Nội: Học viện Công nghệ Bưu chính viễn thông, (2006).
- [3]. Maximo Cobos, "A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks", *Wireless Communications and Mobile Computing*, pp. 1–24, (2017).
- [4]. F.-G. Zeng, K. Nie, G. S. Stickney, Y.-Y. Kong, M. Vongphoe, A. Bhargave, C. Wei, & K. Cao. "Speech recognition with amplitude and frequency modulations". *Proceedings of the National Academy of Sciences*, 102(7), pp. 2293–2298, (2005).
- [5]. G. T. Wang, X. W. Liang, Y. Y. Xue, C. Li, & Q. Ding. "Algorithm Used to Detect Weak Signals Covered by Noise in PIND". *International Journal of Aerospace Engineering*, 2019, pp. 1–10, (2019).
- [6]. Adrián-Martínez, S. et al. "Acoustic Signal Detection Through the Cross-Correlation Method in Experiments with Different Signal to Noise Ratio and Reverberation Conditions". In: Garcia Pineda, M., Lloret, J., Papavassiliou, S., Ruehrup, S., Westphall, C. (eds) *Ad-hoc Networks and Wireless. ADHOC-NOW 2014. Lecture Notes in Computer Science()*, vol 8629. Springer, Berlin, Heidelberg, (2015).
- [7]. Zhang, H., McLoughlin, I., & Song, Y. "Robust sound event recognition using convolutional neural networks". In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 559-563, (2015).
- [8]. Giambattista Parascandolo and Heikki Huttunen and Tuomas Virtanen. "Recurrent Neural Networks for Polyphonic Sound Event Detection in Real Life Recordings". CoRR, abs/1604.00861, (2016).
- [9]. Kay, S. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc, (1993).
- [10]. Ali Mohammad-Djafari. *Non Gaussianity and Non Stationarity modeled through Hidden Variables and their use in ICA and Blind Source Separation*, (2007).
- [11]. Hyvärinen, A., Karhunen, J., & Oja, E. *Independent Component Analysis*. John Wiley & Sons, Inc, (2001).

### ABSTRACT

#### Improving sound event detecting in sound source localization using TDOA method

*This paper presents several research results that enhance TDOA-based sound localization accuracy with the priority of the source of interest. In which, a solution is proposed to improve the quality of audio event detection based on the correlation filter combined with signal preprocessing by the independent component analysis technique ICA. From analysis and discussions are made on that design and using Monte Carlo simulations with the data collected in a real environment, the results show the efficiency of our proposed method in TDOA-based localization.*

**Keywords:** Sound Source Localization; TDOA; ICA.