

XỬ LÝ MÙ TRONG TIỀN XỬ LÝ TIẾNG NÓI

Tiến Anh, Khánh Linh

Hiện nay, phương pháp tách nguồn mù (Blind Source Separation) đang được tập trung nghiên cứu và áp dụng trong nhiều lĩnh vực như tách tín hiệu tiếng nói, tín hiệu Y sinh. Phương pháp này cho phép ước lượng lại các tín hiệu nguồn ban đầu mà chỉ dựa vào những tín hiệu có được tại các sensor thu mà không biết đặc trưng của kênh truyền cũng như các tín hiệu nguồn.

1. Giới thiệu về tách nguồn mù

Gần đây, phương pháp xử lý tín hiệu dùng nhiều micro đang thu hút sự quan tâm của nhiều nhóm nghiên cứu. Tín hiệu có được từ nhiều micro cho phép khai thác được cả thông tin không những về thời gian và tần số của tín hiệu mà còn về tương quan trong không gian giữa chúng. Kết quả của những bài toán xử lý tiếng nói sẽ hiệu quả hơn khi khai thác được nhiều thông tin từ dữ liệu thu được bằng phương pháp này.

Về mặt xử lý tín hiệu, hiện nay thuật toán xử lý mù được nghiên cứu và ứng dụng rộng rãi vì nó có thể giải quyết hiệu quả những bài toán xử lý tín hiệu dùng nhiều đầu thu (sensor, micro). Lý do gọi là "xử lý mù" vì trong quá trình tính toán, khai thác dữ liệu chúng ta chỉ dựa vào những tín hiệu quan sát được ở các đầu thu mà không có thông tin đo được về những tín hiệu ở nguồn phát cũng như kênh truyền. Chính vì vậy, khi áp dụng vào bài toán tách tín hiệu tiếng nói, thuật toán cho phép tách được tín hiệu nguồn mong muốn dựa trên tín hiệu bị trộn lẫn ở đầu thu mà không cần quan tâm tới kênh truyền. Tính chất này mang lại nhiều ứng dụng trong thực tế, đặc biệt khi chúng ta cần xử lý các tín hiệu tiếng nói mà không biết chúng được thu như thế nào.

Nếu trong quá trình truyền từ nguồn phát, tín hiệu tiếng nói (người nói) tới các micro thu môi trường chỉ gây nên sự suy hao về biên độ và có ảnh hưởng của nhiễu cộng thì bài toán tách tín hiệu có thể được giải quyết một cách hiệu quả với điều kiện đã xác định được nhiễu. Trong trường hợp này, kênh truyền không đổi theo thời gian, hay còn gọi là bài toán trộn tức thời. Tín hiệu nguồn được ước lượng thông qua việc ước lượng được ma trận trộn đặc trưng cho kênh truyền. Một số thuật toán có thể cho kết quả tốt trong trường hợp này như phương pháp entropy ngược hay phương pháp dùng thống kê bậc bốn trong.

Mô hình kênh truyền của bài toán đặt ra trong luận văn này đề cập đến vấn đề kênh truyền phức tạp hơn. Cụ thể là, những đặc trưng của kênh truyền thay đổi theo thời gian. Tín hiệu tiếng nói nguồn không chỉ bị suy hao về biên độ, bị làm trễ mà còn bị ảnh hưởng của tiếng vang. Trong trường hợp này, bài toán xử lý mù được gọi là trộn chập (convolutive mixture). Giải quyết hiệu quả bài toán này là vấn đề không đơn giản.

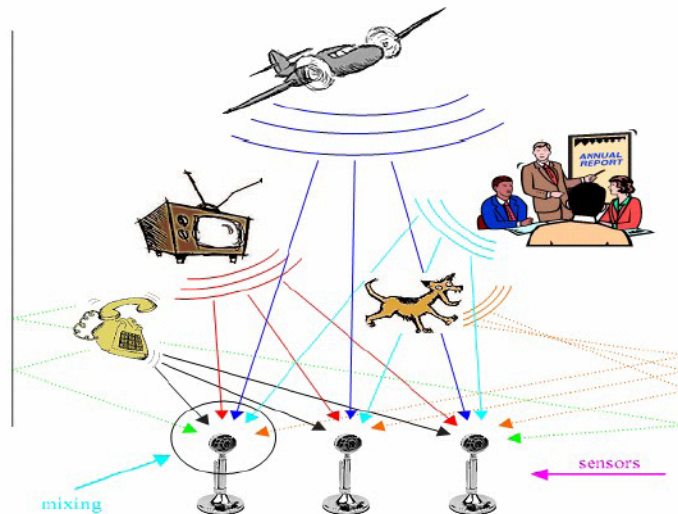
Tuy nhiên, nếu chúng ta biểu diễn mô hình tín hiệu phát và thu trên miền tần số thì bài toán trộn chập trong miền thời gian sẽ được đơn giản hóa thành bài toán trộn tức thời trong miền tần số. Và vì thế, chúng ta có thể khai thác phần nào một số thuật toán đã được phát triển cho trường hợp trộn tức thời.

Kết quả của bài toán tách tín hiệu tiếng nói là cơ sở cho rất nhiều những ứng dụng trong thực tiễn liên quan tới xử lý tiếng nói. Lấy ví dụ trong mô hình nhận dạng tiếng nói, tức là chuyển từ tiếng nói sang dạng text, tín hiệu đưa vào để nhận dạng được lấy từ dữ liệu ở các micro thu. Do đó để có dữ liệu tiếng nói cho quá trình nhận dạng, trước hết cần tách ra được tín hiệu mong muốn từ dữ liệu

bị trộn lẫn và chịu ảnh hưởng của môi trường từ các micro thu để không chỉ lấy được tín hiệu tiếng nói nào đó cần xử lý mà còn làm "sạch" nó khỏi nhiễu. Kết quả của quá trình tách có tốt thì kết quả của nhận dạng mới chính xác được.

2. Mô hình bài toán Cocktail Party

Bài toán dạ hộicocktail thực hiện tách tiếng nói của một người nào đó trong một cuộc dạ hộicocktail có lẫn tiếng ồn và tiếng nói của những người khác. Hình 1 mô tả một thí nghiệm trong đó đầu thu là dãy các micro. Tín hiệu thu chịu ảnh hưởng của nhiễu tín hiệu nhiễu và các hiện tượng méo tín hiệu, phản xạ. Con người có khả năng nghe tiếng nói của một người trong đám đông. Tuy nhiên, với máy tính, khả năng nghe phải dựa vào tính chất của tiếng nói của người, môi trường truyền,...



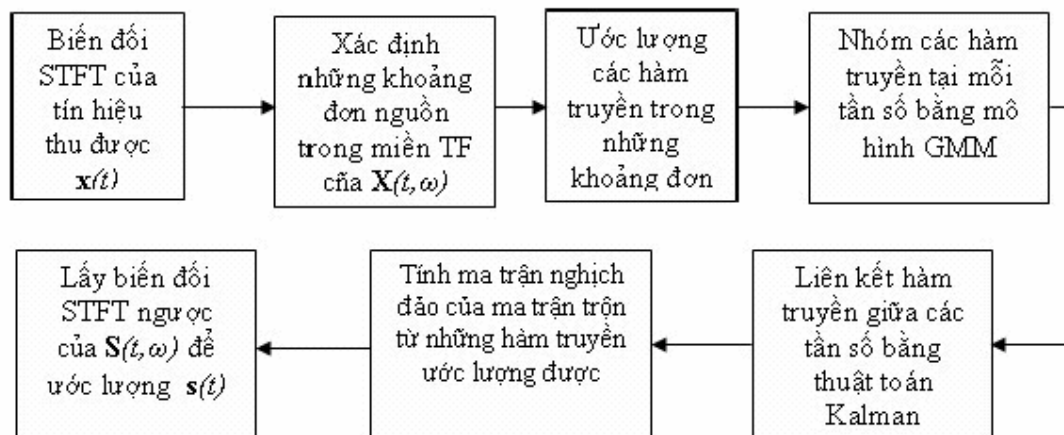
Hình 1. Mô hình bài toán Cocktail Party

Phần này tìm hiểu thuật toán để tách tín hiệu tiếng nói trong điều kiện môi trường có sự xuất hiện của nhiễu tiếng nói đồng thời và chịu ảnh hưởng của nhiễu trắng, trễ trên đường truyền và của tiếng vang do phản xạ. Tín hiệu tại các micro thu là sự trộn lẫn của nhiễu tín hiệu nhiễu và chịu ảnh hưởng của môi trường truyền. Thuật toán có thể tách được tín hiệu nguồn ban đầu mà chỉ dựa vào tín hiệu đo được ở các micro thu sau khi xử lý trong miền thời gian và tần số. Từng thành phần của ma trận trộn hàm truyền trong miền thời gian - tần số ở mỗi tần số được ước lượng trực tiếp từ vector riêng tương ứng với trị riêng lớn nhất tại những khoảng mà ở đó chỉ có một nguồn xuất hiện. Vì ảnh hưởng của kênh truyền nên hàm truyền từ tín hiệu nguồn tới mỗi micro thu tại mỗi tần số ở mỗi thời điểm là khác nhau. Để xác định được hàm truyền tương ứng với mỗi nguồn tại mỗi tần số và liên kết hàm truyền giữa các tần số với nhau sử dụng mô hình trộn Gauss dùng thuật toán EM và thuật toán Kalman.

Thuật toán tách nguồn tiếng nói trong môi trường có ảnh hưởng của trễ trên đường truyền, tiếng vang và nhiễu cộng sử dụng phép phân tích không thời gian tần số được tóm tắt bởi các bước như sau:

- 🌍 Dùng biến đổi STFT, chuyển đổi tín hiệu thu được từ các micro sang miền thời gian - tần số..
- 🌍 Tại mỗi tần số xác định thành phần hàm truyền tương ứng với một nguồn tồn tại trên một khoảng của trục thời gian bằng cách phân tích trị riêng của ma trận tương quan chéo xác định trên khoảng đó.

- Tại mỗi tần số, dùng mô hình trộn Gauss với thuật toán EM để thực hiện việc nhóm các tỉ số biên độ đo được, xác định được các giá trị trung bình và ma trận hiệp phương sai tương ứng với từng nhóm thành phần.
- Dùng bộ lọc Kalman để thực hiện việc liên kết các giá trị trung bình ở đầu ra của mô hình trộn Gauss tại các tần số khác nhau tương ứng với một nguồn lại với nhau. Và xác định được ma trận trộn như là một hàm của tần số.
- Tính ma trận nghịch đảo của ma trận ước lượng được trong miền tần số.
- Thực hiện phép biến đổi STFT ngược để khôi phục tín hiệu trong miền thời gian.



HÌNH 2. Trình tự thực hiện thuật toán

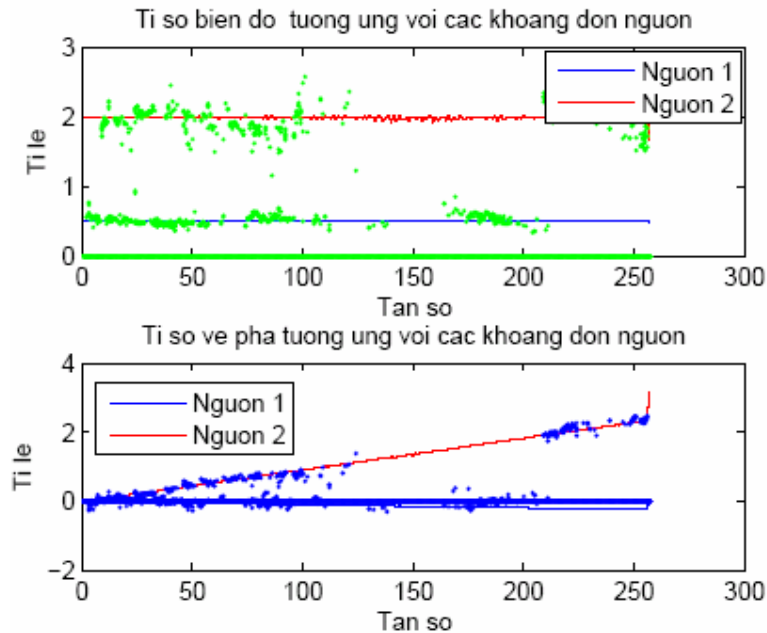
Bài toán trong quá trình thực hiện là với trường hợp 2 nguồn phát và 2 nguồn thu. Tín hiệu thu bởi chương trình Matlab, theo chuẩn một kênh, 16 bit, tần số 11025kHz. Ảnh hưởng của môi trường truyền trong quá trình mô phỏng được thực hiện bằng cách cộng hai tín hiệu nguồn có ảnh hưởng của trễ, lệch trên đường truyền và nhiễu. Trong quá trình mô phỏng, mô hình tín hiệu bị trộn lẫn được giả sử là tín hiệu tại micro 1 chỉ chịu ảnh hưởng của đường truyền trực tiếp. Ở đó có các thông số cần quan tâm như sau:

- $s_1(t)$ và $s_2(t)$ là những tín hiệu nguồn tiếng nói được ghi âm như những điều kiện đã nêu.
- a_{ij} ($i, j = 1, 2$) là những hằng số, là hệ số tương ứng với nguồn thứ i tới micro thu thứ j , đặc trưng cho ảnh hưởng của môi trường truyền.
- $td1$ và $td2$ là thời gian chênh lệch tín từ khi nhận tín hiệu trực tiếp tới khi nhận tín hiệu do phản xạ.
- τ_1 và τ_2 biểu diễn cho ảnh hưởng của độ trễ trên đường truyền từ mỗi nguồn tới micro thu thứ 2. Không mất tính tổng quát, ta giả sử rằng tín hiệu thu được tại micro thứ nhất không bị ảnh hưởng của trễ.
- t_{MP1} và t_{MP2} là thời gian trễ tính từ khi nhận tín hiệu nguồn tới khi nhận được tín hiệu do bị phản xạ.

3. Kết quả tách tín hiệu

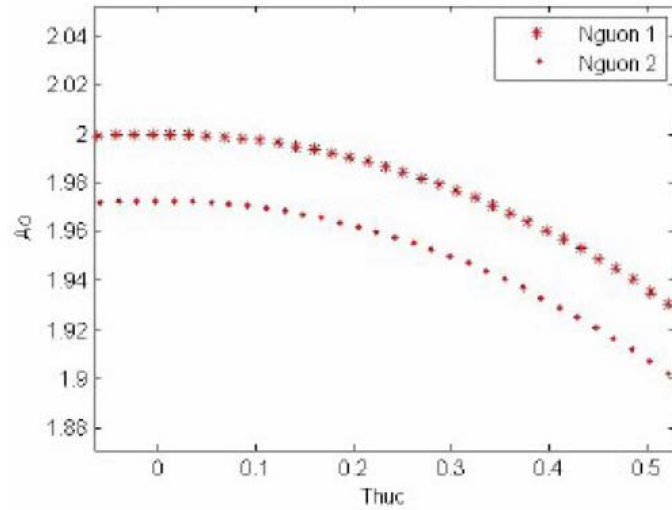
Theo sơ đồ các bước thực hiện thuật toán đã trình bày trong phần 2, các kết quả mô phỏng được chỉ ra trong các hình vẽ dưới đây. Mục đích đặt ra của thuật toán là xác định được những điểm đơn nguồn, từ đó xác định được giá trị về biên độ và pha của thành phần ma trận trộn trong miền tần số.

Trong Hình 3.1 là tập hợp các tỉ số về biên độ và pha của các hàm truyền xác định được trong các khoảng đơn nguồn. Trong quá trình mô phỏng có sử dụng 2 đường lý tưởng làm cơ sở để so sánh (đường màu đỏ và màu xanh đậm), hai đường này biểu diễn tỉ lệ biên độ tín hiệu thu được tại các micro mà không chịu ảnh hưởng nào của môi trường (tỉ lệ tương ứng là 2 và 0.5). Các điểm trên mỗi một trục thẳng đứng biểu diễn cho số lượng các véc tơ của ma trận trộn xác định được trong những khoảng đơn nguồn.



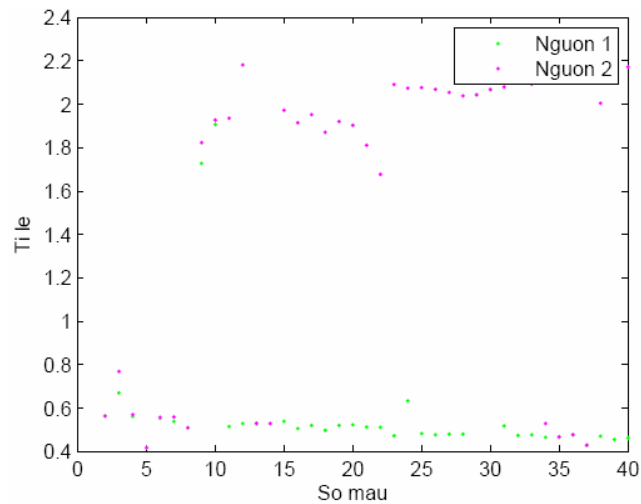
Hình 3.1: Tập hợp những tỉ số biên độ và pha trong những khoảng đơn nguồn với 2 nguồn

Sau khi đã xác định được những khoảng đơn nguồn, xác định được biên độ và pha thì mô hình trộn Gauss sẽ được sử dụng để nhóm các thành phần đó trong một tần số. Tiếp đó, từ kết quả đầu ra của mô hình trộn Gauss, hai bộ lọc Kalman được sử dụng tương ứng để liên kết hàm truyền giữa các tần số với nhau thuộc về một nguồn. Trong hình 3.2 thể hiện hoạt động của hai quá trình này, đường "chấm nhỏ" là của mô hình trộn Gauss và đường "dấu sao" là của lọc Kalman (KF) cho một đoạn nhỏ tương ứng với nguồn một được phóng to để so sánh. Kết quả thực hiện bởi hai quá trình này tương đối gần nhau.

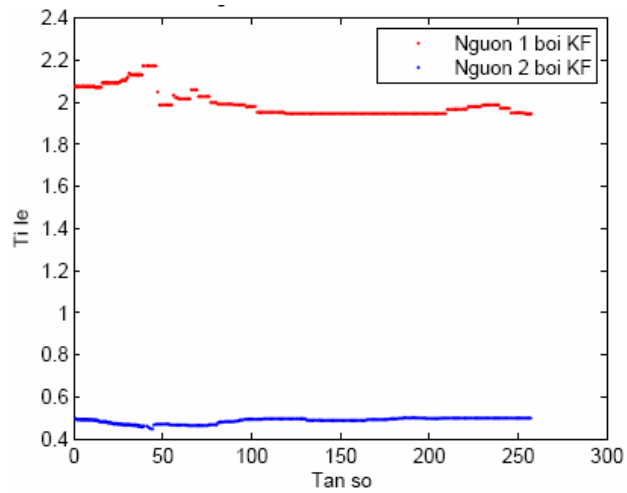


Hình 3.2: Liên kết các giá trị hàm truyền thực hiện bởi mô hình trộn Gauss và bộ lọc Kalman cho nguồn 1 trong một khoảng

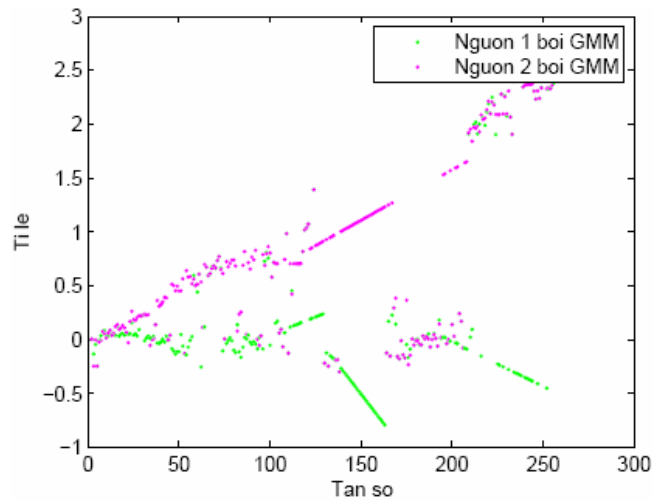
Trong các hình tiếp theo, 3.3, 3.4, 3.5 và 3.6 là các hình tách riêng để so sánh kết quả xác định biên độ và pha của các thành phần ma trận trộn trong những khoảng đơn nguồn trong miền tần số bởi mô hình trộn Gauss và KF. Trong Hình 3.7 là tổng hợp của những hình vẽ trên được mô phỏng trong thuật toán. Từ hình vẽ ta nhận thấy, kết quả thực hiện bởi bộ lọc KF cho các giá trị tương đối sát với hai đường lý tưởng. Trong đường đặc trưng về pha có những nét đứt do tại đó pha không lên tục.



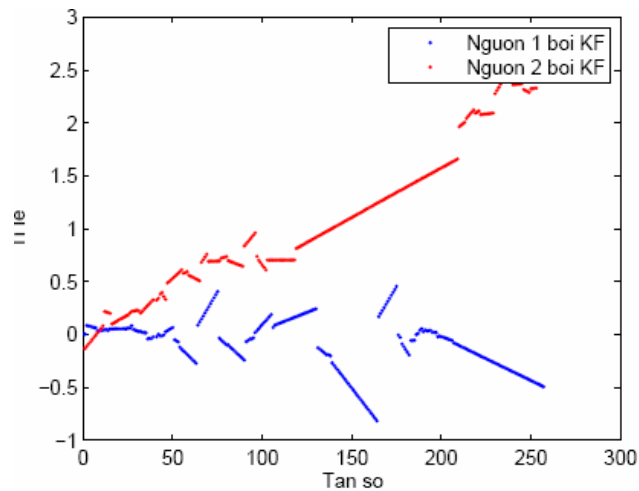
Hình 3.3: Đặc trưng biên độ của hàm truyền tại mỗi tần số xác định bởi mô hình trộn Gauss



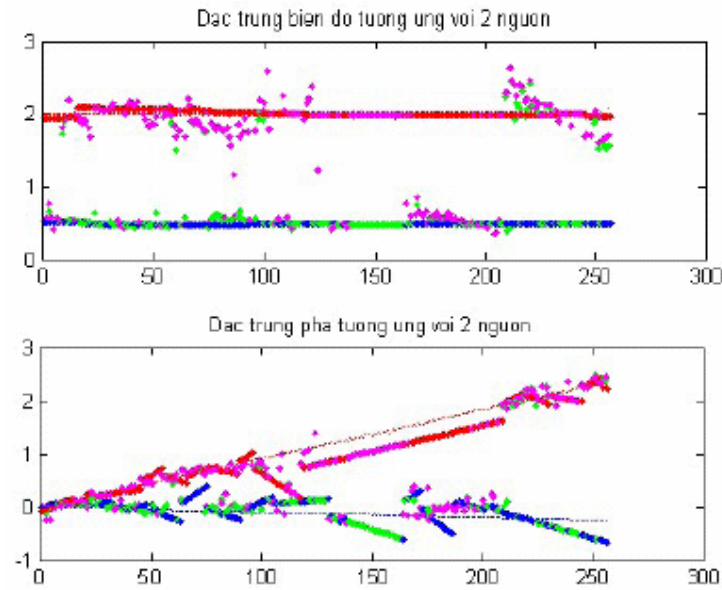
Hình 3.4: Đặc trưng biên độ của hàm truyền giữa các tần số xác định bởi KF



Hình 3.5: Đặc trưng pha của hàm truyền tại mỗi tần số xác định bởi mô hình trộn Gauss

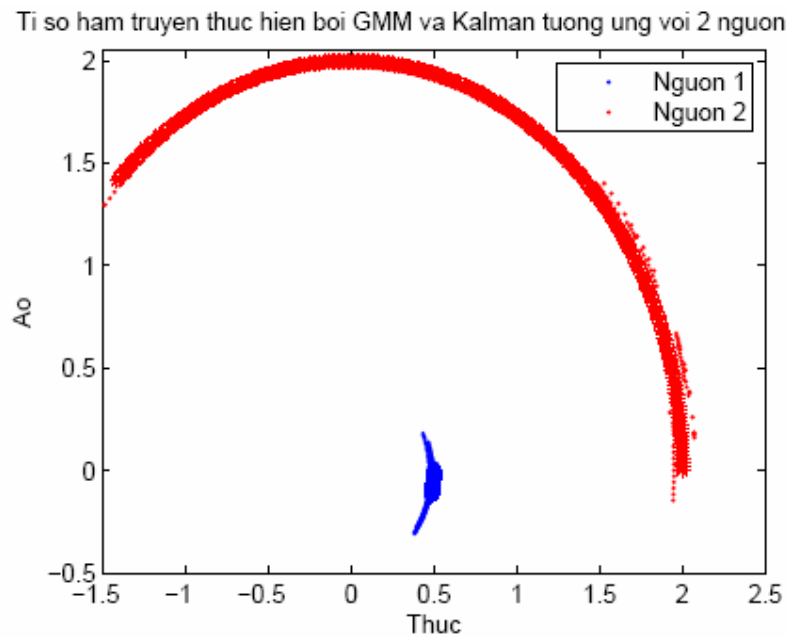


Hình 3.6: Đặc trưng pha của hàm truyền giữa các tần số xác định bởi KF



Hình 3.7: Biên độ và pha của các hàm truyền xác định bởi mô hình trộn Gauss và KF cho 2 nguồn

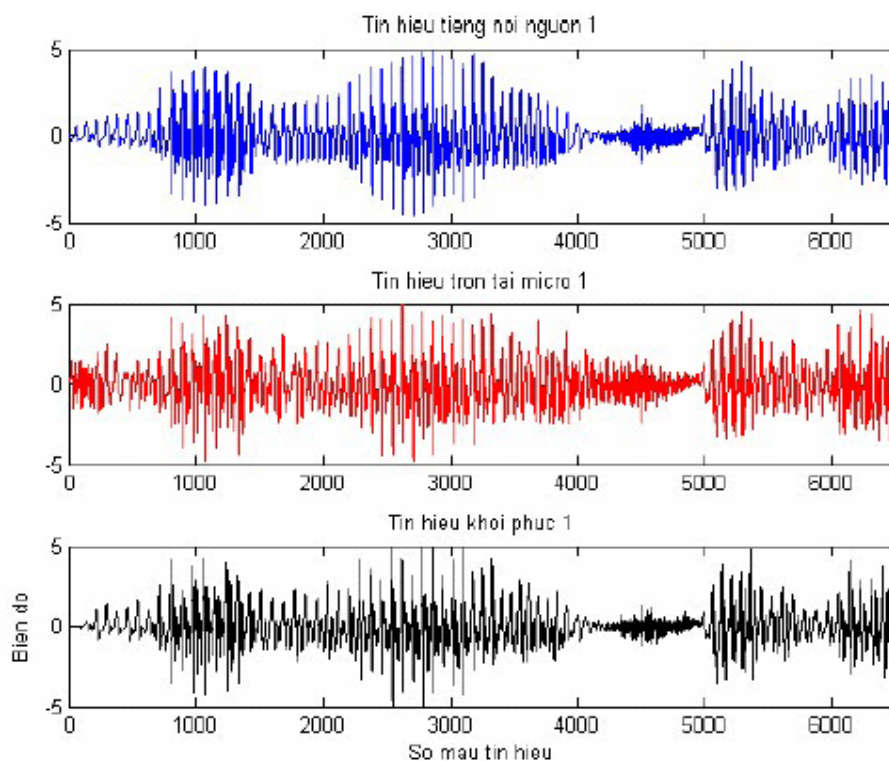
Hình vẽ 3.8 mô tả hoạt động của mô hình trộn Gauss và KF trong việc liên kết hàm truyền tương ứng với cả hai nguồn. Tại mỗi điểm là một giá trị phức (cả biên độ và pha) của thành phần ma trận trộn. Trong mỗi đường đó thực tế gồm có hai đường, một là của mô hình trộn Gauss (GMM) và một là của lọc Kalman (KF). Hình 3.2 là sự phóng đại của hình này trong một khoảng nhỏ.



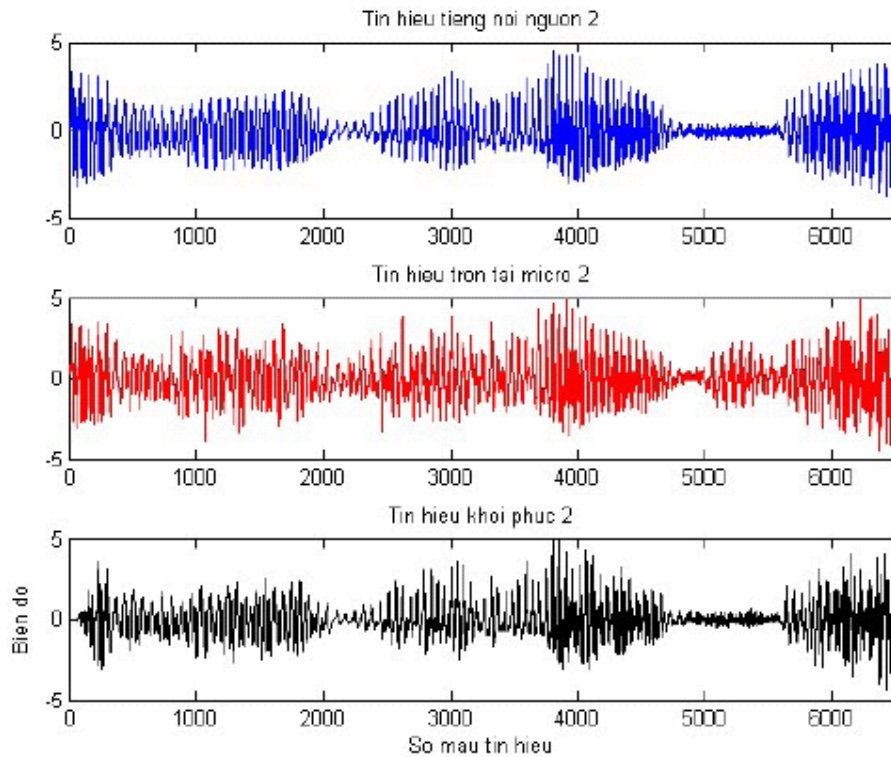
Hình 3.8: Liên kết các giá trị hàm truyền thực hiện bởi mô hình trộn Gauss và bộ lọc Kalman cho 2 nguồn

Tín hiệu sau khi tách thành công:

Kết quả quan trọng nhất của thuật toán này trong việc tách tín hiệu tiếng nói được thể hiện rõ trong các hình 3.9 và hình 3.10. Trên đó vẽ dạng của hai tín hiệu tiếng Việt được sử dụng trong quá trình thử nghiệm. Trong mỗi hình lần lượt là dạng của tín hiệu gốc (thu âm bằng micro), tín hiệu bị trộn lẫn giữa hai nguồn tín hiệu gốc với nhau và có ảnh của môi trường truyền và sau cùng là tín hiệu được khôi phục từ tín hiệu bị trộn lẫn bằng thuật toán nghiên cứu. Từ hai hình vẽ ta nhận thấy, kết quả thực hiện bởi thuật toán là khá chính xác. Dạng của tín hiệu bị trộn lẫn thể hiện rõ sự ảnh hưởng của tín hiệu nguồn khác và của môi trường truyền (với tín hiệu 1 là các khoảng từ 0 - 800; từ 4000 - 5000 và 5600 - 6000; Với tín hiệu 2 là những khoảng từ 2000 - 2500; từ 3400 - 3800 và thấy rõ qua khoảng 4800 - 5800). Độ chính xác của các kết quả này còn được kiểm tra bằng cách phát lại những tín hiệu này. Kết quả tiếng nói phát ra phân biệt khá rõ các quá trình cũng như độ chính xác mà thuật toán thực hiện được.



Hình 3.9: Tín hiệu 1 trong thuật toán: tín hiệu nguồn, tín hiệu trộn và tín hiệu khôi phục qua thuật toán



Hình 3.10: Tín hiệu 2 trong thuật toán: tín hiệu nguồn, tín hiệu trộn và tín hiệu khôi phục qua thuật toán

5. Kết luận

Công trình này nghiên cứu thuật toán tách nguồn bằng phương pháp mù trong môi trường mà tín hiệu bị ảnh hưởng của nhiễu cộng, trễ trên đường truyền và lệch do va đập với vật cản. Kết quả thu âm và xử lý chỉ rõ tính hiệu quả của thuật toán này. Chúng ta đã tách được các tín hiệu nguồn ban đầu với độ chính xác cao. Kết quả được kiểm tra bởi dạng tín hiệu được vẽ ra và phát âm lại để so sánh. Nghiên cứu thành công thuật toán này là cơ sở để mở rộng cho các bài toán phức tạp hơn cũng như ứng dụng vào việc tách tín hiệu trong nhiều lĩnh vực khác.