

## TÁCH ÂM DÙNG PHƯƠNG PHÁP PHÂN TÍCH THÀNH PHẦN ĐỘC LẬP

Trương Tấn Quang, Nguyễn Hữu Phương

Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

(Bài nhận ngày 01 tháng 12 năm 2005, hoàn chỉnh sửa chữa ngày 23 tháng 02 năm 2006)

**TÓM TẮT:** Tại ta thường đồng thời tiếp nhận nhiều nguồn âm (tiếng nói, âm nhạc, nhiễu...) khác nhau nhưng ta vẫn có thể lắng nghe nguồn âm chủ định. Một hệ thống xử lý tiếng hay nhận dạng tiếng cần đạt đến khả năng thông minh như vậy. Bài toán là từ nhiều tín hiệu đã trộn lẫn ta muốn khôi phục các tín hiệu nguồn riêng rẽ.

Trong hơn chục năm qua người ta đã phát triển một phương pháp mới giúp giải bài toán tách nguồn mù (Blind Source Separation – BSS) nêu trên rất hiệu quả, đó là phân tích thành phần độc lập (Independent Component Analysis – ICA). Bài báo trình bày tổng quan tóm lược về ICA và ứng dụng của chúng tôi vào việc tách âm. Kết quả rất tốt. Từ đây chúng tôi có thể tiếp tục phân tích các bài toán tách âm phức tạp hơn dùng ICA.

### 1. GIỚI THIỆU

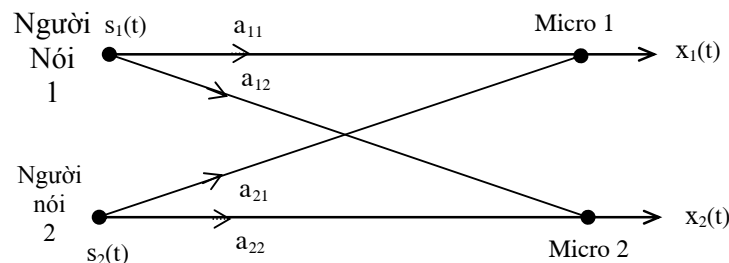
Khi có nhiều tín hiệu ở các khoảng tần số khác nhau được trộn (tổng hợp) lại, ta có thể lọc ra từng tín hiệu riêng biệt như lúc ban đầu. Còn nếu các tín hiệu nằm trong cùng khoảng tần số (ví dụ nhiều người cùng nói, tiếng hát trên nền nhạc, nhiều bản nhạc cùng chơi...), ta không thể dùng phương pháp lọc hay phân tích phổ thông thường. Lúc bấy giờ phải dựa vào các phương pháp thống kê trong đó có phương pháp phân tích thành phần độc lập (Independent Component Analysis - ICA) [1] . . . [5]. ICA có nhiều ứng dụng, bài báo này chỉ đề cập đến ứng dụng tách âm (tiếng nói và âm nhạc) [6] [7]. Do chưa thấy công trình trên các tập san, tạp chí trong nước liên quan đến ICA nên chúng tôi có phần tổng quan hơi là chi tiết (mục 1, 2, 3 và 4).

Ở hình 1 tiếng nói của hai người được thu bởi hai micro:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

trong đó  $a_{ij}$ , với  $i, j = 1, 2$ , là các hệ số trộn phụ thuộc vào nhiều yếu tố như hướng đặt micro, khoảng cách, phản âm của phòng .... Trong mô hình ICA các tín hiệu  $x_i(t)$  và các tín hiệu nguồn  $s_i(t)$  được xem như là các biến ngẫu nhiên, thay vì là các tín hiệu thời gian thực sự, nên trong cách viết chỉ số thời gian  $t$  được bỏ đi.



**Hình 1** Hai tín hiệu nguồn (không quan sát trực tiếp được tức các tín hiệu ẩn) là  $s_1(t)$ ,  $s_2(t)$ ; hai tín hiệu trộn (quan sát được) là  $x_1(t)$ ,  $x_2(t)$

### 1.1 Mô hình ICA cơ bản

Xem trường hợp có  $n$  tín hiệu trộn  $x_i$  được tổng hợp từ  $n$  tín hiệu nguồn  $s_i$  thì

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, \quad i = 1, \dots, n \quad (2)$$

với  $a_{ij}$ ,  $i, j = 1, 2, \dots, n$  là các hệ số trộn được giả sử có trị số thực. Để ý là ta đã giả sử số lượng tín hiệu nguồn và tín hiệu trộn bằng nhau để được đơn giản. Bài toán ICA là ước lượng (phân ly, tách ra) các tín hiệu nguồn  $s_j$  từ các tín hiệu trộn quan sát được  $x_i$  trong lúc các hệ số  $a_{ij}$  không biết (nếu biết các hệ số trộn  $a_{ij}$  ta có thể giải hệ phương trình (2) để tìm các tín hiệu  $s_i$ ). Đây là bài toán “cocktail party” (ý nói rằng tại buổi tiệc có nhiều người nói nhưng ta có thể tập trung nghe một hai người nào đó theo chủ định).

Thường ký hiệu vector – ma trận được dùng. Gọi  $\mathbf{x}$  là vector ngẫu nhiên biểu diễn các phần tử trộn  $x_i$ :  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ ,  $\mathbf{s}$  là vector ngẫu nhiên của các phần tử nguồn  $s_i$ :

$\mathbf{s} = [s_1, s_2, \dots, s_n]^T$ , và  $\mathbf{A}$  là ma trận trộn chứa các hệ số trộn  $a_{ij}$ :  $\mathbf{A} = [a_{ij}]$ . Các vector  $\mathbf{x}$ ,  $\mathbf{s}$  được hiểu là các vector cột. Để ý là khi số lượng tín hiệu trộn và tín hiệu nguồn bằng nhau như đã giả sử thì  $\mathbf{A}$  là một ma trận vuông. Mô hình ICA trở thành:

$$\mathbf{x} = \mathbf{A} \mathbf{s} \quad (3)$$

Phương trình (3) là mô hình ICA tuyến cơ bản (còn gọi ICA chuẩn). Khi cần sử dụng biểu diễn dạng cột của ma trận  $\mathbf{A}$  ta gọi các cột là  $\mathbf{a}_i$  và viết

$$\mathbf{x} = \sum_{i=1}^n \mathbf{a}_i s_i \quad (4)$$

Từ (3) có thể giải ra  $\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$  nếu biết được  $\mathbf{A}$  và nếu  $\mathbf{A}$  là nghịch đảo được, nhưng thực tế không biết  $\mathbf{A}$ . Đây cũng là bài toán tách nguồn (hay tín hiệu) mù (Blind Source/ Signal Separation - BSS) [6]. Các tín hiệu nguồn là các thành phần độc lập (lý do sẽ thấy ở sau).

### 1.2 Các điểm không xác định trong ICA

Mô hình ICA cơ bản (3) và (4) có một số bất định tóm lược như sau:

1. Không thể xác định các phương sai (năng lượng) của các thành phần độc lập. Người ta giả sử mỗi biến có phương sai đơn vị:  $E\{s_i^2\} = 1$ . Tuy nhiên hãy còn bất định về dấu vì khi ta nhân một thành phần độc lập với -1 thì mô hình ICA không thay đổi. Trong hầu hết các ứng dụng yếu tố dấu không có ý nghĩa.
2. Không thể xác định thứ tự của các thành phần độc lập khi phân ly vì cả  $\mathbf{s}$  và  $\mathbf{A}$  đều không biết nên ta có thể thay đổi thứ tự các số hạng trong (4) nên bất cứ thành phần nào cũng có thể là thành phần đầu tiên. Nói rõ hơn thì xác định thứ tự là công việc khó khăn.

## 2. SỰ ĐỘC LẬP THỐNG KÊ

### 2.1 Bất tương quan

Các phân bố xác suất đều giả sử có trị trung bình bằng không. Nếu không phải như vậy thì ta trừ phân bố với trị trung bình của nó, đây là sự qui tâm (centering). Để ý là hiệp phương sai (covariance) chính là tương quan (correlation) khi trị trung bình bằng không. Đối với một vector ngẫu nhiên  $\mathbf{x}$  ma trận hiệp phương sai là

$$\mathbf{C}_{\mathbf{xx}} = E \{ (\mathbf{x} - \mathbf{m}_x) (\mathbf{x} - \mathbf{m}_x)^T \} \quad (5)$$

trong đó  $E \{ \cdot \}$  là toán tử lấy trung bình,  $\mathbf{m}_x$  là vector trung bình. Hiệp phương sai của hai vector ngẫu nhiên  $\mathbf{x}_1, \mathbf{x}_2$  (có trị trung bình bằng không) là:

$$\mathbf{C}_{\mathbf{x}_1\mathbf{x}_2} = E \{ \mathbf{x}_1 \mathbf{x}_2^T \} \quad (6)$$

Khi  $\mathbf{C}_{\mathbf{x}_1\mathbf{x}_2} = 0$  hai vector bất tương quan (uncorrelated). Đối với vector ngẫu nhiên  $\mathbf{x}$  khi các thành phần  $x_i$  của nó bất tương quan hỗ tương thì:

$$\mathbf{C}_{\mathbf{xx}} = \mathbf{D} \quad (7)$$

trong đó  $\mathbf{D}$  là ma trận chéo  $n \times n$ , với các phương sai của các thành phần nằm trên đường chéo.

## 2.2 Độc lập thống kê

Tính bất tương quan nêu trên chưa đủ để phân ly ICA. Ta cần một đặc tính mạnh hơn, đó là sự độc lập thống kê, nghĩa là khi biết một thành phần nào đó ta không thể suy ra các thành phần còn lại. Xem hai vector ngẫu nhiên  $\mathbf{x}_1$  và  $\mathbf{x}_2$  với hàm mật độ xác suất riêng biệt  $p(\mathbf{x}_1)$ ,  $p(\mathbf{x}_2)$  và hàm mật độ xác suất liên kết  $p(\mathbf{x}_1, \mathbf{x}_2)$  là độc lập thống kê nếu và chỉ nếu khi thỏa

$$p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1) p(\mathbf{x}_2) \quad (8)$$

Khi có nhiều vector thì sự thừa số hóa cũng tương tự.

Định nghĩa kỹ thuật ở trên dẫn đến một đặc tính sau của các biến ngẫu nhiên. Xem  $f(\mathbf{x}_1)$  và  $f(\mathbf{x}_2)$  là biến đổi phi tuyến nào đó trên hai vector ngẫu nhiên  $\mathbf{x}_1$  và  $\mathbf{x}_2$  có hàm phân bố đã nói ở trên, thì có thể chứng minh được [1]:

$$E\{f_1(\mathbf{x}_1)f_2(\mathbf{x}_2)\} = E\{f_1(\mathbf{x}_1)\}E\{f_2(\mathbf{x}_2)\} \quad (9)$$

Như vậy sự độc lập là có thể thừa số hóa tương quan phi tuyến. Đây là đặc tính quan trọng vì nó giải thích và nhấn mạnh vai trò các phi tuyến trong ICA. Khi đặt  $f(\mathbf{x}_1) = \mathbf{x}_1$  và  $f(\mathbf{x}_2) = \mathbf{x}_2$  ta thấy là sự độc lập bao gồm luôn sự bất tương quan (nhưng bất tương quan không đương nhiên là độc lập). Cụ thể là ta giả sử  $\mathbf{s}$  ở phương trình (3) là độc lập thống kê nên các tín hiệu nguồn  $s_i$  là các thành phần độc lập. Chính nhờ sự độc lập thống kê mà ta có thể phân ly ra  $\mathbf{s}$  từ (3).

## 2.3 Phi Gauss là độc lập

Mô hình ICA đặt ra một hạn chế là các thành phần độc lập phải có tính phi Gauss (non-gaussianity), tức không có phân bố (hàm mật độ xác suất) là Gauss.

Lý do tính phi Gauss nằm ở chỗ là các biến ngẫu nhiên Gauss được xác định hoàn toàn bởi các thống kê bậc một (trị trung bình) và bậc hai (phương sai), các thống kê bậc cao hơn bằng không. Trong lúc, như sẽ thấy ở sau, mô hình ICA cần các thống kê bậc cao hơn của các thành phần độc lập để thực hiện sự phân ly (ước lượng các thành phần độc lập). Như vậy, sự phi tuyến, tính phi Gauss dẫn đến sự độc lập thống kê.

## 2.4 Các giả sử trong mô hình ICA

Mô hình ICA tuyến cơ bản đặt ra đòi hỏi các giả thiết sau cho việc phân ly (ước lượng) các thành phần độc lập:

1. Các nguồn  $\mathbf{s}$  độc lập thống kê nhau, nghĩa là biết được một nguồn không thể suy ra các nguồn còn lại.
2. Các hàm phân bố xác suất của các nguồn có trị trung bình bằng không.
3. Không có nguồn (thành phần độc lập) nào có phân bố Gauss (thật ra mô hình cho phép có tối đa một thành phần có phân bố Gauss).
4. Ma trận trộn  $\mathbf{A}$  là ma trận vuông tức số lượng nguồn và số lượng trộn bằng nhau. Nếu không phải vậy, bài toán sẽ khó hơn.

## 3. TIỀN XỬ LÝ Ở ICA

Thường trước khi ước lượng ICA cho dữ liệu quan sát được người ta áp dụng một hai tiền xử lý để việc ước lượng ICA được thuận lợi hơn. Ở đây chỉ trình bày hai cách phổ biến áp dụng ở bài báo này.

### 3.1 Qui tâm

Như đã nêu ở trước, các tín hiệu trộn quan sát được  $\mathbf{x}$  phải có trị trung bình  $\mathbf{m} = E[\mathbf{x}]$  bằng không, điều này cũng có nghĩa là các tín hiệu nguồn  $\mathbf{s}$  cũng có trị trung bình là không. Nếu các tín hiệu chưa có trị trung bình là không ta thực hiện phép qui tâm tức trừ phân bố của các biến ngẫu nhiên với các trị trung bình của chúng:

$$\mathbf{x} = \mathbf{x}' - E\{\mathbf{x}'\} \quad (10)$$

trong đó  $\mathbf{x}'$  là vector ngẫu nhiên chưa có trung bình là không. Sau khi đã ước lượng ma trận  $\mathbf{A}$  và các thành phần  $\mathbf{s}$  ta có thể thêm trở lại các trị trung bình của chúng. Khi vector ngẫu nhiên  $\mathbf{x}$  (hoặc  $\mathbf{s}$ ) có trị trung bình là không thì hiệp phương sai và tương quan của nó giống nhau.

### 3.2 Trắng hóa

Sau khi đã qui tâm các biến ngẫu nhiên  $\mathbf{x}$ , ta áp dụng một biến đổi tuyến tính trên  $\mathbf{x}$  để được vector mới là trắng. Sự làm trắng hay trắng hóa (whitening), còn gọi cầu hóa (sphering), có mục đích làm cho dữ liệu bất tương quan. Giả sử ta có vector ngẫu nhiên  $\mathbf{x}$  bất tương quan, tức các xuyến phương sai của các phần tử bằng không, dẫn đến ma trận hiệp phương sai là ma trận chéo có các số hạng chéo tương ứng với các phương sai của các phần tử của  $\mathbf{x}$ . Nếu các phương sai này được cho bằng 1 (mục 1.2), nghĩa là ma trận hiệp phương sai được cho bằng với ma trận đồng nhất  $\mathbf{I}$  thì vector ngẫu nhiên  $\mathbf{x}$  là trắng:

$$E[\mathbf{xx}^T] = \mathbf{I} \quad (\mathbf{x} \text{ là trắng}) \quad (11)$$

Việc làm trắng là một biến đổi tuyến tính

$$\mathbf{z} = \mathbf{V}\mathbf{x} \quad (12)$$

trong đó  $\mathbf{x}$  là dữ liệu cần làm trắng,  $\mathbf{V}$  là ma trận làm trắng,  $\mathbf{z}$  là dữ liệu đã trắng hóa. Cách thường dùng nhất để tìm ma trận làm trắng là thực hiện sự phân ly trị riêng (Eigenvalue Decomposition - EVD) trên ma trận hiệp phương sai:

$$E[\mathbf{xx}^T] = \mathbf{E}\mathbf{D}\mathbf{E}^T \quad (13)$$

trong đó  $\mathbf{E}$  là ma trận trực giao của các vector riêng của  $E[\mathbf{xx}^T]$  và  $\mathbf{D}$  là ma trận chéo của các trị riêng của chúng,  $\mathbf{D} = \text{chéo}(d_1, \dots, d_n)$ ,  $n$  là số lượng nguồn quan sát được  $\mathbf{x}$ . Ma trận làm trắng là:

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \quad (14)$$

trong đó  $\mathbf{D}^{-1/2} = \text{chéo}(d_1^{-1/2}, \dots, d_n^{-1/2})$ . Ma trận làm trắng trên cũng có thể được biểu diễn như:

$$\mathbf{V} = E[\mathbf{xx}^T]^{-1/2} = \mathbf{C}_{\mathbf{xx}}^{-1/2} \quad (15)$$

trong đó  $\mathbf{C}_{\mathbf{xx}} = E[\mathbf{xx}^T]$  là ma trận hiệp phương sai. Vậy

$$\mathbf{z} = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} = \tilde{\mathbf{A}}\mathbf{s} \quad (16)$$

trong đó  $\tilde{\mathbf{A}} = \mathbf{V}\mathbf{A}$  là ma trận trộn đã làm trắng. Mặt khác vì  $\mathbf{z}$  là dữ liệu đã làm trắng nên  $E\{\mathbf{zz}^T\} = \mathbf{I}$ , dẫn đến

$$E\{\mathbf{zz}^T\} = E\{\tilde{\mathbf{A}}\mathbf{s}[\tilde{\mathbf{A}}\mathbf{s}]^T\} = \tilde{\mathbf{A}}E\{\mathbf{ss}^T\}\tilde{\mathbf{A}}^T = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^T = \mathbf{I} \quad (17)$$

Nên nhớ là các thành phần độc lập  $s_i$  đều giả sử có phương sai đơn vị. Vậy ma trận đã làm trắng là trực giao (ở ma trận trực giao nghịch đảo của nó bằng chuyển vị của nó:  $\tilde{\mathbf{A}}^{-1} = \tilde{\mathbf{A}}^T$ ).

Ta biết ma trận  $n \times n$  nếu không trực giao chứa  $n^2$  độ tự do, còn nếu trực giao chỉ chứa  $n(n-1)/2$  độ tự do. Với dữ liệu 2 chiều điều này có nghĩa là độ tự do chỉ là 1 cho một biến đổi trực giao. Khi số chiều lớn, độ tự do của một ma trận trực giao chỉ là phân nửa độ tự do của ma trận không trực giao. Do đó người ta nói làm trắng là đã giải quyết phân nửa bài toán ICA. Bởi vì trắng hóa là một thủ tục đơn giản hơn bất cứ thuật toán ICA nào nên là một tiền xử lý thông dụng.

Sau khi đã ước lượng ma trận trắng hóa  $\tilde{\mathbf{A}}$ , thì từ (16) việc ước lượng các thành phần độc lập  $\mathbf{s}$  trở thành:

$$\mathbf{s} = \mathbf{W}\mathbf{z} \quad (18)$$

trong đó  $\mathbf{W}$  là đảo nghịch của  $\tilde{\mathbf{A}}$ :  $\mathbf{W} = \tilde{\mathbf{A}}^{-1}$ . Với các ma trận vuông việc lấy nghịch đảo rất thuận lợi. Sau khi có được  $\mathbf{W}$ , việc ước lượng ma trận gốc  $\mathbf{A}$  cho bởi

$$\tilde{\mathbf{A}} = \mathbf{V}\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{A} \quad (19)$$

Suy ra:

$$\mathbf{A} = \mathbf{E}\mathbf{D}^{1/2}\tilde{\mathbf{A}} = \mathbf{E}\mathbf{D}^{1/2}\mathbf{W}^{-1} = \mathbf{E}\mathbf{D}^{1/2}\mathbf{W}^T \quad (20)$$

trong đó vì  $\mathbf{W}$  là trực giao nên  $\mathbf{W}^{-1} = \mathbf{W}^T$ . Sau này ta vẫn dùng ký hiệu  $\mathbf{s}$ ,  $\mathbf{x}$ ,  $\mathbf{A}$  dù có làm trắng hoặc không, trừ khi cần làm rõ ra.

Như vậy còn lại là ước tính các hàng của  $\mathbf{W}$ . Có nhiều phương pháp để thực hiện việc ước tính này nhưng ở đây chỉ đề cập phương pháp ICA bằng cách cực đại hóa tính phi Gauss (xem mục 4).

#### 4. ƯỚC LƯỢNG ICA

Ước tính ICA là một công việc khá chi li. Người ta đã phát triển nhiều cách để giải quyết bài toán, một số là [1] . . . [4]

- Cực đại hóa tính phi Gauss (nongaussianity)
- Ước lượng khả năng cực đại (maximum likelihood)
- Cực tiểu hoá thông tin hỗ tương (mutual information)

Trong các phương pháp trước tiên định ra một hàm đối tượng (objective function), còn gọi hàm trị giá (cost function), rồi dùng một thuật toán tối ưu hóa để cực đại hóa hoặc cực tiểu hóa (nói chung là cực đại hóa trị tuyệt đối) hàm đối tượng này để ước lượng các thành phần độc lập. Bài này chỉ trình bày tóm lược phương pháp cực đại hóa tính phi Gauss dẫn đến giải thuật tính toán nhanh và hiệu quả đã được phát triển FastICA.

Theo định lý giới hạn trung tâm (central limit theorem), tổng của nhiều biến ngẫu nhiên có phân bố gần Gauss hơn bất cứ biến ngẫu nhiên gốc nào. Ở mô hình ICA (3) vector ngẫu nhiên  $\mathbf{x}$  gồm các biến ngẫu nhiên là trộn tuyến của các vector biến ngẫu nhiên nguồn  $\mathbf{s}$ . Các nguồn được giả sử độc lập nhau nhưng khi trộn lại (cộng nhau) thì các trộn trở nên gần Gauss hơn. 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 ta các thành phần độc lập.

##### 4.1 Đo tính phi Gauss bằng kurtosis

Đầu tiên là phép đo dựa trên kurtosis của một biến ngẫu nhiên  $y$  có trung bình bằng không là cumulant bậc bốn:

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (21)$$

Thật ra vì ta giả sử  $y$  có phương sai đơn vị, nên kurtosis là  $\text{kurt}(y) = E\{y^4\} - 3$ , tức kurtosis là phiên bản chuẩn hóa của momen thứ tư  $E\{y^4\}$ . Khi  $y$  có phân bố Gauss momen thứ tư bằng  $3(E\{y^2\})^2$  nên kurtosis bằng không đối với các biến ngẫu nhiên Gauss. Hầu hết các biến ngẫu nhiên không phải Gauss kurtosis khác không. Nếu kurtosis là dương biến ngẫu nhiên có phân bố siêu Gauss (supergaussian), còn nếu kurtosis là âm thì biến ngẫu nhiên có phân bố dưới Gauss (subgaussian). Phân bố siêu Gauss không còn dạng hình chuông như Gauss mà tăng nhanh ở trung tâm tương tự như phân bố Laplace, còn phân bố dưới Gauss không nhô lên ở phần giữa như Gauss mà tiến đến phân bố đều với biên độ rất nhỏ ở xa trung tâm.

Việc đo tính phi Gauss bằng kurtosis có vài bất lợi khi các giá trị của nó được tính từ các mẫu quan sát được, vì kurtosis rất bị ảnh hưởng bởi các trị biên (outlier) quan sát được ở hai đuôi của phân bố.

##### 4.2 Đo tính phi Gauss bằng Negentropy

Một số đo tính phi Gauss quan trọng hơn là negentropy. Negentropy là đại lượng dựa trên lý thuyết thông tin gọi entropy vì sai. Entropy của một biến ngẫu nhiên là số đo lượng thông tin trung bình của nó. Entropy (vì sai)  $H$  của vector ngẫu nhiên  $\mathbf{y}$  có hàm phân bố  $f(\mathbf{y})$  định nghĩa như sau:

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (22)$$

Đặc tính quan trọng của entropy là biến ngẫu nhiên Gauss có entropy lớn nhất trong các biến ngẫu nhiên có cùng phương sai. Như vậy entropy, và negentropy định nghĩa theo entropy, có thể dùng để đo tính phi Gauss của một biến ngẫu nhiên. Để có được một số đo tính phi Gauss sao cho bằng không đối với biến Gauss và luôn không âm, người ta định nghĩa negentropy của vector ngẫu nhiên  $\mathbf{y}$ :

$$J(\mathbf{y}) = H(\mathbf{y}_{Gauss}) - H(\mathbf{y}) \quad (23)$$

trong đó  $\mathbf{y}_{Gauss}$  là một vector ngẫu nhiên Gauss cùng ma trận hiệp phương sai (hay ma trận tương quan vì các dữ liệu được giả sử có trung bình là không).

Tuy nhiên tính toán negentropy lại khó khăn. Một số tính toán xấp xỉ đã được phát triển, mà một là:

$$J(y) \approx [E\{G(y)\} - E\{G(y_{\text{Gauss}})\}]^2 \quad (24)$$

Hàm phi tuyến  $G(\cdot)$  có thể chọn theo một hai biểu thức sau:

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \quad (25)$$

$$G_2(y) = -\exp(-y^2/2)$$

với  $1 \leq a_1 \leq 2$  và thường chọn bằng  $a_1 = 1$ .

Trong mô hình ICA, ta muốn tìm các hàng của ma trận  $\mathbf{W}$ . Khi dùng negentropy người ta xây dựng được thuật toán điểm cố định (fixed-point algorithm) [4] sau:

$$\begin{aligned} \mathbf{w} &\rightarrow E\{zg(\mathbf{w}^T \mathbf{z})\} - E\{g'(\mathbf{w}^T \mathbf{z})\} \mathbf{w} \\ \mathbf{w} &\leftarrow \mathbf{w} / \|\mathbf{w}\| \end{aligned} \quad (26)$$

Trong đó  $g(\cdot)$  là đạo hàm của các hàm  $G_1(\cdot)$  và  $G_2(\cdot)$ :

$$\begin{aligned} g_1(y) &= \tanh(a_1 y) \\ g_2(y) &= y \exp(-y^2/2) \end{aligned} \quad (27)$$

Tuy nhiên thường ta không có một thành phần độc lập đơn (chỉ một mà thôi), do đó phải tính nhiều hơn một hàng của  $\mathbf{W}$ . Lúc bấy giờ các dòng  $\mathbf{w}$  khác nhau của ma trận  $\mathbf{W}$  có thể hội tụ đến cùng các cực đại của hàm đối tượng. Để khắc phục vấn đề này, các vector  $w_1, w_2, \dots, w_n$  phải được trực giao hóa sau mỗi lần lặp. Sự trực giao có thể thực hiện theo một số cách.

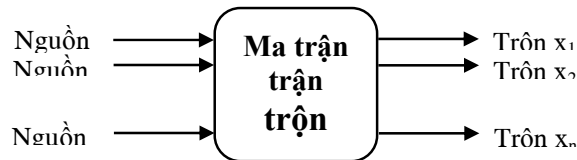
Thuật toán điểm cố định ở trên dẫn đến thuật toán ước lượng fastICA có nhiều ưu điểm so với cách ước lượng dựa vào các thuật toán có cơ sở là gradient. Tài liệu [4] và một số tài liệu khác có ghi chi tiết các thuật toán fastICA. Matlab 7.0 cũng hỗ trợ fastICA.

## 5. ỨNG DỤNG ICA VÀO TÁCH ÂM

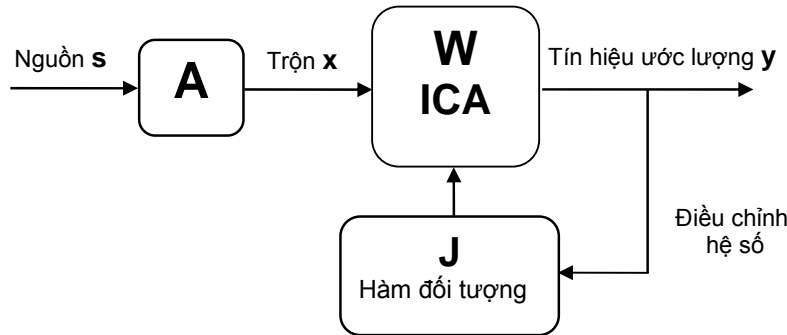
Thực nghiệm tách âm với mô hình ICA, yêu cầu phải thu nhiều nguồn âm cùng lúc. Hình 2 minh họa mô hình trộn âm, trong đó  $s_1, \dots, s_n$  là các nguồn âm thực nghiệm được thu trước vào máy tính. Ma trận trộn  $\mathbf{A}$  được giả lập trên máy tính để thực hiện trộn các nguồn âm, tạo ra các hỗn hợp  $x_1, x_2, \dots, x_n$ . Để ý là số lượng tín hiệu nguồn và số lượng tín hiệu trộn bằng nhau để ma trận  $\mathbf{A}$  là ma trận vuông. Giả sử này làm việc phân tích thuận lợi hơn. Hình 3 minh họa hệ thống tách âm [6] [7]. Tín hiệu âm gốc sau khi được trộn lẫn tạo ra các tín hiệu trộn sẽ đưa đến bộ ICA. Bộ phân tích này thực hiện tách âm bằng giải thuật ước lượng ma trận giải  $\mathbf{W}$  thông qua cực đại tính phi Gauss với hàm đối tượng negentropy  $\mathbf{J}$ . Kết quả ngõ ra bộ phân tích ICA là các âm độc lập ước lượng được của nguồn âm gốc. Nếu  $\mathbf{W}$  là ma trận nghịch đảo của  $\mathbf{A}$  ( $\mathbf{W} = \mathbf{A}^{-1}$ ) thì sự ước lượng là đúng.

Chúng tôi thực nghiệm việc tách âm cụ thể như sau:

- Thu sẵn số nguồn âm thực nghiệm vào máy tính ở tốc độ lấy mẫu 8KHz và 22.5KHz, và mã hoá PCM 16 bit. Chiều dài mỗi đoạn dữ liệu âm là 10s.
- Thực hiện trộn âm từ các nguồn âm đó. Chọn mô hình trộn phù hợp với thực tế (của buổi tiệc, phòng họp,...). Để đơn giản, giả sử theo cách thức trình bày ở mục 1, tức bỏ đi các phản xạ và các trì hoãn của phòng [7].
- Dữ liệu trộn được cho qua bộ phân tích ICA với giải thuật tách âm fastICA (bao gồm quy tâm, trắng hóa và ước lượng). Kết quả là các nguồn âm được tách ra độc lập.



Hình 2. Trộn âm



Hình 3. Tách âm

**6. KẾT QUẢ THỰC NGHIỆM:** Chúng tôi thực hiện nhiều mô phỏng với nhiều loại nguồn âm khác nhau. Sau đây chỉ nêu một số tiêu biểu

**Thực nghiệm 1:** (Hình 4)

Phân tách tiếng nói người, ba nguồn phát  $s$  gồm một giọng nam và hai giọng nữ được trộn để tạo ba hỗn hợp  $x$ . Kết quả tách (ước lượng)  $y$  thu được (đánh giá trên các dạng sóng và nghe qua loa) rất tốt.

**Thực nghiệm 2:** (Hình 5)

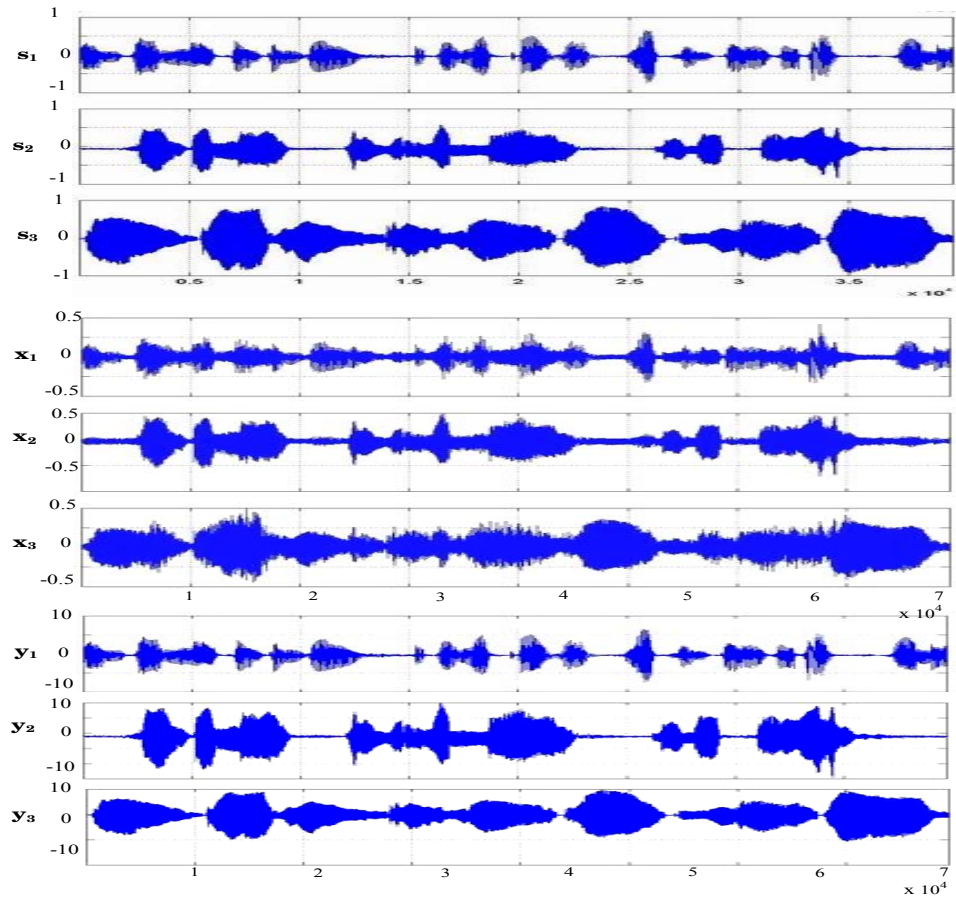
Trên tập dữ liệu lai trộn từ hai nguồn phát là tiếng hát của hai ca sĩ, mỗi nguồn phát 17 giây. Kết quả tách âm rất tốt.

**Thực nghiệm 3:** (Hình 6)

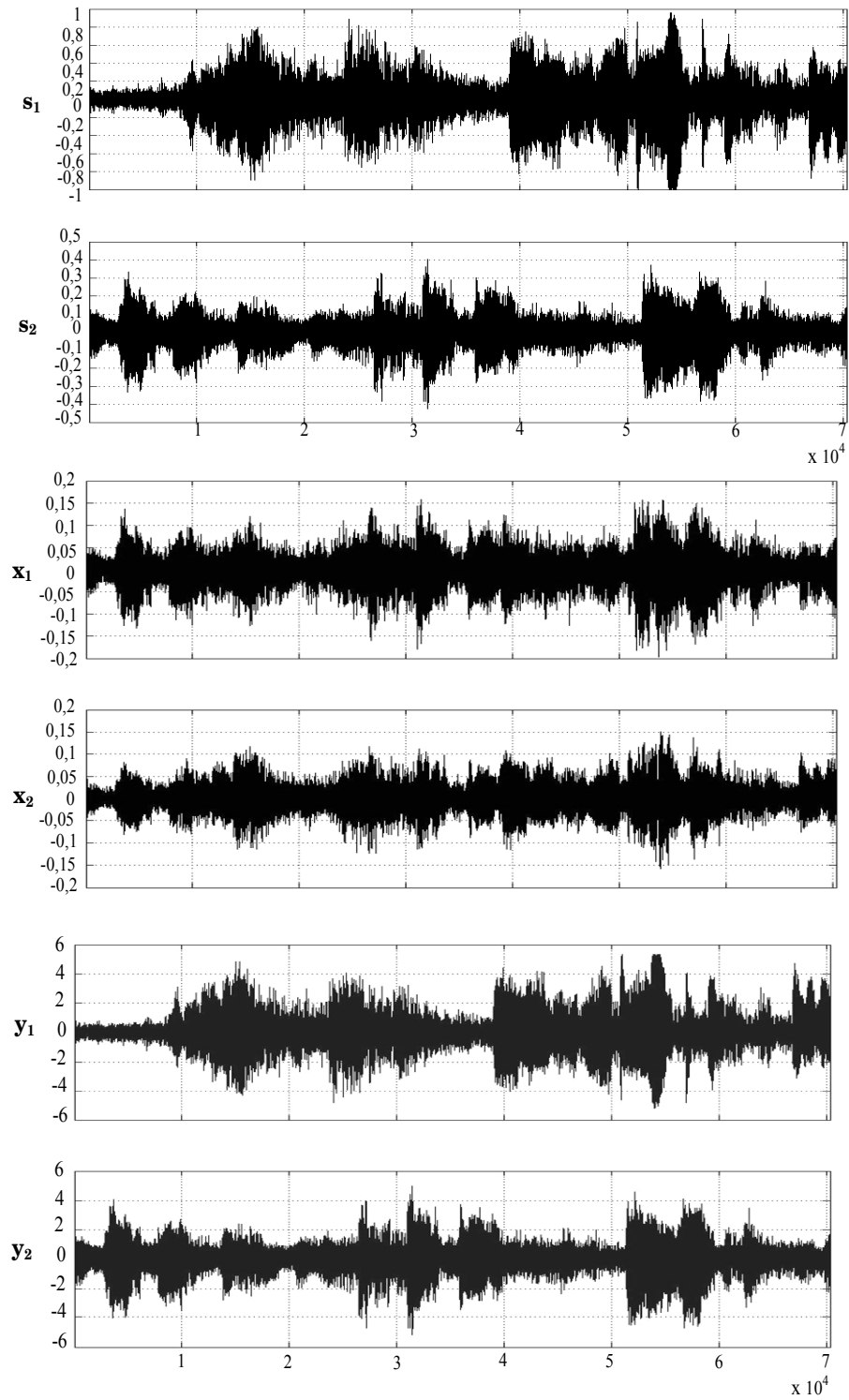
Lặp lại thực nghiệm trên tuy nhiên với hai nhạc phẩm không lời, thời gian phát 10 giây, tần số lấy mẫu 22.5 KHz, mã hóa PCM 16 bit. Kết quả rất tốt.

**Thực nghiệm 4:**

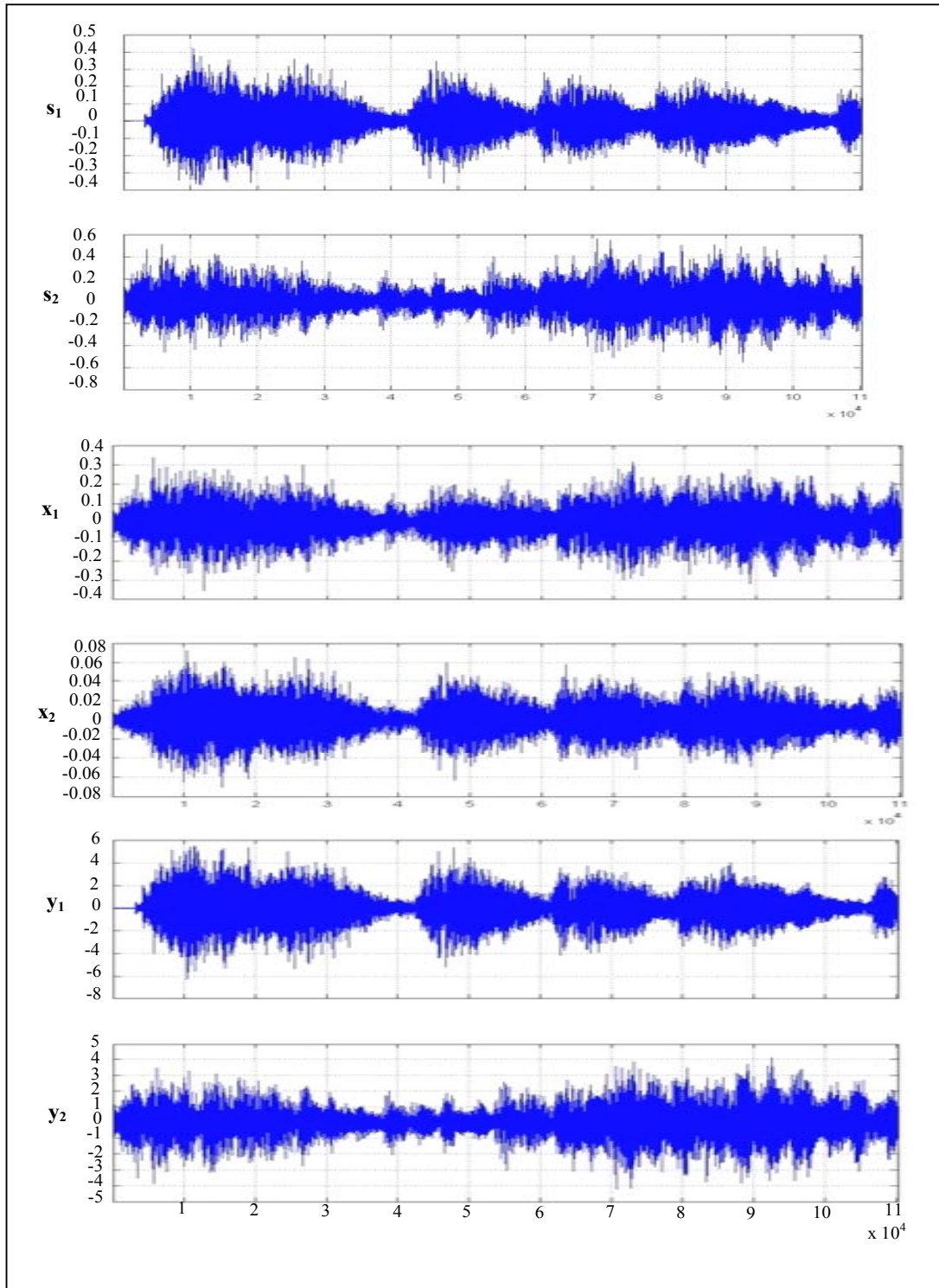
Lai ghép từ tập nhiều loại tín hiệu: tiếng người, tiếng hát, nhạc, tiếng kèn, nhiễu Gauss...; tất cả 10 nguồn. Tín hiệu tách thu được từ hỗn hợp nghe trong và không cảm giác thấy lai trộn. Do chỉ có một nguồn Gauss, nên chấp nhận được trong ICA. Kết quả rất tốt. Vì lý do hạn chế không gian nên không kèm theo hình.



**Hình 4.** Biểu diễn các tín hiệu trong thực nghiệm 1



**Hình 5.** Biểu diễn các tín hiệu trong thực nghiệm 2



**Hình 6.** Biểu diễn các tín hiệu trong thực nghiệm 3

## 7.KẾT LUẬN

Mô hình ICA tuyến cơ bản (mô hình chuẩn) được biểu diễn bởi phương trình (3):  $\mathbf{x} = \mathbf{A}\mathbf{s}$  trong đó  $\mathbf{s}$  là các tín hiệu nguồn (các thành phần độc lập),  $\mathbf{A}$  là ma trận trộn,  $\mathbf{x}$  là các tín hiệu trộn. Ta biết được (quan sát được)  $\mathbf{x}$  và cần xác định  $\mathbf{s}$  trong lúc  $\mathbf{A}$  (và dĩ nhiên cả  $\mathbf{s}$ ) không biết! Thoạt nhìn thì đây là bài toán bất khả giải. Nhưng với một số giả thiết khá thực tế,

đặc biệt là các nguồn s độc lập thống kê hỗ tương. Sự độc lập thống kê đi liền với tính phi tuyến và tính phi Gauss. Tiên xử lý qui tâm và trắng hóa được thực hiện trên  $x$  trước khi ước lượng ICA dùng phương pháp cực đại hóa tính phi Gauss. Cơ sở toán học khá phức tạp nhưng đã có các thuật toán FastICA rất hiệu quả. Nhiều thực nghiệm trên nhiều hỗn hợp tín hiệu nguồn khác nhau cho thấy việc ước lượng ICA rất tốt.

Chúng tôi đã thực hiện nhiều mô phỏng trên máy tính và đánh giá chất lượng tách âm bằng cách so sánh dạng sóng và nghe bằng tai. Thật ra còn một vài cách khác để xem xét như phân tích phổ Fourier, tính sai số xuyên âm . . . Các cách này làm phong phú thêm việc đánh giá và mang tính định lượng hơn nhưng không cho kết quả khác đi.

Nội dung khảo sát của chúng tôi dựa trên giả thiết số nguồn bằng với số trộn (cảm biến) và mô hình ICA chuẩn nên chỉ là cơ sở bước đầu để chúng tôi đi vào các mô hình ICA thực tế và phức tạp hơn như số nguồn khác với số trộn, nguồn có nhiễu, sự lai trộn có trễ, lai trộn chập . . . mà đã có rất nhiều công trình nghiên cứu đã công bố, tuy nhiên vẫn có những nghiên cứu mới về lý luận, sự mô phỏng và ứng dụng.

## SOUND SEPARATION USING THE METHOD OF INDEPENDENT COMPONENT ANALYSIS

**Truong Tan Quang, Nguyen Huu Phuong**  
University of National Sciences, VNU-HCM

**ABSTRACT:** *Our ears often simultaneously receive various sound sources (speech, music, noise . . .), but we can still listen to the intended sound. A system of speech processing or recognizing must achieve the same intelligent level. The problem is that we receive many mixed (combined) signals from many different source signals, and would like to recover the source signals separately.*

*In the last decade or so a method has been developed to help solve the above problem of Blind Source Separation (BSS) effectively, that is the Independent Component Analysis (ICA). This paper gives an brief overview of ICA and our application to sound separation. The results are very good. Henceforth, we will be able to tackle more complex source separation problems using ICA.*

## TÀI LIỆU THAM KHẢO

- [1]. A.Hyvarinen and E.Oja, *Independent Component Analysis: Algorithms and Application*, Neural Networks, 13(4-5): 411-430, 2000.
- [2]. A.Hyvarinen, *Survey on independent component analysis*, Neural Computing Surveys, (2): 94-128, 2000.
- [3]. Hyvarinen, *A. Fast and robust fixed-point algorithm for Independent Component Analysis*, IEEE Transactions on Neural Networks, 2001.
- [4]. A.Hyvarinen, J.Karhunen, and E.Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

- [5]. A.Jelmer and R.Mika, *A Guide To Independent Component Analysis –Theory And Practice*, Report A No 23, March 2004.
- [6]. Ricky Der., *Blind Signal Separation*, Department of Electrical and Computer Engineering McGill University, 2001.
- [7]. Alexander George Westner, *Object –Based Audio Capture : Separating Acoustically-Mixed Sound*. Rutgers University, 1996.