(12) **BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0022711**

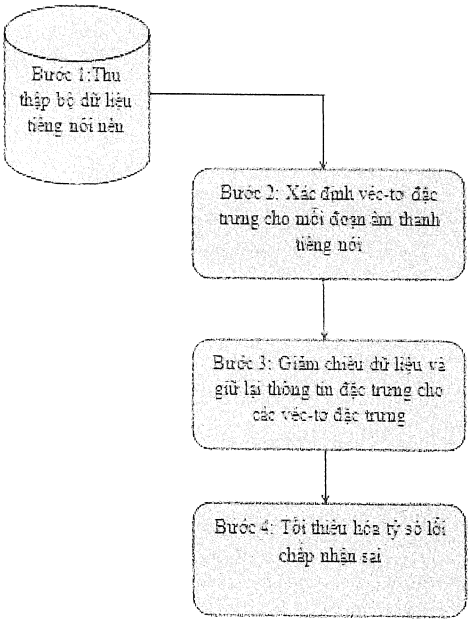
(51)[7] **G10L 15/00**, 17/00

(13) **B**

(21) 1-2017-04750
(22) 27.11.2017
(45) 27.01.2020 382
(43) 26.02.2018 359
(73) TẬP ĐOÀN VIỄN THÔNG QUÂN ĐỘI (VN)
Số 1 đường Trần Hữu Dực, phường Mỹ Đình 2, quận Nam Từ Liêm, thành phố Hà Nội.
(72) Nguyễn Văn Tuấn (VN), Đỗ Ngọc Tuấn (VN), Chu Văn Tạo (VN), Nguyễn Anh Tuấn (VN), Nguyễn Quang Bằng (VN)
(74) Công ty TNHH Tư vấn Quốc Dân (NACI CO., LTD)

# (54) PHƯƠNG PHÁP CHUẨN HÓA VÉC-TƠ ĐẶC TRƯNG TIẾNG NÓI

(57) Phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói dựa trên sự kết hợp giữa phép phân tích phân loại tuyến tính và phép chuẩn hóa theo hiệp phương sai nhằm cải thiện độ tin cậy cho hệ thống nhận dạng người nói đối với dữ liệu tiếng nói được thu nhận từ các hệ thống trinh sát điện tử thông tin liên lạc. Phương pháp này loại bỏ các thành phần ảnh hưởng tới sự biến đổi về tiếng nói, đồng thời giữ lại các thành phần véc-tơ đặc trưng cần thiết cho mỗi người nói cũng như tối thiểu hóa lỗi phân loại; phương pháp này bao gồm bốn bước, cụ thể: bước 1: thu nhận bộ dữ liệu tiếng nói nền; bước 2: xác định véc-tơ đặc trưng tiếng nói; bước 3: giảm chiều dữ liệu và giữ lại thông tin đặc trưng cho các véc-tơ đặc trưng; bước 4: tối thiểu hóa tỷ số lỗi chấp nhận sai.



**Lĩnh vực kỹ thuật được đề cập**

Sáng chế đề cập đến phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói. Cụ thể, phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói được đề cập đến trong sáng chế dựa trên sự kết hợp giữa phép phân tích phân loại tuyến tính và phép chuẩn hóa theo hiệp phương sai. Sáng chế được sử dụng trong bước tiền xử lý véc-tơ đặc trưng tiếng nói trước khi được huấn luyện sử dụng các phương pháp phân loại học máy trong hệ thống nhận dạng người nói.

**Tình trạng kỹ thuật của sáng chế**

Hệ thống nhận dạng người nói với mục đích xác định định danh người nói cho một đoạn âm thanh tiếng nói từ một tập hợp các người nói khác nhau đã được lưu trữ. Hệ thống nhận dạng người nói được áp dụng trong nhiều ứng dụng dân sự và quân sự nhằm xác thực danh tính, đảm bảo an ninh, an toàn.

Trong các hệ thống trinh sát điện tử thông tin liên lạc, để xác định định danh người nói của một đoạn âm thanh tiếng nói bất kỳ, cần phải nghe bằng tai một cách thủ công hoặc một số phương pháp nhận dạng người nói truyền thống như mô hình lượng tử hóa véc-tơ, mô hình đa thành phần Gau-xơ,… Nhược điểm của các phương pháp này là không tính đến sự biến đổi tiếng nói do các điều kiện thu nhận khác nhau như kiểu điều chế, tần số,… của kênh truyền và tình trạng sức khỏe, độ tuổi,…của người nói gây ra. Chính vì vậy, các hệ thống hiện tại không đảm bảo độ tin cậy do sự đa dạng về các kênh thu nhận thông tin liên lạc và số lượng người nói cần theo dõi lớn.

Do đó, nhóm tác giả đề xuất một phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói kết hợp phép phân tích phân loại tuyến tính và phép chuẩn hóa theo hiệp phương sai phù hợp với điều kiện hoạt động thực tế giúp giải quyết bài toán tiền xử lý véc-tơ đặc trưng nhằm mục đích giảm số chiều và tăng sự phân biệt của các véc-tơ đặc trưng giữa các người nói khác nhau, đồng thời, bù lại sự biến đổi do kênh thu nhận tiếng nói gây ra, hướng tới cải thiện độ tin cậy cho hệ thống nhận dạng người nói.

**Bản chất kỹ thuật của sáng chế**

Mục đích của sáng chế là đề xuất phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói kết hợp phép phân tích phân loại tuyến tính và phép chuẩn hóa hiệp phương sai nhằm cải

thiện độ tin cậy cho hệ thống nhận dạng người nói đối với dữ liệu tiếng nói được thu nhận từ các hệ thống trinh sát điện tử thông tin liên lạc.

Để thực hiện được mục đích này, phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói bao gồm bốn bước, cụ thể: bước 1: thu nhận bộ dữ liệu tiếng nói nền; bước 2: xác định véc-tơ đặc trưng tiếng nói; bước 3: giảm chiều dữ liệu và giữ lại thông tin đặc trưng cho các véc-tơ đặc trưng; bước 4: tối thiểu hóa tỷ số lỗi chấp nhận sai.

Theo phương án sáng chế, bằng cách kết hợp phép phân tích phân loại tuyến tính nhằm loại bỏ các thành phần ảnh hưởng tới sự biến đổi về tiếng nói, đồng thời giữ lại các thành phần véc-tơ đặc trưng cần thiết cho mỗi người nói và phép chuẩn hóa hiệp phương sai thực hiện chuẩn hóa các véc-tơ thu được nhằm mục đích tối thiểu hóa lỗi phân loại trước khi làm đầu vào cho quá trình huấn luyện học máy sử dụng phương pháp học có giám sát máy véc-tơ hỗ trợ - Support Vector Machine (sau đây viết tắt là **SVM**).

**Mô tả vắn tắt các hình vẽ**

Hình 1: Hình vẽ thể hiện các bước của phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói.

**Mô tả chi tiết sáng chế**

Sáng chế đề xuất phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói kết hợp phép phân tích phân loại tuyến tính và phép chuẩn hóa hiệp phương sai cho các véc-tơ đặc trưng tương ứng cho mỗi đoạn âm thanh tiếng nói, các bước và trình tự thực hiện của phương pháp kết hợp này được mô tả cụ thể trong hình 1. Chi tiết của từng bước của phương pháp kết hợp này sẽ được trình bày chi tiết cụ thể dưới đây:

Bước 1: Thu nhận bộ dữ liệu tiếng nói nền.

Một trong những yếu tố ảnh hưởng tới độ tin cậy của hệ thống nhận dạng người nói là việc thu thập bộ dữ liệu tiếng nói nền nhằm xây dựng một mô hình tổng quát hóa sự biến đổi cho toàn bộ không gian người nói và các kênh thu nhận tiếng nói khác nhau. Để lựa chọn bộ dữ liệu nền phù hợp cần phải đảm bảo một số ràng buộc sau:

Thứ nhất, bộ dữ liệu tiếng nói nền phải có thời lượng tiếng nói được thu nhận và số lượng người nói khác nhau đủ lớn, thông thường, đối với mỗi người nói khác nhau trong bộ dữ liệu tiếng nói nền, thời lượng thu thập tiếng nói khoảng 1-2 phút, với người nói cần đảm bảo sự đa dạng về đặc điểm giọng nói (vùng miền, giới tính) và số lượng người nói trong bộ dữ liệu nền càng lớn càng tốt (thường lớn hơn 300 người).

Thứ hai, với mỗi người nói, các đoạn âm thanh tiếng nói được thu nhận thông qua một số kênh khác nhau (micro, kênh thoại, kiểu điều chế,…) nhằm phản ánh sự biến đổi về kênh thu nhận ảnh hưởng tới độ tin cậy nhận dạng người nói. Trong trường hợp, hệ thống nhận dạng người nói không phụ thuộc vào giới tính người nói hay loại ngôn ngữ, thì bộ dữ liệu nền phải cân bằng về thời lượng tiếng nói giữa các yếu tố này.

Trong thực tế sử dụng, bộ dữ liệu tiếng nói nền được xây dựng thông qua quá trình thu nhận tín hiệu trinh sát từ hệ thống trinh sát điện tử thông tin liên lạc sóng ngắn và sóng cực ngắn. Với mỗi mục tiêu thông tin liên lạc, các tập tin âm thanh tiếng nói được thu nhận dưới nhiều điều kiện khác nhau (tần số, kiểu điều chế,…) và số lượng người nói thuộc các mục tiêu thông tin liên lạc cần theo dõi rất lớn (khoảng 30-40 người nói khác nhau trong một ngày).

Bước 2: Xác định véc-tơ đặc trưng cho mỗi đoạn âm thanh tiếng nói.

Để xây dựng tập véc-tơ đặc trưng cho mỗi đoạn âm thanh tiếng nói, một không gian véc-tơ tổng hợp sự biến đổi được sử dụng nhằm đánh giá sự biến đổi giữa các người nói và giữa các kênh thu nhận tiếng nói. Việc xác định không gian véc-tơ tổng hợp sự biến đổi so với việc sử dụng không gian siêu véc-tơ trong mô hình đa thành phần Gau-xơ, hay mô hình phân tích các yếu tố ảnh hưởng tới giọng nói giúp giảm số chiều cho mỗi véc-tơ đặc trưng, qua đó giảm chi phí về tính toán. Ngoài ra, so với mô hình phân tích các yếu tố ảnh hưởng tới giọng nói là người nói và kênh thu nhận, không gian véc-tơ tổng hợp sự biến đổi chỉ xác định một không gian véc-tơ đặc trưng duy nhất thể hiện cho sự biến đổi về người nói lẫn các kênh thu thập khác nhau thay vì tương ứng với mỗi yếu tố định nghĩa một không gian véc-tơ riêng biệt. Việc định nghĩa một không gian véc-tơ đặc trưng duy nhất vẫn đảm bảo các đặc trưng về người nói và kênh thu thập tiếng nói là riêng biệt nhưng chi phí tính toán được tiết kiệm đáng kể.

Mô hình tổng hợp sự biến đổi xây dựng một ma trận $\boldsymbol{T}$ thể hiện sự biến đổi tổng quát theo hai yếu tố: (1) sự biến đổi đặc trưng giọng nói giữa các người nói khác nhau; và (2) sự biến đổi của giọng nói được thu nhận từ các kênh khác nhau của cùng một người nói.

Ma trận này chứa các véc-tơ riêng có trị riêng lớn nhất của ma trận hiệp phương sai biến đổi tổng hợp của kênh thu nhận và người nói khác nhau.

Véc-tơ đặc trưng cho đoạn âm thanh $u$ trong không gian véc-tơ biến đổi tổng hợp được xác định theo công thức sau:

$$\boldsymbol{w = (I + T^t \Sigma^{-1} N(u) T)^{-1} . T^t \Sigma^{-1} \widetilde{F}(u)}$$

Trong đó, $N(u)$ và $\tilde{F}(u)$ là các đại lượng có được qua phương pháp thống kê Baum-Welch cho đoạn âm thanh $u$. $\Sigma$ là ma trận hiệp phương sai thu được qua quá trình huấn luyện phân tích các yếu tố kênh thu nhận và người nói.

Bước 3: Giảm chiều dữ liệu và giữ lại thông tin đặc trưng cho các véc-tơ đặc trưng.

Qua không gian véc-tơ tổng hợp sự biến đổi, mỗi véc-tơ đặc trưng cho một đoạn âm thanh của một người nói có số chiều dữ liệu thấp hơn so với các siêu véc-tơ đặc trưng của các phương pháp dựa trên thành phần Gau-xơ truyền thống (số chiều dữ liệu tương ứng là 400 và 19456). Tuy nhiên, không phải tất cả các chiều dữ liệu của véc-tơ đặc trưng đều có ý nghĩa trong quá trình phân loại. Do đó, nhằm cải thiện độ tin cậy phân loại cho các véc-tơ đặc trưng, có thể sử dụng các phương pháp giảm chiều dữ liệu nhưng vẫn giữ lại các thông tin đặc trưng riêng biệt cho mỗi lớp dữ liệu.

Để giảm chiều dữ liệu và giữ lại thông tin đặc trưng, phương pháp phân tích phân loại tuyến tính được sử dụng. Số chiều của dữ liệu mới sau phép phân tích là nhỏ hơn hoặc bằng $C - 1$, trong đó $C$ là số lượng lớp dữ liệu ban đầu. Để thực hiện giảm chiều dữ liệu, phép phân tích sử dụng một ma trận chiếu, sao cho véc-tơ đặc trưng được giảm chiều chỉ giữ lại thông tin đặc trưng cho mỗi lớp dữ liệu, khiến nó không bị lẫn với các lớp dữ liệu khác.

Gọi $D$ là tập hợp chứa tất cả các véc-tơ đặc trưng được trích chọn từ bộ dữ liệu nền, mỗi véc-tơ đặc trưng cho một đoạn âm thanh của một người nói, $w_{s,i}$ là véc-tơ đặc trưng của đoạn âm thanh thứ $i$ của người nói $s$, $n_s$ là số đoạn âm thanh tiếng nói của người nói $s$. Và ký hiệu $S$ là tổng số người nói khác nhau trong tập $D$. Ma trận hiệp phương sai thể hiện phương sai của các véc-tơ đặc trưng của mỗi người nói và ma trận hiệp phương sai thể hiện phương sai giữa các người nói được xác định như sau:

$$S_w = \sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} (w_{s,i} - \bar{w}_s)(w_{s,i} - \bar{w}_s)^T$$

$$S_b = \sum_{s=1}^{S} (\bar{w}_s - \bar{w})(\bar{w}_s - \bar{w})^T$$

Với trung bình của các véc-tơ đặc trưng của mỗi người nói và trung bình của tất cả các véc-tơ đặc trưng trong bộ dữ liệu nền được định nghĩa bởi công thức:

$$w_s = \frac{1}{n_s}\sum_{i=1}^{n_s} w_{s,i}$$

$$\bar{w} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s} w_{s,i}$$

Phép phân tích phân loại tuyến tính cần xác định một phép chiếu sao cho tỷ lệ giữa $S_b$ và $S_w$ là lớn nhất. Giá trị lớn nhất của tỷ lệ này được xác định thông qua đi tìm các trị riêng của biểu thức sau:

$$S_b v = \Lambda S_w v$$

Với $\Lambda$ là ma trận đường chéo chứa các trị riêng. Nếu ma trận $S_w$ là khả nghịch, thì $v$ phải là một véc-tơ riêng của $S_w^{-1}S_b$ ứng với một trị riêng nào đó. Vậy, để tỷ lệ giữa $S_b$ và $S_w$ là lớn nhất thì các véc-tơ riêng $v$ ứng với các trị riêng lớn nhất của $S_w^{-1}S_b$. Gọi ma trận chiếu $A_P$ có mỗi cột ứng với các véc-tơ riêng, ta có, phép biến đổi của véc-tơ đặc trưng được xác định là:

$$\Phi(w) = A_P^T w$$

Bước 4: Tối thiểu hóa tỷ số lỗi chấp nhận sai.

Phép chuẩn hóa theo hiệp phương sai được xây dựng với mục đích tối thiểu hóa tỷ số lỗi chấp nhận sai và từ chối sai trong quá trình huấn luyện SVM. Để tối thiểu hóa tỷ số lỗi này, một tập các cận trên được định nghĩa tương ứng với chỉ số lỗi phân loại. Lời giải tối ưu cho bài toán này là tối thiểu hóa các cận trên hay cũng chính là tối thiểu hóa số lỗi phân loại. Việc tối ưu cho bài toán này, dẫn đến việc định nghĩa một hàm kernel mới cho phương pháp SVM như sau:

$$k(w_1, w_2) = w_1^T R w_2$$

Với ma trận chuẩn hóa $R$ là ma trận đối xứng và nửa xác định dương. Ma trận chuẩn hóa được xác định bởi $R = W^{-1}$, với $W$ là ma trận hiệp phương sai thể hiện phương sai của các véc-tơ đặc trưng sau phép phân tích phân loại tuyến tính của mỗi người nói. Ma trận hiệp phương sai được tính theo công thức:

$$W = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s}\left(w_{s,i}^P - \bar{w}_s^P\right)\left(w_{s,i}^P - \bar{w}_s^P\right)^T$$

Với $\bar{w}_s^P = (1/n_s)\sum_{i=1}^{n_s} w_{s,i}^P$, là trung bình của các véc-tơ đặc trưng sau phép chiếu dựa trên phép phân tích phân loại tuyến tính của mỗi người nói, $S$ là tổng số người nói khác nhau trong tập dữ liệu nền, và $n_s$ là số đoạn âm thanh tiếng nói của người nói $s$. Một hàm ánh xạ đặc trưng $\varphi$ có thể được định nghĩa là:

$$\varphi(w) = B^T w$$

Với ma trận $B$ được tính theo phép phân tích Cholesky của ma trận $W^{-1} = BB^T$. Phép chuẩn hóa hiệp phương sai sử dụng ma trận hiệp phương sai của mỗi người nói để chuẩn hóa hàm kernel cosine nhằm bù cho sự biến đổi giữa các đoạn âm thanh khác nhau của cùng một người nói, nhưng vẫn đảm bảo giữ được các đặc trưng phân biệt giữa các người nói với nhau.

Sáng chế được mô tả chi tiết bằng cách sử dụng các phương án được mô tả ở trên. Tuy nhiên, rõ ràng là đối với người hiểu biết trung bình trong lĩnh vực sáng chế không bị giới hạn ở phương án được mô tả trong phần mô tả sáng chế. Sáng chế có thể được thực hiện ở chế độ cải biến hoặc thay đổi mà không nằm ngoài phạm vi sáng chế được xác định bởi các điểm yêu cầu bảo hộ. Vì vậy những gì được mô tả trong phần mô tả sáng chế chỉ nhằm mục đích minh hoạ, và sẽ không áp đặt bất kỳ giới hạn nào đối với sáng chế.

## Yêu cầu bảo hộ

1. Phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói bao gồm bốn bước, cụ thể:

bước 1: thu nhận bộ dữ liệu tiếng nói nền; bộ dữ liệu tiếng nói nền được thu nhận trong bước này phải có thời lượng tiếng nói được thu nhận và số lượng người nói khác nhau đủ lớn;

bước 2: xác định véc-tơ đặc trưng tiếng nói; mỗi véc-tơ đặc trưng được xác định tương ứng cho một đoạn âm thanh tiếng nói riêng biệt trong bộ dữ liệu tiếng nói nền nhờ sử dụng mô hình tổng hợp sự biến đổi của kênh thu nhận và người nói;

bước 3: giảm chiều dữ liệu và giữ lại thông tin đặc trưng cho các véc-tơ đặc trưng; bước này được thực hiện nhờ sử dụng phép phân tích phân loại tuyến tính;

bước 4: tối thiểu hóa tỷ số lỗi chấp nhận sai; bước này được thực hiện nhờ sử dụng phép chuẩn hóa theo hiệp phương sai.

2. Phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói theo điểm 1, trong đó:

tại bước giảm chiều dữ liệu và giữ lại thông tin đặc trưng cho các véc-tơ đặc trưng được thực hiện như sau:

gọi $\boldsymbol{D}$ là tập hợp chứa tất cả các véc-tơ đặc trưng được trích chọn từ bộ dữ liệu nền, mỗi véc-tơ đặc trưng cho một đoạn âm thanh của một người nói, $\boldsymbol{w_{s,i}}$ là véc-tơ đặc trưng của đoạn âm thanh thứ $\boldsymbol{i}$ của người nói $\boldsymbol{s}$, $\boldsymbol{n_s}$ là số đoạn âm thanh tiếng nói của người nói $\boldsymbol{s}$; và ký hiệu $\boldsymbol{S}$ là tổng số người nói khác nhau trong tập $\boldsymbol{D}$; ma trận hiệp phương sai thể hiện phương sai của các véc-tơ đặc trưng của mỗi người nói ($S_w$) và ma trận hiệp phương sai thể hiện phương sai giữa các người nói ($S_b$) được xác định như sau:

$$\boldsymbol{S_w} = \sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_s} (w_{s,i} - \bar{w}_s)(w_{s,i} - \bar{w}_s)^T$$

$$\boldsymbol{S_b} = \sum_{s=1}^{S} (\bar{w}_s - \bar{w})(\bar{w}_s - \bar{w})^T$$

với trung bình của các véc-tơ đặc trưng của mỗi người nói và trung bình của tất cả các véc-tơ đặc trưng trong bộ dữ liệu nền được định nghĩa bởi công thức:

$$\boldsymbol{w_s} = \frac{1}{n_s} \sum_{i=1}^{n_s} w_{s,i}$$

$$\bar{w} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s} w_{s,i}$$

phép phân tích phân loại tuyến tính cần xác định một phép chiếu sao cho tỷ lệ giữa $S_b$ và $S_w$ là lớn nhất; phép phân tích phân loại tuyến tính là phương pháp đi tìm giá trị lớn nhất cho tỷ lệ giữa $S_b$ và $S_w$; mục đích cuối cùng của bước này chính là đi tìm các trị riêng cho biểu thức sau:

$$S_b v = \Lambda S_w v$$

với $\Lambda$ là ma trận đường chéo chứa các trị riêng, nếu ma trận $S_w$ là khả nghịch, thì $v$ phải là một véc-tơ riêng của $S_w^{-1}S_b$ ứng với một trị riêng nào đó; vậy để tỷ lệ giữa $S_b$ và $S_w$ là lớn nhất thì các véc-tơ riêng $v$ ứng với các trị riêng lớn nhất của $S_w^{-1}S_b$; gọi ma trận chiếu $A_P$ có mỗi cột ứng với các véc-tơ riêng, ta có, phép biến đổi của véc-tơ đặc trưng được xác định là:

$$\Phi(w) = A_P^T w \ .$$

3. Phương pháp chuẩn hóa véc-tơ đặc trưng tiếng nói theo điểm 1, trong đó:

tại bước tối thiểu hóa tỷ số lỗi chấp nhận sai, bước này được thực hiện như sau:

một tập các cận trên được định nghĩa tương ứng với chỉ số lỗi phân loại, dẫn đến việc định nghĩa một hàm Kernel mới cho phương pháp SVM như sau:

$$k(w_1, w_2) = w_1^T R w_2$$

với ma trận chuẩn hóa $R$ là ma trận đối xứng và nửa xác định dương; ma trận chuẩn hóa được xác định bởi $R = W^{-1}$, với $W$ là ma trận hiệp phương sai thể hiện phương sai của các véc-tơ đặc trưng sau phép phân tích phân loại tuyến tính của mỗi người nói; ma trận hiệp phương sai được tính theo công thức:

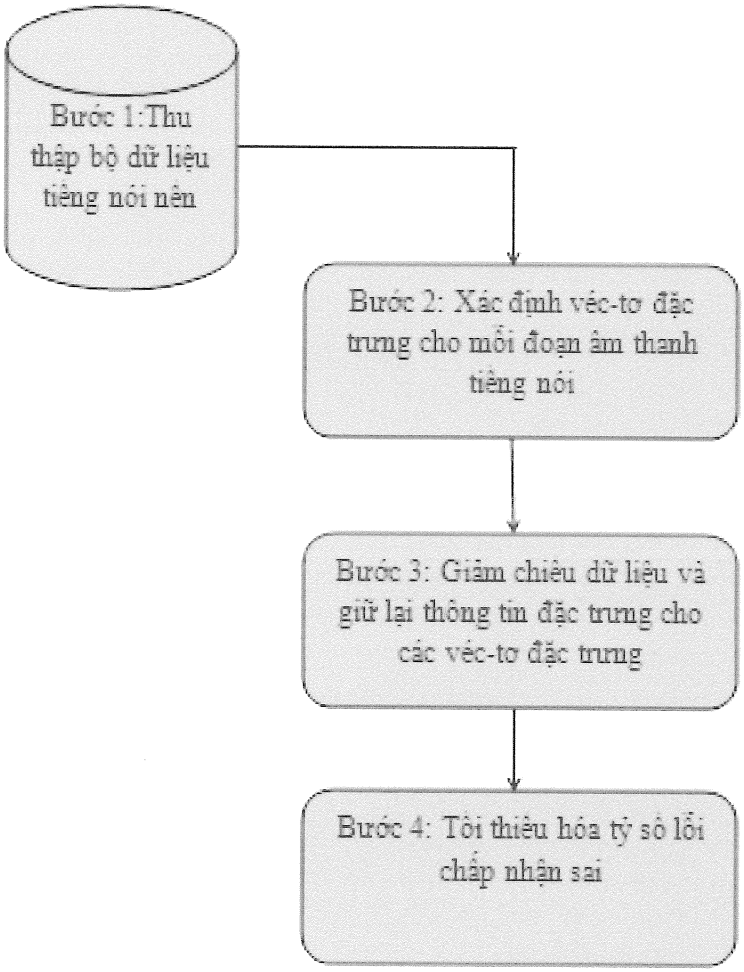
$$W = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{n_s}\sum_{i=1}^{n_s}\left(w_{s,i}^P - \bar{w}_s^P\right)\left(w_{s,i}^P - \bar{w}_s^P\right)^T$$

với $\bar{w}_s^P = (1/n_s)\sum_{i=1}^{n_s} w_{s,i}^P$ là trung bình của các véc-tơ đặc trưng sau phép chiếu dựa trên phép phân tích phân loại tuyến tính của mỗi người nói, $S$ là tổng số người nói khác nhau trong tập dữ liệu này, và $n_s$ là số đoạn âm thanh tiếng nói của người nói $s$; một hàm ánh xạ đặc trưng $\varphi$ có thể được định nghĩa là:

$$\varphi(w) = B^T w$$

với ma trận $B$ được tính theo phép phân tích Cholesky của ma trận $W^{-1} = BB^T$; phép chuẩn hóa hiệp phương sai sử dụng ma trận hiệp phương sai của mỗi người nói để chuẩn hóa hàm kernel cosin nhằm bù cho sự biến đổi giữa các đoạn âm thanh khác nhau của cùng một người nói, nhưng vẫn đảm bảo giữ được các đặc trưng phân biệt giữa các người nói với nhau.

**Hình vẽ**




Hình 1