Kỷ yếu Hội nghị Quốc gia lần thứ X về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), Đà Nẵng, ngày 17-18/08/2017
DOI: 10.15625/vap.2017.00051

MÔ HÌNH GMM ĐỊNH DANH TỰ ĐỘNG MỘT SỐ LÀN ĐIỆU DÂN CA QUAN HỌ BẮC NINH

Chu Bá Thành[1,2], Trịnh Văn Loan[1,2], Nguyễn Hồng Quang[2]

[1] Khoa Công nghệ Thông tin - Đại học Sư phạm Kỹ thuật Hưng Yên

[2] Viện Công nghệ thông tin & Truyền thông - Đại học Bách khoa Hà Nội

ThanhCB.Fit@utehy.edu.vn, LoanTV@soict.hust.edu.vn, QuangNH@soict.hust.edu.vn

TÓM TẮT: *Việc phân loại tự động âm nhạc theo thể loại có tầm quan trọng rất lớn trong việc tự động hoá quá trình lưu trữ, sắp xếp, tìm kiếm nguồn thông tin khổng lồ về âm nhạc. Việt Nam vốn có các làn điệu dân ca rất phong phú cho cả ba miền Bắc - Trung - Nam, trong đó phải kể đến dân ca Quan họ Bắc Ninh. Bài báo trình bày một phương pháp định danh một số làn điệu dân ca Quan họ Bắc Ninh bằng cách sử dụng mô hình GMM (Gaussian Mixture Model) với các tham số của mô hình bao gồm các hệ số MFCC (Mel Frequency Cepstral Coefficients), năng lượng và tần số cơ bản F0. Kết quả cho thấy tỷ lệ định danh chính xác các làn điệu dân ca Quan họ Bắc Ninh dùng trong thử nghiệm phụ thuộc vào số thành phần hỗn hợp của GMM. Việc sử dụng thông tin về tần số cơ bản đã làm tăng đáng kể tỷ lệ định danh chính xác.*

Từ khóa*: Định danh, âm nhạc, Quan họ Bắc Ninh, GMM, MFCC, F0.*

I. GIỚI THIỆU

Nguồn thông tin khổng lồ về âm nhạc hiện nay đòi hỏi cần phải tự động hoá quá trình lưu trữ, sắp xếp, tìm kiếm các thể loại âm nhạc. Việc tự động hoá này có thể được thực hiện dựa trên xử lý tín hiệu âm thanh của các bản nhạc, bài ca tương ứng. Đã có nhiều hướng nghiên cứu theo học máy và kỹ thuật phân tích thống kê được sử dụng để định danh tự động các thể loại âm nhạc trên thế giới [1], [2]. Đối với kho tàng dân ca Việt Nam, việc tự động hoá quá trình định danh các làn điệu dân ca dựa trên xử lý tín hiệu hầu như đang còn ở giai đoạn khởi đầu [3]. Trong số các làn điệu dân ca Việt Nam, dân ca Quan họ Bắc Ninh đã được UNESCO công nhận là di sản văn hoá phi vật thể của nhân loại, có số lượng rất phong phú cả về làn điệu lẫn phương pháp thể hiện. Các phần tiếp theo của bài báo được tổ chức như sau. Phần II trình bày khái quát về các hướng và kết quả nghiên cứu đã đạt được về định danh các thể loại âm nhạc trên thế giới. Thông tin về bộ dữ liệu các làn điệu dân ca Quan họ Bắc Ninh và mô hình GMM dùng cho nghiên cứu định danh sẽ được mô tả trong phần III. Phần IV là kết quả thử nghiệm dùng mô hình GMM với các tham số MFCC, năng lượng, tần số cơ bản để định danh một số làn điệu dân ca Quan họ Bắc Ninh. Cuối cùng, kết luận và hướng nghiên cứu tiếp theo được đưa ra trong phần V.

II. KHÁI QUÁT VỀ ĐỊNH DANH TỰ ĐỘNG CÁC THỂ LOẠI ÂM NHẠC

Việc định danh âm nhạc có thể được thực hiện theo các nhiệm vụ sau: phân lớp thể loại (genre), phân lớp lối thức (mood), định danh nghệ sĩ, nhận dạng nhạc cụ, chú giải âm nhạc [4]. Có nhiều bộ phân lớp khác nhau đã được sử dụng để thực hiện các nhiệm vụ này như: KNN (K-Nearest Neighbors) [5], [27], SVM (Support Vector Machine) [4], [6], GMM [7], ANN (Artificial Neural Networks) [8], [9], [10], [11], LDA (Linear Discriminant Analysis)[12], [13], [14], SRC (Sparse Representation-based Classifier) [15], [16]. Các đặc trưng thường được sử dụng cho các bộ phân lớp này bao gồm các đặc trưng trong miền tần số và các đặc trưng trong miền thời gian. Có thể liệt kê các đặc trưng trong miền tần số như: tỷ lệ biến thiên qua trục không (ZCR), trọng tâm phổ (SC), độ rộng dải tần phổ (SB), đường bao phổ biên độ (ASE), các hệ số sóng con (DWCH), các hệ số MFCC, các hệ số Cepstrum Fourier, các hệ số Cepstrum tiên đoán tuyến tính,… Các đặc trưng trong miền thời gian có thể bao gồm: các mômen thống kê (SM), thông tin điều chế biên độ (AM), các thông tin của mô hình tự hồi quy (ARM),…[4].

Đối với bộ dữ liệu GTZAN [17] gồm có 10 thể loại âm nhạc, mỗi thể loại có 100 đoạn âm thanh và độ dài mỗi đoạn là 30s, độ chính xác phân loại cao nhất đạt được là 92,4% khi sử dụng bộ phân lớp SRC. Cũng với bộ dữ liệu này, độ chính xác phân loại thấp nhất là 60% với bộ phân lớp K-NN, còn nếu dùng bộ phân lớp GMM với các tham số đặc trưng STFT (Short Time Fourier Transform) + MFCC + nhịp + cao độ (pitch) thì độ chính xác đạt được là 61%. Kết quả nghiên cứu trong [18] cho thấy, cùng với bộ dữ liệu GTZAN độ chính xác đạt được cao nhất là 87,4% với bộ phân lớp SVM RBF Kernel.

Đối với bộ dữ liệu âm nhạc Indian Tamil [19] gồm có 216 trích đoạn bài hát với độ dài mỗi đoạn là 30s và tập tham số đặc trưng là MFCC + Skewness + Kurtosis + Flux + Spectral Roll-off, độ chính xác đạt được với bộ phân lớp KNN là 66,23%. Với bộ phân lớp SVM, độ chính xác đạt được là 84,21%.

III. DỮ LIỆU DÙNG CHO THỬ NGHIỆM VÀ MÔ HÌNH GMM

A. Bộ dữ liệu dân ca Quan họ Bắc Ninh dùng cho thử nghiệm

Dân ca Việt Nam là một thể loại âm nhạc cổ truyền Việt Nam do chính người dân lao động sáng tác trong quá trình lao động, sinh hoạt đời thường và được truyền miệng qua nhiều thế hệ nên thường có nhiều dị bản và phần lớn

đều không rõ tác giả. Những làn điệu dân ca được sáng tác ở khắp các vùng miền nên rất đa dạng về giai điệu và phong phú về nội dung. Dân ca gồm nhiều thể thức hát như: hát ru, hát quan họ, hát xẩm, các điệu hò, điệu lý, đồng dao, nói thơ, ngâm thơ hay cả hát sắc bùa, điệu bóng rỗi,… Nét chung nhất của các bài dân ca (ở cả 3 miền Bắc - Trung - Nam) đều thể hiện sự dân dã, mộc mạc, mang âm hưởng tình cảm nhẹ nhàng, nội dung chứa đựng những tâm tư, ước muốn hoặc phản ánh đời sống lao động của người nông dân, ngư dân; tôn vinh những giá trị tình cảm cao đẹp của con người như: lòng thủy chung, hiếu thảo, tình yêu quê hương, thể hiện tình cảm giữa người với người. Để phân định và gọi theo vùng miền hay từng tỉnh, người ta phân định bằng "ca từ", bằng "âm giọng", bằng cách "nhấn nhá", "luyến láy", "ngân nga", "rê giọng",… mà theo đặc tính chỉ vùng miền hay tỉnh đó mới có thể hát được.

Trong kho tàng dân ca Việt Nam, Quan họ Bắc Ninh là thể loại dân ca phong phú nhất về mặt giai điệu. Theo thống kê trong [20], có 213 làn điệu dân ca Quan họ được chia chủ yếu thành 3 giọng: giọng Lề lối, giọng Giã bạn và giọng Vặt. Giọng Lề lối gồm 10 làn điệu và giọng Giã bạn gồm 20 làn điệu. Giọng Vặt chiếm số lượng lớn nhất với 183 làn điệu. Chi tiết hơn về các làn điệu dân ca Quan họ Bắc Ninh và dữ liệu dân ca Quan họ Bắc Ninh được trình bày trong [3].

Dữ liệu dùng cho thử nghiệm là 10 làn điệu Quan họ thuộc giọng Vặt (là giọng phổ biến và nguồn ngữ liệu hiện có phong phú hơn so với 2 giọng còn lại). Mỗi làn điệu gồm 10 file ghi âm với lời khác nhau và do 20 nghệ sỹ thực hiện trong đó có 8 giọng nam, 12 giọng nữ. Toàn bộ có 100 file ghi âm với thời lượng khác nhau và tổng thời lượng khoảng 100 phút. Các file đều có tần số lấy mẫu 16kHz, 16 bit cho một mẫu.

Thực nghiệm phân loại tự động được thực hiện theo phương pháp đánh giá chéo trong đó 80% dữ liệu được dùng cho huấn luyện, 20% dữ liệu còn lại được dùng cho nhận dạng. Dữ liệu dùng cho huấn luyện và nhận dạng là độc lập với nhau. Kết quả nhận dạng cuối cùng là trung bình của 5 lần thử nghiệm.

B. Mô hình GMM

Mô hình GMM đã được sử dụng nhiều trong các nghiên cứu về nhận dạng người nói, nhận dạng ngôn ngữ, định danh phương ngữ, định danh thể loại âm nhạc,…[21]. Mô hình GMM với phân bố Gauss hỗn hợp có thể được xem là xếp chồng tuyến tính của các phân bố Gauss như sau [22]:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathsf{N}(\mathbf{x} \mid \mu_k, \Sigma_k) \tag{3.1}$$

Khi sử dụng mô hình Gauss hỗn hợp để phân lớp dữ liệu, **x** trong (3.1) là vector dữ liệu chứa tập các vector tham số đặc trưng của đối tượng cần biểu diễn, trong đó mỗi phần tử của tập có kích thước D. $\pi_k, k = 1..K$ là các trọng số của hỗn hợp thoả mãn điều kiện $\sum_{k=1}^{K} \pi_k = 1$. Mỗi hàm mật độ Gauss $\mathsf{N}(\mathbf{x} \mid \mu_k, \Sigma_k)$ là một thành phần của hỗn hợp có trung bình là μ_k và hiệp phương sai là Σ_k.

$$\mathsf{N}(\mathbf{x} \mid \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_k|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu_k)^{\mathrm{T}} \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right\} \tag{3.2}$$

Mô hình GMM đầy đủ được mô tả bởi bộ 3 tham số $\lambda = \{\pi_k, \mu_k, \Sigma_k\}, k = 1..K.$

Để định danh một làn điệu dân ca đã được mô hình hoá bởi λ, cần xác định khả hiện (likelihood)

$$p(\mathbf{x}, \lambda) = \prod_{n=1}^{N} p(\mathbf{x}_n \mid \lambda) \tag{3.3}$$

Với *N* là số lượng vector đặc trưng và cũng là số lượng khung của file âm thanh cho một làn điệu nào đó.

Trên thực tế, λ là hàm phi tuyến nên cần dùng thuật giải EM (Expectation Maximization) [22] để xác định λ sao cho $\log p(\mathbf{x} \mid \lambda)$ đạt cực đại.

IV. KẾT QUẢ THỬ NGHIỆM

Các thử nghiệm trong bài báo được thực hiện với mô hình GMM cài đặt trong bộ công cụ ALIZE [23], [24] và các tham số được sử dụng trong mô hình đã được tính toán với các bộ công cụ SPro [25] và Praat [26].

A. Bộ tham số dùng cho thử nghiệm

Dữ liệu dùng cho huấn luyện và thử nghiệm được xử lý, trích chọn đặc trưng để có 2 bộ tham số. Bộ tham số thứ nhất gồm 60 hệ số (19 MFCCs + năng lượng = 20, đạo hàm bậc nhất và đạo hàm bậc hai của 20 hệ số này). Số hệ số MFCC như vậy cũng chính là số lượng baseline của mô hình GMM/UBM của bộ công cụ ALIZE [28]. Bộ tham số

còn lại gồm 60 hệ số trong bộ tham số thứ nhất + 1 hệ số là tần số cơ bản F0. Cả hai bộ tham số này đã được đưa vào mô hình GMM với số thành phần Gauss M thay đổi theo luỹ thừa 2: $M = 2^m, m = 4,5, \dots,13$.

B. Kết quả đạt được

Để thuận tiện trong trình bày kết quả thử nghiệm, các làn điệu dân ca được ký hiệu bằng các chữ cái từ *a* đến *j* như sau:

a - Ăn ở trong rừng
b - Có ai xuôi về
c - Bạn tình ơi
d - Chim khôn đậu ngọn thầu dầu
e - Đêm qua nhớ bạn
f - Lóng lánh
g - Lý thiên thai
h - Nhớ mãi khôn nguôi
i - Ngồi tựa mạn thuyền
j - Se chỉ luồn kim

Thử nghiệm được tiến hành với 10 giá trị của *M*. Sau đây sẽ trích dẫn ma trận sai nhầm cho 2 trường hợp có *M* nhỏ nhất và lớn nhất.

Bảng 1 là ma trận sai nhầm đối với thử nghiệm dùng $M = 16$ cho hai bộ tham số. Trong trường hợp sử dụng bộ tham số thứ nhất, tỷ lệ nhận dạng đúng trung bình đạt 65%. Hai làn điệu *b* và *i* được nhận dạng nhầm lẫn nhau với tỷ lệ cao nhất. Các làn điệu còn lại được nhận dạng nhầm sang *b* với tỷ lệ cao và nhận dạng nhầm sang *i* với tỷ lệ thấp hơn. Có thể giải thích lý do của điều này là hai làn điệu *b* và *i* đều có cùng một nhịp. Hai làn điệu có tỷ lệ nhận dạng đúng cao nhất 90% là *a* và *g*, làn điệu *d* có tỷ lệ nhận dạng đúng thấp nhất là 30%. Trong trường hợp sử dụng bộ tham số thứ 2, tỷ lệ nhận dạng đúng là 62%, giảm 3% so với khi chưa bổ sung tham số F0.

Bảng 1. Ma trận sai nhầm đối với thử nghiệm dùng $M = 16$ với hai bộ tham số

a) Sử dụng bộ tham số thứ nhất

	a	b	c	d	e	f	g	h	i	j
a	**9**	2	2	3	2	2	0	3	0	1
b	1	**6**	1	3	0	2	3	1	5	0
c	0	0	**5**	1	1	1	0	0	1	1
d	2	4	1	**4**	1	1	2	0	1	1
e	2	2	2	3	**4**	3	2	3	0	4
f	2	2	4	1	1	**5**	1	1	0	0
g	0	3	0	0	0	0	**9**	3	3	0
h	1	2	0	0	1	1	4	**8**	0	0
i	0	6	1	2	1	1	4	0	**7**	0
j	1	0	0	2	1	0	2	0	0	**8**

b) Sử dụng bộ tham số thứ hai

	a	b	c	d	e	f	g	h	i	j
a	**7**	2	0	2	2	3	0	2	0	0
b	2	**7**	0	4	1	2	4	2	5	0
c	0	0	**4**	1	1	2	0	0	0	0
d	4	4	2	**4**	2	1	3	0	2	1
e	0	2	2	3	**5**	4	2	0	1	5
f	2	2	2	0	3	**5**	0	2	0	2
g	0	3	0	0	0	1	**8**	2	4	0
h	0	1	0	0	2	1	3	**8**	0	1
i	0	6	0	0	1	2	4	0	**6**	0
j	1	1	0	2	0	0	1	0	0	**8**

Bảng 2 là ma trận sai nhầm đối với thử nghiệm với giá trị lớn nhất của $M = 8192$ cho hai bộ tham số. Tỷ lệ nhận dạng đúng trong trường hợp sử dụng bộ tham số thứ nhất là 61%, có giảm đi so với thử nghiệm với *M* nhỏ nhất cùng bộ tham số. Còn trong trường hợp sử dụng bộ tham số thứ hai, tỷ lệ nhận dạng đúng là 79%, cao hơn 17% so với cùng thử nghiệm với *M* nhỏ nhất. Làn điệu *b* và *i* vẫn có tỷ lệ nhận nhầm sang nhau là lớn nhất.

Bảng 2. Ma trận sai nhầm đối với thử nghiệm dùng $M = 8192$ với hai bộ tham số

a) Sử dụng bộ tham số thứ nhất

	a	b	c	d	e	f	g	h	i	j
a	**6**	4	2	5	2	4	4	3	3	4
b	3	**6**	4	6	6	7	3	4	7	1
c	3	4	**7**	6	5	6	3	3	5	1
d	4	8	5	**8**	4	5	5	3	5	5
e	4	3	4	3	**6**	3	4	2	3	7
f	3	2	5	4	3	**6**	3	5	4	2
g	2	0	3	2	2	2	**4**	3	3	1
h	5	3	5	4	4	5	4	**8**	3	2
i	3	6	6	3	5	4	4	5	**6**	1
j	3	1	2	3	4	2	4	4	3	**4**

b) Sử dụng bộ tham số thứ hai

	a	b	c	d	e	f	g	h	i	j
a	**10**	2	1	5	2	3	0	0	0	2
b	2	**7**	1	4	2	2	4	2	5	0
c	1	0	**8**	1	1	2	0	0	1	2
d	4	6	2	**6**	3	3	3	1	5	4
e	1	3	4	5	**8**	6	1	4	1	5
f	2	2	4	2	3	**7**	0	1	1	1
g	0	2	0	0	0	1	**9**	4	6	0
h	0	1	0	0	2	0	4	**9**	0	0
i	0	6	2	3	1	0	5	0	**7**	0
j	2	2	0	4	3	0	3	0	1	**8**

Đối với các giá trị M còn lại, việc nhận dạng nhầm làn điệu *b* sang *i* và ngược lại cũng vẫn tuân theo quy luật trên tức là, tỷ lệ nhận dạng nhầm sang nhau của hai làn điệu này là lớn nhất.

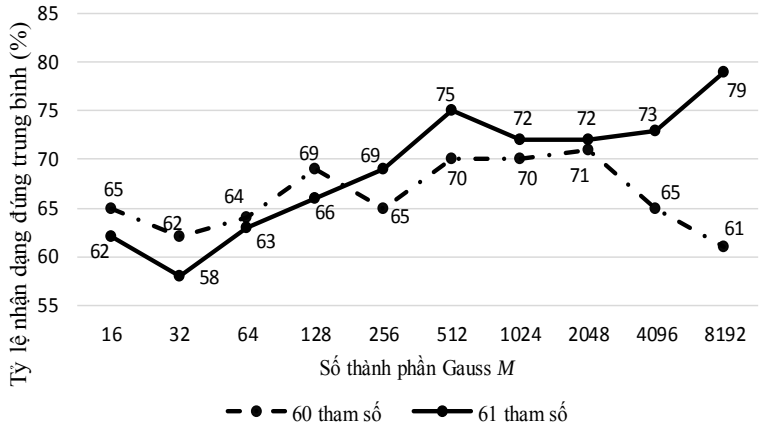
Hình 1 là kết quả thử nghiệm cho 2 bộ tham số với số thành phần Gauss *M* thay đổi theo quy luật $M = 2^m, m = 4,5, \dots, 13$. Với *M* thay đổi từ 16 đến 128, tỷ lệ nhận dạng đúng của bộ tham số thứ nhất cao hơn so với bộ tham số thứ hai. Tuy nhiên, khi *M* thay đổi từ 256 đến 8192 thì tỷ lệ nhận dạng đúng của bộ tham số thứ hai lại cao hơn

so với bộ tham số thứ nhất. Nhìn chung, với bộ 61 hệ số tỷ lệ nhận dạng đúng tăng khi tăng *M*. Với bộ tham số thứ nhất gồm 60 hệ số, tỷ lệ nhận dạng đúng cũng nhìn chung tăng lên theo *M* cho đến $M = 2048$, song tỷ lệ này lại giảm khi *M* tiếp tục tăng từ 4096 đến 8192.

Với bộ tham số thứ hai, khi *M* tăng từ giá trị nhỏ nhất $M = 16$ đến giá trị lớn nhất $M = 8192$, tỷ lệ nhận dạng đúng tăng được 17%. Độ tăng của tỷ lệ nhận dạng đúng trong phạm vi $M = 16$ đến $M = 512$ là lớn và đạt 13%, trong khi đó độ tăng tỷ lệ này lại là nhỏ trong phạm vi $M = 1024$ đến $M = 8192$ và chỉ đạt 4%.

Với bộ tham số thứ nhất, khi *M* tăng từ 16 đến 2048, tỷ lệ nhận dạng đúng tăng 6%. Khi *M* có giá trị lớn nhất *M* = 8192, tỷ lệ nhận dạng lại giảm đi 4% so với trường hợp *M* có giá trị nhỏ nhất $M = 16$. Việc lựa chọn số thành phần Gauss *M* cần phải được cân nhắc tuỳ theo đặc trưng của bộ tham số đưa vào mô hình và độ chính xác nhận dạng cần đạt được. Mặt khác, việc lựa chọn *M* cũng tuỳ thuộc vào giới hạn thời gian tính toán. Nếu *M* càng lớn thì thời gian tính toán càng tăng.

Kết quả thử nghiệm cũng cho thấy ảnh hưởng quan trọng của tham số F0 đến độ chính xác nhận dạng. Khi $M = 8192$, tỷ lệ nhận dạng chính xác tăng lên lớn nhất là 18% nếu sử dụng tham số F0. Ít nhất, tỷ lệ nhận dạng chính xác cũng tăng được 1% khi sử dụng tham số F0 với $M = 2048$.



Hình 1. Kết quả thử nghiệm 2 bộ tham số với số thành phần Gauss $M = 16 \div 8192$

C. So sánh với một số phương pháp khác

Trong một nghiên cứu khác đã được công bố [3], thử nghiệm định danh dùng các phương pháp SMO, MultiLayer Perceptron, MultiClass Classifier đối với một số làn điệu dân ca Quan họ Bắc Ninh đạt tỷ lệ nhận dạng đúng trung bình như trong Bảng 3 [3].

Bảng 3. Tổng hợp kết quả thử nghiệm định danh với SMO, MultiLayer Perceptron và MultiClass Classifier

Phương pháp	**SMO**	**Multilayer Perceptron**	**MultiClass Classifier**
Trung bình tỷ lệ định danh đúng	89%	86%	71%

Bảng 3 cho thấy, trung bình tỷ lệ định danh đúng của hai phương pháp dùng SMO và MultiLayer Perceptron cao hơn so với phương pháp dùng GMM (với *M* = 8192). Tuy nhiên, việc so sánh chỉ mang tính tương đối, do việc sử dụng số lượng tham số trong các phương pháp định danh là khác nhau. Cụ thể:

- Các phương pháp SMO, MultiLayer Perceptron đều sử dụng đầy đủ 384 hệ số, được trích rút từ bộ công cụ OpenSMILE, trong khi phương pháp GMM chỉ sử dụng 61 hệ số.
- Với mô hình GMM, việc trích chọn các tham số được thực hiện theo từng khung, trong khi đó SMO, MultiLayer Perceptron, MultiClass Classifier lấy thống kê 384 hệ số cho toàn bộ mỗi file.

V. KẾT LUẬN

Bài báo đã trình bày các kết quả thử nghiệm sử dụng mô hình GMM để định danh một số làn điệu dân ca Quan họ Bắc Ninh phổ biến. Trong số các làn điệu đã được thử nghiệm, làn điệu *"Ăn ở trong rừng"* đã có tỷ lệ nhận dạng chính xác cao nhất lên đến 100%. Làn điệu *"Chim khôn đậu ngọn thầu dầu"* có tỷ lệ nhận dạng đúng thấp nhất cũng đạt 20%. Tỷ lệ nhận dạng đúng trung bình của các làn điệu đạt được ở mức khả quan so với một số hệ thống phân loại âm nhạc đã được thực hiện. Mặt khác, kết quả thử nghiệm cũng cho thấy ảnh hưởng quan trọng, làm tăng độ chính xác nhận dạng của tần số cơ bản *F0* đối với các làn điệu dân ca Quan họ Bắc Ninh. Hướng tiếp theo của nghiên cứu sẽ là thử nghiệm một số mô hình phân loại âm nhạc cho dân ca Quan họ Bắc Ninh và một số làn điệu dân ca khác trong đó khai thác mô hình học sâu để nhận dạng.

Nhóm tác giả xin chân thành cảm ơn sự hỗ trợ của Trung tâm Nghiên cứu khoa học và Ứng dụng công nghệ - Trường Đại học Sư phạm Kỹ thuật Hưng Yên để hoàn thành bài báo này.

TÀI LIỆU THAM KHẢO

[1] Clark, Sam, Danny Park, and Adrien Guerard. "Music genre classification using machine learning techniques". (2012). http://web.cs.swarthmore.edu/~meeden/cs81/s12/papers/AdrienDannySamPaper.pdf.

[2] Holzapfel, André, and Yannis Stylianou. "A statistical approach to musical genre classification using non-negative matrix factorization". Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on. Vol. 2. IEEE, 2007.

[3] Chu Bá Thành, Trịnh Văn Loan, Nguyễn Hồng Quang. "Định danh tự động một số làn điệu dân ca Việt Nam". Kỷ yếu Hội thảo quốc gia lần thứ XIX: Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông (@), pp.92-97, 2016.

[4] Fu, Zhouyu, et al. "A survey of audio-based music classification and annotation". IEEE transactions on multimedia 13.2 (2011): 303-319.

[5] T. Cover and P. Hart. "Nearest neighbor pattern classification". IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21-27, 1967.

[6] B. E. Boser, I. Guyon, and V. Vapnik. "A training algorithm for optimal margin classifiers". in Proc. ACM Conf. Computational Learning Theory, 1992, pp. 144-152.

[7] R. O. Duda and P. E. Hart, Pattern Classification, 2nd ed. New York: Wiley, 2000.

[8] A. Meng and J. Shawe-Taylor. "An investigation of feature models for music genre classification using the support vector classifier" in Proc. Int. Conf. Music Information Retrieval, 2005.

[9] N. Scaringella and G. Zoia. "On the modelling of time information for automatic genre recognition systems in audio signals" in Proc. Int. Conf. Music Information Retrieval, 2005.

[10] P. Hamel, S. Wood, and D. Eck. "Automatic identification of instrument classes in polyphonic and poly-instrument audio" in Proc. Int. Conf. Music Information Retrieval, 2009.

[11] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman. "A large-scale evaluation of acoustic and subjective music similarity measures" in Proc. Int. Conf. Music Information Retrieval, 2003.

[12] T. Li, M. Ogihara, and Q. Li, "A comparative study of content-based music genre classification" in Proc. SIGIR, 2003.

[13] C.-H. Lin, J.-L. Shih, K.-M. Yu, and H.-S. Lin, "Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features," IEEE Trans. Multimedia, vol. 11, no. 4, pp. 670–682, 2009.

[14] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," EURASIP J. Appl. Signal Process., vol. 2003, no. 1, pp. 5-14, 2003.

[15] I. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: A multilinear approach," in Proc. Int. Conf. Music Information Retrieval, 2008.

[16] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kegl. "Aggregate features and ada boost for music classification". Mach. Learn., vol. 65, no. 2-3, pp. 473-484, 2006.

[17] G. Tzanetakis and P. Cook. "Automatic musical genre classification of audio signals" IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293-302, July 2002.

[18] Lim, Shin-Cheol, et al. "Music-genre classification system based on spectro-temporal features and feature selection". IEEE Transactions on Consumer Electronics 58.4 (2012).

[19] Rajesh, Betsy, and D. G. Bhalke. "Automatic genre classification of Indian Tamil and western music using fractional MFCC". International Journal of Speech Technology 19.3 (2016): 551-563.

[20] Lê Danh Khiêm, Hoắc Công Huynh, Lê Thị Chung, Không gian văn hoá Quan họ, NXB Trung tâm VHTT tỉnh Bắc Ninh, 2006.

[21] Bağcı, Ulaş, and Engin Erzin. "Boosting classifiers for music genre classification." International Symposium on Computer and Information Sciences. Springer Berlin Heidelberg, 2005.

[22] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2013.

[23] Jean-François Bonastre, Frédéric Wils (2005) ALIZE, A FREE TOOLKIT FOR SPEAKER RECOGNITION. IEEE International Conference, pp. I 737 - I 740.

[24] Tommie Gannert (2007) A Speaker Verification System under the Scope: Alize. Stockholm, Sweden School of Computer Science and Engineering.

[25] http://www.irisa.fr/metiss/guig/spro/spro-4.0.1/spro_1.html#SEC1

[26] http://www.fon.hum.uva.nl/praat/downloadwin.html

[27] Scaringella, N.; Zoia, G.; Mlynek, D., "Automatic genre classification of music content: a survey". IEEE Signal Processing Magazine, vol.23, no.2, pp.133,141, March 2006.

[28] Alimohad, Abdennour, Ahmed Bouridane, and Abderrezak Guessoum. "Efficient invariant features for sensor variability compensation in speaker recognition". Sensors 14.10 (2014): 19007-19022.

GMM FOR AUTOMATIC IDENTIFICATION OF SOME QUAN HO BAC NINH FOLK SONGS

Chu Ba Thanh, Trinh Van Loan, Nguyen Hong Quang

ABSTRACT: *Automatic identification of music genre is of great importance in automating the process of storing, organizing and searching for vast amounts of information about music. Vietnam has folk songs that are very rich for all three regions of North - Central - South, including Quan ho Bac Ninh folk songs. This paper presents a method for identifying some of Quan ho Bac Ninh folk songs using the Gaussian Mixture Model (GMM) with model parameters including Mel Frequency Cepstral Coefficients (MFCCs), energy and fundamental frequency F0. The experiment results showed that the exact identification scores depend on the number of mixture components. The use of fundamental frequency information increased considerably the exact indentification score.*