



NHẬN DẠNG PHƯƠNG NGỮ TIẾNG VIỆT TRÊN HỆ THỐNG NHÚNG

Phạm Ngọc Hưng, Nguyễn Thu Hà, Trịnh Văn Loan,
Vũ Xuân Thắng, Phạm Quốc Hùng, Nguyễn Đình Chiến

Trường Đại học Sư phạm Kỹ thuật Hưng Yên

Ngày tòa soạn nhận được bài báo: 19/11/2019

Ngày phản biện đánh giá và sửa chữa: 24/12/2019

Ngày bài báo được chấp nhận đăng: 29/12/2019

Tóm tắt:

Có nhiều yếu tố ảnh hưởng tới nhận dạng tiếng nói trong đó có yếu tố phương ngữ. Tiếng Việt là ngôn ngữ có thanh điệu với nhiều phương ngữ khác nhau. Do vậy, nhận dạng tiếng Việt nói cũng chịu ảnh hưởng đáng kể bởi phương ngữ. Nếu biết được thông tin phương ngữ của tiếng nói trước khi nhận dạng thì hiệu năng hệ thống nhất dạng có thể được cải thiện thông qua việc lựa chọn và sử dụng ngữ liệu phù hợp cho từng phương ngữ khác nhau. Nhận dạng phương ngữ tiếng Việt đã được nghiên cứu và đạt được những kết quả nhất định. Các nghiên cứu thường được thực hiện trên môi trường máy tính có cấu hình mạnh, thuận lợi cho huấn luyện mô hình và tiến hành các thử nghiệm. Các thiết bị nhúng đang được ứng dụng nhiều trong các sản phẩm thực tiễn bao gồm cả nhận dạng tiếng nói và điều khiển thường có cấu hình thấp hơn so với máy tính. Do vậy việc triển khai nhận dạng tiếng nói trên phần cứng đó ít nhiều có những trở ngại. Bài báo này trình bày kết quả nghiên cứu, thử nghiệm nhận dạng phương ngữ tiếng Việt trên hệ thống nhúng sử dụng bộ ngữ liệu VDSPEC với KIT nhúng Raspberry. Nghiên cứu sử dụng mô hình nhận dạng GMM với bộ tham số 13 hệ số MFCC và tham số F0 được chuẩn hóa theo trung bình F0 của mỗi câu cho tỉ lệ nhận dạng phương ngữ trung bình đạt 70%.

Từ khóa: Nhận dạng phương ngữ tiếng Việt, hệ thống nhúng, GMM, MFCC, F0, VDSPEC.

1. Giới thiệu

Tiếng Việt là ngôn ngữ có thanh điệu và có nhiều phương ngữ khác nhau. Sự đa dạng của các phương ngữ là một trong các yếu tố ảnh hưởng tới hiệu năng, tỉ lệ nhận dạng đúng đối với các hệ thống nhận dạng tự động tiếng Việt. Xét về phương diện phát âm, cùng một từ nhưng ở các địa phương khác nhau có thể được phát âm theo cách khác nhau. Với hai phương ngữ khác nhau, có những âm nghe như nhau nhưng nội dung lại được hiểu khác nhau theo từng phương ngữ. Chỉ với yếu tố này cũng đã có thể gây ra nhầm lẫn, ảnh hưởng đáng kể đến kết quả nhận dạng của các hệ thống nhận dạng tiếng Việt nói. Nếu xác định được trước nội dung tiếng nói cần nhận dạng được phát âm theo cách nói của vùng miền nào đó, hay nói khác đi, nếu biết tiếng nói đó thuộc phương ngữ nào thì có thể giúp hệ thống nhận dạng giới hạn phạm vi và sử dụng bộ ngữ liệu phù hợp cho tiếng nói cần được nhận dạng, từ đó tăng hiệu quả nhận dạng.

Nhận dạng phương ngữ nhiều ngôn ngữ trên thế giới đã được nghiên cứu như tiếng tiếng

Thái, tiếng Trung Quốc, Ả rập, tiếng Anh,... Có nhiều nghiên cứu phương ngữ tiếng Ả rập đã được công bố như [1, 2, 3]... Nghiên cứu nhận dạng phương ngữ tiếng Ả rập của tác giả Fadi Biadisy và thành viên nhóm nghiên cứu đã sử dụng câu nhận dạng có độ dài 30 giây. Trong nghiên cứu [1], nhóm tác giả đã sử dụng mô hình ràng buộc âm vị để nhận dạng phương ngữ cho độ chính xác đạt 81,60%, kết hợp ngôn điệu và ràng buộc âm vị cho độ chính xác nhận dạng 86,33% [2]. Nghiên cứu của M. Belgacem và các thành viên sử dụng mô hình GMM nhận dạng tự động phương ngữ tiếng Ả rập cho độ chính xác 73,33%. Trong nghiên cứu này, nhóm tác giả sử dụng bộ công cụ mã nguồn mở ALIZE để thực hiện các thử nghiệm nhận dạng [3]. Nghiên cứu nhận dạng phương ngữ tiếng Việt [4] đạt tỷ lệ nhận dạng đúng phương ngữ tiếng Việt đạt 70,8% khi kết hợp sử dụng bộ tham số MFCC và F0 (tần số cơ bản) chuẩn hóa theo trung bình F0.

2. Phương ngữ tiếng Việt

Phương ngữ tiếng Việt phong phú và được chia thành nhiều vùng khác nhau. Có nhiều quan

điểm khác nhau về việc phân chia các vùng phương ngữ. Có ý kiến chia phương ngữ tiếng Việt thành 2 vùng là phương ngữ Bắc, phương ngữ Nam [5]. Cũng có quan điểm cho rằng phương ngữ tiếng Việt có thể chia thành 4 vùng là phương ngữ Bắc, phương ngữ Trung Bắc, phương ngữ Trung Nam và phương ngữ Nam [6]. Việc phân chia các vùng phương ngữ mang tính tương đối. Chưa có sự thống nhất thực sự về cách phân chia và số vùng phương ngữ tiếng Việt. Tuy nhiên phần nhiều các ý kiến có điểm chung là chia phương ngữ tiếng Việt thành 3 vùng phương ngữ chính đó là phương ngữ Bắc (PNB), phương ngữ Trung (PNT) và phương ngữ Nam (PNN) tương ứng với 3 miền địa lý Bắc, Trung, Nam [5].

Sự khác biệt giữa các phương ngữ tiếng Việt bao gồm cả ngữ âm, từ vựng và ngữ nghĩa [5]. Ví dụ sự khác biệt về từ giữa các phương ngữ thể hiện ở Bảng 1:

Bảng 1: Sự khác nhau giữa các phương ngữ về từ và cách sử dụng

PNB	PNT	PNN
này	ni	nây
thế này	ri	vây
ấy	nớ	đó
thế (ấy)	rừa	vậy
kia	tê	đó
kia	tề	đó
đâu, nào	mô	đâu, nào
soa, thế nào	răng	sao
gì	chì	gì
chúng tôi	bầy tui	tụi tui

Trong PNN, có nhiều từ vay mượn từ tiếng Chăm, tiếng Khơ-me. Trong khi PNB có nhiều từ vay mượn từ tiếng Thái. Có những từ Hán-Việt được sử dụng ở PNB thì PNN lại dùng từ thuần Việt và ngược lại. Dưới đây là một số minh họa cho sự khác biệt này [5]: Khác về phụ âm đầu, thể hiện quá trình xát hóa và hữu thanh hóa như *bui/vui, bá/vá, ban/vai, bo/vo, đả/da, đươi/dưới, ăn phủng/ăn vụng, phở đất/vở đất, phở tay/vở tay, nhà thốt/nhà dột, mưa thâm/mưa dầm, chi/gì, chừ/giờ...*; khác biệt về nguyên âm, có sự biến đổi từ nguyên âm đơn sang nguyên âm đôi như nguyên âm đôi mở dần trong PNB và PNT (*e/iê, a/uơ, o/uo, mêng/miếng, nót/nuốt...*), nguyên âm đôi khép dần (*i/êi, u/ôu, con chi/con chấy, ni/nây, mi/mây...*).

Giữa các phương ngữ cũng có sự khác biệt về thanh điệu. PNB (giọng Hà Nội) có 6 thanh (ngang, huyền, sắc, nặng, hỏi, ngã). PNT và PNN ở một số tỉnh thành chỉ có 5 thanh. Ví dụ, các tỉnh Nghệ An, Hà Tĩnh, thanh ngã trùng với thanh nặng; Huế, thanh ngã lại trùng với thanh hỏi; Đà Nẵng, Thành phố Hồ Chí Minh, thanh hỏi và thanh ngã có hiện tượng pha lẫn với âm điệu mang tính chất trung gian giữa thanh hỏi và thanh ngã Hà Nội.

Sự phong phú về phương ngữ cùng với nhiều khác biệt giữa các phương ngữ tiếng Việt tạo nên những thách thức không nhỏ cho các hệ thống nhận dạng tiếng Việt nói, dễ gây ra sự nhầm lẫn các từ trong phương thức phát âm giữa các phương ngữ khác nhau. Tuy nhiên, sự khác biệt nên trên giữa các phương ngữ cũng có thể được sử dụng như là những tham số để phân biệt các phương ngữ với nhau. Do xử lý ở mức tín hiệu, khi còn chưa biết nội dung tiếng nói cần nhận dạng là gì nên nghiên cứu này chủ yếu xem xét sự khác biệt về phương thức phát âm giữa các phương ngữ (về phương diện xử lý tín hiệu) để nhận dạng phương ngữ tiếng Việt.

3. Bộ ngữ liệu VDSPEC

Để thực hiện các nghiên cứu, thử nghiệm nhận dạng cần thiết phải có bộ ngữ liệu phù hợp. Đã có một số bộ ngữ liệu tiếng Việt được xây dựng phục vụ cho các nghiên cứu nhận dạng tiếng Việt và phương ngữ tiếng Việt. Trong số đó có thể kể đến bộ ngữ liệu VDSPEC (Vietnamese Dialect Speech Corpus) được xây dựng công phu, đặc biệt tính đến các yếu tố liên quan nghiên cứu nhận dạng phương ngữ.

VDSPEC đã được xây dựng bằng phương pháp ghi âm trực tiếp người nói cho ba phương ngữ chính đó là phương ngữ Bắc, phương ngữ Trung và phương ngữ Nam [7]. Trong đó, phương ngữ Bắc lấy giọng Hà Nội làm đại diện, phương ngữ Trung lấy giọng Huế làm đại diện và phương ngữ Nam lấy giọng Thành phố Hồ Chí Minh làm đại diện.

VDSPEC chứa ngữ liệu tiếng nói được ghi âm từ 150 người nói khác nhau. Mỗi người nói đọc 25 câu văn bản đã được chuẩn sẵn và sắp xếp theo chủ đề. Tổng số có 18750 câu đã được ghi âm với tổng thời gian là 45,12 giờ tiếng nói. Bộ ngữ liệu được xây dựng có tính đến sự cân bằng giữa các thanh điệu, chủ đề, các phương ngữ và giới tính. Tiếng nói được ghi âm theo chuẩn

PCM, không nén, tần số lấy mẫu 16kHz, 16 bit mỗi mẫu, ghi trên một kênh (mono). Định dạng này đáp ứng được yêu cầu về dải tần số tiếng nói thông thường đồng thời kích thước tập tin tiếng nói không quá lớn, thuận lợi cho phân tích và trích chọn đặc trưng.

4. Mô hình nhận dạng

Có nhiều mô hình nhận dạng đã được sử dụng để nghiên cứu và thực hiện nhận dạng phương ngữ như HMM, SVM, NN... [1, 8, 11, 12]. Trong đó, mô hình hỗn hợp Gauss đa thể hiện GMM là một trong các mô hình được nhiều nhà nghiên cứu sử dụng cho nhận dạng phương ngữ và cho kết quả nhận dạng đáng ghi nhận [3, 9, 10].

Một mô hình hỗn hợp Gauss đa thể hiện là tổng có trọng số của M thành phần mật độ Gauss như biểu thức (1):

$$p(\mathbf{X}|\lambda) = \sum_{i=1}^M \pi_i g_i(\mathbf{X}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (1)$$

Trong (1), \mathbf{X} là véc tơ dữ liệu (chứa tập các véc tơ tham số của đối tượng cần biểu diễn,

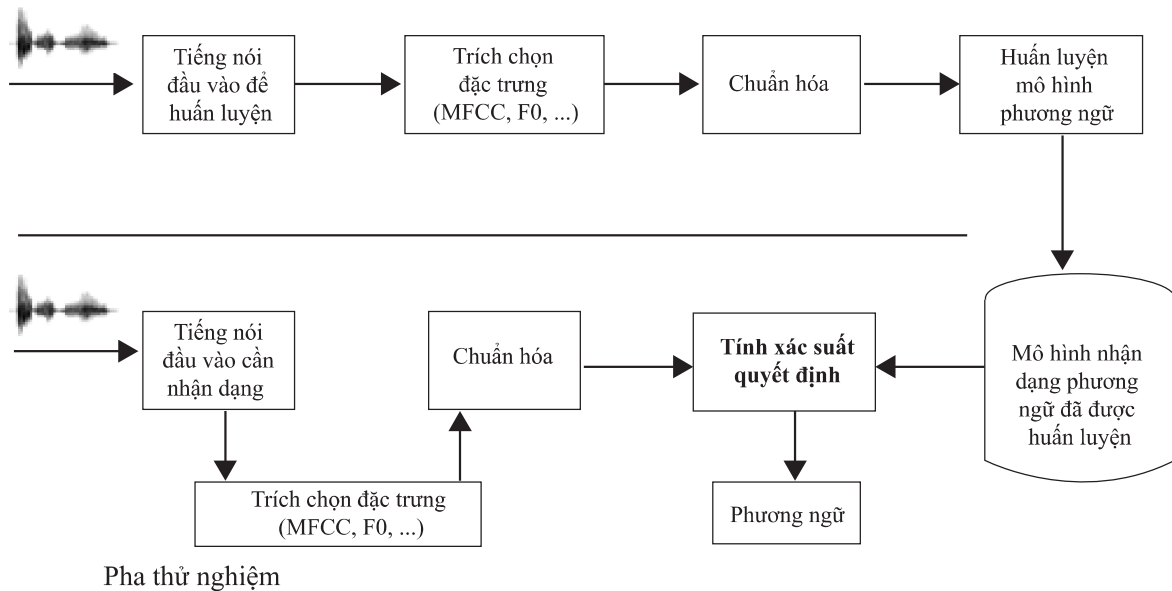
mỗi phần tử của tập giả thiết có kích thước D), $\pi_i, i=1, \dots, M$ là các trọng số của hỗn hợp và $g_i(\mathbf{X}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ là các hàm mật độ Gauss thành phần theo biểu thức (2) với véc tơ trung bình $\boldsymbol{\mu}_i$ của véc tơ kích thước D và ma trận hiệp phương sai $\boldsymbol{\Sigma}_i$ kích thước $D \times D$.

$$g_i(\mathbf{X}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{D/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{X} - \boldsymbol{\mu}_i) \right\} \quad (2)$$

Các trọng số hỗn hợp cần thỏa mãn điều kiện $\sum_{i=1}^M \pi_i = 1$.

Một GMM đầy đủ được tham số hóa bởi véc tơ trung bình, ma trận hiệp phương sai và các trọng số hỗn hợp từ tất cả các thành phần Gauss. Các tham số này có thể được biểu diễn gọn lại theo (3):

$$\lambda = \{\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}, i = 1, 2, \dots, M \quad (3)$$



Hình 1: Sơ đồ hệ thống nhận dạng phương ngữ tiếng Việt

Để định danh phương ngữ, mỗi phương ngữ được biểu diễn bằng một GMM và được tham chiếu bởi mô hình λ của phương ngữ đó. Trong trường hợp dùng MFCC như là véc tơ đặc trưng, đường bao phổ của lớp âm học thứ i được biểu diễn bằng trung bình $\boldsymbol{\mu}_i$ của thành phần thứ

i , còn biến thiên của đường bao phổ trung bình được biểu diễn bằng ma trận hiệp phương sai $\boldsymbol{\Sigma}_i$.

Giả thiết T là số lượng véc tơ đặc trưng hay cũng là toàn bộ số lượng khung (frame) tiếng nói, M là số thành phần Gauss:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\} \quad (4)$$

Khả hiện GMM là :

$$p(\mathbf{X}|\lambda) = \prod_{t=1}^T p(\mathbf{x}_t|\lambda) \quad (5)$$

Biểu thức (5) là hàm phi tuyến đối với λ nên không thể trực tiếp cực đại hóa mà các tham số khả hiện cực đại có thể nhận được bằng cách dùng thuật giải cực đại hóa kỳ vọng EM (Expectation Maximization).

Ý tưởng của thuật giải EM là bắt đầu với mô hình khởi đầu λ , đánh giá mô hình mới $\bar{\lambda}$ sao cho

$$p(\mathbf{X}|\bar{\lambda}) \geq p(\mathbf{X}|\lambda) \quad (6)$$

Mô hình mới lại là mô hình khởi đầu cho bước lặp tiếp theo và quá trình lặp lại cho đến khi ngưỡng hội tụ đạt được. Trên thực tế, thuật giải cực đại hóa kỳ vọng là đi tìm λ sao cho $\log p(\mathbf{X}|\lambda)$ của dữ liệu \mathbf{X} đạt cực đại.

Trong trường hợp thử nghiệm nhận dạng phương ngữ tiếng Việt với mô hình GMM, véc tơ \mathbf{x}_t sẽ chứa các tham số đặc trưng của tín hiệu tiếng nói như hệ số MFCC, các tham số liên quan đến F0...

Nghiên cứu này sử dụng mô hình GMM để thử nghiệm nhận dạng phương ngữ tiếng Việt trên bộ ngữ liệu VDSPEC. Bộ tham số trích chọn tiếng nói sử dụng cho mô hình bao gồm 13 hệ số MFCC và F0 đã được chuẩn hóa theo giá trị trung bình của F0 cho mỗi câu (F0sbM(t)) [4]:

$$F0sbM(t) = F0(t)/\overline{F0(t)} \quad (7)$$

Đây là bộ tham số cho tỉ lệ nhận dạng đúng phương ngữ cao trong nghiên cứu [4] đã công bố.

5. Kết quả thử nghiệm

Thử nghiệm nhận dạng phương ngữ trên hệ thống nhúng được thực hiện trên nền tảng phần cứng là bộ KIT Raspberry Pi. Công cụ nhận dạng là bộ thư viện mã nguồn mở ALIZE.

Ngữ liệu tiếng nói dùng (VDSPEC) được chia làm 5 phần. Bốn phần dùng để huấn luyện. Phần còn lại dùng thử nghiệm nhận dạng. Với mỗi phương ngữ, các file tiếng nói của 10 người nói (bao gồm 5 nam và 5 nữ) được sử dụng để thử nghiệm (test), 40 người nói (20 nam, 20 nữ) dùng cho huấn luyện (training). Số thành phần

Gauss (M) được sử dụng trong mô hình là 20. Bộ tham số bao gồm 13 thành phần MFCC và F0 được chuẩn hóa theo trung bình F0 của câu. Quá trình thử nghiệm được thực hiện theo phương pháp quay vòng.

Bảng 2: Ma trận sai nhầm lượt thử 1

Phương ngữ	PNB	PNT	PNN	Tỷ lệ nhận dạng đúng	TBC %
PNB	1017	165	53	82,3%	71,5%
PNT	161	869	170	72,4%	
PNN	238	249	723	59,8%	

Thời gian nhận dạng: 3 phút 2 giây.

Bảng 3: Ma trận sai nhầm lượt thử 2

Phương ngữ	PNB	PNT	PNN	Tỷ lệ nhận dạng đúng	TBC %
PNB	864	202	153	70,9%	79,1%
PNT	171	915	130	75,2%	
PNN	40	66	1107	91,3%	

Thời gian nhận dạng: 2 phút 58 giây.

Bảng 4: Ma trận sai nhầm lượt thử 3

Phương ngữ	PNB	PNT	PNN	Tỷ lệ nhận dạng đúng	TBC %
PNB	520	519	154	43,6%	71,2%
PNT	116	1030	69	84,8%	
PNN	84	101	1063	85,2%	

Thời gian nhận dạng: 2 phút 55 giây.

Bảng 5: Ma trận sai nhầm lượt thử 4

Phương ngữ	PNB	PNT	PNN	Tỷ lệ nhận dạng đúng	TBC %
PNB	907	168	163	73,3%	73,3%
PNT	191	1016	38	81,6%	
PNN	98	327	788	65,0%	

Thời gian nhận dạng: 3 phút 2 giây.

Bảng 6: Ma trận sai nhầm lượt thử 5

Phương ngữ	PNB	PNT	PNN	Tỷ lệ nhận dạng đúng	TBC %
PNB	868	28	337	70,0%	53,5 %
PNT	501	477	199	39,2%	
PNN	126	415	619	51,3%	

Thời gian nhận dạng: 2 phút 56 giây.

Kết quả các thử nghiệm cho thấy tỉ lệ nhận dạng đúng phương ngữ hầu hết trên 70% từ phần thử thứ nhất đến phần thử thứ 4. Riêng phần thử cuối cho tỉ lệ nhận dạng đúng không cao so với các thử nghiệm trước đó. Tỷ lệ này có thể đang phụ thuộc vào chất lượng tiếng nói được ghi âm thuộc phần thử nghiệm. Tỷ lệ nhận dạng đúng phương ngữ trung bình của 5 thử nghiệm đạt 70%. Thời gian thực hiện nhận dạng trung bình 2 phút 59 giây cho một lượt thử nghiệm thực hiện trên 3750 file tiếng nói cần nhận dạng sau khi đã trích chọn đặc trưng và thực hiện chuẩn hóa.

6. Kết luận

Tiếng Việt là ngôn ngữ có nhiều phương ngữ. Sự đa dạng cùng với các khác biệt trên nhiều phương diện của các phương ngữ là một trong những yếu tố có ảnh hưởng tới hiệu quả nhận dạng của các hệ thống nhận dạng tiếng Việt nói. Bài báo trình bày kết quả nghiên cứu và thử

thử nghiệm nhận dạng phương ngữ tiếng Việt trên hệ thống nhúng sử dụng mô hình GMM với bộ tham số bao gồm 13 hệ số MFCC kết hợp tham số F0 chuẩn hóa theo trung bình F0 trên câu. Kết quả thử nghiệm đạt tỷ lệ nhận dạng đúng phương ngữ là 70% trên bộ ngữ liệu VDSPEC. Việc nhận dạng được phương ngữ có thể ứng dụng để cải thiện hiệu quả nhận dạng cho các hệ thống nhận dạng tự động tiếng Việt nói, đặc biệt là có thể triển khai trên nền tảng phần cứng các hệ thống nhúng có cấu hình không mạnh so với máy tính thông thường và có khả năng hoạt động độc lập, không phụ thuộc kết nối Internet. Điều này giúp phát triển ứng dụng thực tế được thuận lợi hơn.

Lời cảm ơn

Nghiên cứu này được hỗ trợ bởi đề tài khoa học cấp trường của Trường Đại học Sư phạm Kỹ thuật Hưng Yên năm 2019 mã số: UTEHY.L.2019.55.

Tài liệu tham khảo

- [1]. Biadisy, Fadi, Julia Hirschberg, and Nizar Habash (2009) Spoken Arabic dialect identification using phonotactic modeling. Proceedings of the eacl 2009 workshop on computational approaches to semitic languages, pp. 53-61.
- [2]. Fadi Biadisy, Julia Hirschberg (2009) Using Prosody and Phonotactics in Arabic Dialect Identification. Interspeech, vol. 1, pp. 208-211.
- [3]. Mohamed BELGACEM, Georges ANTONIADIS, Laurent BESACIER (2010) Automatic Identification of Arabic Dialects. International Conference on Language Resources and Evaluation (LREC), MALTA, pp. 17-23.
- [4]. Pham Ngoc Hung, Trinh Van Loan, Nguyen Hong Quang (2016) “Automatic identification of Vietnamese dialects”. Journal of Computer Science and Cybernetics, V.32, N.1 (2016), 18-29, DOI: 10.15625/1813-9663/32/1/7905.
- [5]. Hoàng Thị Châu (2009) Phương ngữ học tiếng Việt. NXB Đại học Quốc gia Hà Nội.
- [6]. Nguyễn Kim Thân, Nguyễn Trọng Báu, Nguyễn Văn Tu (2002) Tiếng Việt trên đường phát triển. NXB Khoa học Xã hội
- [7]. Pham Ngoc Hung, Trinh Van Loan, Nguyen Hong Quang (2015) “Building of corpus for Vietnamese dialect identification”, Journal of Science and Technology Technical Universities, No.109-2015. ISSN 2354-1083, pp.49-55.
- [8]. Chen, N. F., Shen, W., Campbell, J. P., & Torres-Carrasquillo, P. A. (2011) Informative dialect recognition using context-dependent pronunciation modeling. Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference, pp. 4396-4399.
- [9]. Chen, Too, Chao Huang, Eric Chang, and Jingehan Wang (2001) Automatic accent identification using Gaussian mixture models. Automatic Speech Recognition and Understanding. ASRU'01. IEEE Workshop, pp. 343-346.
- [10]. Faria, Arlo (2005) Accent classification for speech recognition. In International Workshop on Machine Learning for Multimodal Interaction, pp. 285-293
- [11]. Rao, K. S. (2011) Role of neural network models for developing speech systems. Sadhana, vol. 36, no. 5, pp. 783-836.

- [12]. Rao, K. S., & Koolagudi, S. G. (2011) Identification of Hindi dialects and emotions using spectral and prosodic features of speech. IJSCI: International Journal of Systemics, Cybernetics and Informatics, vol. 9, no. 4, pp. 24-33.

VIETNAMESE DIALECT IDENTIFICATION ON EMBEDDED SYSTEM

Abstract:

Many factors affect speech recognition including dialect factors. Vietnamese is a tone language with many different dialects. Therefore, spoken Vietnamese identification is also significantly influenced by dialect. If the information about dialects is known during the speech recognition process, the performance of recognition systems will be better because the corpus of these systems is normally organized according to different dialects. Vietnamese dialect identification has been studied and achieved certain results. Studies are usually carried out on the computer with a powerful computing environment to train models and tests. Embedded devices are being used extensively in practical products, including speech recognition and control, which often have weaker configurations than computers such as slow speed, low memory. Therefore, the implementation of speech recognition on this hardware is more a little difficult. This article presents the results of the research and experiment of Vietnamese dialect identification at embedded systems using the VDSPEC corpus on Raspberry embedded KIT. The study used the GMM identification model with the parameter set of 13 MFCC coefficients and the F0 parameter was standardized according to the average F0 of each sentence, the average recognition rate of dialects is 70%.

Keywords: *Vietnamese dialects identification, embedded systems, GMM, MFCC, F0, VDSPEC.*