

ISSN 2354-1083

Tạp chí

**KHOA HỌC &
CÔNG NGHỆ**
CÁC TRƯỜNG ĐẠI HỌC KỸ THUẬT

JOURNAL OF
SCIENCE & TECHNOLOGY
TECHNICAL UNIVERSITIES

No.109
2015

MỤC LỤC

1. Nghiên cứu bộ quan sát trượt để ước lượng tốc độ và vị trí rotor động cơ đồng bộ nam châm vĩnh cửu <i>Nguyễn Quang Địch - Trường Đại học Bách khoa Hà Nội</i>	1
2. Fault Location for Non-homogeneous Transmission Lines using Measurement Signals from Two-Ends without Utilizing Line Parameters <i>Nguyen Xuan Vinh^{1,2*}, Nguyen Xuan Tung², Nguyen Duc Huy²</i> ¹ <i>Hanoi University of Science and Technology</i> ² <i>Vinh Long University of Technology Education</i>	8
3. Nghiên cứu thiết kế chế tạo kit phát triển đa năng nhằm xây dựng phòng thí nghiệm điện tử có tính di động cao <i>Nguyen Hoang Dung^{1*}, Nguyen Hoai Giang²</i> ¹ <i>Hanoi University of Science and Technology</i> ² <i>Hanoi Open University</i>	15
4. A Novel two-State ECG Compression Algorithm used in Telemedicine <i>Duong Trong Luong*, Nguyen Minh Duc, Nguyen Tuan Linh, Nguyen Duc Thuan</i> - <i>Hanoi University of Science and Technology, Vietnam</i>	22
5. Toward an Implementable Policy Control Framework in Next Generation Networks <i>Nguyen Tai Hung - Hanoi University of Science and Technology</i>	28
6. Early CU Splitting and Fast Mode Decision for Intra Prediction in HEVC <i>Duong Trinh, Canh Dinh, Toan Nguyen, Trung Ho, Quang Nguyen, Phong Nguyen, Duc Ngo, Thang Nguyen - Hanoi University of Science and Technology</i>	36
7. Development of Features Set for Classification of Imagery Hand Movement – Related EEG Signals <i>Pham Phuc Ngoc, Pham Van Binh*, Vu Duy Hai, Nguyen Duy Tung, Vu Thi Hanh, Nguyen Duc Thuan - Hanoi University of Science and Technology</i>	43
8. Building of Corpus for Vietnamese Dialect Identification <i>Pham Ngoc Hung^{1,2}, Trinh Van Loan^{1*}, Nguyen Hong Quang^{1*}</i> ¹ <i>Hanoi University of Science and Technology</i> ² <i>Hung Yen University of Technology and Education</i>	49
9. Phương pháp xác định tuổi bền của đá mài ché tạo tại Việt Nam thông qua rung động <i>Nguyễn Thị Phương Giang - Trường Đại học Bách khoa Hà Nội</i>	56
10. Mô hình hóa hình học lưỡi cắt kéo mô hình y tế đầu cong <i>Phan Bùi Khôi¹, Bùi Ngọc Tuyên^{1*}, Lê Văn Thắm^{1,2}</i> ¹ <i>Trường Đại học Bách khoa Hà Nội</i> ² <i>Trường Cao đẳng nghề Kỹ thuật Công nghệ</i>	61
11. An Assembled Spectrometer with a CD Reflection Grating and a CCD Detector to Determine Color Specifications of Artificial Lamps In Vietnam <i>Nguyen Thanh Dong*, Nguyen Thi Phuong Mai</i> - <i>Hanoi University of Science and Technology</i>	67
12. Nghiên cứu ảnh hưởng của độ ẩm tương đối tới đặc tính ma sát trong xylyanh – piston khí nén <i>Nguyễn Thị Huyền Dương, Phạm Văn Hùng* – Trường Đại học Bách khoa Hà Nội</i>	73
13. Affect of Chromium Carbide to Microstructure and Microhardness of the Weld Metal During PTA Welding by Eutroloy 16606 Powder Alloys on Carbon Mild Steel <i>Ngo Huu Manh^{1,2}, Bui Van Hanh^{1*}, Nguyen Thuc Ha¹</i> ¹ <i>Hanoi University of Science and Technology</i> ² <i>Saodo University</i>	78

14. Nghiên cứu đề xuất giới hạn cảnh báo mốc ổn định xe bán moóc <i>Võ Văn Hường, Dương Ngọc Khanh*</i> - <i>Trường Đại học Bách khoa Hà Nội</i>	82
15. Nghiên cứu, xác định ngưỡng điều khiển ABS trong hệ thống phanh khí nén <i>Hồ Hữu Hùng^{1,2*}, Dương Ngọc Khanh¹, Lưu Văn Tuấn¹</i> ¹ <i>Trường Đại học Bách Khoa Hà Nội</i> ² <i>Trường Đại học Kinh tế - Kỹ thuật Công nghiệp</i>	87
16. Ảnh hưởng của các thông số vận hành đến trao đổi nhiệt trong lò sôi tuần hoàn <i>Nguyễn Minh Tiên*, Phạm Hoàng Lương</i> - <i>Trường Đại học Bách khoa Hà Nội</i>	92
17. Ảnh hưởng của đất hiếm đến hình thái, hành vi nứt vỡ của các bít M ₇ C ₃ cung tinh trong gang crôm khi chịu va đập <i>Hoàng Thị Ngọc Quyên*, Lê Thị Chiều, Nguyễn Hồng Hải, Phạm Mai Khánh</i> - <i>Trường Đại học Bách khoa Hà Nội</i>	98
18. Nghiên cứu đặc điểm kích thước phần thân dưới cơ thể phụ nữ thành phố Hồ Chí Minh độ tuổi từ 25 đến 45 <i>Nguyễn Thị Thanh Phúc^{1,2}, Trương Thị Diệu^{1,2}, Phan Thành Thảo^{1*}</i> ¹ <i>Trường Đại học Bách Khoa Hà Nội</i> ² <i>Trường Cao đẳng Kinh tế Kỹ thuật Vinatex TP. HCM</i>	102
19. Nghiên cứu đặc tính cách âm của một số tám xơ dùng làm vật liệu cách âm ở Việt Nam <i>Lê Phúc Bình, Hoàng Xuân Hiền, Hồ Phước Lộc, Kiều Tấn Đoàn</i> - <i>Trường Đại học Bách khoa Hà Nội</i>	110
20. Nghiên cứu đánh giá đặc trưng biến dạng nén và tính chất vệ sinh của lót giày đàn hồi <i>Bùi Văn Huấn^{1*}, Đỗ Thị Phương², Cao Thị Kiên Chung²</i> ¹ <i>Trường Đại học Bách khoa Hà Nội</i> ² <i>Trường Đại học Sư Phạm Kỹ thuật Hưng Yên</i>	114
21. Nghiên cứu ảnh hưởng của nồng độ chất khơi mào và monome đến hiệu suất của phản ứng đồng trùng ghép styren lên mạch cao su thiên nhiên đã loại protein <i>Trần Anh Dũng¹, Nguyễn Thị Nhàn¹, Đỗ Quốc Việt¹, Lê Trọng Linh², Nguyễn Huy Tùng¹ Vũ Anh Tuấn¹, Phan Trung Nghĩa¹, Kawahara Seiichi³, Trần Thị Thúy^{1*}</i> ¹ <i>Trường Đại học Bách khoa Hà Nội</i> ² <i>Công ty TNHH một thành viên Môi trường đô thị Hà Nội</i> ³ <i>Trường Đại học Bách khoa Nagaoka</i>	120
22. Investigation of Deinking Possibility of Recycled Newspapers using the Mixture of Fungal Cellulase and Xylanase <i>Do Thi Thanh Binh, Nguyen Van Khang, Nguyen Thi Xuan Sam, Dang Minh Hieu*</i> - <i>Hanoi University of Science and Technology,</i>	124
23. Application of Water Quality Index and Simple Simulation of Pollution Emission for River Water Quality Management in Hai Duong Province, Vietnam <i>Duc-Quang Nguyen¹, Hong-Minh Ta², Trung-Hai Huynh¹</i> ¹ <i>Hanoi University of Science and Technology</i> ² <i>Hai Duong Department of Environment and Natural Resources</i>	128
24. Electrochemical oxidation of Reactive dye using a Pt electrode <i>Nguyen Thi Lan Phuong*, Tran Thi Anh Ngoc, Tran Thi Hien, Nguyen Ngoc Lan</i> - <i>Hanoi University of Science and Technology</i>	134
25. Application of Micro-Membrane Filtration System to Separate Distiller Wet Grains from Whole Stillage of Rice-Based Ethanol Distillery at Lab Scale <i>Do Khac Uan*, Chu Ky Son - Hanoi University of Science and Technology</i>	140

Building of Corpus for Vietnamese Dialect Identification

Xây dựng kho ngữ liệu dùng cho nhận dạng phương ngữ tiếng Việt

Pham Ngoc Hung^{1,2}, Trinh Van Loan^{1*}, Nguyen Hong Quang^{1*}

^{1*} School of Information and Communication Technology, Hanoi University of Science and Technology

² Faculty of Information Technology, Hung Yen University of Technology and Education

Received: March 27, 2015; accepted: October 22, 2015

Abstract

The performance of speech recognition systems will be improved if the corpus are organized in specialized domain and are applied in a consistent way for identifying in specific situations. Vietnamese dialects are various. Building of corpus for Vietnamese dialect is the first step to implement the system of dialect identification used for increasing the performance of Vietnamese recognition in general. This paper presents a method of building corpus for Vietnamese dialect identification. Vietnamese corpus VDSPEC are built with topic-based recording and tonal balance. The duration of corpus is 33,79 hours for 6 topics in total. The basic characteristics and preliminary estimations of the corpus are also described. The results show that there are distinctions of pronunciation modality for Vietnamese tones toward Hue voice and Hanoi voice. These distinctions can be used as important features for identifying these dialects.

Keywords: Vietnamese, corpus, Vietnamese dialect, topic-based recording, tone balance

Tóm tắt

Hiệu năng của các hệ thống nhận dạng tiếng nói sẽ được cải thiện nếu kho ngữ liệu được tổ chức theo các lĩnh vực chuyên biệt và được dùng tương ứng để nhận dạng trong các tình huống cụ thể. Phương ngữ tiếng Việt rất đa dạng. Việc xây dựng kho ngữ liệu phương ngữ tiếng Việt là bước đầu tiên để thực hiện hệ thống định danh phương ngữ được dùng để tăng hiệu năng của hệ thống nhận dạng tiếng Việt nói chung. Bài báo này trình bày phương pháp xây dựng kho ngữ liệu dùng cho nghiên cứu nhận dạng phương ngữ tiếng Việt. Kho ngữ liệu tiếng Việt VDSPEC được xây dựng với việc ghi âm theo chủ đề và có tính tới yếu tố cân bằng thanh điệu. Thời lượng tổng cộng là 33,79 giờ tiếng nói cho 6 chủ đề. Bài báo cũng trình bày các đặc điểm cơ bản của kho ngữ liệu và một số đánh giá, thử nghiệm ban đầu đã thực hiện trên kho ngữ liệu này. Kết quả cho thấy có sự khác biệt trong phương thức phát âm các thanh điệu tiếng Việt đối với giọng Huế và giọng Hà Nội. Sự khác biệt đó có thể được sử dụng như là các đặc trưng quan trọng để định danh các phương ngữ này.

Từ khóa: Tiếng Việt, kho ngữ liệu, phương ngữ tiếng Việt, ghi âm theo chủ đề, cân bằng thanh điệu

1. Introduction

To be able to carry out research on speech recognition in general and in particular on dialect identification, we need a good quality corpus which meet research requirements. For Vietnamese, some corpora exist already such as VNSPEECHCORPUS [1], VOV (Voice of Vietnamese) Corpus [2] or VNBN (United Broadcast News corpus) [3].

The construction of corpus can be done in several different ways. For example, using the available audio sources from radio, television, and then classify, extract audio signals matching requirements, browse and edit the text, respectively [2], [3]. The alternative is to perform recording environments and to select speakers based on record scenario prepared in advance.

Usually, speech recognition systems use corresponding vocabulary. The performance of the recognition system will be improved if the corpus is organized in specialized domain and is applied in a consistent way for identifying in specific situations. In dialect recognition, especially for Vietnamese language, corpus should involve the characteristics of Vietnamese language. The mentioned available corpora do not simultaneously satisfy these requirements. Therefore, building of Vietnamese corpus VDSPEC (Vietnamese Dialect Speech Corpus) was studied to meet the requirements for speech and Vietnamese dialect recognition. It is known that dialect is a form of the language spoken in different regions of the country. These dialects may have distinctions of words, grammar and pronunciation modalities. For Vietnamese, researches on dialects are mainly concentrated on language approach [4]. In our research, we focus only on pronunciation modality for Hanoi and Hue voices and the dialect identification is based on signal

* Corresponding author: Tel: (+844) 3869.6125
Email: loan.trinhvan@hust.edu.vn

processing, hence the corpus does not reflect the difference of dialect words and grammar between these regions.

Section 2 of this paper will present the methods for building Vietnamese corpus in which different topics are recorded to take account of tonal balance in some Vietnamese dialect. Section 3 describes in detail the corpus and section 4 gives conclusions and development in future.

2. Method for building Vietnamese corpus

There are already dialectal corpora for some languages such as English [5], Chinese [6], Arabic [9], Thai [11].... For English, FRED is really a big dialect corpus which cover 8 dialects with 2.45 million words of text and about 300 hours of speech. FRED contains data from 420 different speakers, the age of speakers included in FRED ranges from six years to 102 years. For material included in FRED, it was recorded over 30 years. The corpus permit the investigation of phenomena of non-standard morphosyntax beside analyses of phonetic or phonological details. For Chinese, there are eight major dialectal regions. The authors in [6] have built the corpus for Wu dialect belonging to eight major Chinese dialects and providing information at four levels: phonetic level, lexicon level, language level and acoustic decoder level.

Our corpus is built mainly for the first step research on dialect identification of Vietnamese and the corpus's target is more modest and meets the basic criteria. The corpus is built to cover a relative large range of topics, text contents ensure tonal balance, gender equilibrium for speakers, speakers are selected so that they possess local accent and their voices are steady, low noise for recording environment. For a corpus, there are two ways for recording: spontaneous speech and read speech. To be more active, we have chosen read speech for recording.

The building of Vietnamese corpus is done in two stages. Stage 1 includes compilation, collection and classification of documents by topic; performing adjustments to ensure tone balance in the prepared text. Next, in stage 2, recording is performed using specialized equipment with selected environment. The following is description in detail for these stages.

2.1. Preparation and normalization of documents

For six recording topics, the first topic is intended to estimate tone and fundamental frequency variations of dialects. Therefore, the text content of this topic contains mainly consonants, vowels, single words representing all of the tones.

Five other topics are selected from electronic documents. These documents are stored in UTF-8 encoding uniformly in the entire system. Original text often contains redundant information such as HTML tags, symbols, abbreviations, foreign words, figures, data, date with different format of numbers and letters... To ensure homogeneous, the redundant information is removed and the documents must be normalized such as to transform the numbers into the corresponding text ("9000 đồng" to "chín nghìn đồng"), dates in the usual format "ngày 27/10" into the text "ngày hai bảy tháng mười" or "năm 2003" into "năm hai ngàn lẻ ba"; abbreviations are converted to the corresponding full text to avoid confusion in reading, for example, "tốt nghiệp ĐHMT Hà Nội" is transformed into "tốt nghiệp Đại học Mỹ thuật Hà Nội".

After normalization, text needs to be counted to ensure tone balance. Tone balance means that the appearance probability of six tones is the same (about 717 words for each tone). This procedure is conducted automatically with the support of software or manually.

Finally, each sentence is distinguished by a pair of tags including the opening <s> at the beginning and closing </s> at the end. For example, sentence "As representatives of a joint stock commercial bank" will be saved like "<s> As representatives of a joint stock commercial bank </s>". Each theme is saved into a text file (UTF-8 format) with the file name format "YY.txt" where "YY": code corresponding to the theme (cb: "Basic", ds: "Life," kd: "business", ox: "cars-motor" ...). In the text file of a theme, each passage is to be started by a notation in the format "YYZZZ", where "YY" corresponds to the topic, as mentioned above, ZZZZ is number of paragraphs under the theme "YY". For example, "cb0001" denotes the beginning of paragraph 1 under the theme "cb" (basic). The next line is the text which begins with tag <s> and ends with tag </s>. The length of a paragraph should be chosen so that the time for reading is neither too short nor too long. In our case, each paragraph corresponds to the recording time 10 seconds for normal speaking speed.

2.2. Recording

2.2.1. Recording equipment

The recording is done on computer with high-quality sound card. The type of microphone is suitable for voice recorder (Shure SM48). SM48 has a frequency response from 55Hz to 14000Hz, up 270 Ohms impedance, reaching -57.5 dBV / Pa (1.3 mV) at 1KHz frequency [7]. This type of unidirectional microphone limits background noise and ambient noise sources. Recording studios have low

background noise with signal-to-noise ratio shown later.

2.2.2. Selecting speakers

The selection of speakers have a significant impact on the quality of obtained voice. Speakers are chosen so that they speak with the local accent. The average age of speakers is 21 year old. At this age, voice quality is steady with full features for local voice. The record is also held in different sections to cover the voice variability of human being.

The total number of speakers is 100 people including 50 speakers with northern accent, 50 speakers with middle accent. The number of male voices is the same for female voices.

2.2.3. The software for recording

The software for recording uses the scripting language TCL/TK.

The main functions of the software are management of user information, server for recording, audio files. Besides, this software can give usefull information for user such as: file numbers for each topic and for each speaker, displaying wave form during recording, playback...

2.2.4. Audio recording formats

Recording format has been set in the recording software. Audio is recorded as standard PCM, uncompressed, with sampling frequency of 16KHz, 16 bits per sample using one channel (mono). This format meets the requirements about speech frequency range and not too large file size.

2.2.5. Data archives

To favorize the management and exploitation of corpus, the files are named in a uniform format. Audio files corresponding to each paragraph in the subject are recorded on the disk with file name format "XXYYZZZZ.wav", in which:

- XX: speaker code (ID) including letters, digits. This code is unique
- YY: theme code (as described in Section 2.1)
- ZZZZ: audio clips code (described in Section 2.1)

Speaker information is recorded in a file with file name user.xml and this file contains the following information:

- Code (ID); Full Name
- Address (the local address of a speaker and this location determines the speaker accent), gender, age, contact information

The organisation of corpus is presented in Table 1.

Table 1. Corpus organization

Data	Directory	Content
Speech signal	Wav Folder	File WAV, Sampling frequency: 16000Hz, 16-bit, Mono
Text file of topics	Text Folder	All topics
Speaker information	File user.xml	Basic information of speakers

3. Results

The following is a description of the corpus VDSPEC in detail.

3.1. Text features

The text is organized into six themes. The first theme (basic) is intended for tone studies. The other topics include life sciences, business, law, cars, motorcycles, texts are collected from electronic media Vnexpress. 150 sentences containing 4333 syllables have been collected, classified and selected .

Table 2. The text topics

Topic	Number of sentences	Number of syllables	Sources
Basic	25	349	Compilation
Life	25	855	VnExpress
Science	25	893	VnExpress
Business	25	729	VnExpress
Carsmotor	25	652	VnExpress
Law	25	855	VnExpress
Total	150	4333	

3.2. VDSPEC features

Speech signals are recorded with the sampling frequency 16000 Hz, using one channel (mono) and 16 bits per sample. The corpus consists of 50 male voices and the same for female voices. The average age of speakers is 21. There are two main dialects of Vietnamese for the corpus. The number of northern dialect speaker is 50 and the same speaker number for middle dialect. For each dialect, the number of male voices is equal to the one of female voices. In our case, northern dialect is Hanoi voice and middle dialect is Hue voice. For a topic, each speaker reads 25 sentences in total. The number of recorded sentences is 15000 (100 speakers and 150 sentences for a speaker). The corpus capacity is 3.62GB and total duration is 33.79 hours (Table 3).

Sentence, syllable numbers and corresponding durations for VDSPEC topics are presented in Table 4.

Table 3. VDSPEC features for dialects

Index	Dialect	Number of sentences	Duration (hours)
1	North	7500	16,82
2	Middle	7500	16,97
	Total	15000	33,79

Table 4. Distribution by topic in VDSPEC

Topic	Number of sentences	Number of syllables	Duration (hours)
Basic	25	349	4.73
Life	25	855	6.44
Science	25	893	5.18
Business	25	729	6.48
Carsmotor	25	652	4.70
Law	25	855	6.26
Total	150	4333	33.79

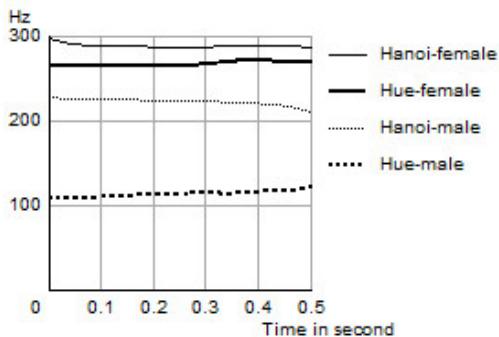


Fig. 1. Level tone (word “*khi*”)

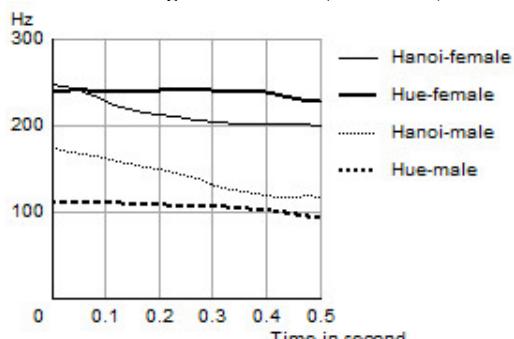


Fig. 2. Low-falling tone (word “*trường*”)

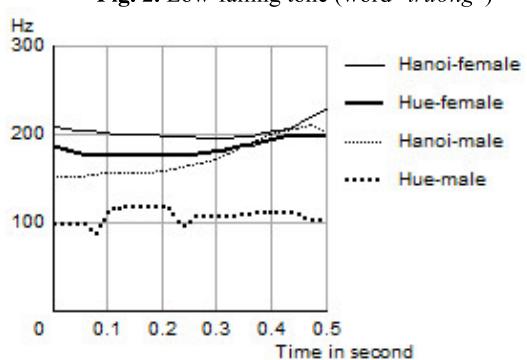


Fig. 3. Rising tone (word “*thuê*”)

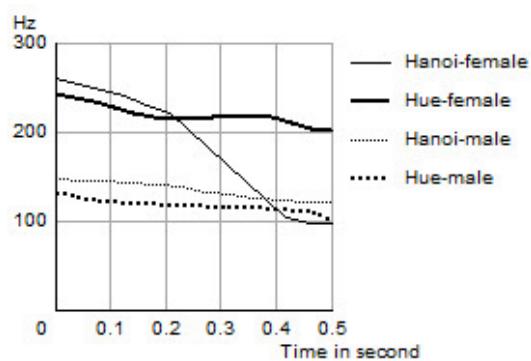


Fig. 4. Heavy tone (word “*mại*”)

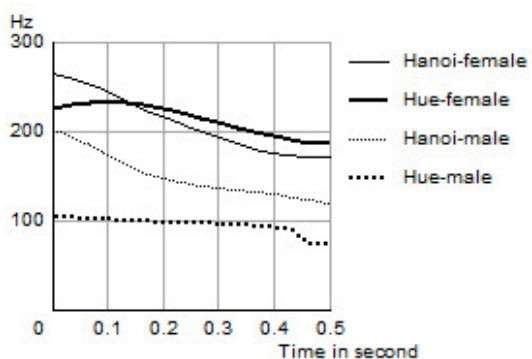


Fig. 5. Asking tone (word “*thú*”)

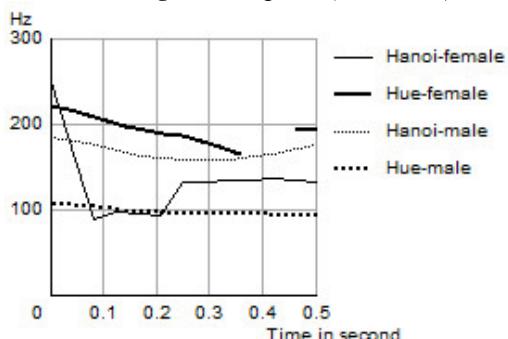


Fig. 6. Broken tone (word “*phẫu*”)

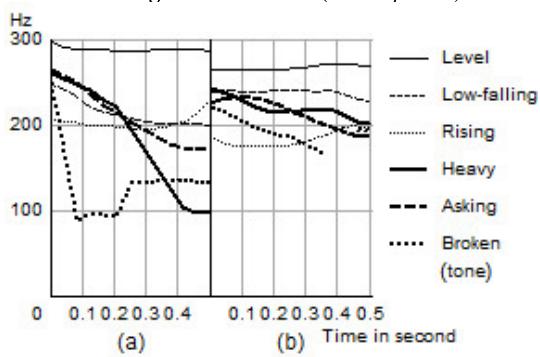


Fig. 7. Variation of 6 tones for female voices.
(a) Hanoi, (b) Hue

Praat [8] was used to estimate fundamental frequency variations for Vietnamese tones in VDSPEC and four representative voices including 2 males and 2 females with two dialects were selected. The durations of the actual tones are usually different. To make the difference more evident, these durations have been normalized by the same time 0.5 seconds. The results are shown in figures from 1 to 8.

For level tone, F0 variation is rather small at around the mid level for both dialects. For Hanoi voice, rising tone starts as mid and then rises but for Hue voice the difference between start and end values for F0 is smaller than the one of Hanoi voice. For low-falling tone, F0 starts low-mid and falls monotonously. With heavy tone, F0 starts mid or low-mid and rapidly falls at the end for Hanoi voice. For asking tone (falling rising tone), F0 goes down and has a tendency to goes up at the end with Hanoi voice. With broken tone, F0 falls down, maybe is broken before going up for Hanoi voice. In conclusion, F0 of tones for Hue voices has tendency to go down monotonously as low-falling or heavy tones for Hanoi voices. In addition, the range of F0 variation for Hue voice is much smaller than Hanoi voice as we can see at the figure 8.

The variation of F0 values for 100 speakers including 50 males and 50 females is also evaluated and is depicted in Figs from 9 to 20. These figures show the statistical results of F0 variations for Hanoi male and female voices (Hanoi), Hue male and female voices (Hue). For each dialect, the number of female voices equals 25 and the same for the number of male voices. The variation of F0 is discretized and normalized by 20 values. In these figures, the abbreviations are the following: Min: minimum value, Avr: average value, Max: maximum value, Med: median value, StD: standard deviation, RS1, RS2: 50% of F0 values are from RS1 to RS2. From Figs 9 and 10, the range of F0 variation for asking tone of Hue voices is smaller than the case of Hanoi voices, nevertheless this range for level tone of Hue voices is larger than Hanoi voices (Figs 15, 16).

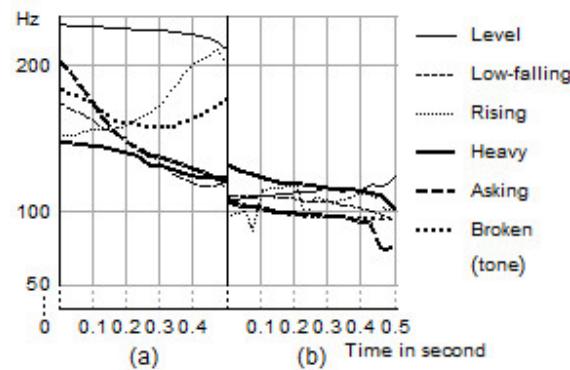


Fig. 8. Variation of 6 tones for male voices

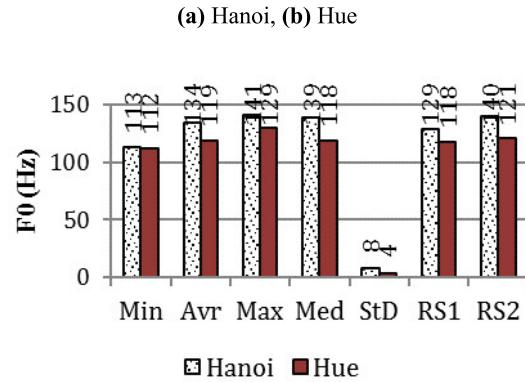


Fig. 9. F0 statistical results for asking tone of male voices.

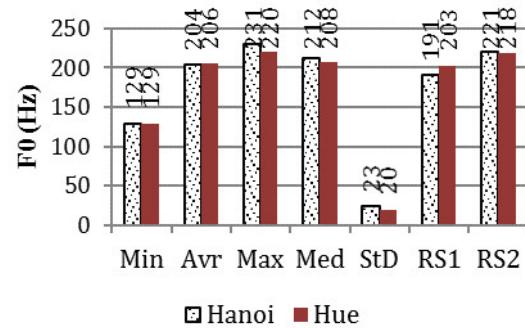


Fig. 10. F0 statistical results for asking tone of female voices.

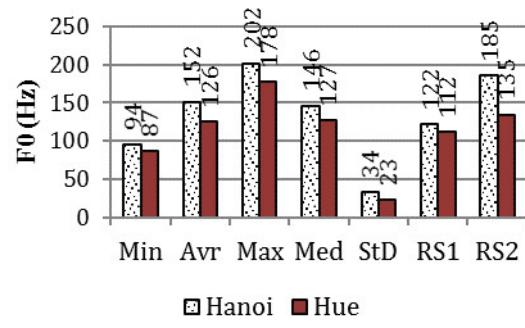


Fig. 11. F0 statistical results for broken tone of male voices.

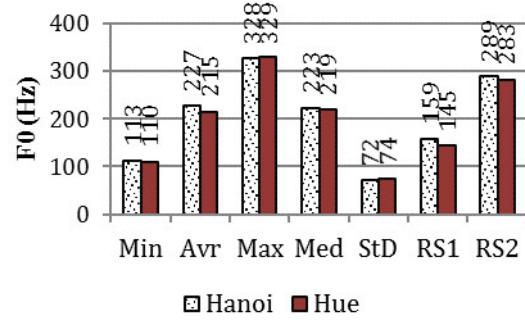


Fig. 12. F0 statistical results for broken tone of female voices.

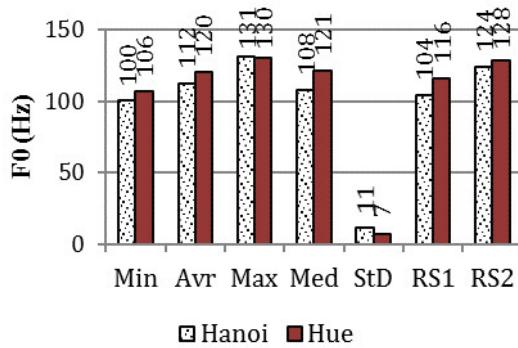


Fig. 13. F0 statistical results for heavy tone of male voices.

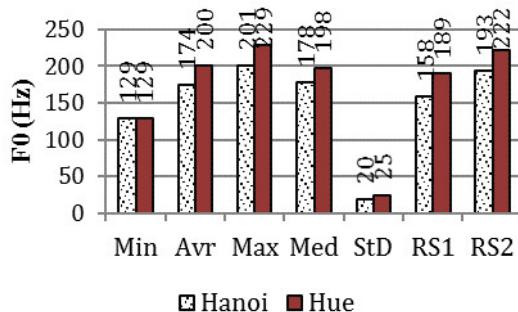


Fig. 14. F0 statistical results for heavy tone of female voices.

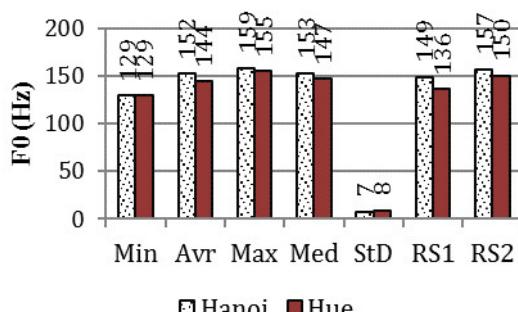


Fig. 15. F0 statistical results for level tone of male voices.

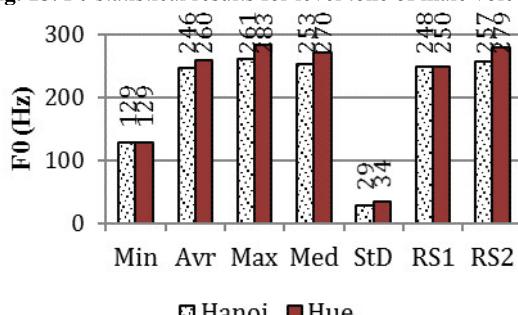


Fig. 16. F0 statistical results for level tone of female voices.

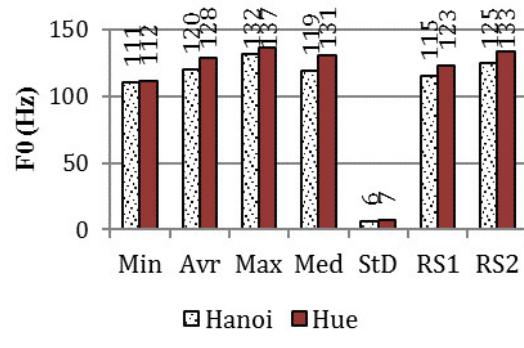


Fig. 17. F0 statistical results for low-falling tone of male voices.

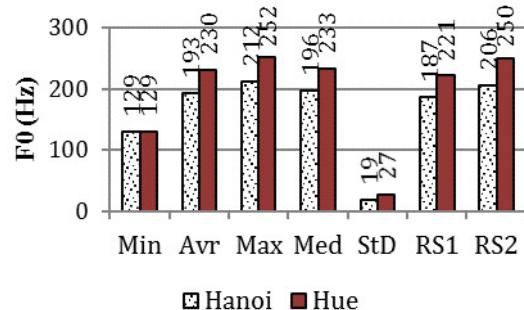


Fig. 18. F0 statistical results for low-falling tone of female voices.

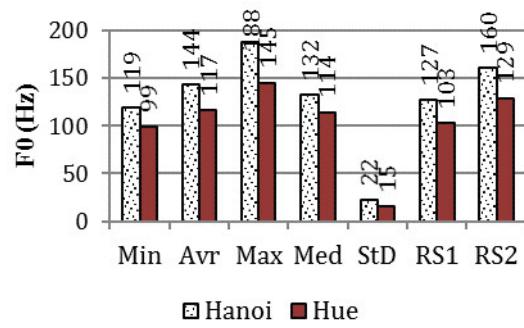


Fig. 19. F0 statistical results for rising tone of male voices.

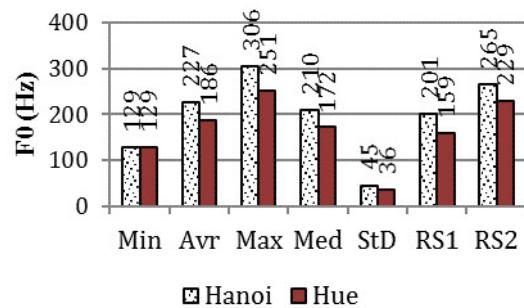


Fig. 20. F0 statistical results for rising tone of female voices.

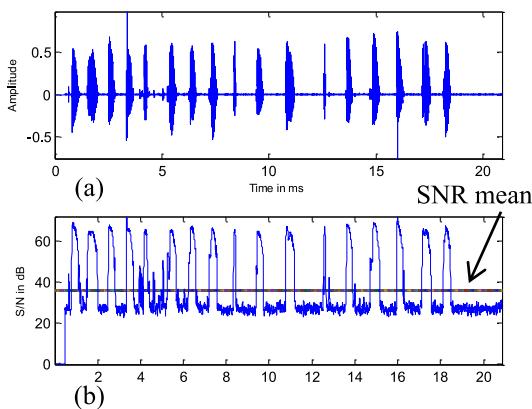


Fig. 21. (a) Signal waveform. (b) Signal-to-noise in dB and the mean value of this ratio.

For broken and rising tones, F0 of Hue voices tends to go down lower in comparison with Hanoi voices as in Figs 11, 12 and 19, 20. In contrast, for heavy and low-falling tones, F0 of Hue voices tends to go up higher than Hanoi voices as we can see from Figs 13, 14 and 17, 18. Generally speaking, the direction and the range of F0 variation for Hue tones tends to be opposed to Hanoi tones. This conclusion is also consistent with the perception in reality of the difference between the pronunciation modality for the tones of Hue voice in comparison with Hanoi voice.

To determine the signal-to-noise ratio of VDSPEC, the influence of background noise on speech signal is assumed to have properties of addition noise. This assumption is consistent with the actual condition in the recording studio. Therefore, the determination of signal-to-noise ratio is the following. During silence, which means no voice and there is only background noise, the noise power will be calculated according to the following formula:

$$P_N = \frac{1}{N} \sum_{n=0}^{N-1} b^2(n)$$

where P_N is short time power for the background noise, N is window length, $b(n)$ is background noise. With the sampling frequency 16000 Hz, N is selected by 256. Being based on assumptions of addition noise, spectrum subtraction method has been implemented. So, the power of clean speech signal is calculated as follows:

$$P_S = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n)$$

Where P_S is short time power of clean speech signal $x(n)$. Finally, the signal-to-noise ratio in dB will be:

$$SNR_{dB} = 10 \log_{10} \frac{P_S}{P_N}$$

According to the mentioned method, the signal-to-noise ratio of the corpus VDSPEC was determined and the average value of this ratio is approximately 35 dB. This value is perfectly appropriate for dialect identification and speech recognition systems.

4. Conclusions and development

This paper presents the methods and results of building a new corpus for Vietnamese taking account of tonal balance for speech recognition and Vietnamese dialect identification. The statistical analysis for the variation of fundamental frequency shows that there are distinctions in pronunciation modality of tones for Hue and Hanoi voices. The distinctions can be used as the important features in combination with other features for identifying these dialects. Our corpus will be served not only for research on dialect identification but also for Vietnamese synthesis. This corpus can be developed more completely by adding different voices and other Vietnamese dialects in the near future.

References

- [1] V.B. Le, D.D. Tran, E. Castelli, L. Besacier, and J-F. Serignat, "Spoken and written language resources for vietnamese," in LREC 2004, Lisbon, Portugal, May 26-28, 2004, vol. II, pp. 599-602
- [2] T.T. Vu, D.T. Nguyen, M.C. Luong, and J-P. Hosom, "Vietnamese large vocabulary continuous speech recognition," in INTERSPEECH 2005, Lisbon, Portugal, September, 2005.
- [3] Vu, Q., Demuyneck, K., Compernolle, D.V., "Vietnamese Automatic Speech Recognition: the FlaVoR Approach", ISCSLP 2006, Kent Ridge, Singapore, 2006.
- [4] Hoàng Thị Châu (2009). Phượng ngữ học tiếng Việt. NXB Đại học Quốc gia Hà Nội.
- [5] A Comparative Grammar of British English Dialects, Bernd Kortmann, 2005, Walter de Gruyter
- [6] Jing Li et al. , "A Dialectal Chinese Speech Recognition Framework, Journal of Comput. Sci. & Technol., Jan.2006, Vol. 21, No. 1, pp. 106-115
- [7] Theatre Supplies and Services [http://adena.co.nz/theatre/products/sound/microphone s-wired/shure/sm-series/shure-sm48.htm](http://adena.co.nz/theatre/products/sound/microphone-s-wired/shure/sm-series/shure-sm48.htm)
- [8] www.praat.org
- [9] Fadi Biadsy, Julia Hirschberg, "Using Prosody and Phonotactics in Arabic Dialect Identification". Interspeech 2009, Vol. 1, pp 208-211
- [10] Jean-Luc Rouas. "Automatic prosodic variations modelling for language and dialect discrimination". IEEE Transactions on Audio, Speech and Language Processing, V. 15, N. 6, p. 1904-1911, 2007.
- [11] Sittichok Aunkaew, Montri Karnjanadecha, Chai Wutiwiwatchai, "Development of a Corpus for Southern Thai Dialect Speech Recognition: Design and Text Preparation", The 10th International Symposium on Natural Language Processing, October 28-30, 2013, Phuket, Thailand