RECOGNIZING FORMANTS AND PITCH PERIODS FOR VIETNAMESE SPEECH BASED ON THE LOCAL MODULUS MAXIMA IN THE WAVELET DOMAIN

T. Le-Tien, H. Tran-Thanh*, D. Tran-Tien, C. Hoang-Dinh
HoChiMinh University of Technology, University of CanTho*,
(Received on February 22nd, 2001)

ABSTRACT: Based on the wavelet transform, an efficient method for recognizing Vietnamese speech has been developed. The continuous wavelet transform (CWT) is used to examine and extract important features of the Vietnamese speech. The formant information is successfully extracted from the local modulus maxima of transformed speech signals (CWT) by the proposed algorithms. In addition, the pitch period which characterizes the voiced sound of speech is also exactly obtained by a simple proposed algorithm based on the phase of CWT of speech. Using the proposed algorithms, a speaker recognition system for the Vietnamese speech has been developed. This system has successfully been experimented with a small dictionary in isolated word mode.

Keywords: Speech Recognition, Continuous Wavelet Transform, Local Modulus Maxima, Phase of Wavelet Transform, Formant, Pitch Period, Vietnamese Speech.

Introduction

Making a machine having capability to listen and speak to human has been attracted by many researchers, particularly it can be addressed to various intelligent systems for robot telecommunications, information technology, production, education, etc. However, building a system having ability of recognizing exactly speech is extremely difficult because the research object, speech, is too complicated. Very wide acoustic variations occur when the same phoneme is spoken by different people, because of differences in the vocal organ [4]. Another problem is the fact that there are no identifiable boundaries between sounds and words. Even the relatively simple task of determining where an utterance begins and ends (called end-point detection) also presents problems and is error-prone, particularly in noisy operating conditions [4]. In addition, the speech signal itself cannot always convey all of the acoustic-phonetic information required to recognize it. Besides, speech also depends on dialect and age of speaker particularly the speech of Northern, Southern and Center people of Vietnam have different features [5] and the influence of six tones of Vietnamese language complicates the process of recognizing speech as well.

In the study on speech recognition methods, many different approaches have been used such as acoustic-phonetic approach, the pattern-matching approach, the artificial intelligence approach [8] such as the waveform processing, the correlation processing and spectral processing approach based on the Fourier Transform [5] are also applied. However, because of the complicated
accent of the Vietnamese speech, these techniques are not effective [9]. Therefore, many researches on the Vietnamese speech recognition have not yet reached satisfactory results.

This research proposes and experiments a new efficient approach to solve this problem. It uses a strong tool that has been developed recently but it has been applied in many fields of research, the wavelet transform, to investigate and recognize the Vietnamese speech. Important features of speech such as formants and pitch periods are extracted by the continuous wavelet transform (CWT) to build a database for speech recognition. From this feature database, the speech recognition is processed by comparing the speech features of the unknown word with the features in the database and choosing the corresponding word to identify the suitable one. The correct words after recognizing can be presented on the screen of a computer used for the signal processing. We also applied successfully the result to control for robots through wireless communications. However due to the limit of the page number, we only address on the method to recognizing the Vietnamese speech with some results of word recognition.

Formant-Extraction method based on CWT

A speech signal is generally a non-stationary signal comprised of many frequency components. A segment of a speech can mathematically be represented to a high accuracy in the Hilbert space as the sum of exponential functions as follows,

$$x(t) = \sum_{i=1}^{N} X_i(t) e^{j\phi_i(t)} \tag{1}$$

where $X_i(t)$ and $\phi_i(t)$ are instantaneous amplitude and instantaneous phase, respectively, of the i^{th} frequency component, which $i=1, 2,..., N$ (suppose that there are N frequency components). Then the instantaneous frequency is defined as follows,

$$\omega_i(t) = \frac{d\phi_i(t)}{dt} \tag{2}$$

The continous wavelet transform (CWT) [1,5] of speech signal x(t) can be written as follows,

$$\begin{aligned} CWT_x(a,b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi^*\left(\frac{t-b}{a}\right) x(t)dt \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \sum_{i=1}^{N} X_i(t) e^{j\omega_i(t).t} \psi^*\left(\frac{t-b}{a}\right) dt \end{aligned} \tag{3}$$

Applying the linear property of CWT, we have

$$CWT_x(a,b) = \frac{1}{\sqrt{a}} \sum_{i=1}^{N} \int_{-\infty}^{\infty} X_i(t) e^{j\omega_i(t).t} \psi^*\left(\frac{t-b}{a}\right) dt \tag{4}$$

The wavelet function, $\psi\left(\frac{t-b}{a}\right)$, limits the speech signal in a short time interval Δt around the time-point $t=b$. Because speech is a signal varying slowly in time, we can assume that in this time interval, the instantaneous amplitude $X_i(t)$ and the instantaneous frequency $\omega_i(t)$ are constants [5,11,12],

$$X_i(t) = X_i(b), \quad \omega_i(t) = \omega_i(b)$$

So eq. (4) can be written again as follows

$$\begin{aligned} CWT_x(a,b) &= \frac{1}{\sqrt{a}} \sum_{i=1}^{N} X_i(b) \int_{-\infty}^{\infty} \psi^*\left(\frac{t-b}{a}\right) e^{j\omega_i(b).t} dt \\ &= \frac{1}{\sqrt{a}} \sum_{i=1}^{N} X_i(b) \int_{-\infty}^{\infty} \left[\psi\left(\frac{t-b}{a}\right) e^{-j\omega_i(b).t}\right]^* dt \\ &= \frac{1}{\sqrt{a}} \sum_{i=1}^{N} X_i(b) \left[\int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) e^{-j\omega_i(b).t} dt\right]^* \end{aligned} \tag{5}$$

(5)

We can see that $\int_{-\infty}^{\infty} \psi\left(\frac{t-b}{a}\right) e^{-j\omega_i(b).t} dt$ itself is the Fourier transform of the wavelet function, $\psi((t-b)/a)$, at instantaneous frequency $\omega_i(b)$ can be expressed as,

$$\psi\left(\frac{t-b}{a}\right)\xleftrightarrow{FT} a.\Psi(a\omega)e^{-j\omega b} \quad (6)$$

where $\psi(\omega)$ is the Fourier transform of the mother wavelet $\psi(t)$. Substitute (6) into (5) gives

$$CWT_x(a,b) = \sqrt{a}.\sum_{i=1}^{N} X_i(b)\Psi^*(a\omega_i(b))e^{j\omega_i(b)b} \quad (7)$$

If the bandpass interval of wavelet function $\Psi(a\omega)$ is narrow enough to contain only the frequency component ω_i, only this component will influences in result $CWT_x(a,b)$ and the modulus of $CWT_x(a,b)$ reaches the maximum value at $a\omega_i = \omega_0$, which ω_0 is the central frequency of wavelet function using in the transform (e.g. Morlet wavelet). Therefore, each instantaneous frequency component ω_i at time-point $t=b$ in the speech signal influences in the result analysis $|CWT_x(a,b)|$ in the neighborhood of scale a_i, which relates to the instantaneous frequency ω_i as follows,

$$a_i = \frac{\omega_0}{\omega_i} \quad (8)$$

From the localization properties of mother wavelet in Fourier domain, the modulus $|CWT_x(a,b)|$ is essentially maximum in the neighborhood of scale a_i, called *ridges* of the CWT. If the Morlet wavelet is used to analyze speech signal, the peaks of *ridges* in CWT will be the position of scales a_i, respectively with frequency component ω_i at analyzing time. If each frequency component is considered as a formant, peaks of *ridges* will correspond to formants in speech signal [5,7].

In the analyzed result, $CWT_x(a,b)$, of speech signal, x(t), peaks of *ridges* can be obtained by local modulus maxima of $|CWT_x(a,b)|$ in scale direction [7,12]. Therefore, just computing the local modulus maxima of *CWT* in the scale direction, we can obtain a picture which represents formants of the speech signal. Combining all of above parts, a proposed method for extracting formant features of speech signals is represented generally in Figure 1.

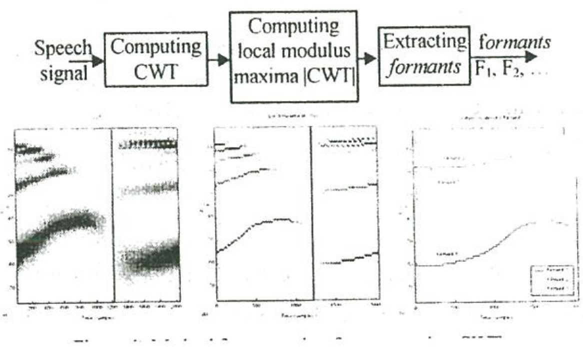


Figure1: Method for extracting formants using CWT

First of all, the analyzed speech signal is transformed in time-frequency domain by the *CWT* to obtained the *ridges* corresponding to its *formants*. Then the analyzed result is computed the local modulus maxima in the scale direction to locate exactly the position of *formants*. At this step, we obtain a picture of formants which are shown by dark horizontal line (figure 1). Finally, a suitable algorithm is used to extract time-dependent amplitudes of these formants from the local maxima [7].

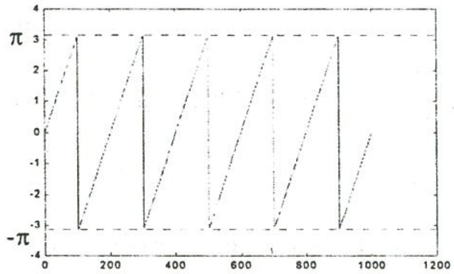


Figure 2: Phase of one frequency component ω

Method for extracting pitch periods

When analyzing a speech signal by CWT by the equation (7), we obtained the result as follows,

$$\begin{aligned} CWT_x(a,b) &= \sqrt{a}.\sum_{i=1}^{N} X_i(b)\Psi^*(a\omega_i(b))e^{j\omega_i(b)b} \\ &= \sqrt{a}.\sum_{i=1}^{N} X_i(b)\Psi^*(a\omega_i(b))e^{j\phi_i(b)} \end{aligned} \quad (9)$$

where $X_i(t)$ and $\phi_i(t)$ are instantaneous amplitude and instantaneous phase, respectively, of the frequency component ω_i of the analyzed speech signal.

The smallest frequency component in the speech signal, ω_1, is the fundamental frequency or pitch frequency. At the scale value a_1 corresponding to this frequency (equation 8), the bandpass interval of the wavelet function $\psi(a_1\omega)$ is very narrow because the frequency localization of the wavelet function at low frequency is very good [1,5]. This bandpass interval is narrow enough to contain only the fundamental ω_1. So only this frequency component have the influence on the analyzed result CWT. Therefore, only the components corresponding to the frequency ω_1 are kept in the sum of equation (9). It can be expressed as

$$CWT_x(a_1,b)=\sqrt{a_1}.X_1(b)\Psi^*(a_1\omega_1(b)).e^{j\phi_1(b)} \quad (10)$$

Module and phase of the transform are determined as follows

$$|CWT_x(a_1,b)|=\sqrt{a_1}.X_1(b)\Psi^*(a_1\omega_1(b)) \quad (11)$$

$$\measuredangle CWT_x(a_1,b)= \phi_1(b) \quad (12)$$

Thus with a fixed scale a_1 corresponding to the fundamental frequency of speech signal; phase of the transform $CWT_x(a_1,b)$, with b changes in existent time interval of signal, is the phase of pitch frequency

$$\measuredangle CWT_x(a_1,t)= \phi_1(t) \quad (13)$$

The fundamental frequency ω_1 itself is a periodic signal with the period $T=2\pi/\omega_1$, so the phase angle $\phi_1(t)$ is also periodic with period T and have value changing in an interval $[-\pi, \pi]$ (Figure 2). Therefore, phase of $CWT_x(a_1,t)$ is also a periodic signal, with the same period as pitch period (1/pitch frequency). We can use points where the phase angle changes suddently from π to $-\pi$ to mark periodic points of pitch frequency or pitch period of the analyzed speech signal.

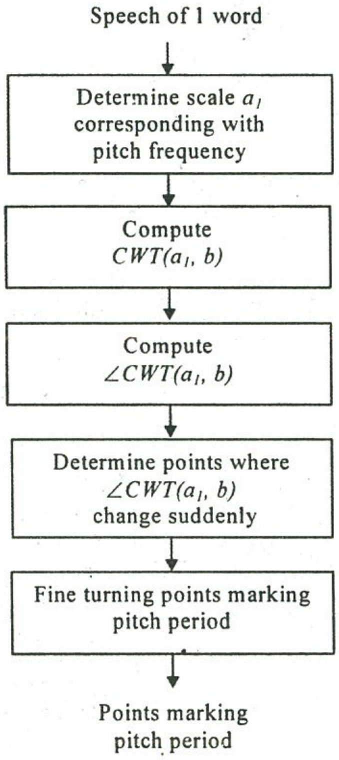


Figure 3: Method for extracting pitch period using phase of CWT

From the above discuss, an efficient approach for extracting pitch periods from speech signals is proposed based on phase of the CWT (Figure 3). The first step is determining the scale value a_l from the value of pitch frequency of the speech (equation 8). Then the wavelet transform of speech signal is taken at only one scale value a_l (1 line in CWT result). Phase of the result will be periodic with the same cycle as pitch period of speech. In practice, points where have suddenly changes in phase always correspond to each main valley of speech waveform. This points are considered as start point of each basis period of speech signal. We just need to make a fine turning to correct this point by the main negative peaks - start point - of speech. The obtained result is speech signal plus point marking start of each period. Now we can extract any pitch period according to the need for recognition between start-point and end-point (next start-point) of this period.

Vietnamese speech recognition system model

Based on the established tools for extracting important features of speech as above, a complete speech recognition system has been built. General block diagram of this system is shown in Figure 4.

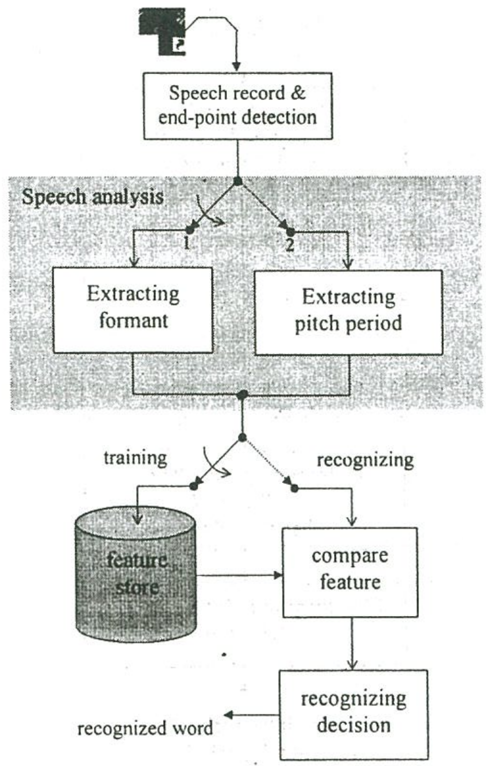


Figure 4: Block diagram of a speech recognition system

This is an isolated word system, dependent-speaker with small vocabulary. In order to recognizing speech, each speaker must train the system by speaking words in recognized vocabulary. Then the recognizing process can be executed. In this system, both proposed methods for extracting formants and pitch periods of speech are used. After speech of speaker has been recorded, a process called end-point detection is done to extract only speech signal in the recorded signal (contain silence and noise). After that, depending on which the training method or recognizing base is selected, the speech is analyzed in respectively way. If the selected recognizing base is addressed to formants, the speech will be analyzed in time-frequency domain by CWT and its formants will be extracted. Otherwise, the recognizing base is addressed to pitch periods, the obtained result will be a period having largest amplitude that corresponds to the position of vowel in speech. In training process, the speaker must pronounce all of words in the dictionary. Each word is analyzed in the same way as above and the obtained features are stored as feature templates in a suitable database. To solve the varying of speech between pronouncing of the same word, each word in the dictionary can be spoken in many times. Then the collected features are used to compute an average feature of each word. After the training process is completed, the system has ability of recognizing speech. To do this task, the speaker needs to pronounce a word in the trained dictionary. The system will analyze the recorded speech and extract important features by the similar way as in the training process. The obtained feature will be compared with all feature templates of trained words in database. Finally, base on the comparing score, the recognizing decision block will select a corresponding word which has the likely feature with the unknown word.

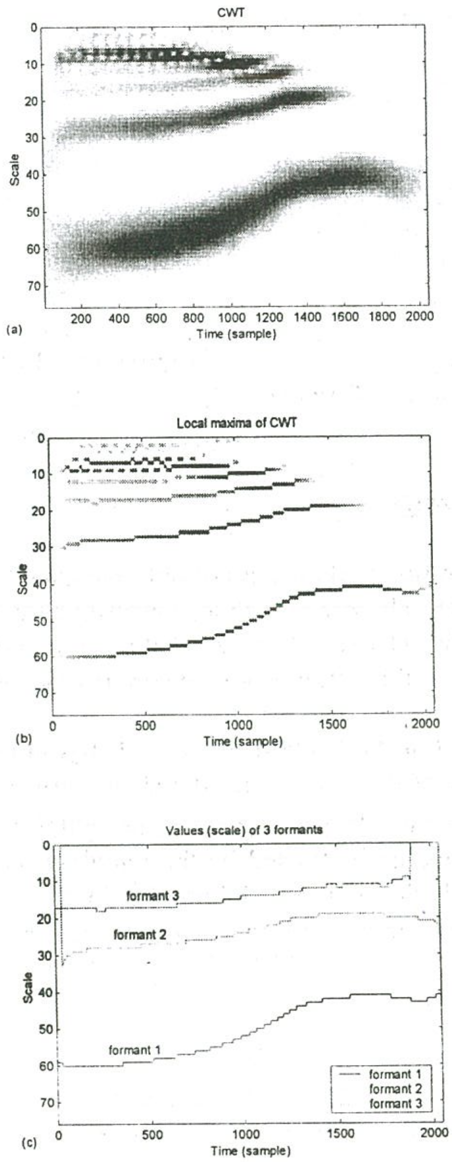
Experimental results

5.1. Extracted formant results

The result of extracting formant based on *CWT* is represented in Figure 5. Speech is sampled at 8 kHz. The mother wavelet is chosen to be the Morlet wavelet with α=2 and the modulated frequency $\omega_0 = 2\pi$. Fisrt, speech signal of a word (e.g. word 'tôùi' -'come') is analyzed by *CWT* to express formants of speech signal (Figure 5(a)). Local maxima modulus of *CWT* is computed to locate exactly the position of formants (Figure 5(b)). From the local maxima, the values and amplitudes of three formants are extracted by a suitable algorithm [7] (algorithm for extracting formants from local maxima of its CWT). In Figure 5(c), the trajectory of values of each extracted formant is similar to the line of respective formant on the local maxima. Besides, amplitude of each obtained formant is correlative with each ridge in CWT, respectively (Figure 5(d)).

5.2. Extracted pitch period results

The result of method based on phase of CWT for extracting pitch period of speech is represented in Figure 6. In this figure, amplitude of signal and phase angle are normalized in interval [-1, 1]. We can see that the phase of CWT is saw-tooth waveform whose falling edges exactly locate main negative peaks (main valleys) on the waveform of speech (Figure 6(a)). Therefore, points marking period which base on down edges of phase CWT are also the exact positions of main valleys. These points which are considered as start-point of each period are shown by vertical lines in Figure 6(b).



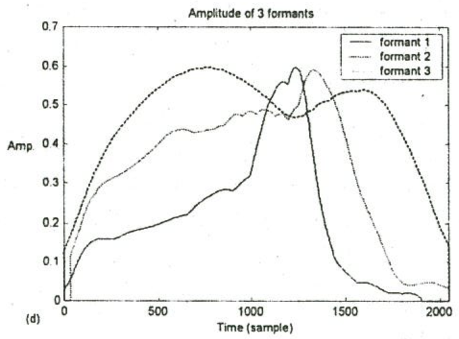
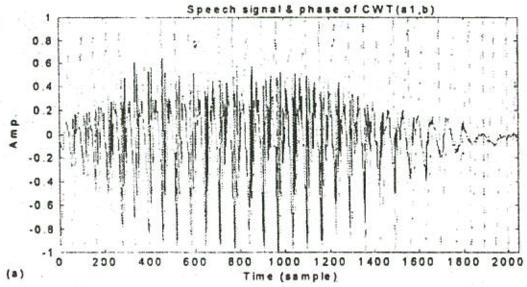


Figure 5: Results of extracting 3 formants from *CWT* ø of word 'tới'



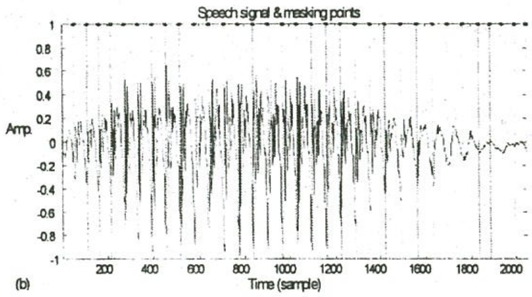


Figure 6: Result of method for extracting pitch periods using phase of *CWT* (word 'a')

5.3. Speech recognizing results

The established speech recognition system can not only recognize whole syllables but also can distinguish between Vietnamese phonemes such as vowels and tones. The base of evaluating the exactness of recognizing is the relative difference between recognized word and other trained words in dictionary. The larger this difference is, the more exact the recognizing process is.

5.3.1. Recognizing the Vietnamese vowels based on pitch periods

To test the reconizing vowels of the system, words whose scripts contain only vowels such as 'a', 'e', 'ê', 'o', ' ô', 'ơ', 'u' and 'y' are chosen to experiment. In the training process, largest periods in speech signal of these words are extracted by the *pitch period method.* A 1.5kHz lowpass filter was used to remove high frequency harmonics in the speech signal (to smooth the waveform). Figure 7 show the difference in waveforms of these periods.

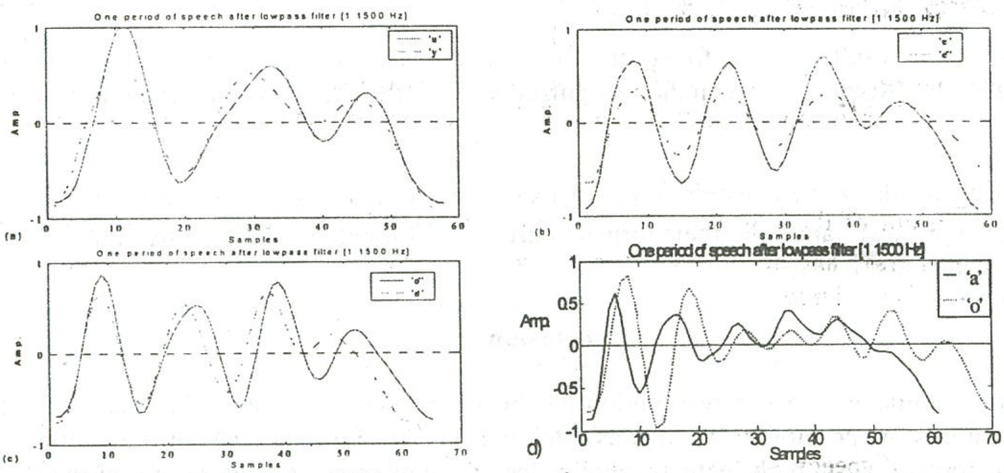


Figure 7 The pitch periods of Vietnamese vowels

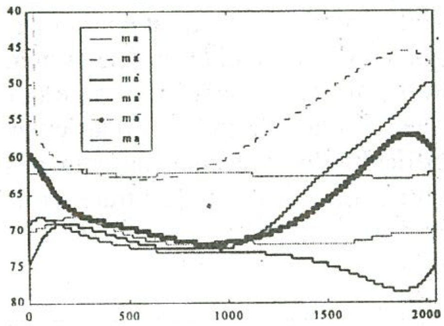


Figure 8: Pitch frequency of words:
'ma', 'má', 'mà', 'mả', 'mã', 'mạ'

The speech recognizing results based on pitch period are represented in Table 1. The recognized word is the one having smallest compared value (AMDF), which is normalized by 0%. So the *relative difference* between this word with other words is larger than 0%. Table 2 shows họw difference in period waveform with respect to this expected value.

5.3.2. Recognizing Vietnamese tones based on formant F1 extracted from the CWT.

In this part, words which contain different tones such as 'ma', 'má', 'mà', 'mả', mã' and 'mạ' are used to experiment the recognizing Vietnamese tones. In Figure 8, the difference of these tones is shown by the formant F1 of the corresponding speech signals obtained in the training process. Therefore, to recognize these tones, only the formant F1 is used. The results are shown in Table 2. For words whose difference in formant F1 are large such as 'ma', 'ma' and 'maø'; the relative difference between the recognized word with other words are quite large (above 70%). In contrast, for words whose formant F1 are less so different such as 'mả' and 'mã', 'mà' and 'mạ'; the relative difference in the recognized results are not large.

5.3.3 Recognizing Vietnamese whole syllable based on formants

To test the recognizing whole syllable, a practical dictionary which contains words 'phải', 'trái', 'tới', 'lui, 'co' and 'duỗi' is used. In this part, the system used three formant F1, F2 and F3 which are extracted by the method based on CWT.

Table 3 shows the result of recognizing process. For words whose difference in formants are large, the relative difference between the recognized word with other words are quite large (over 60%).

So recognizing result of these words is rather exact. In the other hand, for words having the same tone such as 'trái' and 'tới', their formants are less different. So the *relative difference* in recognized results is not large.

Conclusion

The study on Vietmanese speech recognition has been performed by the method based on the important features of speech signals such as pitch periods and formants. Using the continuous wavelet transform (CWT), both features were successfully extracted by the proposed methods. These methods have been implemented effectively to examine the features of Vietnamese speech. Based on these methods, a complete speech recognition system has been established. Some experiment results show that this system has the ability of recognizing rather well speech of words in small dictionary (6 or 8 control words presented above). In addition, the system can also recognize some syllabic structure components. Vowels, the component have large number of element, can be recognized well based on the waveform of pitch periods. In particular, the Vietnamese tones, the component often creates the difficulty for the analysis, can also be recognized successfully by the formant F1 derived from the continuous wavelets transform.

Spoken word	The relative difference with words								**Recognized word**
	'a'	'e'	'ê'	'o'	'ô'	'ơ'	'u'	'y'	
'a'	**0%**	51%	71%	49%	73%	77%	82%	74%	'a'
'a'	**0%**	45%	58%	51%	68%	69%	75%	66%	'a'
'e'	84%	**0%**	87%	94%	88%	94%	95%	94%	'e'
'e'	89%	**0%**	91%	96%	89%	97%	97%	96%	'e'
'ê'	89%	70%	**0%**	94%	65%	94%	94%	94%	'ê'
'ê'	89%	56%	**0%**	94%	55%	93%	95%	94%	'ê'
'o'	59%	78%	88%	**0%**	84%	64%	90%	86%	'o'
'o'	78%	85%	92%	**0%**	90%	80%	94%	92%	'o'
'ô'	73%	42%	51 %	85%	**0%**	85%	88%	88%	'ô'
'ô'	71%	22%	42%	69%	**0%**	73%	88%	89%	'ô'
'ơ'	70%	80%	89%	61%	85%	**0%**	91%	88%	'ơ'
'ơ'	51%	54%	80%	51%	74%	**0%**	82%	79%	'ơ'
'u'	75%	72%	78%	86 %	81%	86%	**0 %**	67%	'u'
'u'	81%	78%	85%	89%	86%	88%	**0%**	73%	'u'
'y'	87%	79%	85%	93%	92%	92%	48%	**0%**	'y'
'y'	98%	97%	98%	99%	99 %	99%	76%	**0%**	'y'

Table 1: The results of recognizing vowels

Speak word	The relative difference with words						Recognized word
	'ma'	'má'	'mà'	'mả'	'mã'	'mạ'	
'ma'	**0%**	94%	82%	93%	93%	94%	'ma'
'ma'	**0%**	81%	77%	73%	83%	70%	'ma'
'má'	62%	**0 %**	93%	87 %	89%	90%	'má'
'má'	98%	**0%**	99%	97%	94%	99%	'má'
'mà'	87 %	93%	**0%**	72%	79%	49%	'mà'
'mà'	90%	97%	**0%**	93%	92%	73%	'mà'
'mả'	90 %	77%	88 %	**0 %**	50 %	74%	'mả'
'mả'	62%	83%	84%	**0%**	24%	74%	'mả'
'mã'	87%	85%	92%	45%	**0%**	87%	'mã'
'mã'	82%	95%	91%	50 %	**0 %**	79%	mã
'mạ'	77 %	95%	53%	86%	83%	**0%**	'mạ'
'mạ'	55%	91%	50%	78%	61%	**0%**	'mạ'

Table 2: The results of recognizing tones, using formant F1

Speak word	The relative difference with words						Recognized word
	'phải'	'trái'	'tới'	'lui'	'co'	'duỗi'	
'phải'	**0%**	69%	46%	92 %	60%	69%	'phải'
'phải'	**0%**	79%	65 %	93 %	68%	60%	'phải'
'trái'	87%	**0%**	42%	97%	82%	95%	'trái'
'trái'	84 %	**0%**	36%	96 %	74 %	94%	'trái'
'tới'	91%	51 %	**0%**	98%	90 %	97%	'tới'
'tới'	84 %	32%	**0%**	95 %	77%	93%	'tới'
'lui'	91 %	89%	73%	**0%**	73%	81%	'lui'
'lui'	93 %	93 %	86%	**0%**	78%	83%	'lui'
'co'	83 %	91%	94%	94 %	**0%**	95%	'co'
'co'	88 %	95%	93%	84%	**0%**	95%	'co'
'duỗi'	82%	89%	75%	83%	88 %	**0%**	'duỗi'
'duỗi'	75%	91%	83%	69 %	84%	**0%**	'duỗi'

Table 3: The results of recognizing whole syllable, using formant F1, F2 and F3

MỘT GIẢI PHÁP NHẬN DẠNG TIẾNG VIỆT HIỆU QUẢ DỰA TRÊN PHÉP BIẾN ĐỔI WAVELETS

Lê Tiến Thường, Trần Thanh Hùng, Trần Tiến Đức, Hoàng Đình Chiến
Đại học Bách Khoa Thành phố Hồ Chí Minh, Đại học Cần Thơ

TÓM TẮT: *Dựa trên phép biến đổi Wavelets , một phương pháp rất hiệu quả để nhận dạng tiếng nói tiếng Việt đã được phát triển. Phép biến đổi Wavlets liên tục được dùng để khảo sát và trích ra những đặc điểm quan trọng của tiếng nói tiếng Việt. Thông tin về formant được trích ra một cách chính xác từ các cực đại địa phương của tiếng nói đã được biến đổi dựa trên một giải thuật đề nghị. Bên cạnh đó, chu kỳ cao độ là một đặc điểm quan trọng của tiếng nói cũng được trích ra một cách chính xác với giải thuật dựa trên pha của biến đổi wavelet liên tục của tiếng nói tiếng Việt. Với giải thuật đề nghị, một hệ thống nhận dạng tiếng nói tiếng Việt phụ thuộc người nói đã được thiết lập. Hệ thống này đã được triển khai thành công với bộ từ vựng nhỏ và độc lập.*

CÁC TỪ KHÓA: Nhận dạng tiếng nói, Phép biến đổi wavelets liên tục, Cực đại địa phương, Pha của phép biến dổi wavelet, Formant, Chu kỳ cao độ, Tiếng nói tiếng Việt.

REFERENCES

[1] M. Vetterli, 'WAVELETS AND SUBBAND CODING', Prentice Hall, 1995, ISBN 0-13-097080-8

[2] M. Misiti, Y. Misiti, G. Oppenheim, J. Michel Poggi, 'MATLAB – Wavelet Toolbox User's Guide', 1997

[3] Robi Polikar, 'The Engineer's Ultimate Guide to Wavelet Analysis", http://www.public.iastate.edu/~rpolikar/WAVELETS/WTtutorial.html, 1996.

[4] F.J. Owens, 'SIGNAL PROCESSING OF SPEECH', MACMILLAN Publisher, 1993, ISBN 0-333-51921

[5] T. Le-Tien, 'A study based on the Continuous Wavelet Transform for the Vietnamese Speech Processing', Proceedings of the 4th International Conference On Neural Information Processing and Intelligent Information Systems ICONIP'97, University of Otago, Dunedin, New Zealand, Nov 1997, pp 1072-1075.

[6] R.K. Balani, 'Talking & Listening Robots or Machines', Proceedings of the International Symposium on Intelligent Robotics, 1993, Bangalore

[7] T.Le-Tien, H. Talhami, D.T. Nguyen, 'Simple algorithm for wavelet maxima modulus extraction in time-scale representation', IEE Electronic Letter, An Internaltional Publication, England, Feb 97, Vol.33.

[8] Many authors, 'Data Signal processing', CRC Press LLC, 1998, ISBN 0-8493-8572

[9] H. Tran - Thanh, 'A STUDY ON VIETNAMESE SPEECH RECOGNITION FOR ROBOT COMMUNICATIONS USING THE WAVELETS TRANSFORM', Honors

Master Thesis, August 2000, Undersupervising of Dr Thuong Le-Tien, HCM Uni. of Technology, Vietnam.

[10] T. Le-Tien, " Some issues of Wavelet Functions for Instantaneous Frequency Extraction in Speech Signals", Proceedings of IEEE International Conference on Speech and Image Technologies for Computing and Telecommunications, TENCON'97, CD-ROM, Queensland University of Technology, Brisbane, Australia, pp 31-34, 2-4 December 1997.

[11] T. Le-Tien, L. Huu Nguyen, "An efficient algorithm for ridge extraction in time-scale and time-frequency representations", Special issue: Research and Development on Informatics and Telecommunications Technology, Vietnam DGPT-Post and Telecommunications Journal, 2-2000, pp17-28.

[12] Carmona, R.A. Hwang, W.L. and Torresani, 'The characterization of signals by the ridges of their wavelet transforms', Technical Report, Department of Mathematics, university of California, March 1995.