

# Phương pháp mới trích chu kỳ cao độ trung bình ứng dụng trong nhận dạng thanh điệu tiếng Việt

## A New Approach Applied to Average Pitch Period Extraction for the Vietnamese Tone Recognition

Lê Tiến Thường, Huỳnh Ngọc Phiên, Trần Tiến Đức

**Abstract:** Vietnamese is a monosyllabic and tonal language. Tone recognition is important because tone affects the lexical identification of words. Tone is the pitch period or fundamental frequency ( $F_0$ ) contour in a syllable. In this paper, we propose a new approach to locate the average pitch period which smooths the tones and a new algorithm to define the average pitch period is also designed. More than 300 lines of poetry from the Truyện Kiều written by Nguyễn Du, where each line is recorded one time only, are used to train the neural networks and 200 lines of poetry another are used to test. The result of recognition which has average accuracy of 95,4% shows that the approach of solving the problem of Vietnamese tone recognition is perfectly suitable.

**Keywords:** Pitch period, fundamental frequency,  $F_0$  contour, tone, Vietnamese speech recognition, neural networks.

### I. ĐẶT VẤN ĐỀ

Bài toán nhận dạng thanh điệu đã được nhiều tác giả trên thế giới quan tâm đặc biệt tại các nước châu Á do ngôn ngữ của họ có hiện tượng thanh điệu chẳng hạn như tiếng Việt, tiếng Thái, tiếng Hán. Bởi vì chức năng của thanh điệu là phân biệt nghĩa của từ nên việc nhận dạng thanh điệu sẽ giảm đáng kể số lượng các âm tiết cần nhận dạng. Tiếng Việt có sáu thanh, tiếng Thái có năm thanh, tiếng Hán phổ thông có bốn thanh. Thanh điệu là sự thay đổi chu kỳ cao độ hay tần số cơ bản  $F_0$ , hay đường biểu diễn thanh điệu trong một âm

tiết là đường vận động chu kỳ cao độ trong âm tiết đó. Vì vậy khi nhận dạng thanh điệu cần phải trích chu kỳ cao độ và yêu cầu đường biểu diễn thanh điệu của một âm tiết phải trơn. Tất cả các nghiên cứu về nhận dạng thanh điệu trong gần mười năm qua cho tiếng Thái, tiếng Hán đều dùng phương pháp trích chu kỳ cao độ bằng hàm tự tương quan [7], [9], hay hàm hiệu biên độ trung bình [4], một biến thể của hàm tự tương quan, hay thuật toán dò theo bộ lọc ngược đơn giản [8] mà về thực chất cũng là hàm tự tương quan. Khuyết điểm vốn có của phương pháp trích chu kỳ cao độ bằng hàm tự tương quan nằm ở chỗ giá trị chu kỳ cao độ không trơn do đó dẫn đến kém chính xác khi nhận dạng, mặc dù trong [7] có áp dụng phương pháp phân tích-tổng hợp thanh điệu khá tinh vi để làm trơn đường vận động chu kỳ cao độ sau khi trích chu kỳ cao độ bằng hàm tự tương quan. Riêng [5] đã đi theo một hướng khác là nhận dạng thanh điệu bằng phương pháp trích chu kỳ cao độ dựa vào các cực trị cục bộ lấy trên phần tín hiệu wavelets theo nghiên cứu của [6], việc xác định chu kỳ cao độ tức thời tuy có độ chính xác cao, nhưng nếu chỉ căn cứ vào chu kỳ cao độ tức thời để suy ra đường vận động chu kỳ cao độ trong toàn bộ âm tiết thì đường này sẽ không trơn nên nhận dạng thanh điệu sẽ kém chính xác.

Trong nghiên cứu này, một phương pháp mới được đề nghị để trích chu kỳ cao độ trung bình làm cho thanh điệu trơn hơn và phù hợp cho nhận dạng thanh

điều, cùng với phương pháp mới này là thuật toán mới để xác định chu kỳ cao độ trung bình. Kết quả được thử nghiệm trên ba tín hiệu là tín hiệu sau lọc thông thấp, tín hiệu chi tiết trên phần wavelets và tín hiệu xấp xỉ trên phần tỷ lệ của biến đổi wavelets, trong đó bộ dữ liệu học và thử tương đối đầy đủ gồm hơn 500 câu thơ trong Truyện Kiều của Nguyễn Du. Bài báo được tổ chức như sau: Phần 2 nêu phương pháp trích chu kỳ cao độ trung bình, Phần 3 là thử nghiệm và Phần cuối là nhận xét và kết luận.

## II. TRÍCH CHU KỲ CAO ĐỘ TRUNG BÌNH

### 1. Định nghĩa 1

Chu kỳ cao độ của tín hiệu tiếng nói là thời gian trôi qua giữa hai xung thanh môn liên tiếp. Việc đo bắt đầu ở một thời điểm xác định trong một chu trình thanh môn, tốt nhất ở thời điểm đóng thanh môn hay nếu thanh môn không đóng hoàn toàn thì ở điểm mà diện tích thanh môn nhỏ nhất [11, trang 475]. Thuật toán phát hiện cao độ của [6] theo định nghĩa này. Nếu chỉ căn cứ vào giá trị tức thời của chu kỳ cao độ để xác định đường vận động chu kỳ cao độ mà không loại bỏ các chu kỳ cao độ bị biến dạng thì đường này sẽ không trơn nên nhận dạng sẽ kém chính xác.

### 2. Định nghĩa 2

Chu kỳ cao độ là độ dài trung bình của một vài chu kỳ, là thời gian trôi qua trung bình của một số ít chu trình kích thích liên tiếp. Xác định giá trị trung bình như thế nào và trên bao nhiêu chu kỳ phụ thuộc từng phương pháp trích chu kỳ cao độ [11, trang 477]. Các thuật toán xác định chu kỳ cao độ trung bình theo hàm tự tương quan, hàm hiệu biên độ trung bình theo định nghĩa này.

Đối với bài toán nhận dạng thanh điệu yêu cầu chu kỳ cao độ của một khung phải trơn, do đó chu kỳ cao độ trung bình theo Định nghĩa 2 phù hợp hơn so với chu kỳ cao độ tức thời của Định nghĩa 1. Như vậy khung phải đủ dài để đảm bảo có nhiều chu kỳ trong khung đó. Chúng tôi theo Định nghĩa 2, nhưng quan điểm trung bình do chúng tôi đề nghị làm cho chu kỳ cao độ trung bình trơn hơn so với quan điểm trung

bình theo hàm tự tương quan hay trung bình hiệu biên độ. Trước tiên, ta hãy nêu quan điểm trung bình của hàm tự tương quan, hàm hiệu biên độ và của chúng tôi, đồng thời xem xét ví dụ minh họa để làm sáng tỏ sự khác biệt này.

### 3. Hàm tự tương quan

Hàm tự tương quan - AutoCorrelation Function (ACF) trên khung tiếng nói có độ dài  $N$  [2][11] là

$$r(p) = \sum_{n=0}^{N-1-p} s(n)s(n+p) \quad (1)$$

trong đó  $p$  được giới hạn trong vùng tần số cơ bản, đối với người nam vùng tần số cơ bản thường từ 80 Hz đến 200 Hz. Nếu tần số lấy mẫu tín hiệu tiếng nói là 16 kHz, thì giá trị của  $p$  sẽ từ 80 đến 200. Chu kỳ cao độ  $P_0$  được chọn sao cho  $r(P_0)$  đạt giá trị lớn nhất.

### 4. Hàm hiệu biên độ trung bình

Hàm hiệu biên độ trung bình - Average Magnitude Difference Function (AMDF) trên khung tiếng nói có độ dài  $N$  [2][11] là

$$d(p) = \sum_{n=0}^{N-1-p} |s(n) - s(n+p)| \quad (2)$$

Chu kỳ cao độ  $P_0$  được chọn sao cho  $d(P_0)$  đạt giá trị nhỏ nhất.

Hàm hiệu biên độ trung bình khác với hàm tự tương quan ở chỗ nó dùng phép trừ thay cho phép nhân nên tốc độ tính toán rất nhanh. Khi nhận dạng thanh điệu tiếng Việt, phương pháp trích chu kỳ cao độ tức thời của [6] không thể tốt bằng phương pháp trích chu kỳ cao độ trung bình theo AMDF vì đường vận động chu kỳ cao độ của AMDF mang tính trung bình nhiều hơn nên trơn hơn.

### 5. Hàm chu kỳ cao độ trung bình mới

Khi khảo sát các đỉnh cực trị cục bộ theo [6], chúng tôi nhận thấy hiện tượng có một số chu kỳ cao độ bị biến dạng. Điều này dẫn đến ý nghĩ là phải loại bỏ các chu kỳ bị biến dạng này. Hàm chu kỳ cao độ trung bình - Average Pitch Period Function (APPF) trong một khung được định nghĩa theo đúng toán học là khoảng cách trung bình của các đỉnh cực đại hay cực tiểu giả tuần hoàn hợp lý nhất (nhiều nhất). Hàm như

vậy có khả năng loại bỏ những chu kỳ biến dạng vì ta chỉ tính chu kỳ cao độ trung bình trên các đỉnh giả tuần hoàn,

$$P0 = \frac{1}{M} \sum_{m=0}^{M-1} d_m \quad (3)$$

trong đó  $d_m$  là khoảng cách giữa hai đỉnh cực đại hay cực tiểu giả tuần hoàn,  $M$  là số lượng các khoảng cách trên. Tư tưởng xét chu kỳ cao độ dựa trên đỉnh cực đại hay cực tiểu cũng được Reddy (1967) [11, trang 198] bàn đến, nhưng điều chúng tôi quan tâm là xác định khoảng cách trung bình toán học của chu kỳ một cách thích nghi.

## 6. Ví dụ

Ví dụ này minh họa các điểm cực đại hay đỉnh giả tuần hoàn hợp lý nhất và xác định chu kỳ cao độ trung bình. Cho khung dữ liệu  $s = [5 \ 0 \ 0 \ 6 \ 0 \ 0 \ 7 \ 0 \ 0 \ 0 \ 8 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 7 \ 0 \ 0 \ 8 \ 0 \ 0 \ 0 \ 0 \ 0 \ 9 \ 0 \ 0 \ 0 \ 0 \ 0 \ 8]$ . Các đỉnh được gọi là giả tuần hoàn nếu khoảng cách giữa hai đỉnh không lệch quá một so với khoảng cách giữa hai đỉnh khác và chu kỳ cao độ  $p$  từ 3 đến 7. Như vậy ta có các đỉnh giả tuần hoàn thứ nhất tại  $s[0]$ ,  $s[3]$ ,  $s[6]$ ,  $s[10]$ ,  $s[21]$ ,  $s[24]$  và các đỉnh giả tuần hoàn thứ hai tại  $s[24]$ ,  $s[30]$ ,  $s[37]$ .

Khoảng cách giữa  $s[0]$  và  $s[3]$  là  $d_0 = 3 - 0 = 3$

Khoảng cách giữa  $s[3]$  và  $s[6]$  là  $d_1 = 6 - 3 = 3$

Khoảng cách giữa  $s[6]$  và  $s[10]$  là  $d_2 = 10 - 6 = 4$

Khoảng cách giữa  $s[21]$  và  $s[24]$  là  $d_3 = 24 - 21 = 3$

Có tất cả 4 khoảng cách của các đỉnh giả tuần hoàn thứ nhất. Tương tự

Khoảng cách giữa  $s[24]$  và  $s[30]$  là  $d_0 = 30 - 24 = 6$

Khoảng cách giữa  $s[30]$  và  $s[37]$  là  $d_1 = 37 - 30 = 7$

Có tất cả 2 khoảng cách của các đỉnh giả tuần hoàn thứ hai.

Các đỉnh giả tuần hoàn hợp lý nhất là các đỉnh có số lượng khoảng cách nhiều hơn, đó là các đỉnh giả tuần hoàn thứ nhất. Theo quan điểm về chu kỳ cao độ trung bình của chúng tôi thì

$$P0 = \frac{d_0 + d_1 + d_2 + d_3}{4} = \frac{3 + 3 + 4 + 3}{4} = 3.25$$

Trong khi đó theo (1) thì  $r(3) = 128$ ,  $r(4) = 56$ ,  $r(5) = 0$ ,  $r(6) = 107$ ,  $r(7) = 120$ . Giá trị lớn nhất là  $r(3) = 128$  hay  $P0 = 3$  (xấp xỉ với  $P0$  của các đỉnh giả tuần hoàn thứ nhất).

Theo (2) thì  $d(3) = 67$ ,  $d(4) = 83$ ,  $d(5) = 97$ ,  $d(6) = 71$ ,  $d(7) = 62$ .

Giá trị nhỏ nhất là  $d(7) = 62$  hay  $P0 = 7$  (xấp xỉ với  $P0 = \frac{d_0 + d_1}{2} = \frac{6 + 7}{2} = 6.5$  của các đỉnh giả

tuần hoàn thứ hai). Rõ ràng giá trị  $P0$  của chúng tôi mang tính trung bình nhiều hơn nên tròn hơn. Nói cách khác, chúng tôi đã áp dụng bộ lọc trung bình di chuyển thích nghi trên các đỉnh giả tuần hoàn hợp lý nhất.

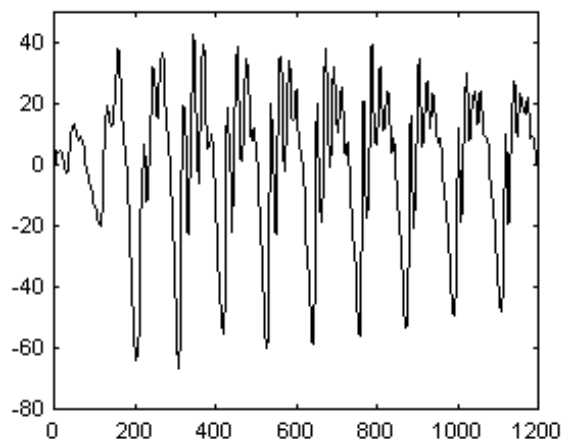
## 7. Thuật toán tính hàm APPF

Tiếng nói được lấy mẫu ở tần số 16 kHz, độ dài khung 1200 mẫu. Khung tín hiệu được lọc bằng bộ lọc thông thấp có tần số cắt 500 Hz và đáp ứng tần số bằng phẳng nhất rồi xén tín hiệu theo ngưỡng động của [3] để giữ lại các đỉnh lớn. Ta lần lượt xử lý những đỉnh dương rồi đến những đỉnh âm. Các đỉnh dương được phát hiện bằng cách xét dấu đạo hàm cấp một, đỉnh chỉ được ghi nhận khi khoảng cách giữa hai đỉnh lớn hơn 40 nhằm loại bỏ đỉnh của các formant hay tần số cao. Xác định khoảng cách giữa các đỉnh, khoảng cách này được chọn nếu nó nằm trong phạm vi tần số cơ bản là lớn hơn 80 và nhỏ hơn 200. Sắp xếp các khoảng cách này, điều này tương đương với việc chuyển khoảng cách sang miền tần số, nếu sắp khoảng cách theo thứ tự tăng thì tần số cao đứng trước, tần số thấp đứng sau. Tính giá trị trung bình giữa các khoảng cách giả tuần hoàn, khoảng cách được gọi là giả tuần hoàn nếu độ lệch giữa hai khoảng cách không vượt quá 5. Những đỉnh âm cũng được xử lý tương tự. Cuối cùng ta chọn  $P0$  có số lượng khoảng cách giữa các đỉnh giả tuần hoàn hợp lý nhất (nhiều nhất), rồi suy ra  $F0 = 16000 / P0$ .

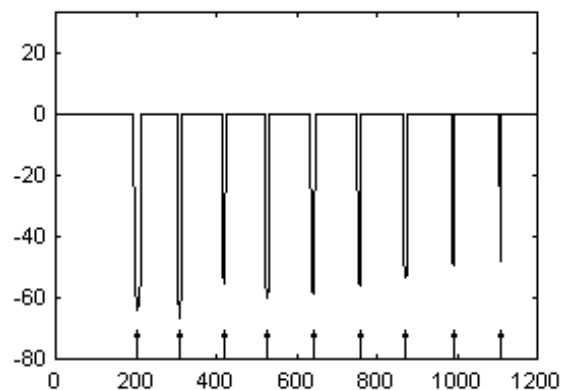
## 8. Độ phức tạp tính toán

Khi thực hiện hàm AMDF ta phải duyệt khung (1200 mẫu)  $200 - 80 = 120$  lần, trong khi hàm APPF duyệt khung 1 lần khi tính đạo hàm cấp một và xét dấu đạo hàm để xác định đỉnh dương, thời gian thực hiện các bước phía sau không đáng kể do số lượng các đỉnh trong một khung thường nhỏ hơn 12. Tổng cộng cho các đỉnh dương và đỉnh âm, số lần duyệt khung là 2. Nếu thời gian thực hiện các phép tính có thể bỏ qua, thì thời gian thực hiện hàm APPF nhanh hơn thời gian thực hiện hàm AMDF  $120 / 2 = 60$  lần.

Hình 1 là tín hiệu xấp xỉ ở mức 5. Tín hiệu này có được khi ta thực hiện biến đổi wavelet đôi - Dyadic Wavelet Transform (DyWT) [6] cho khung đầu tiên của âm tiết *trăm* có độ dài 1200 mẫu. Mục đích của việc chọn tín hiệu ở mức 5 là để giảm tối đa ảnh hưởng của nhiễu và các formant, nhưng vẫn giữ đầy đủ phạm vi tần số cơ bản. Lấy ngưỡng tín hiệu này, ta được tín hiệu trên Hình 2 gồm nhiều đỉnh lớn. Áp dụng thuật toán APPF cho tín hiệu trên Hình 2, ta sẽ được chu kỳ cao độ trung bình.



Hình 1. Tín hiệu xấp xỉ lấy ở mức 5 của DyWT cho khung đầu tiên trong âm tiết “trăm” gồm 1200 mẫu.



Hình 2. Tín hiệu thu được sau khi lấy ngưỡng tín hiệu ở Hình 1 gồm nhiều đỉnh lớn.

## III. THỬ NGHIỆM

Tiếng nói được phát âm bằng giọng nam của người miền nam không phân biệt thanh ngã và hỏi trong môi trường văn phòng làm việc bình thường có tần số lấy mẫu 16 kHz, độ dài khung 1200 mẫu, độ dời 300 mẫu cho âm tiết dài và 150 mẫu cho âm tiết ngắn, đối với âm tiết quá ngắn, độ dài khung là 400 mẫu. Dữ liệu học gồm 300 câu thơ đầu tiên trong Truyện Kiều của Nguyễn Du và thêm 120 thanh nặng được chọn từ các câu thơ thứ 1000 trở đi, mỗi câu thơ chỉ ghi một lần. Tổng cộng gồm 708 thanh ngang, 559 thanh huyền, 209 thanh ngã-hỏi, 387 thanh sắc và 339 thanh nặng. Dữ liệu thử gồm 200 câu thơ từ câu thứ 301 đến 500 trong Truyện Kiều, mỗi câu thơ chỉ ghi một lần, gồm 467 thanh ngang, 372 thanh huyền, 154 thanh ngã-hỏi, 246 thanh sắc và 146 thanh nặng.

Trình tự các bước thử nghiệm được tiến hành theo [3] gồm lọc thông thấp, xén tín hiệu theo ngưỡng động, xác định  $F_0$  bằng cách thực hiện hàm AMDF hay hàm APPF, hiệu chỉnh các khung có  $F_0 = 0$ , lọc trung vị hay lọc trung bình di chuyển, chuẩn hóa chiều dài đường biểu diễn thanh điệu  $F_0$  của âm tiết thành cố định với chiều dài  $L = 5$ , nội suy bằng đa thức bậc hai, chọn  $L$  điểm nội suy cách đều và  $L$  đạo hàm cấp một, chuyển sang thang ê-xi-ben để tăng độ phân biệt, cuối cùng đưa vào mạng nơ-ron phân tầng gồm 3 mạng nơ-ron để nhận dạng 5 thanh điệu; mạng

thứ nhất dùng để nhận dạng 3 thanh: ngang-sắc, huyền-nặng và ngã-hỏi; mạng thứ hai dùng để nhận dạng 2 thanh: ngang và sắc sau khi mạng thứ nhất đã nhận dạng mẫu thử là thanh ngang-sắc; mạng cuối cùng dùng để nhận dạng 2 thanh: huyền và nặng sau khi mạng thứ nhất đã nhận dạng mẫu thử là thanh huyền-nặng.

Kết quả nhận dạng tính bằng tỷ lệ phần trăm và được trình bày theo hàng ngang, chẳng hạn khi thử nghiệm thanh ngang, ta thấy tỷ lệ nhận dạng đúng thanh ngang là 99,2%, nhầm sang thanh sắc là 0,4%, nhầm sang thanh nặng là 0,4%.

- **Thử nghiệm 1.** Thực hiện hàm AMDF trên tín hiệu lọc thông thấp [3].

*Bảng 1. Kết quả Thử nghiệm 1*

Thanh	Ngang	Huyền	Ngã-Hỏi	Sắc	Nặng
Ngang	99,2			0,4	0,4
Huyền		91,6	0,6		7,8
Ngã-Hỏi	1,0		96,1	1,9	1,0
Sắc	8,5	0,1	0,3	91,1	
Nặng	0,6	14,3			85,1
Độ chính xác trung bình: 92,6					

- **Thử nghiệm 2.** Thực hiện hàm APPF trên tín hiệu lọc thông thấp.

*Bảng 2. Kết quả Thử nghiệm 2*

Thanh	Ngang	Huyền	Ngã-Hỏi	Sắc	Nặng
Ngang	99,3			0,6	0,1
Huyền	0,2	92,7	0,4		6,7
Ngã-Hỏi		0,4	97,4	1,8	0,4
Sắc	2,9			97,1	
Nặng	0,6	12,9			86,5
Độ chính xác trung bình: 94,6					

- **Thử nghiệm 3.** Thực hiện hàm APPF trên tín hiệu chi tiết của DyWT ở mức 5 theo [6].

*Bảng 3. Kết quả Thử nghiệm 3*

Thanh	Ngang	Huyền	Ngã-Hỏi	Sắc	Nặng
Ngang	99,0			0,6	0,4
Huyền		91,9	0,6		7,5
Ngã-Hỏi	0,5		96,1	1,4	2,0
Sắc	3,3		0,4	96,3	
Nặng		8,1			91,9
Độ chính xác trung bình: 95,0					

- **Thử nghiệm 4.** Thực hiện hàm APPF trên tín hiệu xấp xỉ của DyWT ở mức 5.

*Bảng 4. Kết quả Thử nghiệm 4*

Thanh	Ngang	Huyền	Ngã-Hỏi	Sắc	Nặng
Ngang	99,1			0,6	0,3
Huyền		93,1	0,2		6,7
Ngã-Hỏi		0,6	94,8	1,3	3,3
Sắc	2,8		0,3	96,9	
Nặng		7,0			93,0
Độ chính xác trung bình: 95,4					

## IV. NHẬN XÉT VÀ KẾT LUẬN

Trong bốn thử nghiệm trên, nhận xét có thể thấy độ chính xác của thanh huyền và nặng kém nhất, điều này là do đường biểu diễn thanh điệu của thanh huyền và nặng cùng thoải thoải đi xuống nên dễ lẫn vào nhau. Các Thử nghiệm 2, 3, 4 dùng hàm APPF cho kết quả tốt hơn so với hàm AMDF do loại bỏ các chu kỳ biến dạng, nên chu kỳ cao độ trung bình trong một khung trơn hơn dẫn đến đường biểu diễn thanh điệu trong toàn bộ âm tiết gồm nhiều khung sẽ trơn hơn, do đó kết quả nhận dạng sẽ chính xác hơn. Thử nghiệm 3 có độ chính xác cao hơn Thử nghiệm 2 vì biến đổi wavelet dùng hàm wavelet là đạo hàm cấp một của một hàm trơn [6] có khả năng phát hiện chính xác các đỉnh cực trị địa phương, đặc biệt đối với tín hiệu trong vùng chuyển tiếp do đóng thanh môn gây ra. Thử nghiệm 4 có độ chính xác cao nhất do tín hiệu xấp xỉ

được khai triển trên cơ sở địa phương wavelets nên độ phân biệt giữa các đỉnh giả tuần hoàn rõ hơn và mang tính trung bình nhiều hơn so với các đỉnh lấy trên tín hiệu chi tiết của biến đổi wavelet.

Chúng tôi đã đề nghị hàm APPF trơn hơn hàm ACF và AMDF khi xác định chu kỳ cao độ trung bình trong một khung tiếng nói, đồng thời thuật toán tính hàm APPF cũng được xây dựng. Kết quả nhận dạng có độ chính xác trung bình 95.4% khi áp dụng hàm APPF cho tín hiệu xấp xỉ trên cơ sở wavelets ở mức 5 cho thấy hướng giải quyết vấn đề của bài viết phù hợp với nhiệm vụ nhận dạng thanh điệu tiếng Việt.

## TÀI LIỆU THAM KHẢO

- [1] ĐOÀN THIÊN THUẬT, *Ngữ âm tiếng Việt*, NXB Đại học và Trung học Chuyên nghiệp, 1980.
- [2] F. J. OWENS, *Signal processing of speech*, Macmillan, London, 1993.
- [3] LÊ TIẾN THƯỜNG, TRẦN TIẾN ĐỨC, *Nhận dạng thanh điệu tiếng nói tiếng Việt bằng mạng nơ-ron phân tầng*, Tạp chí Tin học và Điều khiển học, 2003.
- [4] N. THUBTHONG, B. KIJSIRIKUL, *Tone Recognition of Continuous Thai Speech Under Tonal Assimilation and Declination Effects Using Half-Tone Model*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 9, No. 6, 2001, World Scientific Publishing.
- [5] QUOC-CUONG NGUYEN, E. CASTELLI, NGOC-YEN PHAM, *Tone Recognition for Vietnamese*, herakles.imag.fr/castelli/masters\_ts/Eurospeech.pdf
- [6] S. KADAMBE, G. F. BOUDREAU-BARTELS, *Application of the Wavelet Transform for Pitch Detection of Speech Signals*, IEEE Transactions on Information Theory, Vol. 3, No. 7, March 1992.
- [7] S. POTISUK, M. P. HARPER, J. GANDOUR, *Classification of Thai Tone Sequences in Syllable-Segmented Speech Using the Analysis-by-Synthesis Method*, IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 1, January 1999.
- [8] S.-H. CHEN, Y.-R. WANG, *Tone Recognition of Continuous Mandarin Speech Based on Neural Networks*, IEEE Transactions on Speech and Audio Processing, Vol. 3, No. 2, March 1995.
- [9] T. LEE, P. C. CHING, *Cantonese Syllable Recognition Using Neural Networks*, IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 4, July 1999.
- [10] TRINH ANH TUẤN, *Nghiên cứu các đặc trưng để phân tích và tổng hợp tín hiệu âm tần*, Luận án tiến sĩ kỹ thuật, Học viện Công nghệ Bưu chính Viễn thông, Hà Nội, 2000.
- [11] W. HESS, *Pitch Determination of Speech Signals*, Springer-Verlag Publisher, 1983.

**Ngày nhận bài: 15/12/2003**

---

## SƠ LƯỢC TÁC GIẢ

### LÊ TIẾN THƯỜNG

Sinh năm 1957 tại TP. Hồ Chí Minh.

Đã nhận bằng kỹ sư năm 1981, tiến sĩ năm 1998 tại Đại học Tasmania, Australia ngành Điện tử-Viễn thông. Được phong Phó Giáo sư năm 2002.

Hiện công tác tại Khoa Điện - Điện tử, Đại học Bách Khoa TP. HCM.

Lĩnh vực nghiên cứu: Xử lý tín hiệu, Thông tin số, xử lý tín hiệu radar, wavelets và ứng dụng, neural-fuzzy systems.

### HUỖNH NGỌC PHIÊN

Sinh năm 1944 tại Quảng Ngãi.

Đã nhận bằng cử nhân giáo khoa toán năm 1968, cử nhân khoa học toán năm 1970 tại ĐH Huế, thạc sĩ năm 1976 và tiến sĩ năm 1978 tại Học viện Công nghệ Châu Á - AIT, Thái Lan.

Hiện đang là giáo sư và hiệu trưởng của Trường Công nghệ Tân tiến - Học viện Công nghệ Châu Á.

Lĩnh vực nghiên cứu: Ứng dụng mạng nơ-ron, mô hình mô phỏng, thống kê ứng dụng.

**TRẦN TIẾN ĐỨC**

Sinh năm 1962 tại Quảng Ngãi.

Nhận được bằng kỹ sư điện tử tại Đại học Bách khoa TP.HCM, năm 1987, thạc sĩ tin học tại Đại học Khoa học Tự nhiên TP. HCM năm 1998.

Hiện đang là giảng viên Khoa Công nghệ thông tin, Đại học Sư phạm Kỹ thuật TP.HCM.

Lĩnh vực nghiên cứu: Xử lý tiếng nói, mạng nơ-ron, wavelets.