

## PHƯƠNG PHÁP NÂNG CAO CHẤT LƯỢNG TIẾNG NÓI BẰNG CÁCH TRIỆT NHIỀU THÀNH PHẦN XẤP XỈ VÀ THÀNH PHẦN CHI TIẾT TRÊN MIỀN WAVELET

Đỗ Huy Khôi, Nguyễn Thành Trung, Trịnh Văn Hà\*

Trường Đại học Công nghệ Thông tin & Truyền thông – ĐH Thái Nguyên

### TÓM TẮT

Kỹ thuật triệt nhiễu và nâng cao chất lượng tiếng nói dùng phép biến đổi Wavelet đã được nghiên cứu nhiều trên thế giới. Hầu hết các nghiên cứu tập trung vào cách ước lượng và đặt ngưỡng toàn cục cho toàn bộ tín hiệu. Bài báo này trình bày phương pháp triệt nhiễu nâng cao chất lượng tiếng nói bằng biến đổi Wavelet cho tín hiệu tiếng nói thành hai thành phần hệ số chi tiết và hệ số xấp xỉ, sau đó áp dụng kỹ thuật trừ phổ và kỹ thuật ước lượng bình phương tối thiểu MMSE (*minimum mean square error*) của Ephraim/Malah cho các hệ số đó. Các kết quả mô phỏng cho thấy tiếng nói có nhiễu được triệt nhiễu bằng phương pháp đề xuất có SNR cao hơn các phương pháp trừ phổ, phương pháp MMSE và phương pháp Wavelet của Dohono.

**Từ khóa:** wavelet, triệt nhiễu, phổ trừ, MMSE, PSNR.

### TỔNG QUAN VỀ TRIỆT NHIỀU TÍN HIỆU TIẾNG NÓI

Nhiều ảnh hưởng nhiễu đến hiệu quả xử lý tín hiệu. Vì vậy, triệt nhiễu và nâng cao chất lượng tín hiệu là bước quan trọng trong các hệ thống xử lý tín hiệu thời gian thực [3].

Mô hình chung của tín hiệu có nhiễu là:

$$x_k = s_k + n_k, k = 0, \dots, K - 1 \quad (1)$$

Trong đó  $s_k$  là tín hiệu tiếng nói sạch,  $n_k$  là nguồn nhiễu độc lập với phương sai  $\sigma_k$  ( $\sigma_n^2 = 1$ ) (giả sử  $n_k$  là nhiễu trắng).

Gọi  $\hat{s}$  là giá trị ước lượng của tín hiệu tiếng nói sạch. Mục đích của các phương pháp triệt nhiễu tín hiệu tiếng nói là tối thiểu sai số bình phương trung bình  $E(|\hat{s}, s|^2)$

$$E \left[ \|\hat{s} - s\|^2 \right] = \sum_{k=0}^{K-1} E(\hat{s}_k - s_k)^2 \quad (2)$$

### PHƯƠNG PHÁP BIẾN ĐỔI WAVELET CHO TÍN HIỆU TIẾNG NÓI CÓ NHIỄU.

#### Phép biến đổi wavelet

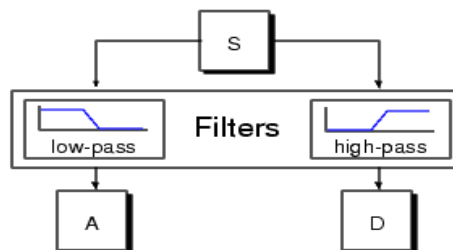
Biến đổi wavelet WT của một tín hiệu  $x(t)$  được định nghĩa

$$WT(b,a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (*)$$

Trong đó  $\psi\left(\frac{t-b}{a}\right)$  là hàm wavelet mẹ,  $\bar{\psi}(t)$

là hàm wavelet phức của  $\psi(t)$ ,  $b$  là toán tử dịch và  $a$  là toán tử tỉ lệ.

Trong thực tế biến đổi wavelet thường dùng là wavelet rời rạc DWT (Discrete Wavelet Transform) được thực hiện bằng cấu trúc đa phân giải MRA (Multiresolution Analysis) phân tích tín hiệu ra hai thành phần chi tiết (detail) và xấp xỉ (approximation). Thành phần chi tiết là thành phần tần số cao và chứa nhiễu nên các phương pháp triệt nhiễu bằng wavelet đều dựa trên ý tưởng đặt ngưỡng cho các thành phần chi tiết và loại bỏ các thành phần chi tiết nhỏ hơn ngưỡng.



**Hình 1.** Biến đổi wavelet rời rạc dùng đa phân giải Biến đổi Wavelet rời rạc áp dụng cho tín hiệu tiếng nói.

DWT được xây dựng dựa trên cấu trúc đa phân giải MRA [6]. Tín hiệu nguyên thủy  $S$  đi

\* Tel: 0983 454755, Email: hatvhit@gmail.com

qua 2 bộ lọc có đặc tính bù nhau và phân tách thành 2 tín hiệu, cA là các hệ số xấp xỉ, cD là các hệ số chi tiết.

Quá trình phân tách có thể được lặp lại, với xấp xỉ hoàn toàn được tách ra, do đó một tín hiệu được tách thành nhiều thành phần phân giải thấp hơn. Nó được gọi là cây phân tách Wavelet.

**PHƯƠNG PHÁP TRIỆT NHIỀU THAM KHẢO**

**Phương pháp trừ phổ**

Ý tưởng chung của phương pháp trừ phổ [1, 5] là chọn một mức phổ sản tương ứng với phổ của nhiễu nền và tách ra khỏi phổ tín hiệu lần nhiễu. Giả thiết nhiễu  $n_k$  là quá trình ngẫu nhiên dừng trong khoảng thời gian một khung tiếng nói và không tương quan với tín hiệu tiếng nói.

Từ (1), sau khi cửa sổ hoá ta được:

$$x_w(k) = s_w(k) + n_w(k) \quad (3)$$

Phổ của tín hiệu lần nhiễu là

$$|X_w(w)|^2 = |S_w(w)|^2 + |N_w(w)|^2 + S_w(w)N_w^*(w) + S_w^*(w)N_w(w) \quad (4)$$

Nếu chúng ta cho rằng  $n(k)$  có trung bình bằng 0 và không tương quan với  $s(k)$  thì  $S_w(w)N_w^*(w) + S_w^*(w)N_w(w)$  tiến tới 0. Do vậy ta có :

$$|\bar{S}(w)|^2 = |X(w)|^2 - E[|N(w)|^2] \quad (5)$$

$$|\bar{S}(w)|^2 = |X(w)|^2 \left( 1 - \frac{E[|N(w)|^2]}{|X(w)|^2} \right) \quad (6)$$

$$|\bar{S}(w)|^2 = |X(w)|^2 .G(w) \quad (7)$$

Gọi  $G(w)$  là hệ số trọng số phổ. Áp dụng biến đổi Wiener và đơn giản hóa bằng hàm biến đổi trọng số theo [1] ta có:

$$G = \text{Max} \left( 1 - \sqrt{\alpha \frac{N_{PSD}(w)}{X_{PSD}(w)}}, \beta \right) \quad (8)$$

Với  $\alpha$  là hệ số ước lượng trên và  $\beta$  là sàn phổ được chọn tương ứng.

**Phương pháp MMSE (minimum mean square error) của Ephraim/Malah**

Trong phương pháp MMSE của Ephraim/Malah [7], các thành phần phổ của tiếng nói và nhiễu được mô hình thành các biến ngẫu nhiên Gaussian.

Phân khung băng con tiếng nói thứ  $i$  thành các khung có độ dài bằng nhau. Ngưỡng nhiễu ước lượng trong khung thứ  $p$  và băng con thứ  $i$  là  $\lambda^{i,p}$  được xác định theo Jansen [4].

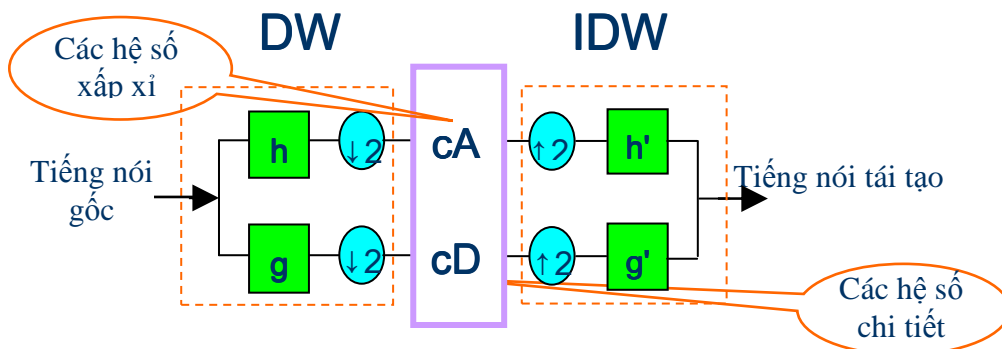
$(R_m^i)^{\text{post}}$  và  $(R_m^i)^{\text{priori}}$  là các tỉ lệ hệ số trên ngưỡng CTR (Coefficient to Thershold Ratio) tiền nghiệm và hậu nghiệm:

$$(R_m^i)^{\text{priori}} = \frac{|c_m^i|}{\lambda_m^i} \quad (9)$$

Các ngưỡng nhiễu đối với từng hệ số  $c_m^i$  là  $\lambda_m^i$  được ước lượng giống nhau trong từng khung. Nói cách khác trong khung  $p$   $\lambda_m^i = \lambda^{i,p}$ .

CTR hậu nghiệm tương ứng

$$(R_m^i)^{\text{post}} = \alpha \frac{|c_m^i|}{\lambda_m^i} + (1-\alpha) \max[0, (R_m^i)^{\text{priori}} - 1] \quad (10)$$



Hình 2. Phân tích Wavelet rời rạc tiếng nói

Với  $\alpha$  là một hệ số có thể thay đổi  $0 < \alpha < 1$ . Với CTR tiên nghiệm và hậu nghiệm xác định như trên ta có công thức biểu diễn bộ lọc đặt ngưỡng mềm Ephraim/Malah:

$$H_m^i = \frac{(R_m^i)^{post}}{1 + (R_m^i)^{post}} \left[ \frac{1}{(R_m^i)^{priori}} + \frac{(R_m^i)^{post}}{1 + (R_m^i)^{post}} \right] \quad (11)$$

Áp dụng bộ lọc này cho các hệ số phân rã  $c_m^i$  với từng bank  $i$ , các hệ số phân rã đã được triệt nhiễu được tính như sau:

$$\hat{C}_m^i = H_m^i \cdot C_m^i \quad (12)$$

**Mô hình triệt nhiễu đề xuất**

Trong tín hiệu tiếng nói thành phần nhiễu “musical noise” được loại bỏ ở phương pháp MMSE là ước lượng ngưỡng nhiễu đối với từng hệ số phân rã trong khung nên giảm thiểu sai số phổ. Kết quả thực nghiệm trong [5, 7] cho thấy hiệu quả triệt nhiễu của phương pháp này không cao so với phương pháp trừ phổ và phương pháp triệt nhiễu bằng Wavelet của Dohono đối với trường hợp nhiễu có cường độ lớn (PSNR của tín hiệu lần nhiễu nhỏ).

Mặt khác thành phần nhiễu “musical noise” là các thành phần chi tiết trên miền Wavelet. Thuật toán triệt nhiễu như sau:

Bước 1. Phân rã DWT cho toàn bộ tín hiệu tiếng nói có nhiễu

Bước 2. Sử dụng bộ lọc để lấy các hệ số chi tiết và các hệ số xấp xỉ.

Bước 3. Áp dụng phương pháp trừ phổ cho hệ số xấp xỉ.

Bước 4. Áp dụng phương pháp MMSE cho hệ số chi tiết.

Bước 5. Tái tạo wavelet và thu được tín hiệu đã được triệt nhiễu.

**Các tham số thực nghiệm.**

Giá trị phổ sàn trên  $\alpha = 0.9$ , sàn dưới  $\beta = 0.5$  ứng với khoảng có tiếng nói,  $\alpha = 1.2$ ,  $\beta = 0.1$  ứng với khoảng lặng [1]. DWT được thực hiện với thuật toán FWT, Wavelet mẹ được chọn là Deubechies 8, số mức phân rã là 3.

**KẾT QUẢ TRIỆT NHIỄU THỰC NGHIỆM TRÊN MATLAB**

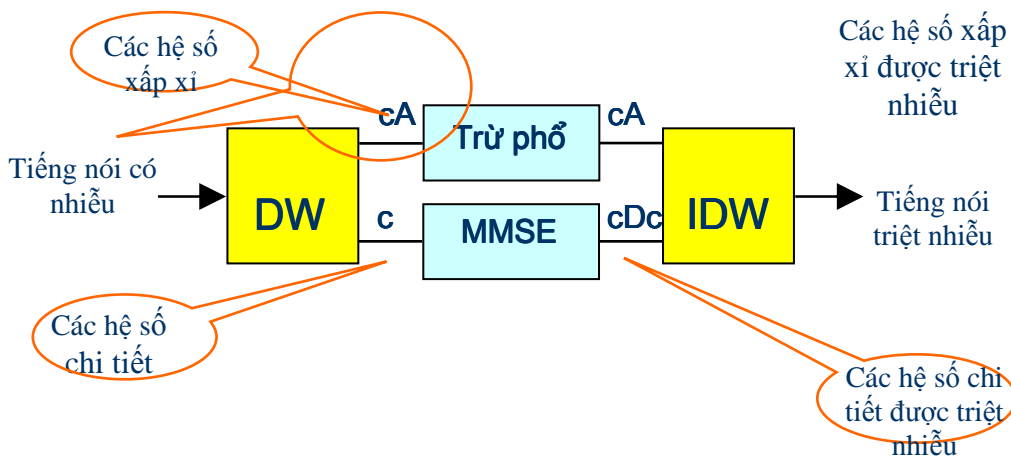
Chúng tôi đã thử nghiệm hệ thống triệt nhiễu đề nghị với đầu vào là tiếng nói sạch cộng với nhiễu Gauss trắng nhân tạo.

Phương pháp chung để xác định và đánh giá hiệu quả triệt nhiễu của các phương pháp triệt nhiễu trong [2, 4, 5, 7] là thông qua giá trị PSNR được xác định như sau:

$$MSE = \frac{1}{N} \sum_{i=1}^N (S_i - \hat{S}_i)^2 \quad (13)$$

$$PSNR = 10 \log_{10} \frac{[\max(S_i)]^2}{MSE} \quad (14)$$

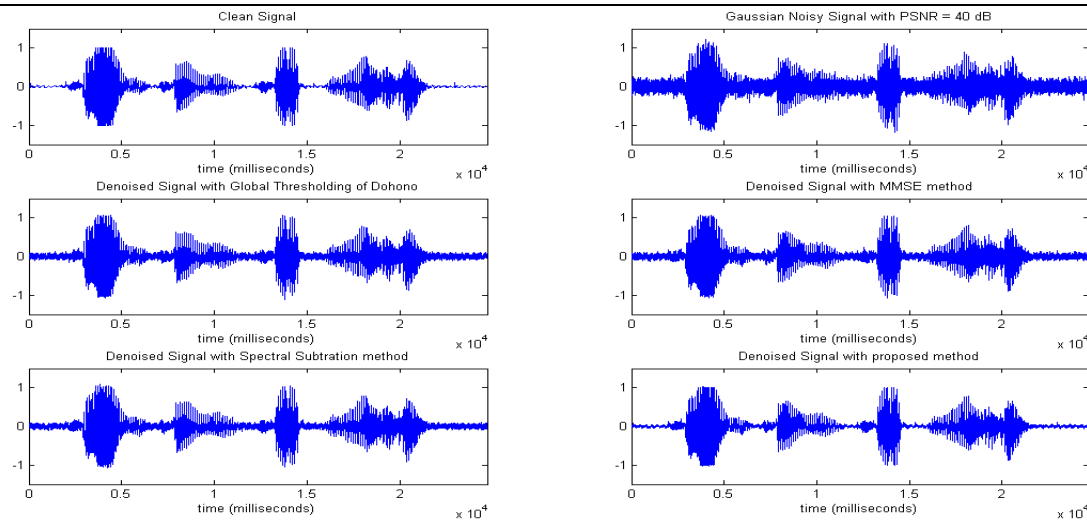
$S_i$  là tín hiệu nhiễu tại thời gian  $i$  còn  $\hat{S}_i$  là tín hiệu đã triệt nhiễu tại thời gian  $i$ .



**Hình 3.** Mô hình triệt nhiễu sử dụng

**Bảng 1.** So sánh PSNR của các phương pháp triệt nhiễu

Tín hiệu nhiễu	Phương pháp trừ phổ	Phương pháp trung bình tối thiểu MMSE	Phương pháp ngưỡng toàn cục của Dohono	Phương pháp kết hợp trên miền Wavelet
47.1320	58.0015	52.6225	55.5868	63.4432
52.1753	58.4312	54.3736	59.9119	65.6679
57.4413	59.8643	58.4647	64.1735	67.4762
62.5753	63.1262	64.6353	69.0072	70.2716
67.2824	69.7463	71.4373	71.3361	71.9635
72.2251	72.2601	73.6573	72.8592	73.6253



**Hình 4.** Hình ảnh âm thanh trước và sau khi triệt nhiễu

Để so sánh hiệu quả của phương pháp đề xuất với các phương pháp khác, chúng tôi thử nghiệm triệt nhiễu và tính toán các giá trị PSNR (priori signal to noise ratio) của phương pháp trừ phổ, phương pháp MMSE của Malah, phương pháp triệt nhiễu Wavelet của Dohono [2] và phương pháp đề xuất.

Kết quả PSNR tính được với tín hiệu có nhiễu và đã triệt nhiễu được cho trong bảng 1. Qua kết quả PSNR nhận được và chất lượng âm thanh kiểm nghiệm trực tiếp ta thấy phương pháp đề xuất có kết quả tốt nhất, nhiễu gần như được loại bỏ mà chất lượng tiếng nói giảm không đáng kể. Trong thực nghiệm nếu PSNR của tín hiệu lần nhiễu nhỏ hơn 20 dB thì cả 3 phương pháp đều không cho hiệu quả triệt nhiễu đáng kể. Khi PSNR đầu vào nhỏ (40-60 dB) phương pháp đề xuất cho kết quả cao hơn hẳn, tuy nhiên khi PSNR tăng dần (> 70 dB) thì sự phân biệt của các phương pháp là không đáng kể.

**KẾT LUẬN**

Dựa trên các kết quả nghiên cứu chúng tôi đánh giá phương pháp triệt nhiễu sử dụng biến đổi Wavelet thành các hệ số chi tiết, hệ số xấp xỉ và áp dụng triệt nhiễu cho từng thành phần hệ số với kỹ thuật trừ phổ và kỹ thuật MMSE của Ephraim/Malah là một kỹ thuật triệt nhiễu hiệu quả và có thể áp dụng trong khối tiền xử lý của các hệ thống xử lý tiếng nói thời gian thực như mã hóa, nhận dạng tiếng nói thời gian thực.

**TÀI LIỆU THAM KHẢO**

[1]. Hà Đình Dũng, Nguyễn Kim Quang, (2003), “Xây dựng bộ giảm nhiễu sử dụng phương pháp trừ phổ ứng dụng trong hệ thống nhận dạng tiếng nói”, Báo cáo hội thảo quốc gia CNTT, Thái Nguyên.  
 [2]. Donoho, D. L, (1995), “Denoising via soft thresholding”, IEEE Trans. Information Theory.  
 [3]. Gibert Strang, Truong Nguyen, (1996), Wavelet and Filter Banks, Weliesley- Cambridge Press, The United States of America.

- [4]. Jansen M. , (2001), *Noise Reduction by Wavelet Thresholding*, Springer-Verlag, New York.
- [5]. S.F. Boll, “*Suppression of Acoustic Noise in Speech Using Spectral Subtraction*”, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, April 1979, pp. 113-120.
- [6]. Stéphane Mallat, (1999), *A Wavelet Tour of Signal Processing*, Second Edition.
- [7]. Y. Ephraim and D. Malah, (1985), “*Speech enhancement using a minimum mean square error log-spectral amplitude estimator*” IEEE Trans. on ASSP, pp. 443-445.

#### SUMMARY

#### **SPEECH DENOISING AND ENHANCEMENT BASED ON DENOISING COMPONENT APPROXIMATIONS AND DETAILS IN THE WAVLET DOMAIN**

**Do Huy Khoi, Nguyen Thanh Trung, Trinh Van Ha\***

*College of Information and Communication Technology - TNU*

There are many researches about the methods of speech denoising and enhancement using wavelet in the world. Most of researches study the methods to estimate the global or sub band dependent threshold overall signal.

In this paper, we present a speech denoising approach using discrete Wavelet transform ,with approximation coefficients based on spectral subtraction method and details coefficients based on MMSE methods. The simulation results show that the noisy speech denoised by our proposed method has higher SNR than the spectral subtraction denoising, the MMSE denoising and the Wavelet denoising of Dohono.

**Keyword:** *Wavelet, denoising, spectral subtraction , MMSE, PSNR (priori signal to noise ratio) speech processing.*

Ngày nhận bài:05/10/2012 , ngày phân biệ: 20/10/2012, ngày duyệt đắg:10/12/2012

---

\* Tel: 0983 454755, Email: hatvhit@gmail.com