ĐÁNH GIÁ GIẢI PHÁP GIẢM NHIỄU CHO TÍN HIỆU TIẾNG NÓI SỬ DỤNG CÁC PHÉP BIẾN ĐỔI WAVELET

EVALUATING SOLUTIONS TO NOISE REDUCTION FOR SPEECH USING WAVELET TRANSFORMATIONS

Dương Ngọc Pháp, Võ Thị Diệu Hạnh

Trường Cao đẳng Công nghệ Thông tin, Đại học Đà Nẵng; dnphap@cit.udn.vn, dieuhanh3107@gmail.com

Tóm tắt - Bài báo tập trung trình bày kỹ thuật giảm nhiễu nâng cao chất lượng cho tín hiệu tiếng nói sử dụng các phép biến đổi trong miền Wavelet. Phương pháp xác định nhiễu sử dụng bộ lọc phần trăm (PF: Percentile Filter) [6] được sử dụng để thực hiện tiền xử lý cho các kỹ thuật nén nhiễu, bao gồm hàm nén nhiễu ngưỡng cứng, nén nhiễu ngưỡng mềm và nén nhiễu thích nghi sử dụng phương pháp lọc Wavelet thống kê có tính cảm quan (PSWF). Phần mô phỏng sẽ trình bày kết quả đánh giá hiệu quả giảm nhiễu của ba phương pháp trên trong các môi trường nhiễu khác nhau, đồng thời cũng sẽ thực hiện so sánh, đánh giá với các phương pháp giảm nhiễu trong miền tần số đã được nghiên cứu. Kết quả cho thấy các kỹ thuật trong miền Wavelet cho hiệu quả giảm nhiễu khá tốt, đặc biệt đối với nhiễu màu.

Từ khóa - Wavelet; giảm nhiễu; tiếng nói; PSWF, miền Wavelet

Abstract - The article focuses on demonstrating noise reduction techniques for enhancing speech signaling using transformations in the Wavelet domain. Noise determination with Percentile Filter is used to perform preprocessing for interference suppression techniques including hard-edge interference compression, soft-band interference compression, and noise correction using sensory Wavelet filtering (PSWF). The simulation will show the results of the noise reduction of the three above methods in different noise environments, and will also make comparisons with the noise reduction methods in the research frequency domain. The results show that the techniques in the Wavelet domain for noise reduction are good, especially for color noise.

Key words - Wavelet; noise reduction; speech; PSWF; Wavelet domain

1. Đặt vấn đề

Trong cuộc sống con người, phương thức giao tiếp chủ yếu với nhau là tiếng nói. Hầu hết chất lượng tiếng nói trong các hệ thống thông tin liên lạc đều bị suy giảm do tác động bởi nhiễu. Vì vậy, việc nghiên cứu và đưa ra các kỹ thuật nhằm loại bỏ nhiễu đóng vai trò quan trọng trong việc đảm bảo chất lượng và tính trung thực của tín hiệu tiếng nói trong các hệ thống thông tin liên lạc.

Việc nâng cao chất lượng tiếng nói bao gồm việc cải thiện chất lượng, tính dễ hiểu và giảm sự khó chịu cho người nghe bằng cách giảm tối đa nhiễu tác động vào tiếng nói. Các kỹ thuật đã được sử dụng để giảm nhiễu cho tín hiệu tiếng nói bao gồm trừ phổ (SS) [2], ước lượng MMSE [3], log MMSE [12], lọc Wiener (WF) [4], biến đổi Wavelet…

Nội dung bài báo sẽ tập trung đánh giá các hàm nén nhiễu trong miền Wavelet bao gồm hàm nén nhiễu ngưỡng cứng [8], nén nhiễu ngưỡng mềm [9] và nén nhiễu thích nghi sử dụng phương pháp lọc Wavelet thống kê có tính cảm quan (PSWF) [1], [7] dựa trên việc xác định ngưỡng nhiễu với bộ lọc phần trăm PF [6]. Trong miền Wavelet, tín hiệu tiếng nói được phân tích thành các sóng con, ứng với mỗi hệ số Wavelet đóng góp vào mức nhiễu với phương sai σ^2 cho phép xác định chính xác các hệ số nhiễu. Kết quả thực hiện sẽ được so sánh, đánh giá với các kỹ thuật đã đề xuất cho thấy rằng, các hàm nén nhiễu trong miền Wavelet tỏ ra khá hiệu quả với các loại nhiễu khác nhau trong môi trường thực.

2. Phép biến đổi Wavelet

Gọi f(x) là tín hiệu ban đầu, phép biến đổi Wavelet của f(x) sử dụng hàm Wavelet ψ_0 được biểu diễn [5]:

$$W(s,b) = \frac{1}{\sqrt{s}}\int_{-\infty}^{+\infty} f(x).\psi_0^*\left(\frac{x-b}{s}\right)dx \quad (1)$$

Trong đó:

- W(s,b) là hệ số biến đổi Wavelet của f(x), với s là tỉ lệ (nghịch đảo của tần số) và b là dịch chuyển đặc trưng vị trí.

- $\psi_0^*(x)$ là hàm liên hiệp phức của Wavelet, $\psi_0(x)$ được gọi là hàm Wavelet phân tích.

Biểu thức (1) có thể được viết lại dưới dạng tích nội như sau:

$$W(s,b) = \langle f(x), \psi_{0(s,b)}(x)\rangle \quad (2)$$

Trong đó: $\psi_{0(s,b)}(x) = \frac{1}{\sqrt{s}}\left(\frac{x-b}{s}\right)$ (3)

Hình 1. *Miền phân tích của phép biến đổi Wavelet*

Mỗi một tín hiệu được phân tích thành hai thành phần: thành phần xấp xỉ A "*tương ứng với thành phần tần số thấp*" và thành phần chi tiết D "*tương ứng thành phần tần số cao*" thông qua hai bộ lọc thông thấp và thông cao, trong đó, bộ lọc thông cao sử dụng hàm Wavelet $\Psi(x)$ và bộ lọc thông thấp sử dụng hàm tỉ lệ (scaling function) $\Phi(x)$.

Mối quan hệ giữa hàm tỉ lệ và hàm Wavelet được cho bởi:

$$\Phi(x) = \sum_{k=0}^{N-1} c_k.\Phi(2x-k) \quad (4)$$

$$\Psi(x) = \sum_{k=0}^{N-1}(-1)^k c_k.\Phi(2x+k-N+1) \quad (5)$$

Trong đó, $\Psi(x)$ là hàm tỉ lệ cho hàm Wavelet Φ và c_k là các hệ số Wavelet.

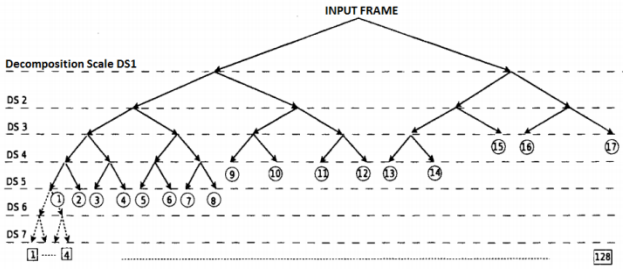
3. Mô hình nhiễu cộng trong miền Wavelet

Do tính chất tuyến tính của biến đổi Wavelet rời rạc DWT (*Discrete Wavelet Transform*), mô hình nhiễu cộng trong miền tần số cũng đúng trong miền Wavelet [7]:

$$Y_{m,i}^{k}(n) = X_{m,i}^{k}(n) + D_{m,i}^{k}(n) \quad (6)$$

trong đó, $Y_{m,i}^{k}(n)$, $X_{m,i}^{k}(n)$ và $D_{m,i}^{k}(n)$ mô tả các chuỗi hệ số gói Wavelet của các tín hiệu bị nhiễu, tín hiệu sạch và tín hiệu nhiễu tương ứng được tính tại tỷ lệ thứ m cho frame thứ *i* ($k = 1, 2, 3, \ldots 2^m$).

Trong nghiên cứu này ứng dụng phân tích gói Wavelet nhị phân WPD (Wavelet Packet Decomposition) thực hiện tại tỷ lệ phân tích $m = 7$. Hình 2 mô tả cấu trúc phân chia gói Wavelet.



Hình 2. *Cấu trúc cây phân chia gói*

4. Kỹ thuật chọn ngưỡng

Giải pháp khử nhiễu Wavelet được xem như là phương pháp ước lượng thống kê không có tham số. Nguyên lý thực hiện dựa trên việc định ngưỡng hoặc nén các hệ số Wavelet mà giá trị của chúng thấp hơn ngưỡng nhiễu. Việc chọn ngưỡng sẽ tác động đến sự làm trơn hoặc làm cho khớp với tín hiệu được khử nhiễu.

Cụ thể là ngưỡng giá trị nhỏ sẽ dẫn đến sự khử nhiễu nhẹ, và hệ quả là nhiễu còn tồn đọng nhiều sau xử lý. Ngược lại, ngưỡng với giá trị lớn hơn sẽ nén nhiều hệ số Wavelet, dẫn đến làm trơn tín hiệu nhưng cũng đồng thời làm suy hao các thành phần unvoiced (ví dụ âm /s/) của tiếng nói, và gây méo tín hiệu tiếng nói sau khi khử nhiễu. Trong công bố [10], tác giả đã đề xuất công thức tính ngưỡng nhiễu toàn cục UT (Universal Threshold) dựa trên nguyên tắc tối thiểu hóa hàm rủi ro (risk function) giữa tín hiệu sạch mong muốn và tín hiệu được khử nhiễu:

$$E\{R(T)\} = E\left\{\left\|E\{\hat{X}_k(n)\} - E\{X_k(n)\}\right\|^2\right\} + E\{\left\|\hat{X}_k(n) - E\{\hat{X}_k(n)\}\right\|^2\}$$

$$E\{R(T)\} = E\left\{\left\|E\left\{\hat{X}_k(n)\right\} - E\{X_k(n)\}\right\|^2\right\} + E\left\{\left\|\hat{X}_k(n) - E\left\{\hat{X}_k(n)\right\}\right\|^2\right\} \quad (7)$$

Trong đó E{.} là toán tử trung bình thống kê, $\hat{X}_k(n)$là hệ số Wavelet đã được cải thiện. Với giả thiết về tính trực giao của biến đổi Wavelet và phân bố nhiễu với phương sai σ^2, thì ngưỡng UT tỷ lệ thuận với độ lệch chuẩn σ và chiều dài N của chuỗi hệ số Wavelet. Để ước lượng tốt nhiễu màu, UT được tính toán cho mỗi gói Wavelet độc lập tại mỗi khung tín hiệu như sau:

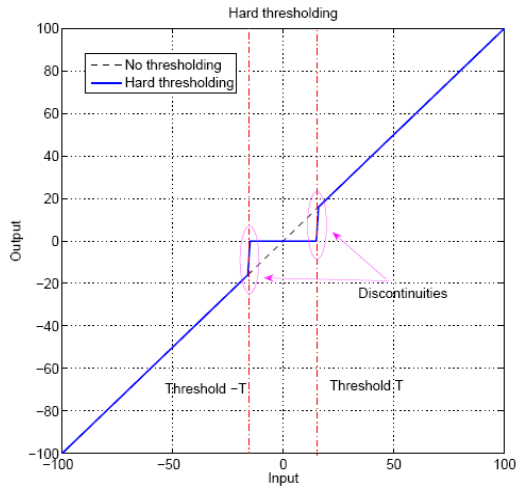
$$T_{k,i} = \sigma_{k,i}^{MAD}\sqrt{2logN_{k,i}} \quad (8)$$

Trong đó, $\sigma_{k,i}^{MAD} = \frac{1}{0.6745} Median(|Y_{k,i}(n)|)$ là ước lượng bền vững của độ lệch chuẩn bằng cách tính Median Absolute Deviation (MAD) của chuỗi các hệ số.

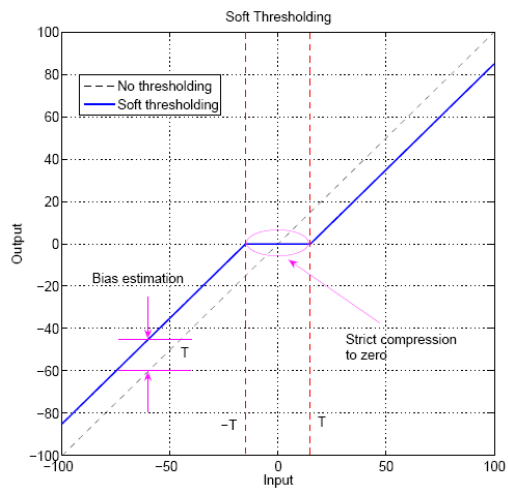
5. Thuật toán giảm nhiễu trong miền Wavelet

5.1. Kỹ thuật nén nhiễu ngưỡng cứng và ngưỡng mềm

Kỹ thuật khử nhiễu trong miền Wavelet xuất phát từ nguyên tắc: Mỗi hệ số Wavelet đóng góp vào mức nhiễu với phương sai là σ^2, nhưng chỉ có một vài số hệ số Wavelet ảnh hưởng tới tín hiệu [10]. Điều này cho phép thay thế các hệ số nhiễu bằng zero. Các kỹ thuật nén ngưỡng cứng và nén ngưỡng mềm là các hàm khử nhiễu khá đơn giản nhưng chưa tối ưu. Các Hình 3 và 4 mô tả mối quan hệ giữa ngõ vào và ngõ ra của đặc tuyến hàm nén nhiễu ngưỡng cứng và ngưỡng mềm trong miền Wavelet.



Hình 3. *Đặc tuyến hàm nén nhiễu ngưỡng cứng trong miền Wavelet [7], [8]*



Hình 4. *Đặc tuyến hàm nén nhiễu ngưỡng mềm trong miền Wavelet [7], [9]*

Hàm nén hard-threshold G^H (T) nén tất cả các hệ số Wavelet nhỏ hơn ngưỡng nhiễu về tối thiểu trong khi những hệ số còn lại thì không tác động gì.

$$\hat{X}_k^H(n) = G^H(T, Y) = \begin{cases} Y_k(n) & , \text{if } |Y_k(n)| > T \\ 0 & , \text{if } |Y_k(n)| \le T \end{cases} \quad (9)$$

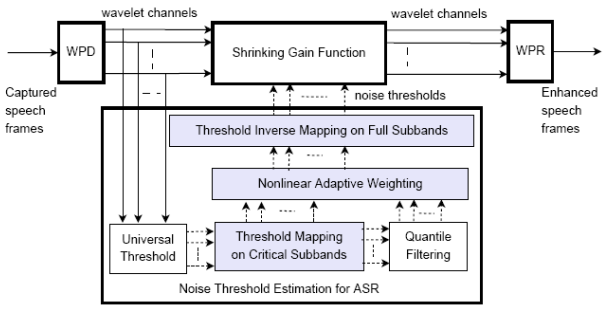
Với hàm nén ngưỡng mềm được đề xuất trong [9], những điểm gián đoạn trong đặc tuyến ngõ ra của hàm nén ngưỡng cứng được loại bỏ bằng cách nén các hệ số Wavelet hơn mức ngưỡng 1 giá trị, bằng chính giá trị ngưỡng T được ước lượng:

$$\hat{X}_k^S(n) = G^S(T,Y) = \begin{cases} sgn(Y_k(n))(|Y_k(n)| - T) & ,if\ |Y_k(n)| > T \\ 0 & ,if\ |Y_k(n)| \le T \end{cases} \quad (10)$$

Hàm nén ngưỡng mềm dù đã cải thiện nhưng vẫn chưa đạt tối ưu vì quy luật nén tuyệt đối các hệ số dưới ngưỡng về 0, nó có thể phá hủy những hệ số Wavelet của tín hiệu tiếng nói vô thanh, do sự tương tự của tiếng nói vô thanh và một số mức nhiễu. Hàm nén ngưỡng cứng dẫn đến phương sai lớn hơn vì tính chất không liên tục của hàm gain, trong khi ngưỡng mềm thì dẫn đến mức độ dịch chuyển lớn hơn, do tất cả những hệ số nào lớn hơn ngưỡng đều giảm đi một lượng bằng giá trị ngưỡng T.

5.2. Thuật toán giảm nhiễu dùng phương pháp lọc Wavelet thống kê có tính cảm quan PSWF

Phần này trình bày phương pháp giảm nhiễu Wavelet tối ưu (Optimal Shrinkage) dùng thuật toán phân tích gói Wavelet WPD (Wavelet Packet Decomposition), bởi thuật toán lọc Wavelet thống kê có tính cảm quan PSWF (Perceptually Statistical Wavelet Filter) với mức nhiễu toàn cục UT (Universal Threshold) gọi tắt là Shrinking-UT-PF, với sơ đồ khối như Hình 5.



Hình 5. *Sơ đồ của thuật toán PSWF [1], [7]*

Trước hết, ngưỡng $T_{k,i}$ được tính từ các hệ số Wavelet $Y_{k,i}(n)$ tại mỗi gói. Sau đó, ngưỡng nhiễu cảm quan $P_{j,i}$ được tính cho mỗi CWS (critical Wavelet subband) thông qua bộ ánh xạ ngưỡng. Tiếp theo, áp dụng kỹ thuật lọc percentile thống kê để ước lượng ngưỡng nhiễu percentile $\tilde{\Gamma}_{j,i}$ cho mỗi CWS một cách thích nghi.

Các ngưỡng đã được ước lượng này lại tiếp tục được tinh chỉnh bởi trọng số hóa trong miền thời gian và miền tần số. Cuối cùng, sau khi qua bộ ánh xạ ngược ngưỡng từ CWS sang toàn bộ các băng con tương ứng các gói phân tích Wavelet, các ngưỡng $\tilde{\Gamma}_{j,i}$ này được đưa vào hàm nén nhiễu thích nghi để khử nhiễu các hệ số Wavelet. Các mẫu tiếng nói được khôi phục bởi phép khôi phục gói Wavelet (WPR).

Thuật toán sử dụng một hàm Wavelet nén nhiễu được cải tiến, mà về bản chất chính là hàm ngưỡng cứng được làm trơn theo cơ sở luật μ. Quy tắc nén được thể hiện bởi hàm ước lượng sau (posteriori function) của tỷ số tín hiệu trên ngưỡng được phân đoạn trước $\gamma_{k,i}$ như sau:

$$H_{k,i} = \begin{cases} 1, & if\ \gamma_{k,i} < 1 \\ sgn\left\{Y_{k,i}(p)\dfrac{(1+\mu_{k,i})^{\gamma_{k,i}}-1}{\mu_{k,i}\gamma_{k,i}}\right\}, & if\ \gamma_{k,i} < 1 \end{cases}$$

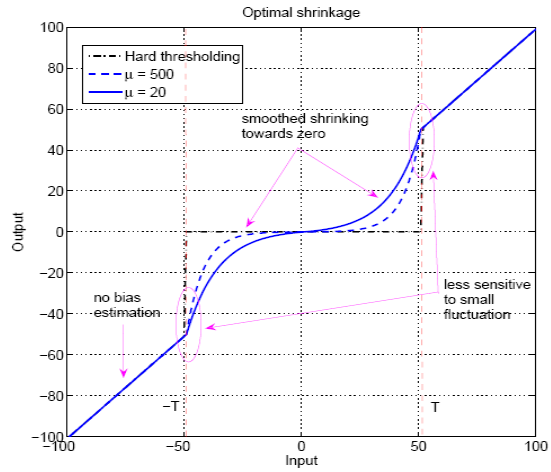
$$H_{k,i} = \begin{cases} 1, & if\ \gamma_{k,i} > 1 \\ \mathrm{sgn}\{Y_{k,i}(p)\}\dfrac{(1+\mu_{k,i})^{\gamma_{k,i}}-1}{\mu_{k,i}\gamma_{k,i}}, & if\ \gamma_{k,i} \le 1 \end{cases} \quad (11)$$

Ở đây, $\gamma_{k,i} \triangleq \frac{|Y_{k,i}(p)|}{\Gamma_{k,i}}$ và thông số thích nghi $\mu_{k,i}$ được định nghĩa [1]:

$$\mu_{k,i} = \exp(\beta \frac{\tilde{\Gamma}_{k,i}}{\max_i\{\tilde{\Gamma}_{k,i}\}})\frac{\max_p\{|Y_{k,i}(p)|\}}{\tilde{\Gamma}_{k,i}}$$

$$\mu_{k,i} = \exp\left(\beta\frac{\tilde{\Gamma}_{k,i}}{\max_i\{\tilde{\Gamma}_{k,i}\}}\right)\frac{\max_p\{|Y_{k,i}(p)|\}}{\tilde{\Gamma}_{k,i}} \quad (12)$$

Trong đó, phần mũ exp tự thích nghi với chính nó bởi ngưỡng nhiễu được làm trơn và chuẩn hóa $\tilde{\Gamma}_{k,i}$với hằng số độ dốc $\beta = 5{,}8$.



Hình 6. *Đặc tuyến hàm nén nhiễu thích nghi trong miền Wavelet [7]*

Hình 6 mô tả mối quan hệ giữa ngõ vào và ngõ ra của đặc tuyến hàm nén nhiễu thích nghi trong miền Wavelet. Kỹ thuật nén theo luật μ biểu diễn một sự thỏa hiệp giữa ngưỡng cứng và ngưỡng mềm. Trong khi hàm ngưỡng cứng có phương sai lớn hơn nhưng bias nhỏ hơn, thì hàm ngưỡng mềm có bias lớn hơn và phương sai nhỏ hơn. Nói cách khác, ngưỡng cứng có xu hướng giữ dạng gốc của tín hiệu, còn ngưỡng mềm có xu hướng làm mịn tín hiệu hơn [9]. Một ưu điểm lớn của nén luật μ so với các luật khác là nó không thiết lập tất cả hoặc một phần của các hệ số Wavelet, mà trị tuyệt đối của các hệ số này đều thấp dưới ngưỡng, về bằng 0 như được thực hiện bởi ngưỡng cứng hoặc mềm.

6. Thực hiện giảm nhiễu và đánh giá kết quả

6.1. Cơ sở dữ liệu

Cơ sở dữ liệu đánh giá được lấy từ thư viện NOIZEUS [13] gồm 30 câu thoại được ghi âm trong phòng thí nghiệm

theo chuẩn của IEEE là tín hiệu thoại sạch. Mỗi câu dài trung bình khoảng 2 s.

Năm loại nhiễu được chọn để nghiên cứu là nhiễu ô tô (Car), nhiễu đám đông (Babble), nhiễu trắng (White), nhiễu từ tàu hỏa (Train) và nhiễu đường phố (Street).

6.2. Tiêu chí đánh giá

Các phương pháp đánh giá khách quan được khảo sát để thực hiện đánh giá chất lượng tín hiệu tiếng nói qua các tiêu chí đánh giá, bao gồm:

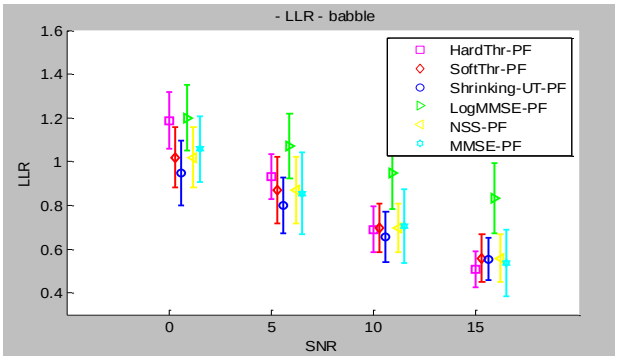
- Segmental Signal-to-Noise Ratio (SegSNR);
- Log Likelihood Ratio (LLR);
- Cepstrum Distance (CEP);
- Perceptual Evaluation of Speech Quanlity (PESQ);
- Weighted Spectral Slope (WSS).

6.3. Kết quả đánh giá và phân tích

Bài báo sẽ thực hiện đánh giá, so sánh 3 kỹ thuật đã nghiên cứu là NSS-PF, MMSE-PF và LogMMSE-PF ước lượng nhiễu sử dụng bộ lọc phần trăm (PF) với 3 kỹ thuật giảm nhiễu trong miền Wavelet là HardThr-PF, SoftThr-PF, Shriking-UT-PF.

6.3.1. Đánh giá theo phương pháp LLR

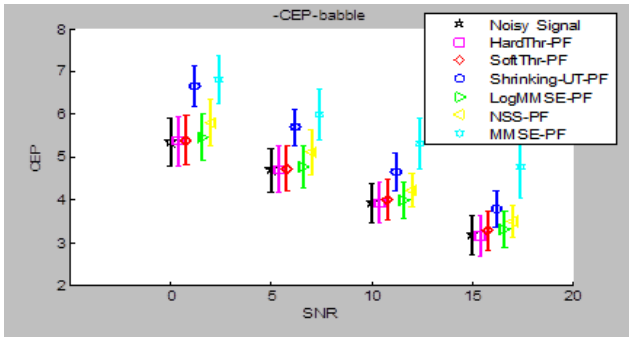
Phương pháp đo LLR sẽ được lựa chọn để đánh giá 6 kỹ thuật giảm nhiễu với 5 mức nhiễu khác nhau, kết quả được thực hiện với nhiễu tiếng ồn đám đông như ở Hình 7.



Hình 7. *Kết quả đánh giá LLR của 6 thuật toán tăng cường chất lượng tiếng nói với loại nhiễu tiếng ồn đám đông (Babble)*

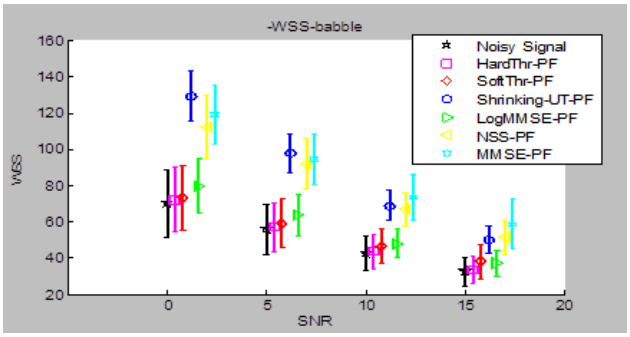
Kết quả so sánh chỉ số LLR chỉ ra rằng lọc nhiễu dùng các thuật toán trong miền Wavelet cho kết quả tốt hơn các thuật toán được so sánh.

6.3.2. Đánh giá theo phương pháp CEP



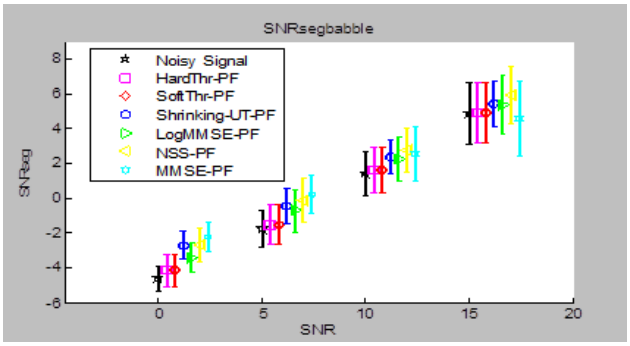
Hình 8. *Kết quả đánh giá CEP của 6 thuật toán tăng cường chất lượng tiếng nói với loại nhiễu tiếng ồn đám đông (Babble)*

6.3.3. Đánh giá theo phương pháp WSS



Hình 9. *Kết quả đánh giá WSS của 6 thuật toán tăng cường chất lượng tiếng nói với loại nhiễu tiếng ồn đám đông (Babble)*

6.3.4. Đánh giá theo phương pháp SegSNR

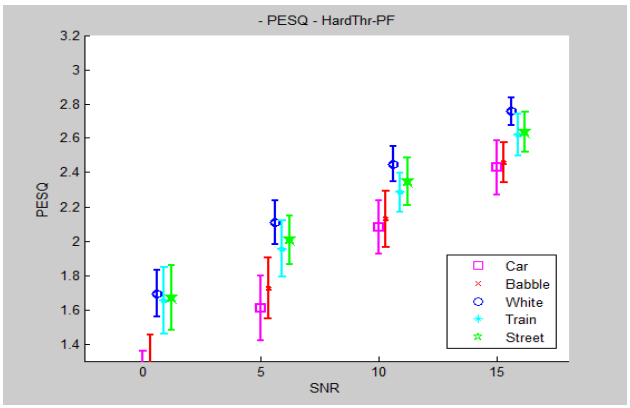


Hình 10. *Kết quả đánh giá SegSNR của 6 thuật toán tăng cường chất lượng tiếng nói với loại nhiễu tiếng ồn đám đông (Babble)*

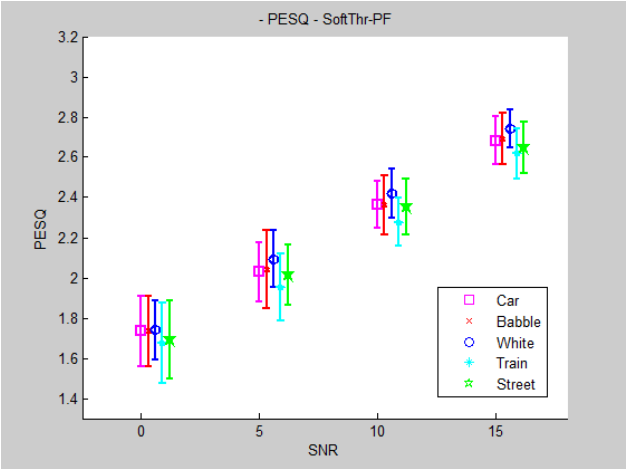
6.3.5. Đánh giá theo phương pháp cảm quan PESQ

Thực hiện đánh giá PESQ cho kỹ thuật giảm nhiễu sử dụng 3 thuật toán trong miền Wavelet ứng với 5 môi trường nhiễu khác nhau cho kết quả như ở Hình 11, Hình 12, và Hình 13.

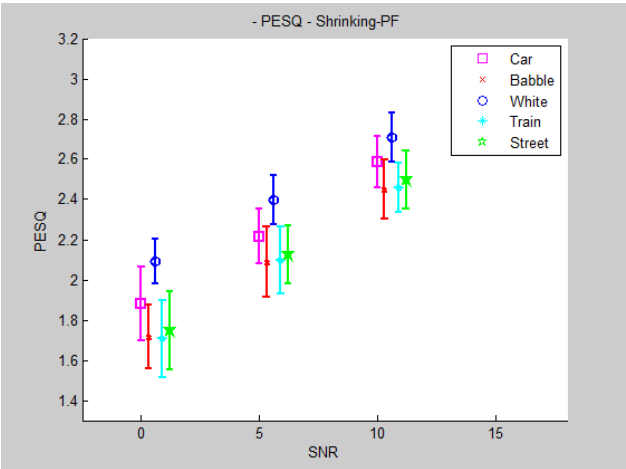
Kết quả đánh giá cho thấy hàm nén nhiễu Wavelet PSWF cho chỉ số PESQ tốt hơn hẳn hai kỹ thuật Wavelet còn lại, vì hàm ánh xạ ngưỡng tích hợp 128 mức ngưỡng toàn cục vào 17 giá trị ngưỡng tương ứng với 17 băng con thứ yếu (CWS) [7], thay vì sử dụng trực tiếp các giá trị ngưỡng toàn cục bởi phép phân tích gói, trong đó, chỉ số này ứng với loại nhiễu trắng (white) là cao hơn hẳn so với các loại nhiễu còn lại.



Hình 11. *Kết quả đánh giá hàm nén ngưỡng cứng cho 5 môi trường nhiễu*

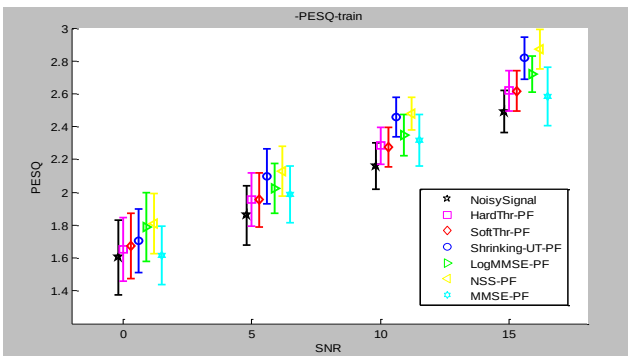


Hình 12. *Kết quả đánh giá hàm nén ngưỡng mềm cho 5 môi trường nhiễu*



Hình 13. *Kết quả đánh giá hàm Wavelet PSWF cho 5 môi trường nhiễu*

Ngoài ra, đánh giá cảm quan PESQ trên các môi trường nhiễu khác nhau còn được so sánh giữa các kỹ thuật giảm nhiễu với nhau như Hình 14 được đánh giá trên nhiễu từ tàu hỏa (train).



Hình 14. *Kết quả phương pháp đánh giá PESQ của 6 thuật toán tăng cường chất lượng tiếng nói với nhiễu tiếng ồn tàu hỏa*

Kết quả đánh giá cho thấy các thuật toán trong miền Wavelet cho kết quả xử lý khá thấp khi mức nhiễu cao. Tuy nhiên thuật toán giảm nhiễu Shrinking-UT-PF cho kết quả khá tốt khi mức nhiễu thấp.

7. Kết luận

Kết quả đánh giá dữ liệu tiếng nói sau khi được tăng cường sử dụng các thuật toán đã trình bày, cho thấy các tiêu chí đánh giá khác nhau có sự thay đổi nhất định trong các môi trường nhiễu khác nhau và bởi các kỹ thuật giảm nhiễu khác nhau.

Tuy nhiên, nhìn chung tín hiệu tiếng nói bị tác động bởi nhiễu trắng cho kết quả sau tăng cường tốt hơn, vì có mật độ phổ phẳng.

Trong nhóm các thuật toán nén nhiễu miền Wavelet, thuật toán Shrinking-PF cho kết quả xử lý tốt hơn, khả năng nén nhiễu cao hơn, tín hiệu sau xử lý ít bị phá hủy, và vẫn đảm bảo tính dễ nghe của tín hiệu. Kết quả trên hoàn toàn tương đồng sau khi nghe thử các mẫu tín hiệu đã xử lý.

Trong số các phương pháp đánh giá đã tìm hiểu, phép đánh giá PESQ và SNRseg cho kết quả đáng tin cậy hơn, tương đồng cao với cảm nhận nghe chủ quan.

TÀI LIỆU THAM KHẢO

[1] Phạm Văn Tuấn, Hoàng Lê Uyên Thục, "Giải pháp giảm nhiễu trong miền Wavelet để nâng cao hiệu suất nhận dạng tiếng nói tự động", *Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng*, số 4(39), 2010.

[2] Boll, S.F., "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans, Acoust. Speech Signal Process.*, 27(2), 113-120, 1979.

[3] Ephraim, Yariv, and David Malah, *Speech enhancement using optimal non-linear spectral amplitude estimation*, Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83, Vol. 8, IEEE, 1983.

[4] Cohen, Israel, "Speech enhancement using a noncausal a priori SNR estimator", *IEEE signal processing letters*, 11.9 (2004): 725-728.

[5] Graps, Amara, "An introduction to Wavelets", *IEEE computational science and engineering*, 2.2 (1995): 50-61.

[6] Pham T.V., Gernot Kubin, "WPD-based Noise Suppression Using Nonlinearly Weighted Threshold Quantile Estimation and Optimal Wavelet Shrinking", Proc. Interspeech, Lisboa, Portugal, 4-8 Sep., 2005.

[7] Van Pham, Tuan, *Wavelet Analysis For Robust Speech Processing and Applications*, Diss. Ph. D. Thesis, 2007.

[8] Donohol, M. Johnstone, and M. Johnsone, Ideal spatial adaptation via Wavelet shrinkage, Biometrika 12.8 (1994): 430-445.

[9] Donoho D. L., "De-noising by soft thresholding", *IEEE Trans. Information Theory*, 41:613–627, 1995.

[10] Pham T.V., Gernot Kubin, Erhard Rank, *Robust Speech Recognition Using Adaptive Noise Threshold Estimation And Wavelet Shrinkage*, Proc. IEEE ICCE, Hoi An, Vietnam, 04-06 Feb., 2008.

[11] Ephraim Y. and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33:443–445, 1985.

[12] Ephraim, Yariv, and David Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator", *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33.2 (1985): 443-445.

[13] http://ecs.utdallas.edu/loizou/speech/noizeus/ truy cập lần cuối 20/03/2017.

(BBT nhận bài: 03/05/2017, hoàn tất thủ tục phản biện: 07/06/2017)