

Wavelet Packets Best Tree 4 Points Encoded (BTE) Features

Amr M. Gody¹
Fayoum University

Abstract

The research aimed to introduce newly designed features for speech signal. The newly developed features are designed to normalize the dynamic structure of best tree decomposition of wavelet packets. The 4 points encoded vector is a full of information just like the original best tree's structure. It is a loss less encoding system that grants 100% reconstruction of the original best tree. The encoding process for BTE features vector is developed as such to minimize the distance based on frequency adjacency. The implied scoring system makes BTE suitable for recognition problems.

1. Introduction

It is known that human speech is decomposed of short time duration's unites called phonemes. Each phoneme contributes in specific piece of information. We can assume it as the characters that construct the whole word in any written language. Information in each phoneme is encoded into the frequency domain. Simply the information is a pattern of frequency components [1]. Features are extracted from the speech signal to best represent such information.

It is believed that human hearing system is the best recognition system. By trying to simulate human hearing system, good practical results may be achieved. Speech signal is processed in this research in such a manner that low frequency components have more weights than high frequency components [2]. The human ear responds to speech in a manner such that as indicated by Mel scale in figure 1. This curve explains a very important fact. Human ears cannot differentiate between different sounds in high frequency scale while it can do this in low frequency scale. Mel scale is a scale that reflects what human can hear. As shown by figure 1, a change in frequency from 4000(HZ) to 8000 (HZ) makes only 1000 (Mel) change in Mel scale. This is not the case in the low frequency range starts at 0(HZ) and ends by 1000 (HZ). In this low frequency range it is appeared that 1000(Hz)'s change is equivalent to 1000(Mel) change in Mel scale. This explains that human hearing is very sensitive for frequency variation in low range while it is not the case in high range.

Wavelets are short duration waveforms that can express any function by scaling and shifting of certain mother signal that is called mother wavelet [5]. Wavelet algorithm is acting as a filter banks on the input signal. The output of the filter banks are the wavelet signal's amplitudes.

¹ Department of Electrical Engineering, Email: amg00@fayoum.edu.eg

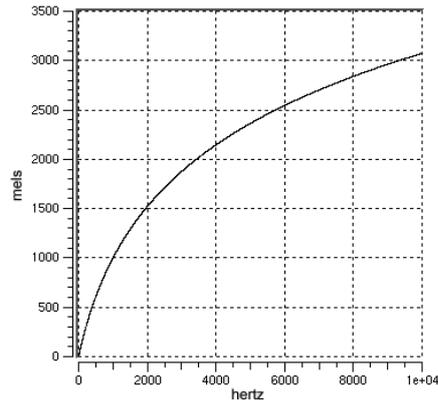


Figure 1: Mel scale curve that models the human hearing response to different frequencies [3].

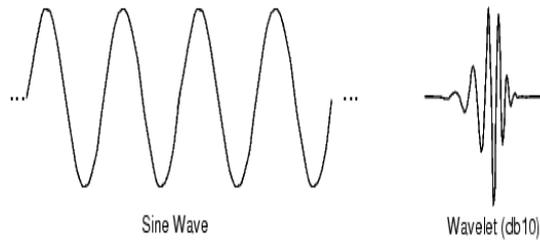


Figure 2: Sin wave is used for Fourier representation of the signal while wavelet function is used in wavelet representation for Daubechies 10 points filter. Sin wav is infinite in time but finite in frequency domain while wavelet is finite in both time and frequency domains [5].

Figure 2 indicates a very important property of wavelet function. Wavelet function is a finite in time. It is also finite in frequency [4]. This is not the case of "Sine" basis functions (harmonic functions) used for Fourier analysis. All derived wavelets are orthogonal. This makes each wavelet acts as an identifier of the signal in a certain band. Figure 3 gives a brief comparison between different possible spaces to express certain function [5].

Here's what this looks like in contrast with the time-based, frequency-bas

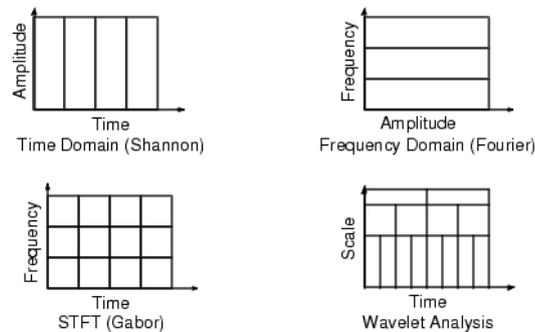


Figure 3: Comparison between different signal spaces [5].

Wavelet packets are an extension to wavelet transform. It includes the high frequency parts in the analysis for more signal resolution of the frequency spectrum as shown in figure 4.

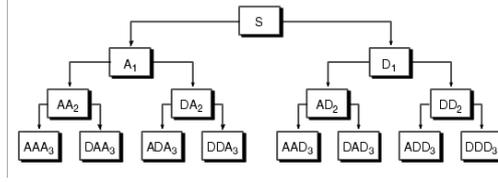


Figure 4: Signal decomposition using wavelet packets [5].

To simplify the subject, let us discuss Fourier series as a signal representation tool.

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right) \quad (1)$$

Equation 1 indicates the Fourier series representation of function $f(x)$. By the same approach, $f(x)$ may be expressed using wavelet packets as in equation 2.

$$f(x) = \sum_j \sum_{n=0}^{2^j-1} b_{jn} W_{j,n}(x) \quad (2)$$

"b" is wavelet coefficients and "W" is wavelet packet. Let us start with the two filters of length $2N$, where $h(n)$ and $g(n)$, corresponding to the wavelet filters.

$$W_{2n}(x) = \sqrt{2} \sum_{k=0}^{2N-1} h(k) W_n(2x - k) \quad (3)$$

$$W_{2n+1}(x) = \sqrt{2} \sum_{k=0}^{2N-1} g(k) W_n(2x - k) \quad (4)$$

$g(k)$ and $h(k)$ are filter banks. Where:

$W_0(x) = \phi(x)$ is called the scaling function.

$W_1(x) = \psi(x)$ is called wavelet function.

Where:

$$W_{jnk}(x) = w_n(2^{-j}x - k) \quad (5)$$

$$n \in N \text{ and } (j, k) \in Z$$

K is not a dynamic parameter after the decomposition of the signal rather it is a constant value for each wavelet packet W . This makes it much better to abstract (5) as :

$$W_{j,n} = w_n(2^{-j}x - k) \quad k \in Z \quad (6)$$

Hence:

$$W_{0,0}(x) = \Phi(x - k) \quad (7)$$

$$W_{1,1}(x) = \Psi\left(\frac{x}{2} - k\right) \quad (8)$$

The idea is explained by figure 5. Scaling " ϕ " and wavelet " Ψ " functions are used to generate W functions that cover all the frequency-scale space. The parameter k is used to indicate the time location of certain W function. K is chosen to best fit the original function to be expressed by wavelet packets while the scaling and wavelet functions are designed such that all W functions to be orthogonal.

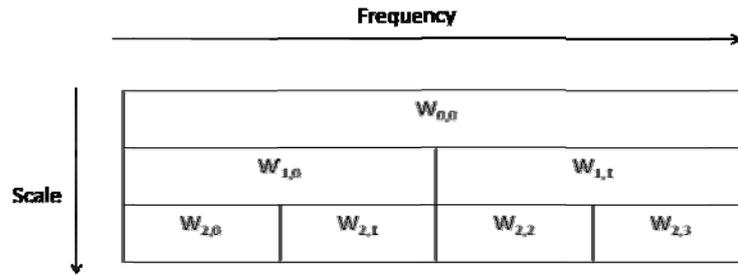


Figure 5: Frequency-Scale space for wavelet packets.

Many researchers deal with the best way to optimize the full binary tree in such that best describe the contained information [6]. Different entropy functions may be used in such optimization [7,8].

The objective of this paper is to introduce new features for speech signal. Features are developed from the wavelet packets best tree decomposition of speech signal. This research aims to explain the proposed features in details. Also it targets to introduce the benefits of using the proposed feature in speech recognition problems.

2. Feature extraction

In this section the process of feature extraction will be explained. Best Tree 4 point Encoded features (BTE) will be explained now. Wavelet packets process is very similar to filter banks. Both of them are filter banks in nature. The wavelet packets method is a generalization of wavelet decomposition that offers a richer signal analysis. Wavelet packet atoms are waveforms indexed by three naturally interpreted parameters: position, scale (as in wavelet decomposition), and frequency. For a given orthogonal wavelet function, we generate a library of bases called wavelet packet bases. Each of these bases offers a particular way of coding signals, preserving global energy, and reconstructing exact features. The wavelet packets can be used for numerous expansions of a given signal. We then select the most suitable decomposition of a given signal with respect to an entropy-based criterion [9].

The first step in BTE is to align the neighboring bands. This is very important for a good scoring process. Scoring process tries to score adjacent bands in such that minimizing the distance. For our case of best tree by Matlab, adjacent bands are indexed not in sequence.

	L3	L2	L1	L0
12.5	7	3	1	0
25	8			
37.5	9	4		
50	10			
62.5	11	5	2	
75	12			
87.5	13	6		
100	14			

Figure 6 : Wavelet packet tree analysis chart to figure out adjacent bands.

The objective is to remap node indices in such that adjacent node indices lay in adjacent frequency bands. To explain this subject considers the following table that represents the indices in a typical wavelet packet tree for 4-levels decomposition. Figure 6 represents band indexes in Matlab wavelet packets for 3 levels decomposition. Node indices are written inside the boxes that represent the nodes in the wavelet tree decomposition. As shown in figure 6 that node 7 and node 6 are too far in frequency while they are subsequent nodes as wavelet packets indexing system. This problem needs to be altered in such that adjacent frequency bands are listed as contiguous numbers. This way we will ensure that indexing system reflects frequency scale. This property may be used in the scoring system. Information in figure 6 is tabulated in table 1 to make it simple to figure out adjacent bands. Traversing tree as Left → Right → Center will be very logical to make good criteria for adjacency. Figure 7 explains the new indexing system.

Now we are ready to apply the best tree algorithm to optimize the full binary tree shown in figure 7. The optimization minimizes the number of tree nodes such that it best fit the information included in the speech signal. The entropy is used in the optimization algorithm.

Now we can apply the encoding by considering clusters of 7 bands. Each cluster will be encoded in 7 bits in such that each bit is associated to a certain band. Figure 11 explains the clusters.

Table 1 : Bandwidth distribution over wavelet packet decomposition bands.

Filter bank's Upper Limit with respect to total bandwidth (%)	Filter Bank's Node- index according to wavelet packet indexing system
100	0
50	1
100	2
25	3
50	4
75	5
100	6
12.5	7
25	8
37.5	9
50	10
62.5	11
75	12
87.5	13
100	14

	L3	L2	L1	L0
12.5	0	2	6	14
25	1			
37.5	3	5		
50	4			
62.5	7	9	13	
75	8			
87.5	10	12		
100	11			

Figure 7 : Proposed indexing to solve the adjacency problem due to wavelet packet’s indexing system.

In figure 11, clusters are surrounded by bold black boxes. Bits are ordered as in figure 11. The least Significant Bit (LSB) is assigned to band number 0 and the Most Significant Bit (MSB) is assigned to band number 6.

Band width		Level 4	Level 3	Level 2	Level 1	Level 0
0-25%	V0	0	2	6	Group 1	Base signal
		1				
		3	5			
		4				
25-50%	V1	0	2	6		
		1				
		3	5			
		4				
50-75%	V2	0	2	6	Group 2	
		1				
		3	5			
		4				
75-100%	V3	0	2	6		
		1				
		3	5			
		4				

Figure 8 : Clustering chart to explain the 4 points encoding algorithm.

As shown in figure 8, each cluster will be encoded by 7 bit valued number. The number is formed such that it reflects the tree structure within the cluster. Trees that cover the same bands will be almost adjacent trees. This property will be utilized in the scoring system. By considering all clusters, a vector of 4 components will be formed. Each vector’s component represents a certain cluster. And each cluster covers a certain area in the total bandwidth. This is the 4 point encoded method that construct BTE features vector.

Figure 9 introduces a simple example to explain features encoding for a frame of speech signal. Circles in figure 9 represent leaf nodes in the best tree decomposition.

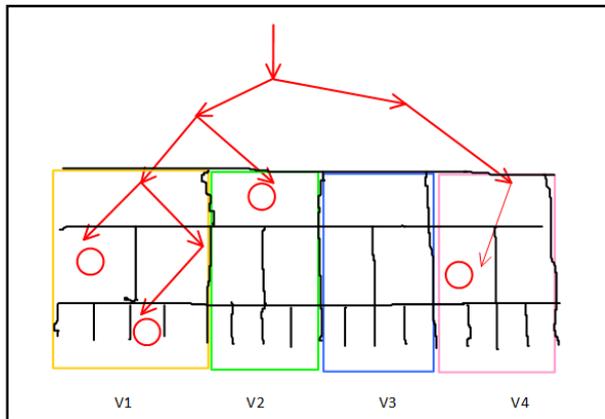


Figure 9: Best tree 4 point encoding example.

The indicated tree structure in figure 9 will be encoded into features vector of 4 elements as shown in table 2.

Table 2 : Best tree 4 point encoding evaluation.

Element	Binary Value	Decimal value	Frequency Band
V1	0001100	12	0 - 25 %
V2	1000000	64	25% - 50%
V3	0000000	0	50%-75%
V4	0000100	4	75%- 100%

Features vector for this example speech frame will be:

$$F = \begin{pmatrix} 12 \\ 64 \\ 0 \\ 4 \end{pmatrix} \quad (9)$$

Matlab is used to implement BTE features extraction. The following code snippet is the core part of Matlab function to implement BTE features extraction.

```
function [res] = BTE (frame, depth)
    nbIn = nargin;
    nbout = nargout;
    if nbIn < 1 , error('Not enough input arguments.');
```

```
elseif nbIn == 1, level = 4;
elseif nbIn == 2, level = depth;
end;
if nbout < 1 , error('Not enough output arguments.');
```

```
end;
t = wpdec(frame,level,'db4','shannon');
u = leaves (t);
```

```

bt = besttree(t);
v = leaves (bt);
% res = score(v,0,4)/1000;
res = box4encoder(v);

end

```

The function "box4encoder" in the above code snippet is responsible for encoding Best tree as indicated in table 2. Matlab functions needed for this research are all packaged into a Class Library². This step makes it easy to call Matlab functions from within the C# development environment³ that is being used as Business and Cue Logic⁴ "BCL". The following Matlab command is used to invoke the packaging tool in Matlab:

Deploytool

Figure 10 explains the deploy tool utility that is available in Matlab 7.5. This is a very useful tool that enables calling for all Matlab functionalities from other more advanced software development environments.

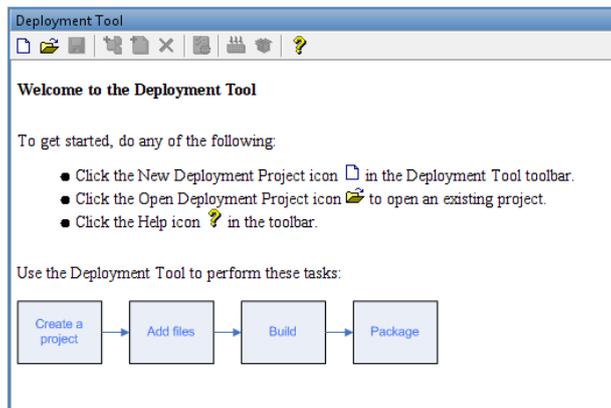


Figure 10: Deployment tool for packaging Matlab functions into Class Library suitable for calling from C# development environment.[5]

The Matlab function called "wav2BTE" is developed in Matlab. Part of the code of "wav2BTE" is indicated in the following code snippet.

```

[y fs] = wavread(file);
S = 20e-3*SamplingRate;
F = framing(y,S,0,0);
A = BTE (F(:,1));
for i = 2:n
    A = [A BTE (F(:,i))];
end;
version = uint32([3 1]);
Frame = uint32(20);
wpdepth =uint32( 4);
fid = fopen(outfile, 'wb');
fwrite(fid,version,'int32');
fwrite(fid,Frame,'int32');
fwrite(fid,wpdepth,'int32');

```

² Class library is the name of the entity used by Microsoft in the dot net framework to package functions and procedure. By packaging all needed functions into class library, we can reuse the functions from any dot net programming language for further use.

³ Dot net programming language by Microsoft Corporation.

⁴ Business and Cue Logic "BCL" is a name for all program snippets that is being written to control program sequencing. This includes loops, conditions, input and outputs.

```

fwrite(fid,uint32(fs),'int32');
fwrite(fid,size(A),'int32');
fwrite(fid,2,'int32');
fwrite(fid, uint16(A),'int16');
status = fclose(fid);

```

3. Testing BTE scoring system

This section is dealing with testing the scoring system of BTE features. As indicated before the scoring system is designed as to minimize the distance based on frequency coverage. Signals that has similar frequency spectrum are close and signals that have different frequency component are far. Figure 11 introduce the score of 4 BTE feature vectors. Check marks mark the frequency bands being covered by leaf's nodes. FV is the abbreviation for Feature Vector. As it shown in figure, Vectors A, B, C and D are almost identical vectors. They just differed in 19% and 25% Bandwidth components of wavelet packets. The scoring makes B and C are too close while A and D are too far. This is logical as vector A has no resolution in level 4 while B and C have adjacent components in level 4. Also D has no component at all in 19% and 25%. Also C is equally distant from D and B. This is also logical as the 19% component at level 4 for vector C is in the middle between the 13% component at level 4 for vector D and the 25% component at level 4 for vector B.

Bandwidth	6%	13%	19%	25%	31%	38%	44%	50%	56%	63%	69%	75%	81%	88%	94%	100%
A				1			<input checked="" type="checkbox"/>						1			<input checked="" type="checkbox"/>
FV=	1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>					1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>				
	35				64				36				64			
B				1			<input checked="" type="checkbox"/>						1			<input checked="" type="checkbox"/>
FV=	1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>					1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>				
	19				64				36				64			
C				1			<input checked="" type="checkbox"/>						1			<input checked="" type="checkbox"/>
FV=	1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>					1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>				
	11				64				36				64			
D				1			<input checked="" type="checkbox"/>						1			<input checked="" type="checkbox"/>
FV=	1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>					1	<input checked="" type="checkbox"/>	1	<input checked="" type="checkbox"/>				
	3				64				36				64			

Figure 11: Scoring sheet that explains the scoring of 4 different feature vectors.

The above discussion explains that the scoring hold information that we can rely in the recognition system. The above discussion is summarized in figure 12. As it is indicated in figure 12, vector C are in the middle path between B and D. Vector A and D are at the far limits.

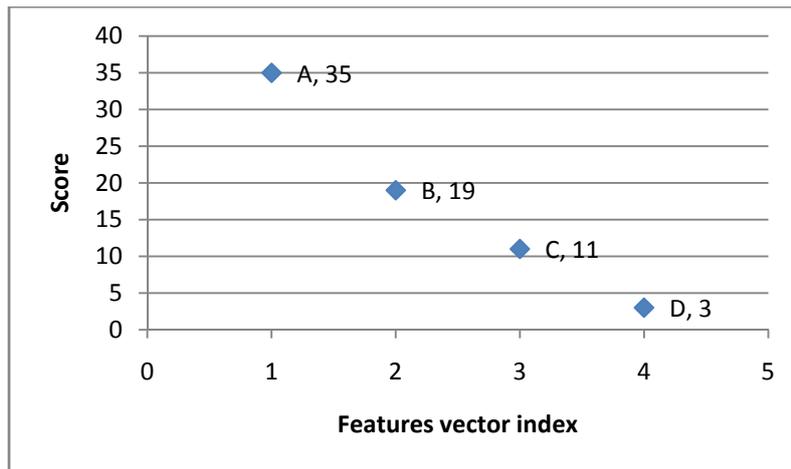


Figure 12: Summary results of scoring system

4. Conclusions

Wavelet packets make a similar processing on speech signal as the Filter banks method. It is much smarter than filter banks in that the number of filters is adapted by considering signal entropy to find the best tree. The problem of having dynamic size feature vectors is solved by considering the 4 points encoding algorithm. The proposed encoding system grants that minimizing distance between feature vectors based on adjacency in frequency domain. This adjacency based on frequency domain of feature vectors distance calculation makes (BTE) features are highly promising in speech recognition systems.

5. References

- [1] Amr M. Gody, "Natural Hearing Model Based On Dyadic Wavelet", The Third Conference on Language Engineering CLE'2002, Page(s): 37-43, October 2002
- [2] Alessia Paglialonga, "Speech Processing for Cochlear Implants with the DiscreteWavelet Transform: Feasibility Study and Performance Evaluation", Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006
- [3] Mel scale, http://en.wikipedia.org/wiki/Mel_scale
- [4] Gilbert Strang, "Wavelets and filter banks", Wellesley-Cambridge Press, ISBN: 0-9614088-7-1, pp. 37-86, ©1996.
- [5] MatLab, http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/ch06_a11.html.
- [6] Coifman, R.R.; M.V. Wickerhauser (1992), "Entropy-based algorithms for best basis selection," IEEE Trans. on Inf. Theory, vol. 38, 2, pp. 713-718.
- [7] Hai Jiang, Meng Joo Er and Yang Gao, " Feature Extraction Using Wavelet Packets Strategy", Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii USA, December 2003
- [8] http://en.wikipedia.org/wiki/Information_entropy.
- [9] Coifman, R.R.; M.V. Wickerhauser (1992), "Entropy-based algorithms for best basis selection," IEEE Trans. on Inf. Theory, vol. 38, 2, pp. 713-718.