2000년도 대한전기학회 하계학술대회 논문집 2000.7.17-20

Wavelet Packet을 이용한 Network 상의 음성 코드에 관한 연구

백한욱, 정진현
광운대학교 제어계측 공학과

A Study of Speech Coding for the Transmission on Network by the Wavelet Packets

Han Wook Baek, Chin Hyun Chung
Dept. of Control & Instrumentation Engineering, Kwangwoon Univ.

Abstract

In general, a speech coding is dedicated to the compression performance or the speech quality. But, the speech coding in this paper is focused on the performance of flexible transmission to the network speed. For this, the subbanding coding is needed, which is used the wavelet packet concept in the signal analysis.
The extraction of each frequency-band is difficult to general signal analysis methods, after coding each band, the reconstruction of these is also a difficult problem. But, with the wavelet packet concept(perfect reconstruction) and its fast computation algorithm, the extraction of each band and the reconstruction are more natural.
Also, this paper describes a direct solution of the voice transmission on network and implement this algorithm at the TCP/IP network environment of PC.

1. Introduction

The speech coding algorithm in this paper is aim to the speech transmission which flexibly changes the speech coefficient's resolution and then, with this, a adaptive service for the network environments is available. For this, the decomposition method of wavelet packets is used in the signal analysis, the data extracted to each frequency-band with this is coded to the speech coefficients by "Bit-Allocation & Quantization". The method's application to each frequency-band data extracted by the wavelet packets is first of all needed the determination of its optimum processing parameters, which can reduce the reconstruction error to the minimum and simultaneously must consider each frequency-band's weight in voice. Also, the usage of this algorithm to the real network communication is described by the implementations.
At first, it describes the theorem of background theories for this algorithm.

2. Wavelet Packet

A wavelet packet is a generalization of a wavelet in that each octave frequency band of the wavelet spectrum is further subdivided into finer frequency bands by using the two-scale relations repeatedly. In other words, the development of wavelet packets is a refinement of wavelets in the frequency domain and is based on a mathematical theorem proven by Daubechies[3](**splitting trick**). The theorem is stated as follows:

If $f(\cdot - k)|_{k\in z}$ forms an orthonormal basis and

$$F_1(x) = \sum_k h_0[k] f(x-k) \quad (1)$$

$$F_2(x) = \sum_k h_1[k] f(x-k) \quad (2)$$

then $\{F_1(\cdot - 2k), F_2(\cdot - 2k); k\in Z\}$ is an orthonormal basis of $E = span\{f(\cdot - n); n\in Z\}$.
This theorem is obviously true when f is the scaling function φ since the two scale relations for φ and the wavelet ψ give

$$V_j \ni \varphi(2^j t) = \sum_k h_0[k]\varphi(2^{j+1}t - k) \quad (3)$$

$$W_j \ni \psi(2^j t) = \sum_k h_1[k](2^{j+1}t - k) \quad (4)$$

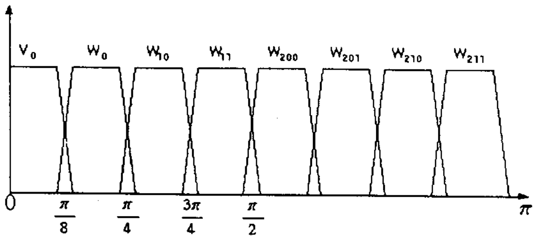
If we apply this theorem to the W_j spaces, we generate wavelet packet subspaces.
The general recursive formulas for wavelet packet generation are

$$\mu_{2l}(t) = \sum_k h_0 \mu_l(2t-k) \quad (5)$$

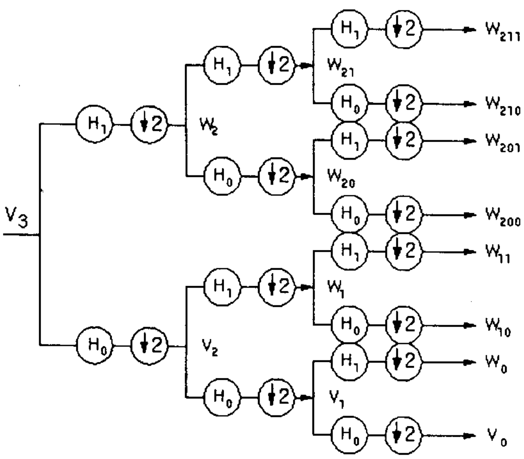
$$\mu_{2l+1}(t) = \sum_k h_{1[k]}\mu_{l(2t-k)} \quad k\in Z \quad (6)$$

where $\mu_{0=\varphi}$ and $\mu_{1=\psi}$ are the scaling function and the wavelet, respectively.
The translations of each of these wavelet packets form an orthogonal basis and the wavelet packets are orthogonal to one another within the same family generated by an orthonormal scaling function. Then, we can decompose a signal into many wavelet packet components and remark that a signal may be represented by a selected set of wavelet packets without using every wavelet packet for a given level of resolution.

The decomposition tree for wavelet packets uses the same decomposition block of two parallel filtering channels followed by decimation by 2 ($\downarrow 2$) as in the wavelet algorithm. Any



(Fig. 1) Frequency Response for the Two-Ban Wavelet Packet Filter Bank



(Fig. 2) The full binary tree for the three-sca wavelet packet transform

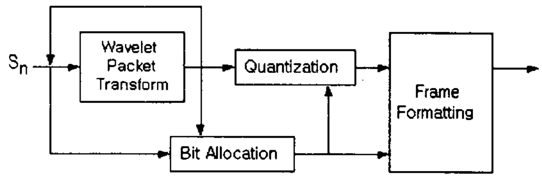
coefficient set in the tree (Fig.2) may be processed by this block. In the wavelet decomposition tree, only the approximation coefficient sets (V_j) are processed for different resolutions, while the wavelet coefficient sets (W_j) are outputs of the algorithm. In the wavelet packet decomposition, the wavelet coefficient sets are also processed by the same building block to produce wavelet packet coefficient sets.

The number of coefficient sets is 2^m if the original coefficient set is processed for m resolutions.

It is important to keep track of the indices of the wavelet packet coefficients in the decomposition algorithm. To achieve perfect reconstruction, if a coefficient set has been processed by $h_0[n]$ and ($\downarrow 2$), the result should be processed by $g_0[n]$ and ($\uparrow 2$). The same order is applicable to $h_1[n]$ and $g_1[n]$. $g_0[n]$ and $g_1[n]$ are the reconstruction wavelet filters.

3. Speech Coding

In this paper, it use total energy and a weighted value of each band a criterion to carry out bit allocation. The weighted value of each band in speech is decided by its information quantity of speaker's intention: valid voiced/unvoiced. Thus, each band is applied different bit allocation. And, if the energy of a band is high, that band is important and certain bits are allocated to it. This process is repeated until all the available bits are used up. The total available bits are determined according to the required bit rate



(Fig. 3) Diagram of Speech Coding

After bit allocation, the coefficients in each band that has non-zero bit allocation are normalized with the maximal one in that band as a scalefactor, and then quantized according to the number of bits allocated to that band with simple linear quantizer. In this way, a segment of a signal is now represented by the coded coefficients, the index of the decomposition level, the indexes of band that have non-zero bit allocation, and scalefactors. All the information is then formatted into a frame for transmission.

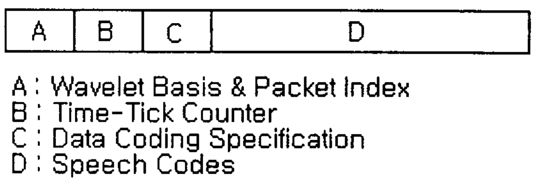
4. Implementation Algorithm

Since the speech coding in this paper is finally aim to speech transmission adaptively to network speed, the transmission data stream's structure is also important. Thus, the structure of data stream must include the following properties.

- Time-tick marker data for synchronous processing and checking network speed
- Used wavelet basis marker data for the reconstruction
- Applied encode-processing specification for decoding speech factors
- Different socket port for select-processing adaptively to network speed

The fourth of upper items is in order that if network speed is slow, data from more weighted band's port are preceded.

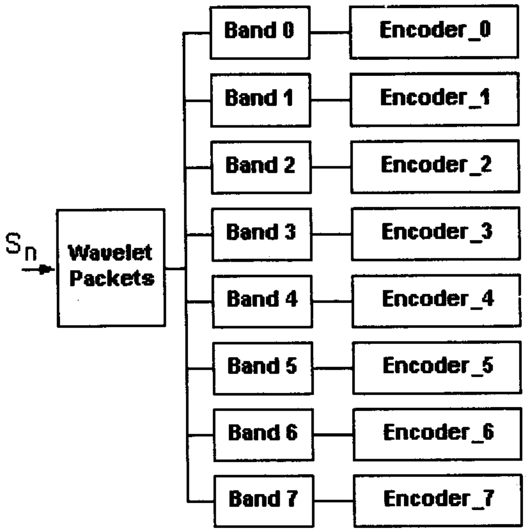
Fig. 4 describes the implemented structure of data stream



(Fig. 4) Data Stream

hen speech signals are used, the reconstructed signals are slightly more blurred, especially for high frequency parts, it is the cause that high frequency band is assigned little bits in bit allocation. The number of each band's bit assignment is decided by the band's assumed weight and energy, which relate to speaker's valid frequency band of intention communication.

Since the received part checks network speech with successive time-tick counter data, select-processing can be available. Since each band is formatted at different socket's port, handling of each band is more easy by the simple extraction of band data.



(Fig. 5) Process Diagram

5. Conclusion

In this paper, each frequency band analyzed by wavelet packet is encoded its data stream, and is formated at different network socket port. The division of each band's transmission port is available to select-decoding adaptively with checking network speed. Speech signals are encoding in different bands according to their characteristics. An adaptive speech transmission service can be achieved by the combination of sub-banding and efficient usage of network protocol. The advantages of speech transmission scheme in this paper are more adaptive to network speed and available select-processing.

[참 고 문 헌]

[1] Haitao Guo / C.Sidney Burrus, *Introduction to Wavelets and Wavelet Transforms*, Prentice-Hall

[2] L.R.Rabiner, *Digital Processing of Speech Signals*, Page 120 Prentice-Hall

[3] I.Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Ser.Appl.Math.61.Philadelphia: SIAM, 1992, 326 pp

[4] Lawrence Rabiner / Biing-Hwang Jung, *Fundamentals Of Speech Recognition*, Prentice-Hall

[5] Richard A. Roberts / Clifford T. Muilis, *Digital Signal Processing*, Addison Wesley

[6] Mark Kahrs / Karlheinz Brandenburg, *Application Of Digital Processing To Audio And Acoustics*, KAP

[7] Jaideva C. Goswami & Andrew K. Chan, *Fundamentals of Wavelets* pp.171,281-, Wiley Interscience

[8] W. Richard Stevens, *TCP/IP Illustrated Volume I*,Addison-Wesley