Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI:10.21474/IJAR01/9712
DOI URL: http://dx.doi.org/10.21474/IJAR01/9712
ISSN NO. 2320-5407

RESEARCH ARTICLE

TEXT INDEPENDENT SPEAKER RECOGNITION SYSTEM USING WAVELET FILTERS.

R. A. Deshpande, Dr. B. B. Musmade and Vaishnavi V Pawar.
D Y Patil College of Engineering Akurdi, Pune.

Manuscript Info

Manuscript History
Received: 08 July 2019
Final Accepted: 10 August 2019
Published: September 2019

Key words:-
feature extraction, discrete wavelet transform, vector quantization.

Abstract

This paper depicts a new speech feature extraction technique for use in automatic speaker recognition (ASR). Wavelets have been shown to be successful front end processors for speaker recognition. Front-end or feature extractor is the first component in an automatic speaker recognition system. Feature extraction transforms the raw speech signal into a compact but effective representation that is more stable and discriminative than the original signal. Since the front-end is the first component in the chain, the quality of the later components (speaker modeling and pattern matching) is strongly determined by the quality of the front-end. In other words, classification can be at most as accurate as the features.

Wavelet Transform coefficients can be utilized to feature parameters in various forms. An appropriate transformation base is also important for the feature extraction. To use the coefficients directly as the feature is the simplest way to exploit the wavelet transform characteristics. We thus introduce a simple feature extraction model based on the result of DWT. In order to parameterize the speech signal, we should first decompose the signal in the dyadic form using the Mallat algorithm

Speaker Recognition System in text independent mode has been developed and tested on set of 100, 50 and 25 speakers. The performance of system was tested with 2 sec, 3 sec and 5 sec speech samples which were different from the text used for training (i.e., text independent mode) respectively. A speaker was said to be identified if the Euclidean distance is minimum for VQ.

Introduction:-

Wavelet transform has been successfully applied to various fields such as speech and image processing. In particular, wavelet transforms have been successfully applied to image coding, providing high compression ratio. And various filter banks have been proposed and evaluated for image compression [2]. On the other hands, this wavelet decomposition has been increasingly applied in speaker recognition [3,4], largely due to its ability to produce features that contains information of narrow and wide bands without assuming a stationary signal. In contrast to other transforms, the wavelet transform can provide local information of speech features that might be useful for speech recognition. And it is expected that the performance of wavelet transform would be influenced by the choice of the structure of filter banks and the type of wavelet filters. In this paper, we propose to use symmetric octave band filter banks for speaker recognition and evaluate various orthogonal and biorthogonal wavelet filters.

Corresponding Author:-R. A. Deshpande.
Address:-D Y Patil College of Engineering Akurdi, Pune.

Wavelet Decomposition

In this section, we will briefly discuss the wavelet transform. A typical 2-channel filter bank is shown in Fig. la. Signal X(n) is applied to analysis filters H_0 (low pass filter) and H_1 (high-pass filter). The outputs of these filters, after down-sampling by 2, constitute the reference signal $r_1(n)$ and detail signal $d_1(n)$ for a one-level decomposition. Then, $r_1(n)$ and $d_1(n)$ are formulated as follows [5,6,7]:

$$z[r_1(n)] = \frac{1}{2}\left[H_0(z^{1/2})X(z^{1/2}) + H_0(-z^{1/2})X(-z^{1/2}) \right] \quad (1)$$

$$z[d_1(n)] = \frac{1}{2}\left[H_1(z^{1/2})X(z^{1/2}) + H_1(-z^{1/2})X(-z^{1/2}) \right] \quad (2)$$

Where Z[ ] represents the Z-transform operator. For perfect reconstruction, the following conditions should be met

$$G_0(z)H_0(z) + G_1(z)H_1(z) = 2z^{-1} \quad (3)$$

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0 \quad (4)$$

where G_0 and G_1 are synthesis filters [5]. In orthogonal filters, the following orthogonal condition is satisfied,

$$< g_i[n-2k].g_j[n-2l] >= \delta[i-j]\delta[k-l] \quad (5)$$

And in biorthogonal filters, the following biorthogonal condition is satisfied,

$$< h_i[-n].g_j[n-2l] >= \delta[i-j]\delta[l] \quad (6)$$

In order to construct multi channel filter banks, we can cascade two channel filter banks. For instance, the octave band filter bank is shown in Fig. 1 b.

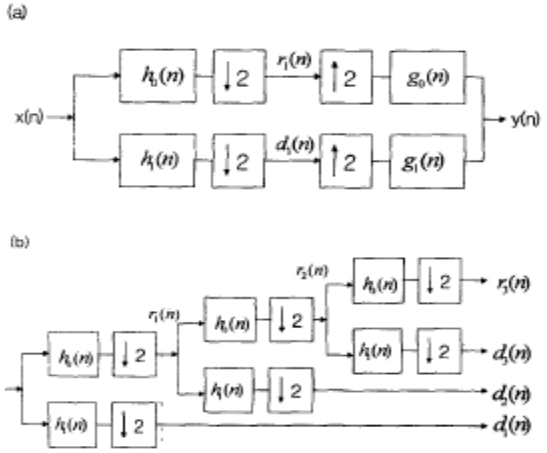

Fig 1:-(a) 2 channel filter bank, (b) 3-level octave band filter bank.

Symmetric Octave Filter Bank

Structure of filter, banks determines time and frequency resolutions. For example, full tree structured filter banks yield an equal division of the spectrum, which is similar to the short-time Fourier transform. On the other hand, the octave band filter bank provides time-frequency resolution as shown in Fig. 2a. A problem with the octave band filter bank is the relatively wide bandwidths of the high frequency spectrum. In speech recognition, in order to distinguish between consonants, we need detailed information of high frequency components of speech signal. Apparently, the octave band filter bank fails to provide such information. In order to overcome this problem, we propose a symmetrical octave band structured filter bank as shown in Fig. 2b. The symmetrical octave band structured filter bank provides better high frequency resolutions than the octave band filter bank and produces fewer features than the full tree structured filter bank. In this paper, we will use a 5-level symmetric octave structure filter bank.

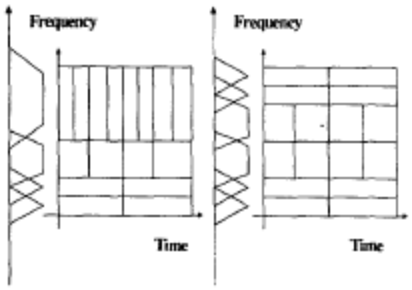

Fig 2:-(a) Wavelet tiling of octave band structured filter bank, (b) wavelet tiling of symmetric octave band structured filter bank.

Feature Extraction

We now describe the feature extraction method based on the symmetrical octave band structured filter bank. First, speech signal S(n) is divided into frames by applying a window function

$$S(n) \rightarrow [S_0(n), S_1(n), ..., S_{i-1}(n)] \tag{7}$$

And each frame Si(n) is decomposed by the symmetrical octave band structured filter bank. And we compute each subband energy of Si(n) as follows:

$$e_j = \sum_n \left| S_{i,j}(n) \right|^2 \tag{8}$$

And we propose to use $\{e_j\}$ as a new feature for speaker recognition. Thus, for each frame, we obtain a feature vector whose elements are $\{e_j\}$. And we normalize the feature vectors by dividing the total input energy. The feature extraction procedure is illustrated in Fig. 3. Since we use the HMM as a recognizer, we applied the vector quantization to the feature vectors [I].

Evaluation of Wavelet Filters

In order to evaluate various, wavelet filters for speaker recognition, we performed experiments with

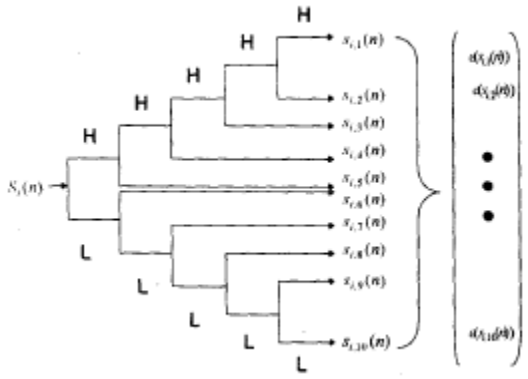

Fig 3:-Extracting procedure of 10 dimension feature by 5 level symmetric octave band filter bank.

English digit words. The sampling rate was 11 kHz. The speech data consists of 100,50,25 speakers. And each speaker spoke each digit 10 times.. The Hamming window with the length of 20 ms was used. We used the 5-level symmetric octave structure filter bank in order to obtain a feature vector from each frame. And we tested various Daubechies' orthogonal and biorthogonal filters. Classifier using Vector Quantization with codebook size of 16 is

used. Percentage error is calculated for a set of 100, 50 and 25 speakers. Table 1 summarizes the percentage errors for training duration of 20 sec and test durations of 5 sec. Sym-8 gives comparatively good results which are shown as (DWTC).

Table 1:-Percentage error rate

Sr. No.	No. of Speakers	% Error of 'SIS With Features of'
		DWTC
1	100	13.33 @ 5 sec
2	50	6.67% @ 5 sec
	25	2.33 @ 5 sec

@ Indicates Test Time

Summary & Conclusions:-

As per the practical experimental results,

1. Wavelet based features are more noise roboust, Wavelet based feature extraction is best applicable when the recording is done in noisy environment. However the performance varies by wavelet to wavelet (Harr, Sym-N, Daubechies-N, morlet, Mexican hat, etc.). Sym-8 gives comparatively good results.
2. Speaker Identification System works excellent for long duration training session (20 sec) and testing session (5 sec).

In this paper, we explored the possibility to use wavelet filters for feature extraction in speech recognition and evaluated the various Daubechies' wavelet filters. Although experimental results are not conclusive, the wavelet transform may provide valuable features for speech recognition with a reasonable computation cost.

References:-

1. Igor Bisio, et al., "Smart and Robust Speaker Recognition for Context-Aware In-Vehicle Applications", IEEE Transactions on Vehicular Technology , Vol. 67 , No. 9 , 2018.
2. J. Liu, N. Kato, H. Ujikawa, K. Suzuki, "Device-to-device communication for mobile multimedia in emerging 5G networks", **ACM Trans. Multimedia Comput. Commun. Appl.**, vol. 12, no. 5 s, 2016.
3. J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, "Internet of things (IoT): A vision architectural elements and future directions", **Future Gener. Comput. Syst.**, vol. 29, no. 7, pp. 1645-1660, 2013.
4. D. Zeng, S. Guo, Z. Cheng, "The web of things: A survey", **J. Commun.**, vol. 6, no. 6, pp. 424-438, 2011.
5. C.J.Long C. J., Datta, S., "Wavelet Based Feature Extraction for Phoneme Recognition", IEEE Trans. Acoustics, Speech, signal Proc., 2003, pp.264-267.