Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)

Shiv Kumar, Aditya Shastri

Abstract—A good amount of work has been done in the field of compression, voice signal detection, and spectrum analysis which has been generated a number of results in the past few decades. In this research, following three important problems have been identified:

1. *To distinguish between constitutional and unconstitutional Voice: It is an important task to identify authenticity of recorded voice of the specific person. Here it has been tried to develop a Simulator which identifies constitutional and unconstitutional voice.*
2. *To identify words sequence:It is an important task to recognize words sequence in the recorded voice. Sometimes voice may be recorded fast, clear, or loud. Here it has been tried to develop a simulator to checkout whether recorded words are in proper sequence are not.*
3. *To develop a simulator which does not change file extension and quality of voice signal after compression: Normally, after compression, file extension is changed and quality of the voice signal is deteriorated. Here it has been tried not to change extension of the file after compression with minor distortion in voice signal.*

As per review of above three problems, it is being considered a simulator may be designed which may resolve above problems. With this view, the research title is chosen as "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)" which is suitable for pervasive computing, voice signal detection, and spectrum analysis. AVSDC is divided into following two parts:

1. *Automatic Voice Signal Detection (AVSD)*
2. *Automatic Voice Signal Compression (AVSC)*

Automatic Voice Signal Detection (AVSD) is used to identify constitutional and unconstitutional voice signal automatically which is performed on the basis of frequency, pitch value, formant value, and sequence of words in the voice signal for several samples of the same voice. An underline purpose of AVSD is to identify fake voice in the security system. Frequency is being mapped to the frequency domain by computing its DFT using the FFT algorithm. Sequence of words is computed by continuously computing difference between absolute averages of two adjacent significant windows and comparing it to a predefined threshold.

Word Identification System is part of AVSD which is designed to checkout whether recorded words in proper sequence are not. Normally, sometimes spoken words of voice may be recorded very fast, smoothly, or loudly. The main idea behind the word identification system is to first train it with several versions of the same word, thus yielding a "reference fingerprint". Then, subsequent words can be identified based on how close they are to this fingerprint. The whole idea is evaluated on the basis of Euclidean distance theory.

Automatic Voice Signal Compression (AVSC) takes .wav stereo file as an input and compress 50 to 60 percent of the source file at about 45 kbps with high quality voice signal by taking the help of adaptive wavelet packet decomposition and psychoacoustic model. AVSC takes .wav stereo file as an input and creates .wav mono file after compression. After compression minor distortion is also possible. The main feature of AVSC is that file extension does not change after compression. In other words, compression is done from .wav to .wav extension. AVSC takes .wav stereo file as an input and after compression it creates .wav mono file as an output. AVSC also computes entropy and SNR (Signal to Noise Ratio) of the source file during the compression.

Index Terms— MatLab7.0, Euclidean Distance Theory, Wavelet, Frequency Volue, Pitch Value, Average Significant Window

I. INTRODUCTION

Today, everyone wants to present their data in the form of multimedia which contains text, image, sound, graphics, and so forth. Although, multimedia data take large space in storage devices as compared to traditional data because traditional data did not use more images, graphics, sounds etc. Now a day, all academics, libraries, banking services, digital network services, conferences, and even security devices use multimedia data. Thus it is a challenge to the multimedia technologies which provides compressed and secure data to take less space in the storage device and provide secure and authentic service.

Though current generation uses multimedia technology. It is the need of hour to provide the challenging part of the multimedia where compression, authentication, and security are the main features. Number of efforts has been tried where some success were achieved, however it is felt that more attempts are required in the field of compression, authentication, and security. As a result, a greater emphasis is being placed in the design of new and efficient coders for voice communication and transmission.

It has always been an important task to identify authenticity of a recorded voice for an individual person. In the field of security and authenticity, several standards have been developed to serve many agencies, banking services, libraries, and lockers where voice signals, finger prints, eye prints, play a major role. If one is talking about the traditional work on voice signal, it may

Revised Manuscript Received on March 2012.

Shiv Kumar, Research Scholar, Department of Computer Science & Engineering, Banasthali University, Tonk (Rajasthan), India, (Email: shivksahu@rediffmail.com)

Prof. (Dr.) Aditya Shastri, Vice Chancellor Banasthali University, Tonk (Rajasthan), India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

provide better security and authenticity but not with complete success. A fake voice may unlock the security parameter. Thus it is felt that some more research is required to distinguish the fake and legal voice. There is still a need to develop a security parameter which symbolizes the constitutional and unconstitutional voice signal. This security parameter has been developed in this research which differentiates constitutional and unconstitutional voice signal with the help of frequency, pitch value, formant value, spectrum analysis, and distance between spoken words of voice signal.

In the field of compression, different software has been developed to compress voice signal which obtained very high compression ratios. However, after compression file extension is changed and quality of voice is also deteriorated. While analyzing further, after compression, it is observed that file extension is changed and to read compressed file, specific software is required. For example, if one is trying to read .mp3 file in the global network then it may be read only by mp3 player. It means, to read .mp3 file, mp3 player must be installed in the respective device which may create following drawbacks:

- Specific software is required for decompression.
- Large space needs to be enhanced.
- Higher data processing time is needed.
- Overall cost of the process will be higher.

Above shortcomings may be resolved by .wav extension. Wave file (supported to .wav extension) is a file that can run in any multimedia software. Here if voice is recorded in .wav extension and compression is also done with same extension, above shortcomings may be minimized. This is one of the research areas on which present research work is also focused by taking the help of adaptive wavelet packet decomposition and psychoacoustic model which minimize above shortcoming:

- Did not need specific software for decompression.
- Memory requirement is minimized.
- Data processing time is minimized.
- Overall cost the process is reduced.

This research work titled "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)" is divided in two parts:

1. ***Automatic Voice Signal Detection (AVSD)***
2. ***Automatic Voice Signal Compression (AVSC)***

Automatic Voice Signal Detection (AVSD) is used to identify constitutional and unconstitutional voice signal automatically which is performed on the basis of frequency, pitch value, formant value. The word identification is also part of AVSD which is designed to checkout whether spoken words of voice in proper sequence are not. Word identification is done by taking the help of Euclidean distance theory.

Automatic Voice Signal Compression (AVSC) takes .wav file as an input and compress 50 to 60 percent of the source file as an output at about 45 kbps with high quality voice signal by taking the help of adaptive wavelet packet decomposition and psychoacoustic model. After compression some minor distortion may be possible. AVSC takes .wav stereo file as an input and after compression it creates .wav mono file as an output. AVSC also compute entropy, SNR (Signal to Noise Ratio) of the source file during the compression.

II. LITERATURE SURVEY AND PROBLEM IDENTIFICATION

Jalal Karam [1] has been explained different techniques for wavelet compression in which speech processing for compression and recognition was addressed. Various methods and paradigms based on the time-frequency and time-scale domains representation for the purpose of compression and recognition were discussed along with their advantages and disadvantages. Level dependent and global threshold compression schemes were also examined in details. Work was good but wavelet family is not defined, how much data is compressed, and what will be file extension after compression [1], it is not defined. According to this previous work and drawback, the work is further extended in this research wok, where dubechies wavelet family 10 (dB10) is used for transparent compression at about 45 kbps. Compression is done 50 to 60 percent of the source file at about 45kbps with same extension i.e. .wav.

Sarantos Psycharis [2] has defined a technique for image compression by using lossy compression. The encoding of coefficients is done using run-length-encoding of the zeros. This previous work is done exclusively for image compression where after compression extension is changed. This work was good but after compression file extension is changed. To read this compressed file, specific software is required. What would be happen if one is to compress a file without changing the extension? This work can be further extended and approached in this research work where after compression file extension does not change (i.e. .wav to .wav).

P R Deshmukh [3] has talked about the Multi-wavelet Decomposition for Audio Compression in which he describes the performance of different types of wavelets for composing the transient audio signal. The wavelets that were investigated for this purpose are dubechies family of wavelets, called as wavelet packets and multi-wavelets. The performances of various wavelets are compared, based on compression ratio and the signal to noise ratio (SNR) value of the reconstructed signal. One of the main challenges to the application of multi-wavelets is the problem of multi-wavelet initialization (or better known as pre-filtering). In the case of scalar wavelets, the given signal data is usually assumed to be the scaling coefficients which are sampled at a certain resolution, and hence, multi-resolution decomposition can directly be applied on the given signal. Unfortunately, the same technique cannot be employed directly in the multi-wavelet setting. Some preprocessing has to be performed on the input signal, prior to multi-wavelet decomposition.

In this implementation work [3] that was based on multi-wavelet, gives good results in comparison to the daubechies bases. On the basis of the experiment done, it can therefore be concluded that multi-wavelets, with an appropriate choice of pre-filtering method, seem to represent a promising substitute for scalar wavelets in audio data compression problems. In this previous research work, how to analyze whether recorded voice is audible? i.e. Is the recorded voice in the range of human nervous system? This problem is resolved by in this research work which computes whether the recorded signal is in the range of human nervous system or not.

As per research work of Stefan Wabnik [4], voice recognition technology can be used more efficiently in car security systems. Now if voice recognition is done by .wav file then it can be better but if voice recognition is done by compressed .wav file then it will be best. To get compressed .wav file, this research work may be used.

Shaleena Jeeawoody [5] developed a more efficient car security system using the technology voice recognition. The function of this voice recognition car security system is to unlock only when it recognizes a password spoken by the password holder. As per results show that among the 15 words tested, no two voices overlapped. For a given word, the voice spectrum differs from one person to another. The limitation of words, here work is approached for N-bits voice detection.

Khalid Saeed [6], explained voice recognition where it is not defined that which type of file will recognize and how? While this thesis is used to identify constitutional and unconstitutional voice signal automatically which is performed on the basis of frequency, pitch value, formant value, and distance between words of voice signal for several samples of the same sound.

According to Maria Markaki[7], a novel feature set for the detection of singing voice in old and new musical recordings but what about if fake voice is recorded for voice detection. In this case, this problem cannot be solved. But, It is resolved in this research work where fake voice can be easy identified automatically which is performed on the basis of frequency, pitch value, formant value, and distance between words of voice signal for several samples of the same sound.

Trung Nghia [8] has calculated the bit rate results of approximately 25 kbps. Different decomposition levels and auditory masking models are defined where six-level decomposition produces a 64-band wavelet packet tree. For coefficients quantization, novel method is used. This previous work is further extended in this thesis where compression is done 50 to 60 percent of the source file at about 45 kbps with same extension. To do this, N level decomposition, hard threshold, Global Threshold, Tonality Computation, 16-bit compression scheme, 16 bit dynamic quantization (narrow range quantization) is used.

Frank A Russo [9] computed tonality of the recorded voice signal. If tonality is computed then further work is also possible which is done in this thesis. After tonality computation, Entropy, SNR is also computed and further compression is also done.

According to Mohamed Cherif Amara Korba [10], Daubechies Wavelet 8 (dB8) was used to compute SNR and decomposition was implemented by an efficient 5 level tree structure. If SNR is computed then further work can also extend for the compression. According to drawbacks of this previous work, work is further extended in this research where Daubechies Wavelet 10 (dB10) is used which for better result than dB8. Work is also further extended for the compression in this research study.

According to Grigor Marchokoy [11], speech signals were sampled at 8 KHz at 8 bits/sample and the results indicate 31% to 37 % compression of the source file which was good. This previous work is further extending where 50 to 60 percent compression is done at about 45 kbps with high quality audible voice signal with same extension.

III. PROPOSED METHODOLOGY

A. Block Diagram for AVSDC

The Proposed Block Diagram for "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)" is given in Figure 1. As mentioned in the Figure 1, AVSDC is divided into two major parts: first is AVSC (Automatic Voice Signal Compression) which takes .wav stereo file as an input and create .wav mono file after compression. It compresses 50 to 60 percent of source file at about 45 kbps with .wav extension by taking the help of adaptive wavelet packet decomposition and psychoacoustic model. The second is AVSD (Automatic Voice Signal Detection) which identifies constitutional and unconstitutional voice signal with the help of frequency, pitch value, and formant value. Word identification is also part of AVSD which is designed to checkout whether recorded words are in proper sequence are not. Normally, sometimes words may be recorded very fast, smoothly, or loudly. The main idea behind the word identification system is to first train it with several versions of the same word, thus yielding a "reference fingerprint". Then, subsequent words can be identified based on how close they are to this fingerprint. The whole idea is evaluated on the basis of Euclidean distance theory.

AVSC takes .wav stereo file as an input. Before execution of input file, some variables and parameters are decided like wavelet family = dB10, frame size = 2048 = number of points for each FFT, number of FFT = 5, psychoacoustic model=on, wavelet compression=on, quantization = on, heavy compression = off, etc. Now decomposition is done for N equal frames with the help of wavelet then wavelet compression scheme is decided with the help of threshold and masking technique. Psychoacoustic model computes Power Spectral Density (PSD), Power spectral density of tone masker (PTM), tone computation, entropy, and signal to noise ratio (SNR) of the voice signal. If recorded voice signal is in the range of human auditory system then voice signal is a tone and it will be a process for voice compression otherwise this processing will be interrupted. Now quantization and default number of bits allocation per frame is computed, offset value (Header Value) is calculated and it is put in the form of chunk of data.

This chunk of data is called encoded bit stream. This encoded bit stream is decoded and the frames are reconstructed for N number of equal frame which develops compressed .wav mono file.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

After compression, minor distortion may be possible which is computed in this phase.

AVSD takes compressed .wav file to differentiate constitutional and unconstitutional voice while word identification system is used to checkout whether words are sequenced in proper ways or not. It can take both mono and stereo .wav file for purpose of execution but compressed .wav file takes minimize the execution time.

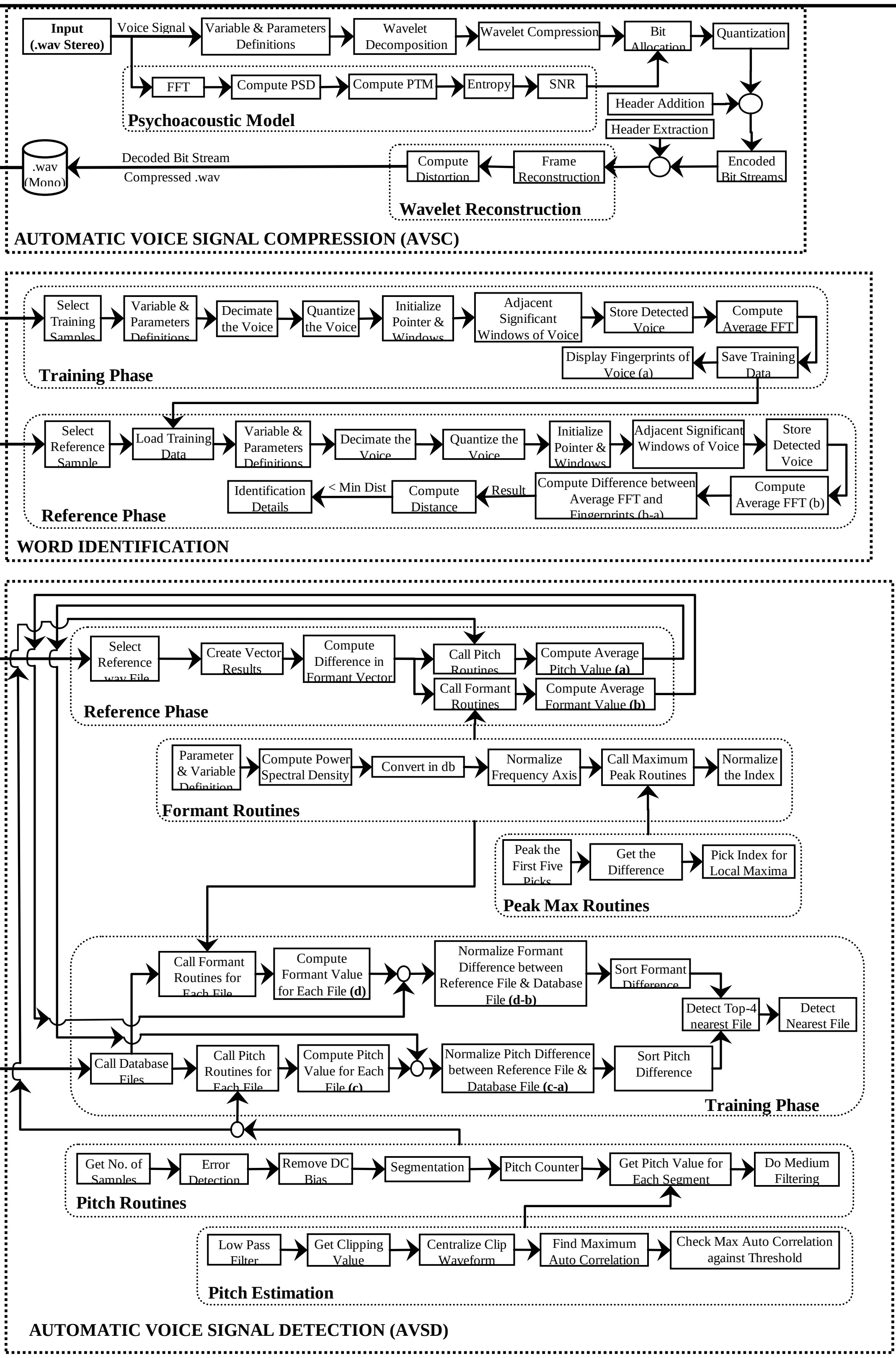


Fig.-1: Proposed Block Diagram for "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)"

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The word identification is done in two the phases: first is Training Phase and second is Reference Phase. In the Training Phase, some recorded .wav files are taken which are more than one. Before execution of Training Phase, some variables and parameters definition are fixed like number of points for each FFT = 1024, number of FFT = 7, signal length = 1.024 second, beginning of voice threshold = 0.05, window length = 1000, number of bits for quantization = 8, down sampling factor = 4, etc. Now voice signal is decimated by down sampling factor. It is done with the help of low pass filtering process. Voice is now quantized. After detection beginning word of voice, pointer and average window is initialized, and then adjacent significant window is computed.

This detected voice signal is stored and FFT value is evaluated which is saved in the temporary database for the purpose of identification in the Reference Phase. In the Reference Phase, only single .wav file is taken for the purpose of word identification along with Training Phase. The stored data of Training Phase is reloaded. Later, reloaded data is decimated, quantized, initialized the pointer and then adjacent significant window is estimated. This process is repeated for reference file. The detected voice signal is stored and FFT is computed. Now the difference between adjacent significant window for trainee and reference file is computed. The comparison between trainee and reference file is done by taking the Euclidean distance between them. To calculate this value, they are considered as five 1024-dimensional vectors (one for each matrix row), and the average of their respective Euclidean distance is computed. This is shown in Equation-1:

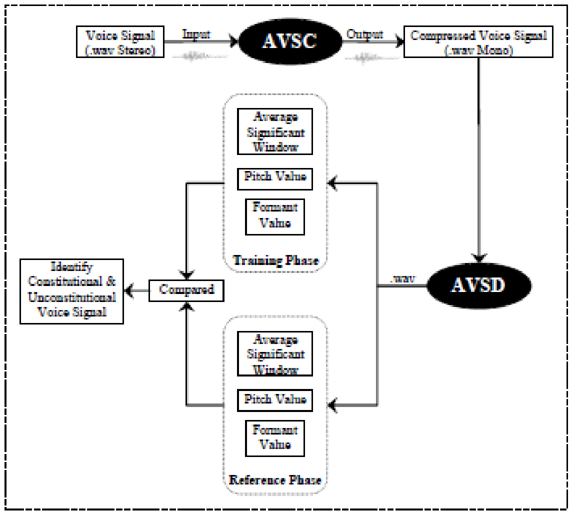
$$D = \frac{1}{5}\sum_{n=1}^{5}\sqrt{\sum_{i=1}^{1024}\left(a_{ni}-b_{ni}\right)^{2}} \text{......................(1)}$$

Where D is the distance, and ani and bni are the ith components of the voice signal. The n index points to each of the five vector pairs.

If the distance is less than maximum distance, then analyzed word is considered as to "identify the reference word". In this implementation, maximum distance is considered 140 which may vary.

AVSD differentiates constitutional and unconstitutional voice signal which is implemented into two phases: First phase is for Reference Phase while second is for Training Phase. At first, a reference .wav compressed file is loaded through which their pitch value and formant value is computed with the help of pitch routines and formant routines. Similarly, pitch value and formant value are also computed for all stored .wav files in the database by Training Phase. Now the difference between pitch value and formant value for Reference Phase and Training Phase is computed and calculated values are shortlisted. According to these shortlisted values, top matched .wav file is detected.

B. 0- Level DFD for AVSDC



[Fig.-2: 0-Level DFD for "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC)]

C. Algorithm for AVSDC

The proposed algorithm for "Design of Simulator for Automatic Voice Signal Detection and Compression (AVSDC) is defined in three parts: First part for "Automatic Voice Signal Compression (AVSC)", second for Word Identification, and third part for "Automatic Voice Signal Detection (AVSD) which are given below:

Algorithm for AVSC:

AVSC takes .wav stereo file as an input and compress it 50 to 60 percent of source file at about 45 kbps with .wav mono type extension. The algorithm for AVSC is given in following steps:

1. Recorded .wav stereo file is loaded
2. Define variable and parameter definition
3. Fixed Decomposition level = N
4. Break loaded voice signal into N frame
5. Define wavelet compression scheme
6. Fixed wavelet coefficients
7. Compute masking thresholds
8. Calculate power spectrum density (PSD)
9. Compute tonality
10. If tonality is within human nervous system then go to next step otherwise exit
11. Evaluate energy of the voice signal
12. Determine entropy of the voice signal
13. Calculate Signal to Noise Ratio (SNR)
14. Define quantization level
15. Compute offset value to shift memory location of entire partition
16. Reconstruct signal based multilevel wavelet decomposition structure
17. Evaluate distortion of the voice signal
18. Define wavelet expander scheme for reconstructed signal
19. Write compressed file in .wav mono type extension

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Algorithm for Word Identification System:

Word Identification System is part of AVSD which is designed to checkout whether recorded words are in proper sequence are not. Normally, sometimes words may be recorded very fast, smoothly, or loudly. The main idea behind the word identification system is to first train it with several versions of the same word, thus yielding a “reference fingerprint”. Then, subsequent words can be identified based on how close they are to this fingerprint. The whole idea is evaluated on the basis of Euclidean distance theory. The algorithm for word identification is given in following steps:

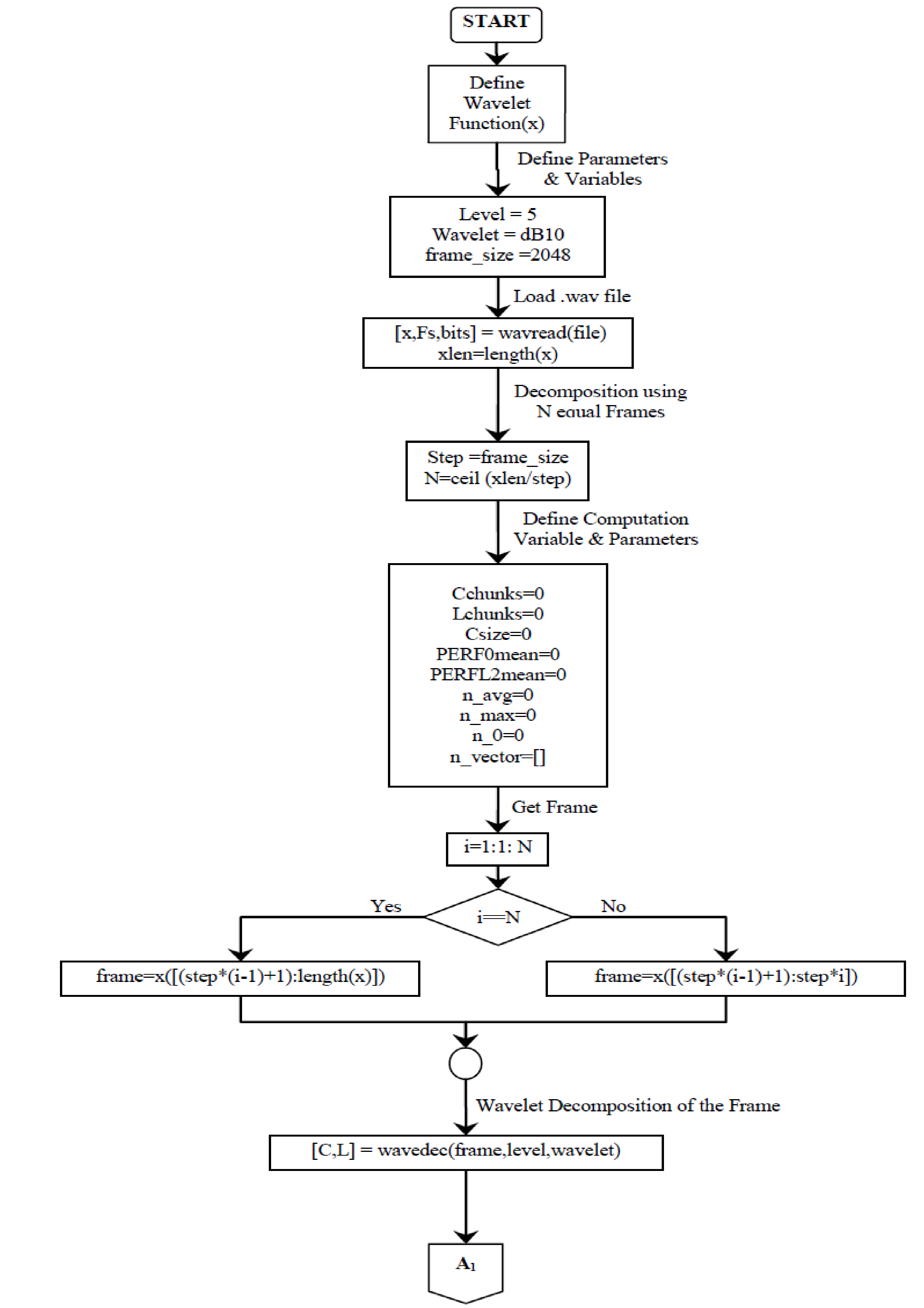
1. Select Training Voice Signal (more than one .wav file)
2. Define Variable and Parameters Definitions
3. Decimate the voice
4. Quantized the voice
5. Initialize Pointer
6. Initialize Windows
7. Compute Adjacent Significant Windows
8. Store detected Voice
9. Compute Average FFT
10. Save Training data
11. Compute Fingerprints (a)
12. Select Reference Voice Signal (Only one .wav file)
13. Load Training Data
14. Define Variable and Parameters Definitions
15. Decimate the voice
16. Quantized the voice
17. Initialize Pointer
18. Initialize Windows
19. Compute Adjacent Significant Windows
20. Store detected Voice
21. Compute Average FFT (b)
22. distance c = (b-a)
23. If distance ≤ 140 then word is identified
24. Otherwise word is not identified

Algorithm for AVSD:

AVSD differentiates constitutional and unconstitutional voice signal which is implemented into two phases: First phase is for Reference Phase while second is for Training Phase. At first, a reference .wav compressed file is loaded through which their pitch value and formant value is computed with the help of pitch routines and formant routines. Similarly, pitch value and formant value are also computed for all stored .wav files in the database by Training Phase. Now the difference between pitch value and formant value for Reference Phase and Training Phase is computed and calculated values are shortlisted. According to these shortlisted values, top matched .wav file is detected. The algorithm for AVSD is given in following steps:

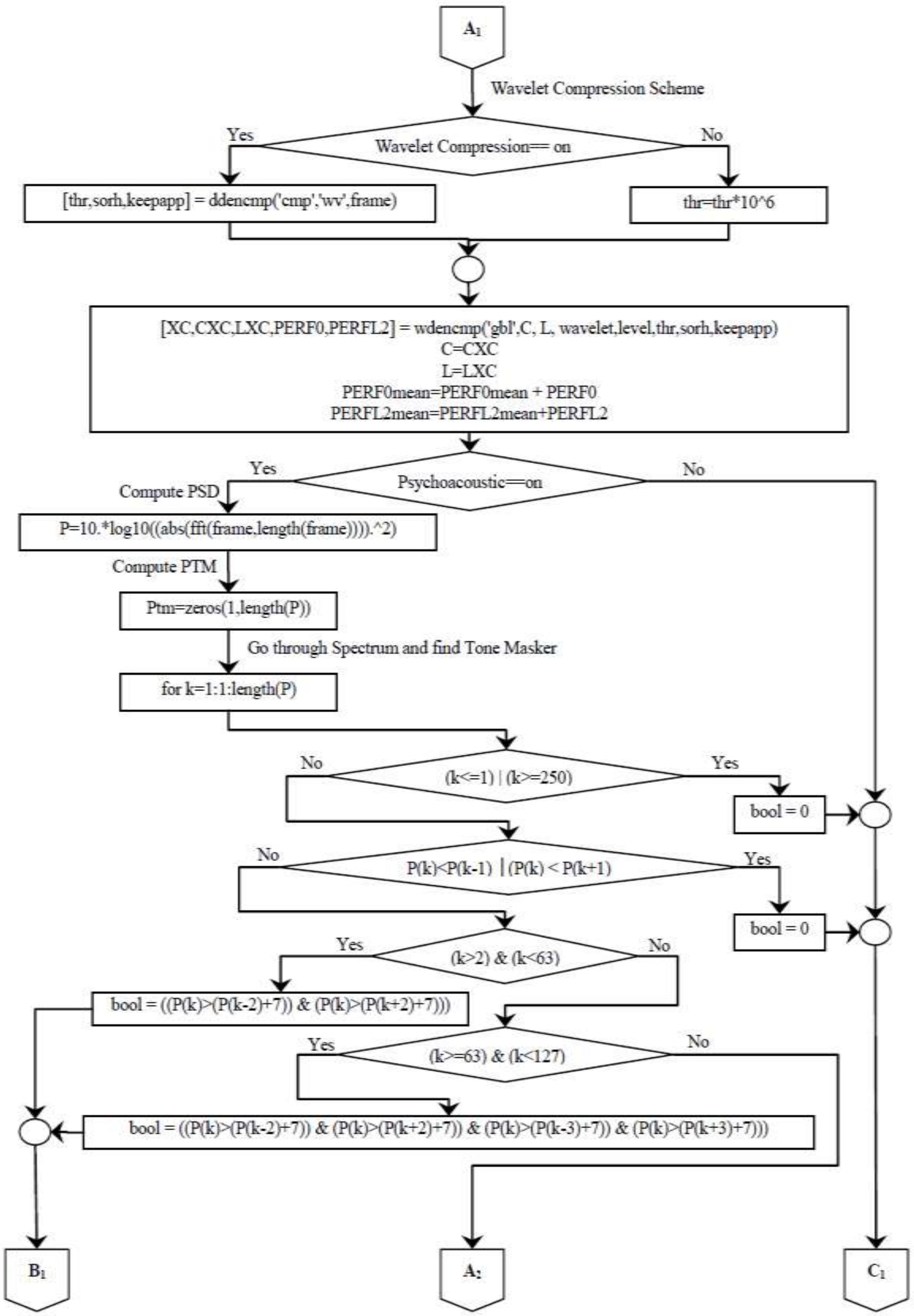
1. Select Training Voice Signal (more than one .wav file)
2. Compute Vector Results
3. Compute difference in Formant Vectors
4. Call Pitch Routines
   a. Get Number of Samples
   b. Error Detection
   c. Remove DC Bias
   d. Do Segmentation
   e. Set Pitch Counter
   f. Get Pitch Value for each segment
      i. Do Low Pass Filtering
      ii. Get Clipping Value
      iii. Centralize Clip Waveform
      iv. Find Maximum Auto Correlation
      v. Check Maximum Auto Correlation Against Threshold
   g. Do Medium Filtering
5. Compute Average Pitch Value (a)
6. Call Formant Routines
   a. Define Parameters and Variable Definitions
   b. Compute Power Spectral Density
   c. Convert in db
   d. Normalize Frequency Axis
   e. Call Maximum Peak Routines
      i. Peak the first five picks
      ii. Get the Difference
      iii. Pick Index for Local Maxima
   f. Normalize Index
7. Compute Average Formant Value (b)
8. Call Database files
9. Call Pitch Routines for each file
   a. Get Number of Samples
   b. Error Detection
   c. Remove DC Bias
   d. Do Segmentation
   e. Set Pitch Counter
   f. Get Pitch Value for each segment
      i. Do Low Pass Filtering
      ii. Get Clipping Value
      iii. Centralize Clip Waveform
      iv. Find Maximum Auto Correlation
      v. Check Maximum Auto Correlation Against Threshold
   g. Do Medium Filtering
10. Compute Pitch Value for Each File (c)
11. Call Formant Routines for Each File
   a. Define Parameters and Variable Definitions
   b. Compute Power Spectral Density
   c. Convert in db
   d. Normalize Frequency Axis
   e. Call Maximum Peak Routines
      i. Peak the first five picks
      ii. Get the Difference
      iii. Pick Index for Local Maxima
   f. Normalize Index
12. Computer Formant Value (d)
13. Normalize Pitch Difference between Reference file and Database File (a-c)
14. Normalize Formant Difference between Reference file and Database File (b-d)
15. Sort Pitch and Formant Difference
16. Get Nearest File

D. Flow Chart for AVSC



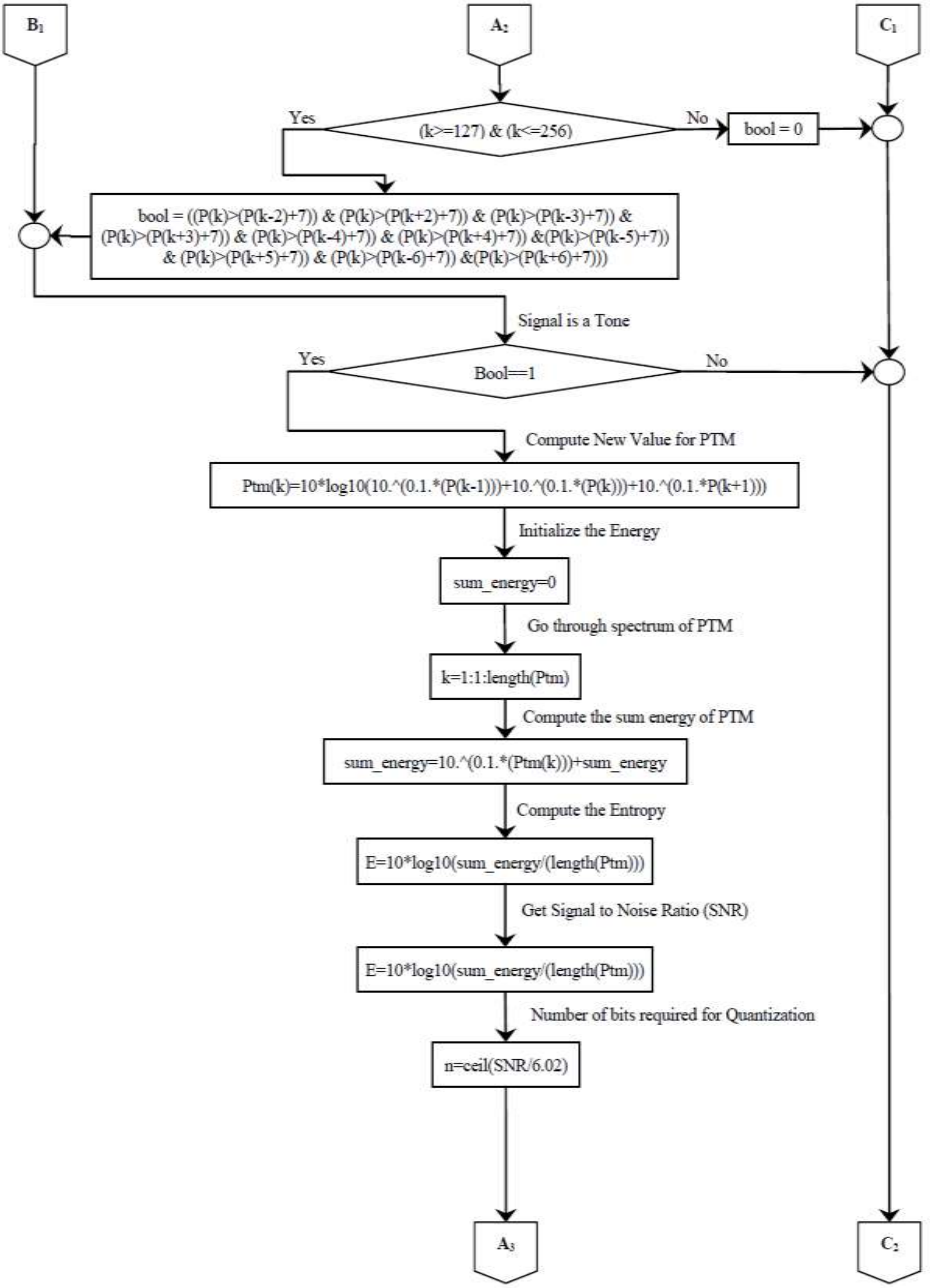
[Fig.-3: Flow Chart for AVSC]

Continue…



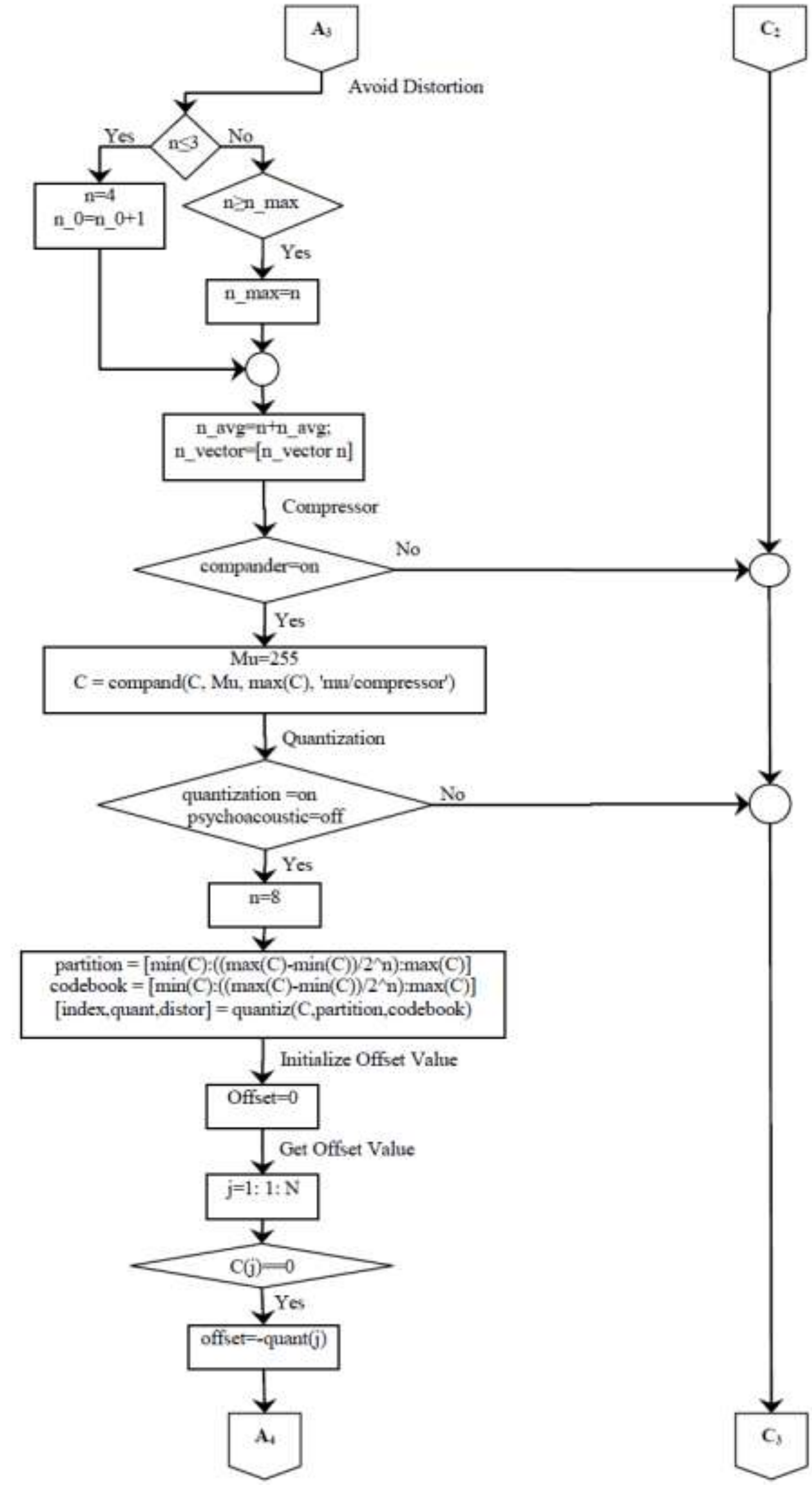
[Fig.-3: Flow Chart for AVSC]

Continue…



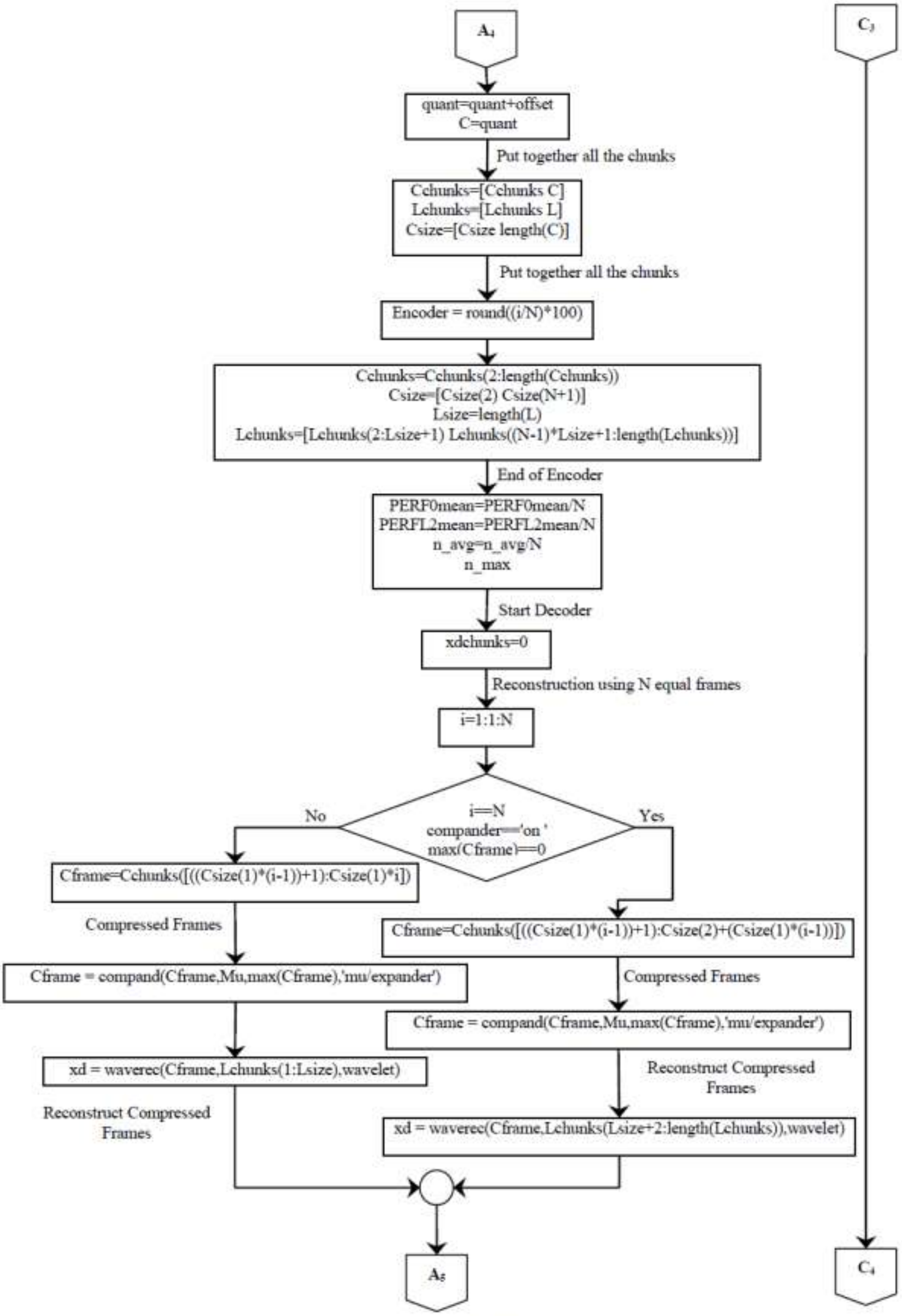
[Fig.-3: Flow Chart for AVSC]

Continue…



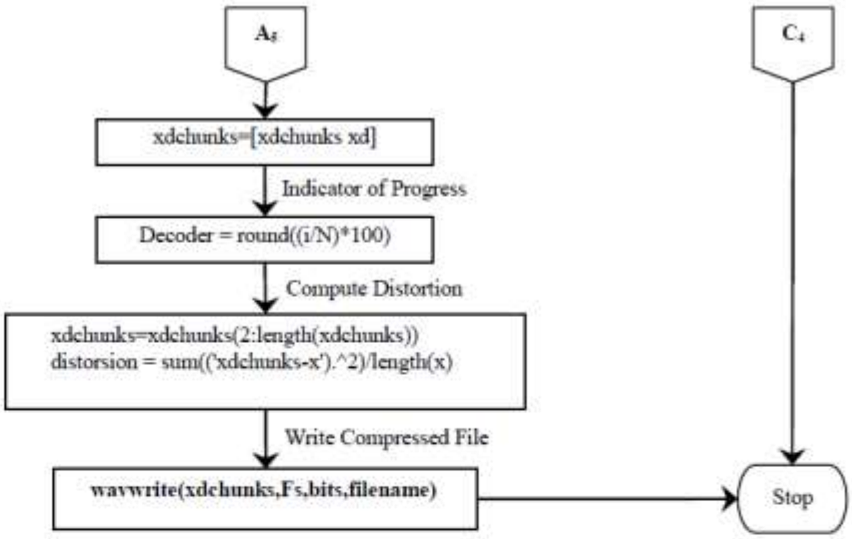
[Fig.-3: Flow Chart for AVSC]

Continue…



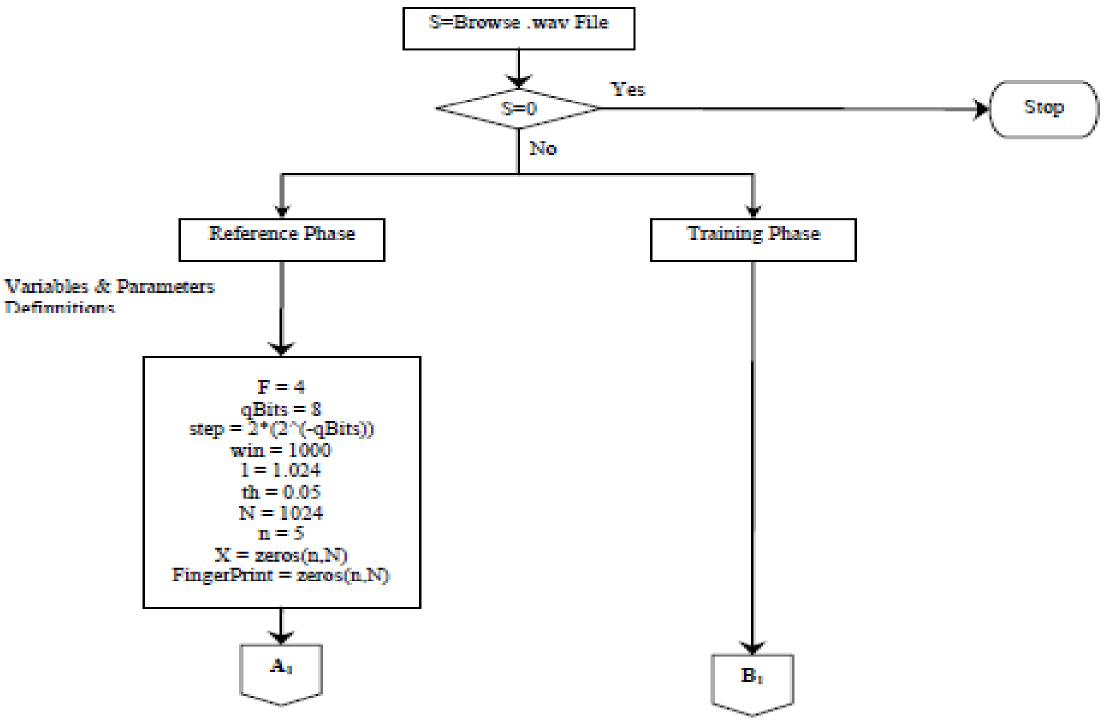
[Fig.-3: Flow Chart for AVSC]

Continue…



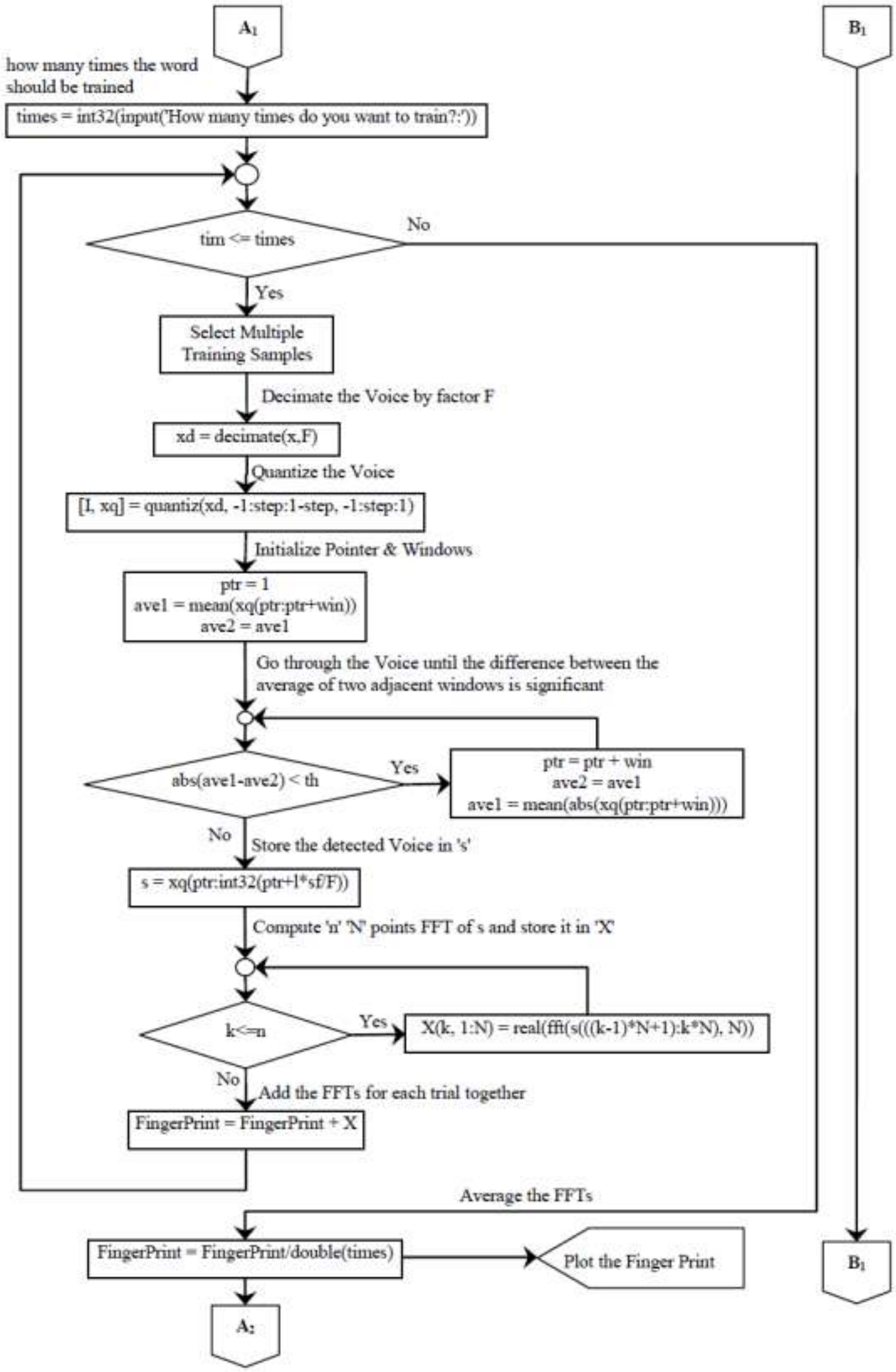
[Fig.-3: Flow Chart for AVSC]

E. Flow Chart for Word Identification System



[Fig.-4: Flow Chart for Word Identification]

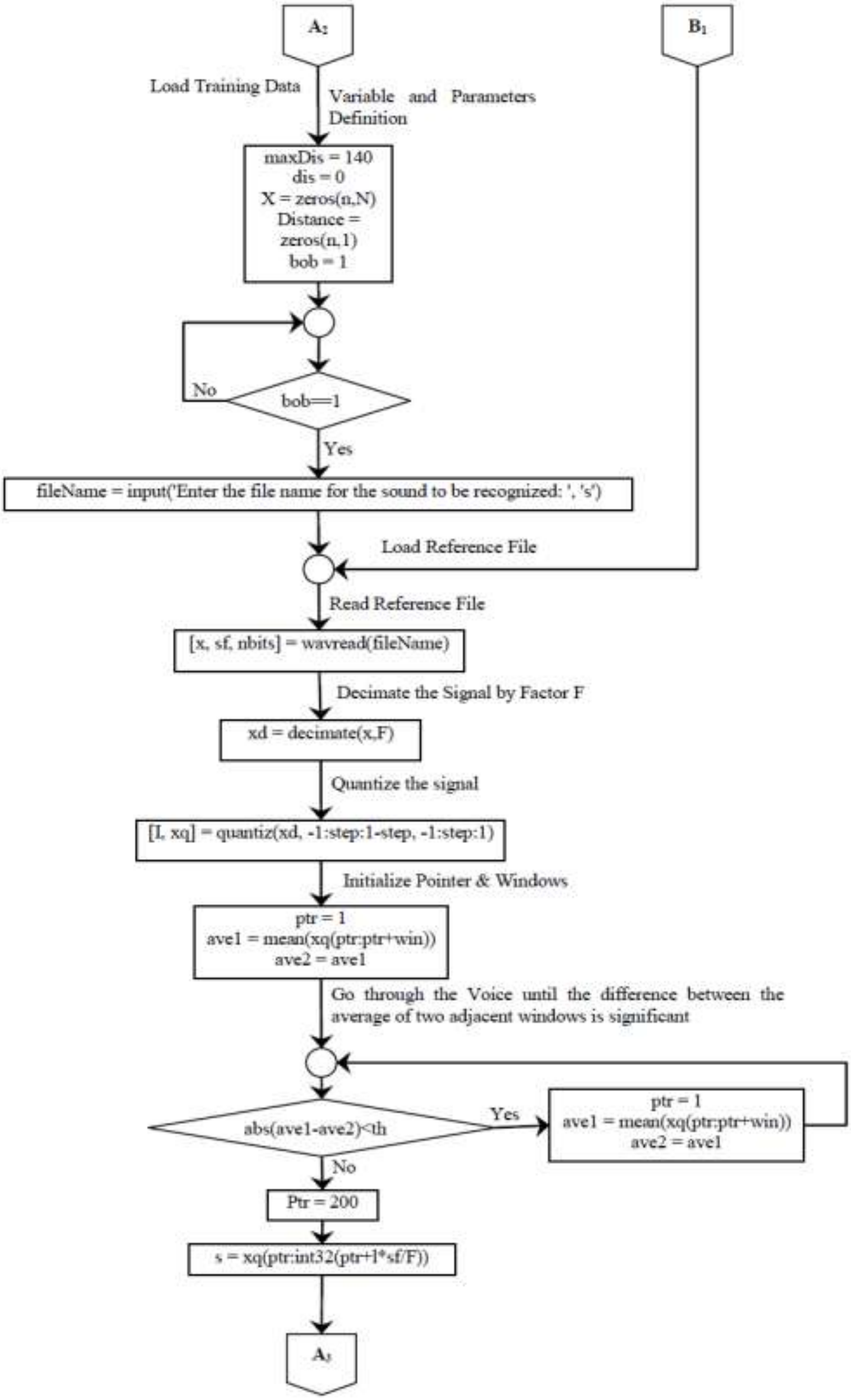
Continue…



[Fig.-4: Flow Chart for Word Identification]

Continue…

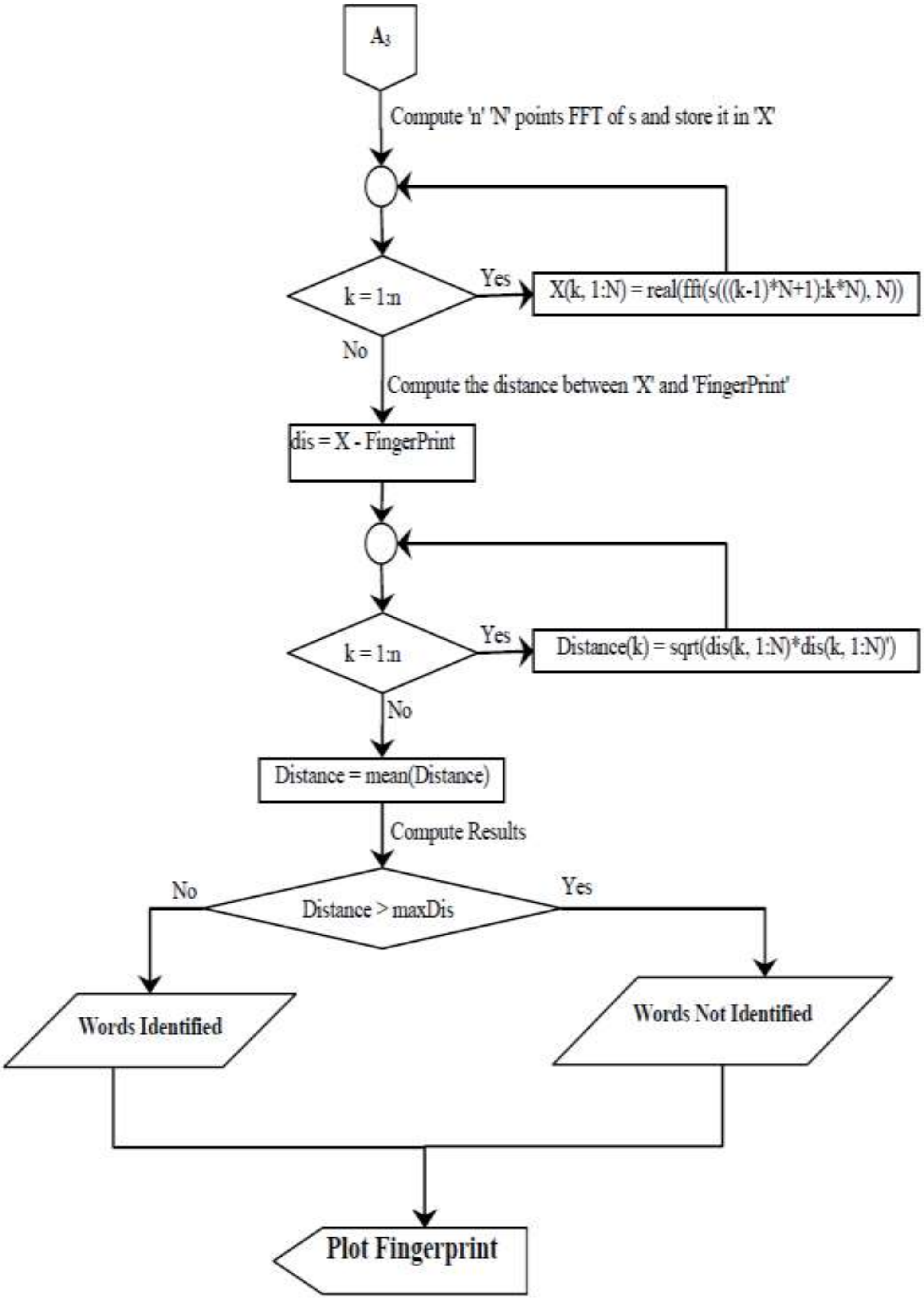
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



[Fig.-4: Flow Chart for Word Identification]

Continue…

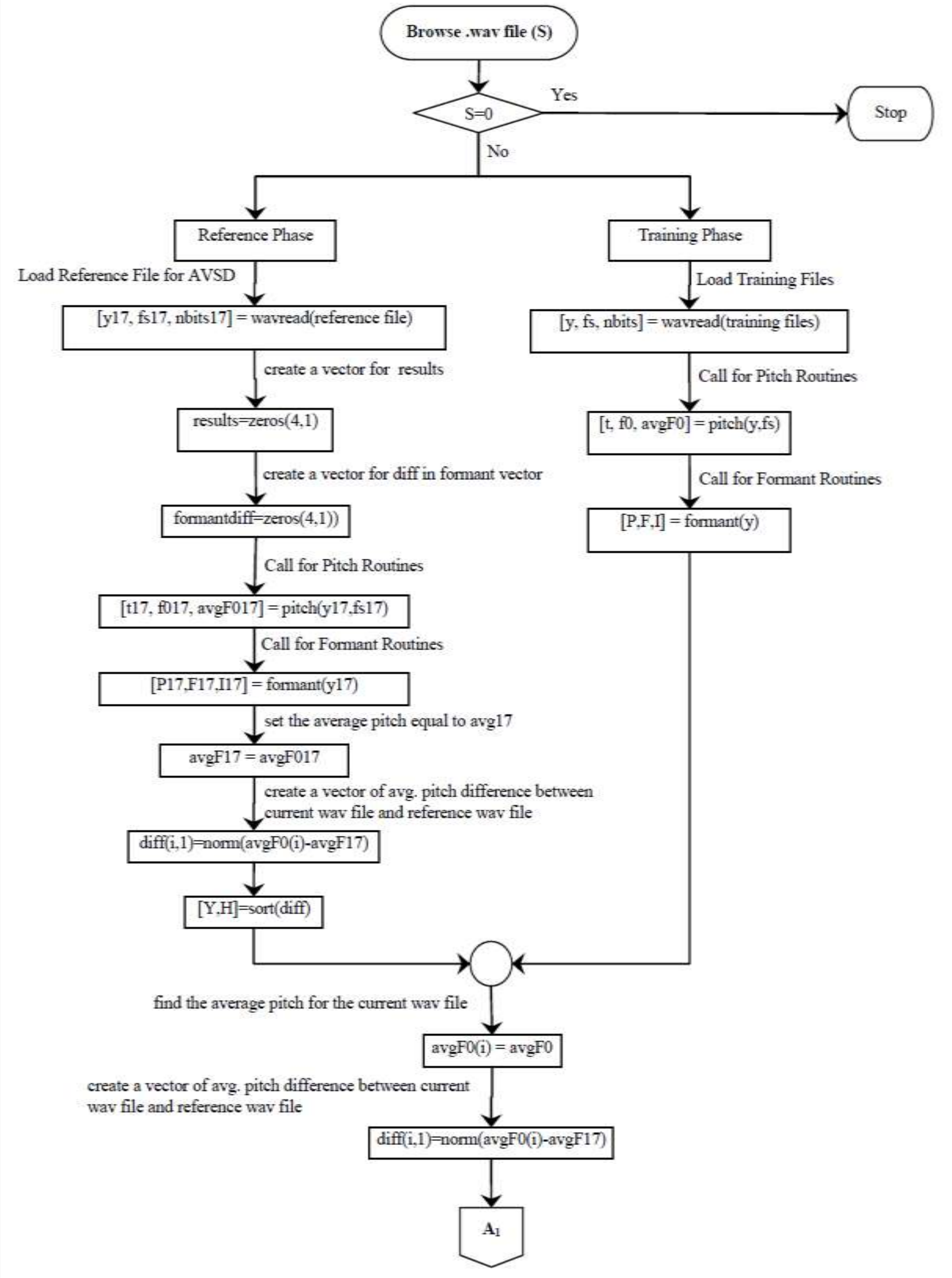
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



[Fig.-4: Flow Chart for Word Identification]

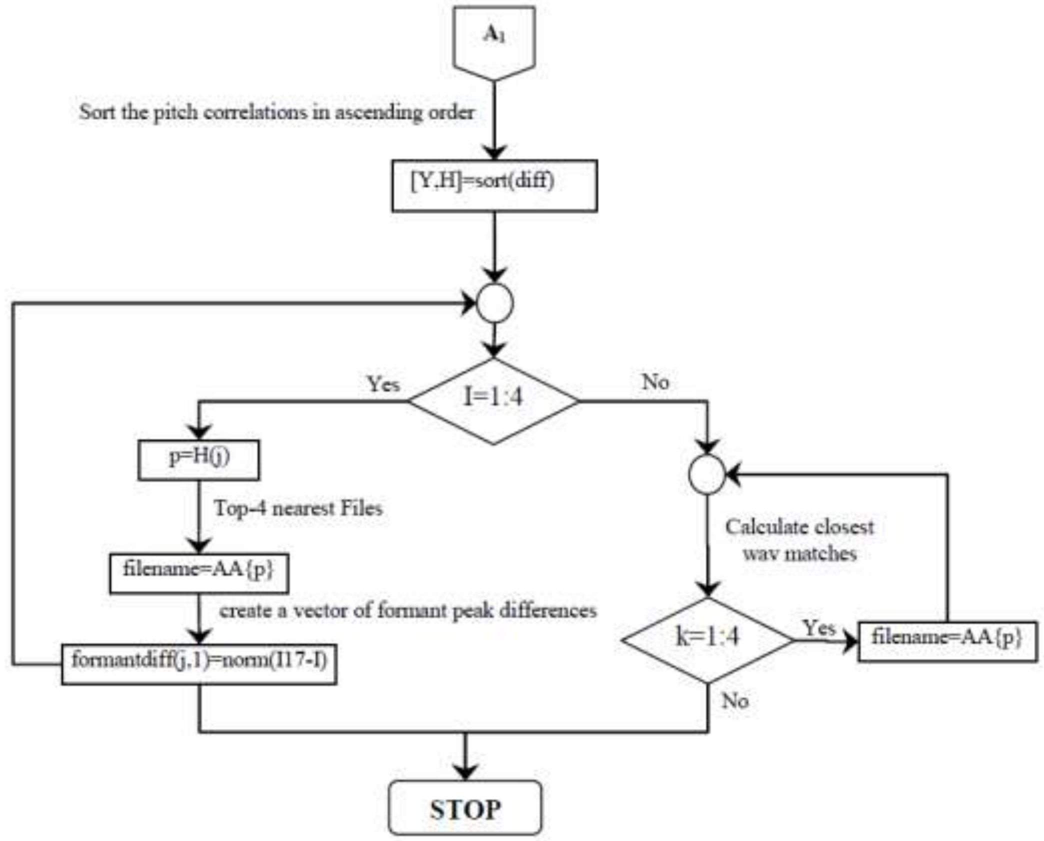
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

F. Flow Chart for AVSD



[Fig.-5: Flow Chart for AVSD]

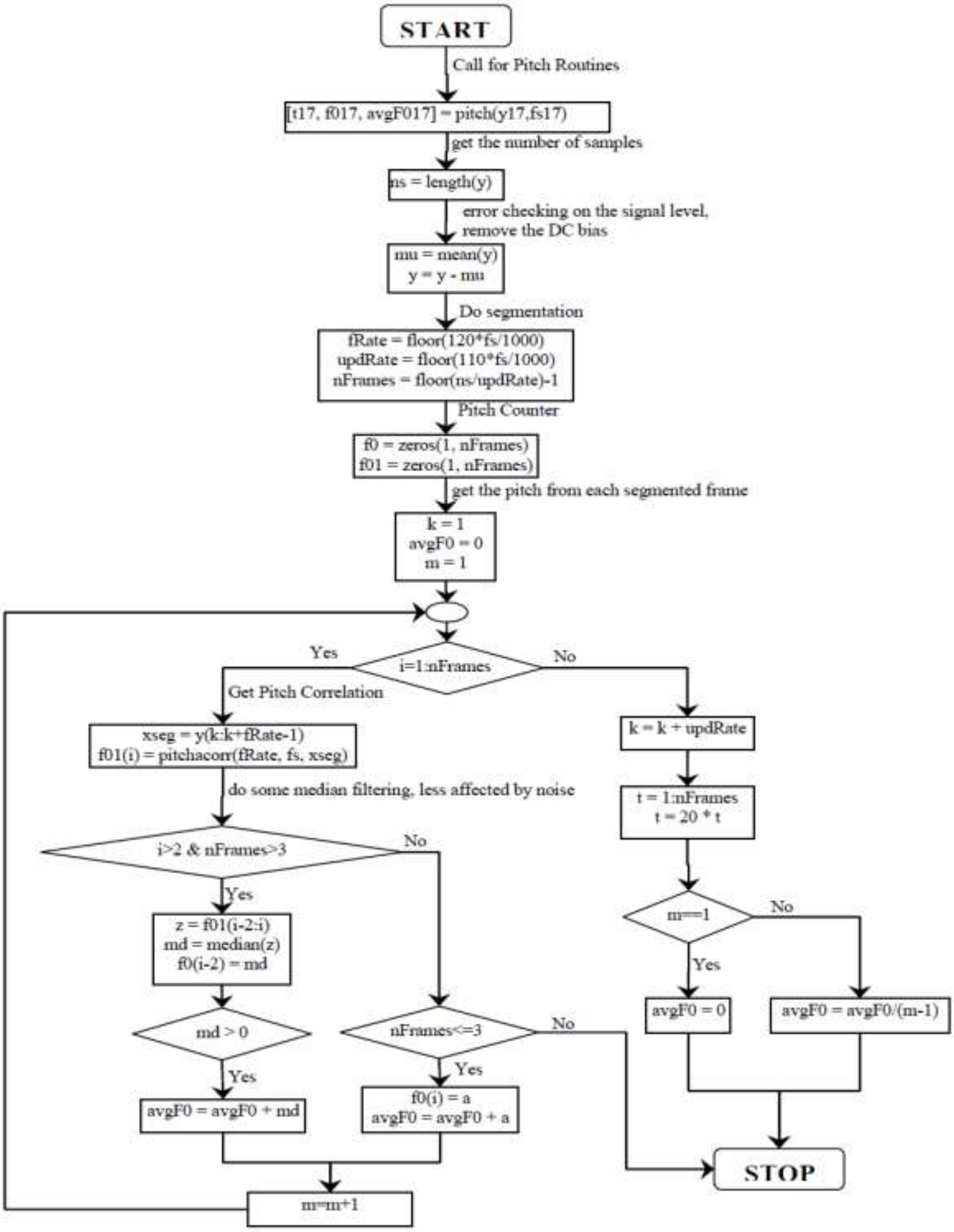
Continue…



[Fig.-5: Flow Chart for AVSD]

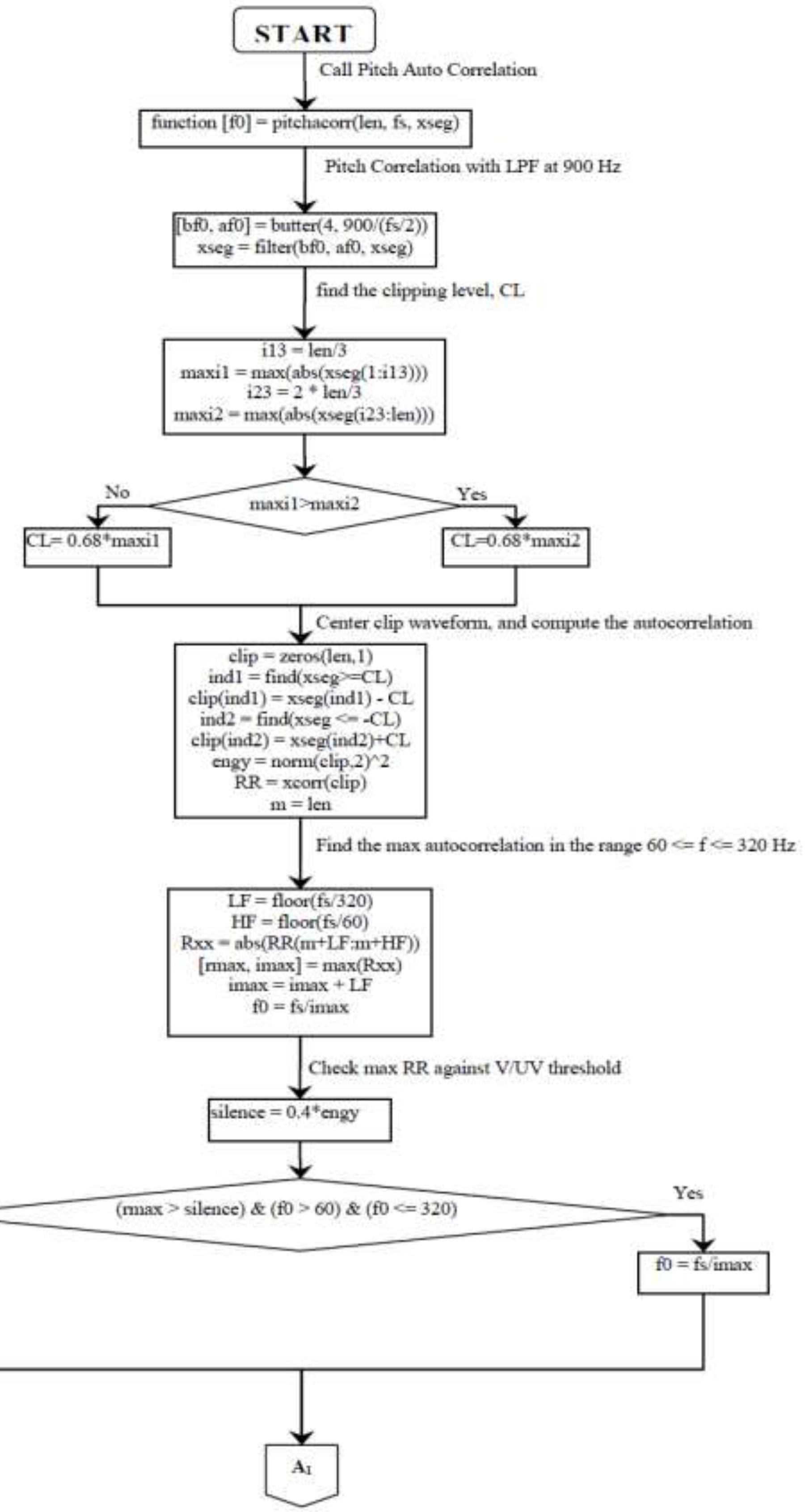
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

G. Flow Chart for Pitch Routines: Function



[Fig.-6: Flow Chart for Pitch Routines: Function]

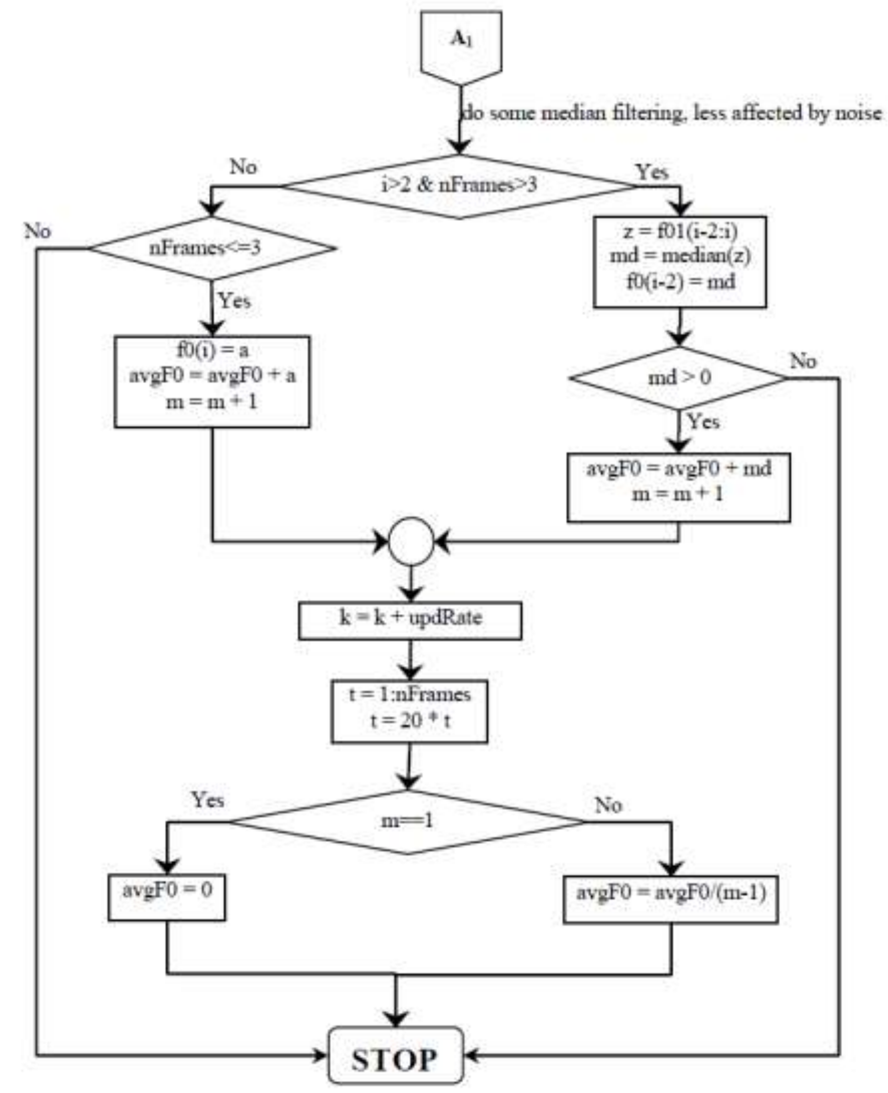
H. Flow Chart for Pitch Correlation: Function



[Fig.-7: Flow Chart for Pitch Correlation]

Continue…

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



[Fig.-7: Flow Chart for Pitch Correlation]

IV. RESULT ANALYSIS

The result analysis of AVSDC has been explained by taking the help of Intel (R) Core (TM) i5 CPU, M 460@ 2.53GHz, 3.00GB RAM, DELL INSPIRON N4010 Laptop, and MatLab 7.0. To get better analysis, voice is recorded in same environment for both genders male and female.

A. Result Analysis for AVSC

AVSC takes .wav stereo file as an input and compress it 50 to 60 percent of the source file as an output at about 45 kbps with high quality voice signal by taking the help of adaptive wavelet packet decomposition and psychoacoustic model. AVSC takes .wav stereo type file as an input and compress it .wav mono type file as an output.

For the result analysis of 'Automatic Voice Signal Compression (AVSC)', 30 number of .wav stereo files are recorded in the same environment from different people for both male and female categories. The recorded files for male category are dj3avsdc.wav, dj4avsdc.wav, dj5avsdc.wav, rb1avsdc.wav, rb2avsdc.wav, rb3avsdc.wav, rb5avsdc.wav, rb1topic.wav, rb2topic.wav, rb3topic.wav, rb4topic.wav, and rb5topic.wav while recorded files for female category are dsp1avsdc.wav, dsp2avsdc.wav, dsp3avsdc.wav, dsp4avsdc.wav, dsp5avsdc.wav, dsp2topic.wav, dsp3topic.wav, dsp4topic.wav, dsp5topic.wav, mb1avsdc.wav, mb2avsdc.wav, mb3avsdc.wav, mb4avsdc.wav, mb5avsdc.wav, mb1topic.wav, mb2topic.wav, mb3topic.wav, and mb5topic.wav.

These .wav stereo files are compressed, the details of which are given in Table 1.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

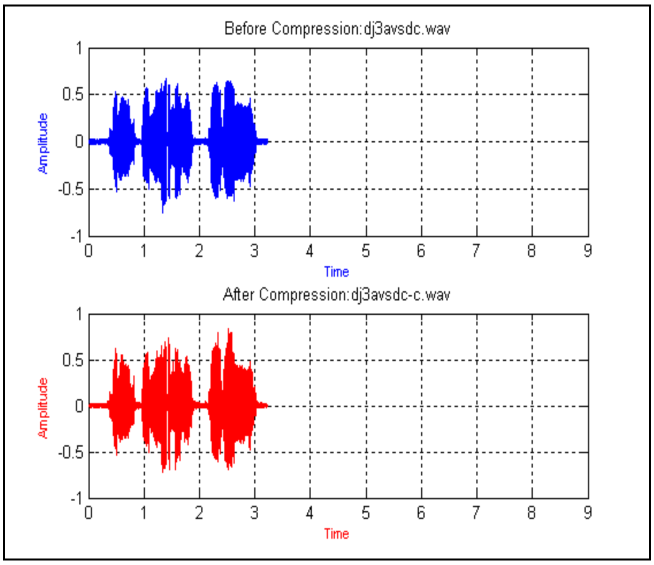
Table 1: Results for AVSC Compressed File

S. No.	Before Compression (.wav File)			After Compression (.wav File)				
	File Name	File Type	Size (Bytes)	File Name	File Type	Size (Bytes)	Distortion (Bytes)	Compression (Percentage)
1	dj3avsdc	Stereo	**569650**	dj3avsdc-c	Mono	284840	0.785861	**50.0026**
2	dj4avsdc	Stereo	**509674**	dj4avsdc-c	Mono	254852	0.878348	**50.0029**
3	dj5avsdc	Stereo	**677254**	dj5avsdc-c	Mono	338642	0.66099	**50.0022**
4	dsp1avsdc	Stereo	**539662**	dsp1avsdc-c	Mono	269846	0.829534	**50.0028**
5	dsp2avsdc	Stereo	**454990**	dsp2avsdc-c	Mono	227510	0.983927	**50.0033**
6	dsp3avsdc	Stereo	**444406**	dsp3avsdc-c	Mono	222218	1.00736	**50.0034**
7	dsp4avsdc	Stereo	**474394**	dsp4avsdc-c	Mono	237214	0.943677	**50.0036**
8	dsp5avsdc	Stereo	**481450**	dsp5avsdc-c	Mono	240740	0.929845	**50.0031**
9	dsp2topic	Stereo	**798970**	dsp2topic-c	Mono	399500	0.560287	**50.0019**
10	dsp3topic	Stereo	**747814**	dsp3topic-c	Mono	373922	0.598618	**50.0020**
11	dsp4topic	Stereo	**682546**	dsp4topic-c	Mono	341288	0.655865	**50.0022**
12	dsp5topic	Stereo	**737230**	dsp5topic-c	Mono	368630	0.607212	**50.0020**
13	mb1avsdc	Stereo	**504382**	mb1avsdc-c	Mono	252206	0.887564	**50.0030**
14	mb2avsdc	Stereo	**522022**	mb2avsdc-c	Mono	261026	0.857569	**50.0029**
15	mb3avsdc	Stereo	**527314**	mb3avsdc-c	Mono	263672	0.848961	**50.0028**
16	mb4avsdc	Stereo	**555538**	mb4avsdc-c	Mono	277784	0.805826	**50.0027**
17	mb5avsdc	Stereo	**469102**	mb5avsdc-c	Mono	234566	0.954324	**50.0032**
18	mb1topic	Stereo	**802498**	mb1topic-c	Mono	401264	0.557824	**50.0019**
19	mb2topic	Stereo	**742522**	mb2topic-c	Mono	371276	0.602884	**50.0020**
20	mb3topic	Stereo	**966550**	mb3topic-c	Mono	483290	0.463139	**50.0016**
21	mb5topic	Stereo	**952438**	mb5topic-c	Mono	476234	0.470001	**50.0016**
22	rb1avsdc	Stereo	**626098**	rb1avsdc-c	Mono	313064	0.715002	**50.0024**
23	rb2avsdc	Stereo	**599638**	rb2avsdc-c	Mono	299834	0.746556	**50.0025**
24	rb3avsdc	Stereo	**636682**	rb3avsdc-c	Mono	318356	0.703115	**50.0024**
25	rb5avsdc	Stereo	**544954**	rb5avsdc-c	Mono	272492	0.821478	**50.0028**
26	rb1topic	Stereo	**888934**	rb1topic-c	Mono	444482	0.50358	**50.0017**
27	rb2topic	Stereo	**913630**	rb2topic-c	Mono	456830	0.489967	**50.0016**
28	rb3topic	Stereo	**867766**	rb3topic-c	Mono	433898	0.515865	**50.0017**
29	rb4topic	Stereo	**929506**	rb4topic-c	Mono	464768	0.481598	**50.0016**
30	rb5topic	Stereo	**837778**	rb5topic-c	Mono	418904	0.534331	**50.0018**

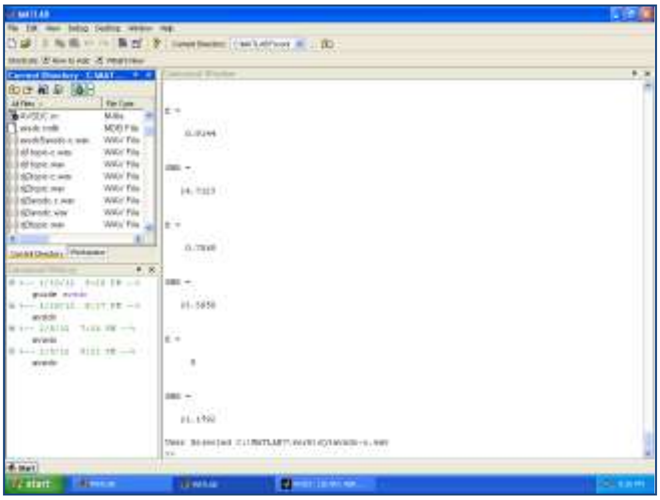
The voice compression is based on threshold masking concept in which unwanted signal (noise) is masked with conceptual signal. The execution time for compressed signal and uncompressed signal will be same. For example if x is original file and its total running time is t, and if y is compressed file of x then its total running time will be t.

During the compression, minor distortion may be possible. Distortion of any file is depending on the size of the file and contained noise with the file. Distortion analysis has been given in the Table 1. This minor distortion may not be read by human auditory system. If both compressed and uncompressed file is played at the same time and listened carefully, it may be found that human auditory system may not differentiate both compressed and uncompressed file due to minor distortion but spectrum analysis may differentiate it. Here, some spectrum analysis for respective each file and respective screening results are given in Figure 8 to Figure 10. Reality of the results may be seen in screenshot of the respective file while background execution may be seen by command window processing where entropy (E), and SNR of each frame for respective files are determined.

[Fig.- 8: Screenshot for dj3avsdc.wav File]



[Fig.- 9: Spectrum Analysis for dj3avsdc.wav File]



[Fig.- 10: Command Window Processing for dj3avsdc.wav File]

B. *Result Analysis for Word Identification*

Word Identification System is part of AVSD which is designed to checkout whether spoken words are in proper sequence are not. Normally, sometimes words may be recorded fast, clear, or loud. The main idea behind the word identification system is to first train it with several versions of the same word, thus yielding a "reference fingerprint" then subsequent words may be identified based on how near they are to this fingerprint. The whole idea is evaluated on the basis of Euclidean distance theory.

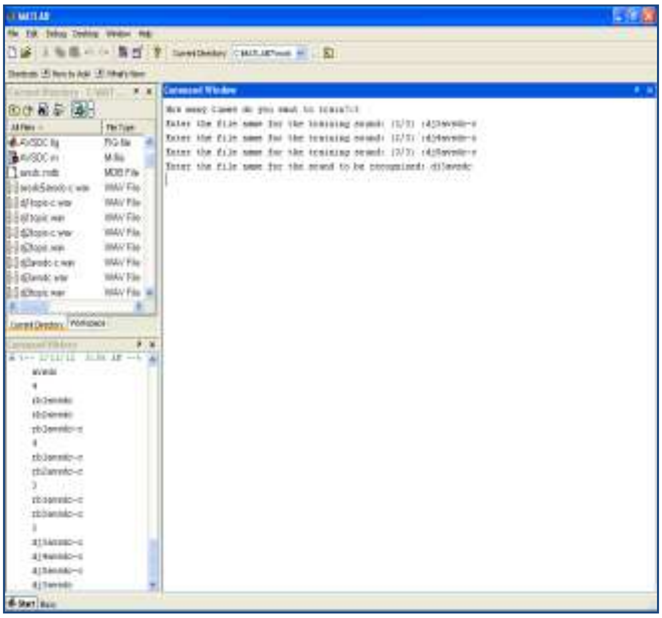
Word Identification may be started by just clicking 'Identify' button which is mentioned in word identification phase of AVSD. After clicking 'Identify' button, word identification process is moved to command window. Here a constant number of train file is required. As per given input constant number, name of .wav trainee have to be written. After that, name of a reference .wav file has to be written in the command window. When trainee and reference files are successfully processed then average finger print of trainee files and reference file is plotted for which result is computed in the AVSDC frame work.

Result details of word identification system are given in the Table 2 while plotted results and screenshots of some respective files are given at Figure 11 to Figure 14.

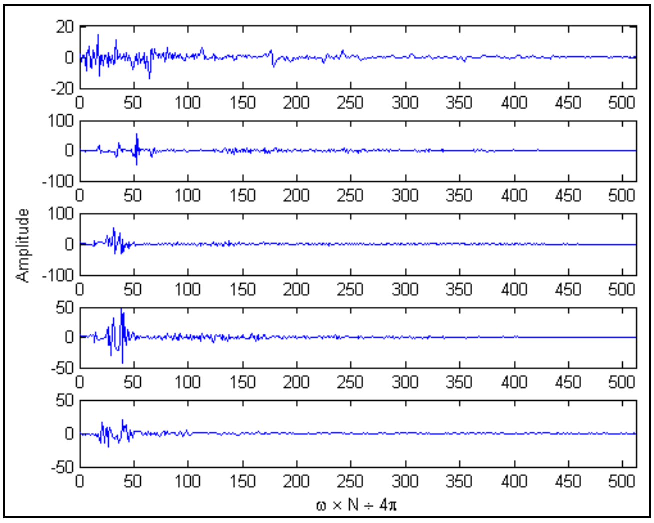
TABLE 2: Result Details for Word Identification System

S.No.	Trainee Files (.wav)	Reference File (.wav)	Results
1	dj3avsdc-c	dj3avsdc-c	**words not identified**
	dj4avsdc-c		
	dj5avsdc-c		
2	dsp1avsdc-c	dsp1avsdc-c	**words identified**
	dsp2avsdc-c		
	dsp3avsdc-c		
	dsp4avsdc-c		
3	mb1topic-c	mb4topic-c	**words not identified**
	mb2topic-c		
	mb3topic-c		

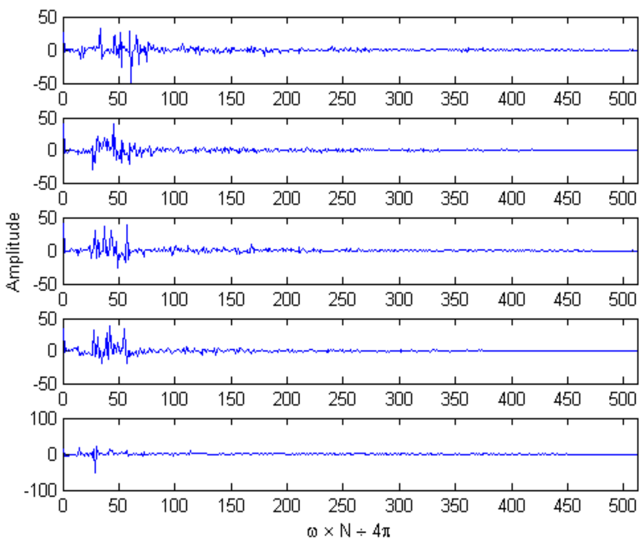
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



[Fig.- 11: Word Identification Processing Details in Command Window for djavsdc-c.wav Files]



[Fig.- 12: Fingerprint for Trainee Files djavsdc-c.wav]



[Fig.- 13: Fingerprint for Reference File dj3avsdc-c.wav]

[Fig. - 14: Screenshot of Word Identification for djavsdc-c.wav files]

C. Result Analysis for AVSD

Automatic Voice Signal Detection (AVSD) is used to identify constitutional and unconstitutional voice signal automatically which is performed on the basis of frequency, pitch value, and formant value.

AVSD database files may be called by clicking ‘AVSD Database’ button. After loading AVSD database, a reference file is picked from the AVSD database, the details of which are uploaded in the AVSDC framework. Rests of the files in the AVSD database are counted as trainee files. Now, after clicking the ‘AVSD’ button, AVSD processing is started where pitch and formant value of reference file and trainee files is computed. On the basis of this pitch value and formant value, comparison is done in between reference file and each trainee file. Top 4 files are selected on the basis of lowest pitch and formant value. Now selection process will repeat from top-4 file to select nearest file.

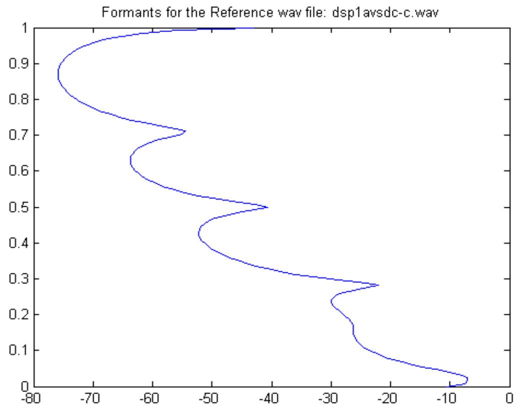
AVSD brief results are given at Tables 3 respective screening results are given in Figure 14 to Figure 20.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

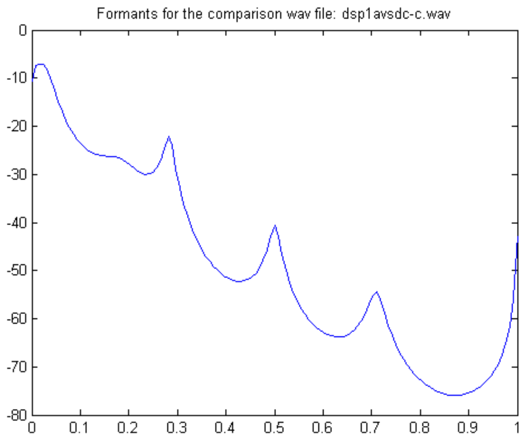
Table 3: AVSD Results for dsp1avsdc-c.wav

Reference File (.wav)		Trainee Files (.wav)		Pitch Differences (b-a)	Top-4 Matched Files (.wav)	Formant Differences	Very Nearest File (.wav)
File Name	**Pitch Value (a)**	**File Name**	**Pitch Value (b)**				
dsp1avsdc-c	272.351	dsp2topic-c	276.06873	3.71773	**dsp1avsdc-c**	0	**dsp1avsdc-c**
		dsp3topic-c	282.4804	10.1294	**dsp5topic-c**	0.776823	
		dsp4topic-c	281.08408	8.73308	**dsp2topic-c**	0.992311	
		dsp5topic-c	274.54634	2.19534	**dsp2avsdc-c**	0.0110485	
		mb1avsdc-c	304.2249	31.8739			
		mb2avsdc-c	317.4178	45.0668			
		mb3avsdc-c	293.3996	21.0486			
		mb4avsdc-c	307.9445	35.5935			
		mb5avsdc-c	293.1547	20.8037			
		mb1topic-c	295.0616	22.7106			
		mb2topic-c	294.6076	22.2566			
		mb3topic-c	300.7211	28.3701			
		mb5topic-c	323.9582	51.6072			
		rb1avsdc-c	368.2477	95.8967			
		rb2avsdc-c	385.133	112.782			
		rb3avsdc-c	342.1935	69.8425			
		rb5avsdc-c	345.7603	73.4093			
		rb1topic-c	372.0271	99.6761			
		rb2topic-c	379.252	106.901			
		rb3topic-c	368.0342	95.6832			
		rb4topic-c	387.942	115.591			
		rb5topic-c	386.367	114.016			
		dsp1avsdc-c	272.351	0			
		dsp2avsdc-c	276.66115	4.31015			
		dsp3avsdc-c	287.8316	15.4806			
		dsp4avsdc-c	287.1671	14.8161			
		dsp5avsdc-c	280.35412	8.00312			

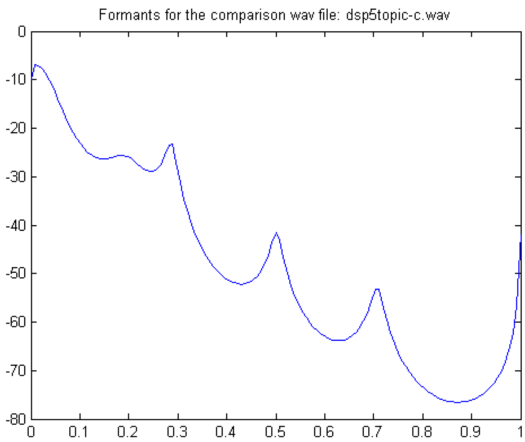
[Fig.- 15: Screenshot for dsp1avsdc-c.wav]



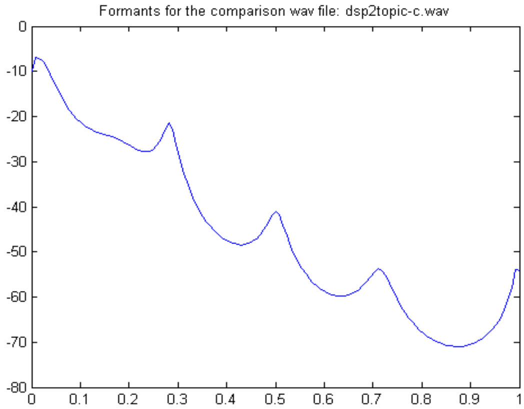
[Fig.-16: Formant for Reference File dsp1avsdc-c.wav]



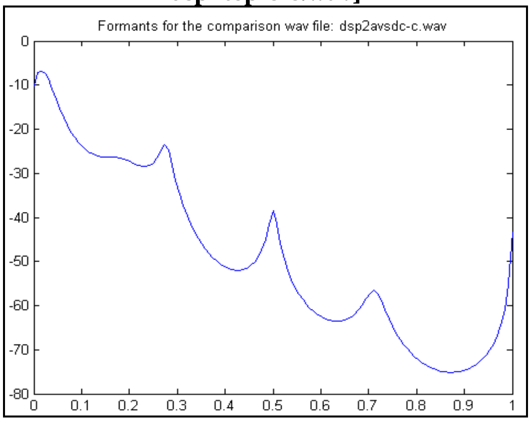
[Fig.- 17: Formant difference with dsp1avsdc-c.wav and dsp1avsdc-c.wav]



[Fig.- 18: Formant difference with dsp1avsdc-c.wav and dsp5topic-c.wav]



[Fig.-19: Formant difference with dsp1avsdc-c.wav and dsp2topic-c.wav]



[Fig.-20: Formant difference with dsp1avsdc-c.wav and dsp2avsdc-c.wav]

V. CONCLUSION

The purpose of AVSDC is to (i) compress 50 to 60 percent of the source file (ii) identify word sequence and (iii) verify constitutional and unconstitutional voice signals. AVSDC is divided in two parts: AVSD and AVSC. Word identification system is part of AVSD. For the result analysis of AVSDC, 30 number of .wav files are recorded for both male and female categories.

Automatic Voice Signal Compression (AVSC) takes .wav stereo file as an input and compress it 50 to 60 percent of the source file at about 45 kbps with high quality voice signal by taking the help of adaptive wavelet packet decomposition and psychoacoustic model. AVSC takes .wav stereo file as an input and creates .wav mono file after compression as an output. After compression minor distortion is also possible. The main feature of AVSC is that file extension does not change after compression. In other words, compression is done from .wav to .wav extension. AVSC also computes entropy and SNR (Signal to Noise Ratio) of the source file during the compression. AVSC results are given in Table 1 and some respective screening results are given in Figure 8 to Figure 10 at result analysis part while their methodology is defined in methodology part.

Word identification system is part of AVSD which is designed to checkout whether recorded words in proper sequence are not. Normally, sometimes spoken words of voice may be recorded very fast,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

smoothly, or loudly. The main idea behind the word identification system is to first train it with several versions of the same word, thus yielding a "reference fingerprint". Then, subsequent words can be identified based on how close they are to this fingerprint. The whole idea is evaluated on the basis of Euclidean distance theory. The executed results for word identification system are given in Table 2 and some respective screening results are given in Figure 11 to Figure 14.

Automatic Voice Signal Detection (AVSD) is used to identify constitutional and unconstitutional voice signal automatically which is performed on the basis of frequency, pitch value, formant value, and sequence of words in the voice signal for several samples of the same voice. An underline purpose of AVSD is to identify fake voice in the security system. Frequency is being mapped to the frequency domain by computing its DFT using the FFT algorithm. Sequence of words is computed by continuously computing difference between absolute averages of two adjacent significant windows and comparing it to a predefined threshold. The computed results for AVSD are given in Table 3 and some respective screening results are given in Figure 15 to Figure 20.

REFERENCES

1. Jalal Karam, "Various Speech Processing Techniques For Speech Compression And Recognition", proceedings of world academy of science, engineering and technology volume 26 December 2007 ISSN 1307-6884, © 2007 waset.org
2. Sarantos Psycharis "The Didactic Use Of Digital Image Lossy Compression Methods For The Vocational Training Sector", University of Agean, Proceeding of Current Developments in Technology-Assisted Education 2006(FORMATEX 2006)
3. P R Deshmukh, "Multi-wavelet Decomposition for Audio Compression", IE(I) Journal-ET, Vol 87 , July 2006
4. Background on the Psychoacoustic Model "W.A.V.S. Compression"
5. Stefan Wabnik, Gerald Schuller, Ulrich Kr ¨amer and Jens Hirschfeld, "Frequency warping In Low delay Audio Coding", ICASSP 2005, 0-7803-8874-7/05/$20.00 ©2005 IEEE
6. Shaleena Jeeawoody, "Voice Analysis and Recognition as a Car Theft Deterrent", California State Science Fair 2008, Project No: J1307, Ap2/08
7. Khalid Saeed, "Sound and Voice Verification and Identification a Brief Review of Töeplitz Approach", Znalosti 2008, pp. 22-27, ISBN 978-80-227-2827-0.FIIT STU Bratislava, .stav informatiky a softvÈrovÈho inæinierstva, 2008.
8. Maria Markaki, Andre Holzapfel, Yannis Stylianou "Singing Voice Detection using Modulation Frequency Features", Computer Science Department, University of Crete, Greece
9. Hongwu YANG, Dezhi HUANG, Lianhong CAI, "Perceptually Weighted Mel–Cepstrum Analysis Of Speech Based On Psychoacoustic Model", IEICE TRANS. INF. & SYST., VOL.E89-D, No.12 December 2006
10. Samar Krimi, Kais Ouni, Noureddine Ellouze, "Realization of a Psychoacoustic Model for MPEG-1 using Gammachirp WaveLet", Proceeding in EUSIPCO,2005
11. Tsung-Han Tsai, Yi-Wen Wang, Shih-Way Hung "An MDCT-Based psychoacoustic model co-processor design for MPEG-2/4 AAC audio encoder" Proceeding in the 7th International Conference on Digital Audio Effects (DAFx'04), Naples, Italy, October 5-8, 2004
12. Ganesh K Venayagamoorthy, Viresh Moonasar, Kumbes Sandrasegaran, "Voice Recognition Using Neural Networks",Institute of Information Science and Technology (IIST), Massey University, New Zealand, Published in IEEE in 1998, Val. No. 0-7803-5054-5.0029
13. Andress Holzinger with assistance from G. Searle, "Multimedia Basics", Volume 1: Technology, ISBN: 81-7008-243-9, Firewall Media Publication
14. Electronic, Electrical & Computer Engineering, University of Birhingham
15. Audio Signal Processing and Coding "Andress Sparias Ted Painter Venkatraman
16. Johan F. Koegel Buford, "Multimedia System", ISBN: 81-7808-162-8, Tenth Indian Reprint, Pearson Publication
17. J.N. Holmes, Speech Synthesis and Recognition, Chapman & Hall, London, 1988
18. Math Work on MatLab
19. Amara Graps, Institute of Electrical & Electronics Engineers, "An Introduction to Wavelet", Published by the IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, CA 90720, USA, Vol. 2, and Num. 2.
20. Engin Avci, "An Automatic System for Turkish Word Recognition Using Discrete Wavelet Neural Network Based on Adaptive Entropy", International Conference-Elazig, Turkey-Dec 06, 2007 , The Arabian Journal for Science and Engineering, Volume 32, Number 2B, Paper Received 5 September 2005; Revised 17 February 2006; Accepted 20 December 2006
21. Trung Nghia Phung, Ngoc Phan Vu, "A Low Bit Rate Wide-Band Speech Coder In The Perceptual Wavelet Packet Domain", International Symposium on Electrical & Electronics Engineering 2007 - Oct 24, 25 2007 - HCM City, Vietnam
22. Bal´azs Bank, Federico Avanzini, Gianpaolo Borin, Giovanni De Poli, Federico Fontana, Davide Rocchesso, "Physically Informed Signal Processing Methods for Piano Sound Synthesis: A Research Overview", EURASIP Journal on Applied Signal Processing 2003:10, 941–952, ©2003 Hindawi Publishing Corporation
23. N. R. Chong, I. Burnett†, J. F. Chicharo,"A new waveform interpolation coding scheme based on pitch synchronous wavelet transform" decomposition, IEEE Transactions On Speech And Audio Processing, Vol. 8, No. 3, May 2000
24. Frank A Russo, Lola L Cuddy, Alexander Galembo, William Forde Thompson, "Sensitivity to tonality across the pitch range", Perception, 2007, volume 36, pages 781-790, DOI: 10.1068/p5435, ISSN 0301-0066 (print)
25. Kazunori Mano, "Design of a Toll-Quality 4-KBIT/S Speech Coder Based on Phase-Adaptive PSI-CELP",IEEE Proceeding 1997
26. Sorin Dusan, James Flanagan, Amod Karve, Mridul Balaraman, "Speech Coding Using Trajectory Compression and Multiple Sensors", Center for Advanced Information Processing (CAIP)- Rutgers University, Piscataway, NJ, U.S.A.
27. Murali Mohan Deshpande, K.R.Ramakrishnan, "A Novel BWE Scheme Based On Spectral Peaks In G.729 Compressed Domain", Indian Institute of Science, Department of Electrical Engineering, Bangalore, India
28. V. Vijaya Kumar1, U.S.N.Raju2, M. Radhika Mani3 , A.L.Narasimha Rao 4, "Wavelet based Texture Segmentation methods based on Combinatorial of Morphological and Statistical Operations", Proceeding in IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008
29. N. Ruiz Reyes, M. Rosa Zurera, F. López Ferreras, D. Martínez Muñoz, J.M. Villafranca, " A New Cost Function To Select The Wavelet Decomposition For Audio Compression", SPAIN
30. Marcus Purat, Peter Noll, "Audio Coding with a dynamic wavelet packet decomposition based on frequency-Varying Modulated Lapped Transforms" Proceeding in IEEE 0-7803-3192-3/96 @1996
31. Mohamed Cherif Amara Korba, Djemil Messadeg, Rafik Djemili, Hocine Bourouba, "Robust Speech Recognition Using Perceptual Wavelet Denoising and Mel-frequency Product Spectrum Cepstral Coefficient Features", Proceeding in International Conference Informatica-32 (2008) 283-288
32. Jong-Tzy Wang, Ming-Shan Lai, Kai-Wen Liang, Pao-Chi Chang, "Adaptive wavelet quantization Index Modulation Technique for audio watermarking", National Central University-Tiwan
33. N. Ruiz Reyes, M. Rosa Zurera, F. Lopez Ferreras, D. Martínez Munoz, "A New Perceptual Entropy-based Methods to Achieve Signal Adpted Wavelet Tree in A low Bit Rate Perceptual Audio Coder", SPAIN
34. N. Ruiz Reyes, M. Rosa Zurera, F. Lopez Ferreras, P. Jarabo Amores, P. Vera Candeas, "On the Coding Gain of Dynamic Huffman Coding Applied to a Wavelet Based Perceptual Audio Coder", Escuela Universitaria Politénica 23700 Linares - Jaén (SPAIN)
35. Grigor Marchokov, Atanas Gotchev, Zdravko Nikolov, "A Wavelet-Packet Based Speech Coding Algorithm" Ministry of Education and Science, National Science Council, under the Grand No: MU-I-003/96, Proceeding in 2001. Sofia
36. Frank Kurth, Michael Clausen, "Filter Bank Tree and M-Band Wavelet Packet Algorithms in Audio Signal Processing", Universit¨at Bonn, Institut f¨ur Informatik V, R¨omerstr. 164, D-53117 Bonn, Germany.
37. O. Farooq, S. Datta, J. Blackledge, "Blind Tamper Detection in Audio using Chirp based Robust Watermarking", Proceeding in WSEAS TRANSACTIONS on SIGNAL PROCESSING, ISSN: 1790-5052, Issue 4, Volume 4, April 2008
38. Latha Pillai, "Quantization", XAPP615 (v1.1) June 25, 2003, 1-800-255-7778

39. George Tzanetakis, Georg Essl, Perry Cook, "Audio Analysis using the Discrete Wavelet Transform", Computer Science Department *also Music Department
40. HE Dong-Mei and GAO Wen, "Wideband Speech And Audio Coding Based On Wavelet Transform And Psychoacoustic Model", Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001
41. Ciprian Doru Giurc¢aneanu , Ioan T¢abus¸ and Jaakko Astola, "Integer wavelet Transform Based Lossless Audio Compression", Signal Processing Laboratory, Tampere University of Technology-Finland
42. David Darlington, Laurent Daudet and Sandler, "Digital Audio Effects in the wavelet domain", Proc. of 5th Int. Conference on Digital Audio Effects (DAFX-02), Hamburg, Germany, September 26-28, 2002
43. Claudia Schremmer, Thomas Haenselmann , Florian Bömers, "Wavelets In Real-Time Digital Audio Processing: A Software For Understanding Wavelets In Applied Computer Science", Department of Praktische Informatik IV, University of Mannheim, Germany
44. F. Abramovich and B.W. Silverman, "Wavelet Decomposition Approaches to statistical Inverse Problem", University of Bristol (U.K.)
45. Gavriel Yarmish, "The Simplex Method Applied to Wavelet Decomposition", Proceedings of the 10th WSEAS International Confenrence on Applied Mathematics, Dallas, Texas, USA, November 1-3, 2006
46. Deepa Kundur , Dimitrios Hatzinakos, "Digital Watermarking Using Multi-resolution Wavelet Decomposition", Natural Sciences and Engineering Research Council (NSERC) of Canada and by Communications and Information Technology Ontario (CITO).
47. Caroline Chaux, Laurent Duval2 and Jean-Christophe Pesquet,"2D Dual-Tree M-Bandwavelet Decomposition", Universit de Marne-la-Vall¥ee, Champs-sur-Marne, France
48. Chalermchon Satirapod, Clement Ogaja, Jinling Wang and Chris Rizos, "An Approach to GPS Analysis incorporating Wavelet Decomposition", School of Geomatic Engineering, The University of New South Wales, Sydney NSW 2052, AUSTRALIA
49. V. A. Baturin and I. V. Mironova, "Low-Degree Solar Oscillation Spectrum with Wavelet Decomposition", PACS numbers : 96.60.Ly, 95.75.Wz, 95.75Pq, DOI: 10.1134/S1063773706020071
50. Sina Jahanbin, Hyohoon Choi, Alan C. Bovik, Kenneth R. Castleman, "Three Dimensional Face Recognition Usingwavelet Decomposition Of Range Images", Laboratory for Image and Video Engineering, The University of Texas, Austin, Texas
51. K. Kwak and W. Pedrycz, "Review on “Face Recognition Using Fuzzy Integral and Wavelet Decomposition Method” IEEE Trans. Syst. Man Cyb. 34, 1 (Aug. 2004)
52. A.S. RYBAKOV, "Wavelet Decomposition And Fractal Analysis For Joint Measurements Of Laser Signal Delay And Amplitude", Automatic Control and Computer Sciences, Vol.35, No.3, pp.11-19, 2001, Avtomatika I Vychislitel’naya Tekhnika, Vol.35, No.3, pp.14-24, 2001
53. W. Kosek And W. PopiNski, " Forecasting Pole Coordinates Data By Combination Of The Wavelet Decomposition And Auto-covariance Prediction", Space Research Centre, Polish Academy Of Sciences, Warsaw, Poland
54. G. A. Blackburn , J. G. Ferwerda, "Improving The Quantification Of Leaf Biochemistry And Water Content Through Wavelet Decomposition Of Reflectance Spectra", Department of Geography, Lancaster University, Lancaster, LA1 4YB, UK -
55. Rade Kutil and Andreas Uhl, "Optimization of 3-D Wavelet Decomposition on Multiprocessors", Journal of Computing and Information Technology - CIT 8, 2000, 1, 31–40
56. R. Martinez-Noriega, H. Kang, B. Kurkoski, K. Yamaguchi and M. Nakano-Miyatake, "Audio Watermarking Based on Wavelet Transform and Quantization Index Modulation", National Polytechnic Institute of Mexico
57. Needeljko Cvejic, Tapio Seppanen, "A Wavelet Domain LSB Insertion Algorithm for high capacity audio Stegnography", Media Team Oulu, Information Processing laboratory, Finland
58. S.Manikandan, "Speech Enhancement Based On Wavelet De-noising", Academic Open Internet Journal, Valume17, 2006, ISSN-1311-4360
59. Rachid Moussaoui, Jean Rouat, Roch Lefebvre, "Wavelet Based Independent Component Analysis For Multi-Channel Source Separation", Departement de genie electrique et de genie informatique, Universite de Sherbrooke, Quebec, Canada
60. Pavel Rajmic and Jan Vlach, "Real-Time Audio Processing Via Segmented wavelet Transform" Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15, 2007
61. Bastiaan Kleijn , “Speech Signal Processing”, Speech, Music and Hearing TMH/KTH Annual Report 2001
62. Turkish Word Recognition Using Discrete Wavelet Neural Network Based on Adaptive Entropy”, International Conference-Elazig, Turkey-Dec 06, 2007, the Arabian Journal for Science and Engineering, Volume 32, Number 2B.
63. Murtaza Bulut, “Mult-level emotional speech analysis resynthesis”, Proposed thesis for Ph.D. in Electrical Engineering, University of Southern California (USC), Los Angeles, CA, (graduation expected in November 2007).
64. Michael I Mandel, “Recognition and Organization of Speech and Audio (LabROSA)”, Ph.D. candidate in the department of Electrical Engineering at Columbia University, Proposed work for 2009 (Jan) exp.
65. Graziano Bertini, Federico Fontana, Diego Gonzalez, Lorenzo Grassi, Massimo Magrini “Voice Transformation Algorithms With Real Time Dsp Rapid Prototyping Tools”, Proceeding In International Conference –Eusipco 2005
66. Marcos Faundez-Zanuy, “Non Liner speech Processing”, Proceeding in ISCpad 66, 19-22 April-2005, Barcelona, Spain
67. Keikichi HIROSE, “Speech Prosody 2004”, Proceeding in ISCApad 56, International Conference: Speech Prosody 2004 on March 23 -26, 2004-Japan,
68. Kyogu Lee,"Pitch Perception: Place Theory, Temporal Theory, and Beyond", IEE 391 Special Report (Autumn 2004), Center for Computer Research in Music and Acoustics (CCRMA), Music Department, Stanford University
69. Douglas A. Reynolds, Larry P. Heck, “Automatic Speaker Recognition” Presented at the AAAS 2000 Meeting Humans, Computers and Speech Symposium 19 February 2000, Nuance Communications United States Air Force.
70. Matti Karjalainen, Tero Tolonen,"Multi-Pitch and Periodicity Analysis Model for Sound Separation and Auditory Scene Analysis", Laboratory of Acoustics and Audio Signal Processing-Helsinki University of Technology
71. Brett A. St. George, Ellen C. Wooten, Louiza Sellami “Speech Coding and Phoneme classification Using MATLAB and NeuralWorks” Department of Ellectrical Engineering,U.S. Naval Academy-Annapolis, MD 21402
72. John-Paul Hosom, Lawrence Shriberg, Jordan R. Green, “Diagnostic Assessment of Childhood Apraxia of Speech Using Automatic Speech Recognition (ASR) Methods”, Oregon Health & Science University –Beaverton, Accepted for Publications by NIH Public access.
73. Dalibor Mitrovic, Matthias Zeppelzauer, Christian Breiteneder, “ Discrimination and Retrieval of Animal Sounds”,
74. Viresh Moonsasar, Ganesh k. Venayagamoorty,"Artificial Neural Network Based Automatic Speaker Recognition using a hybrid technique for feature extraction", Proceedings, IEEE Trans on Acoustics, Speech and Signal,
75. Fracarro Radioindustrie, CitecVoice, Alpikom, “Robust Speech Recognition”, Copyright © Partners of the DICIT consortium.
76. Nachiappan, PM Abdul Manan Ahmad “Speech Coding Effects On Recognition Accuracy For Timit”, Postgraduate Annual Research Seminar 3-4 July 2007,(PARS-07)
77. M. Herrera Martinez, "Evaluation of Audio Compression Artifacts", Acta Polytechnica Vol. 47 No. 1/2007
78. Nadine E. Miner, Thomas P. Caudell, "Using Wavelets to Synthesize Stochastic-based Sounds for Immersive Virtual Environments", Dept. of Electrical and Computer Engineering, University of New Mexico
79. D Kumar, P Carvalho, M Antunes, J Henriques, M Maldonado, R Schmidt, J Habetha, "Wavelet Transform And Simplicity Based Heart Murmur Segmentation", ISSN 0276-6547 Computers in Cardiology 2006;33:173-176
80. Dmitriy Genzel and Eugene Charniak, “Entropy Rate Constancy in Text”, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 199-206.
81. Rong Zhang1, Rongshan Yu, Qibin Sun, Wai-Choong Wong, “A New Bit-Plane Entropy Coder for Scalable Image Coding” 0-7803-9332-5/05/$20.00 ©2005 IEEE
82. Hariharan Subramanian, “Audio Signal Classification” M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November2004
83. M. Bank, S. Podoxin, V. Tsingouz, “Estimation of non distortion audio signal compression”, Department of Communication Engineering, Center for Technological Education Holon.
84. Guillaume Gravier, Scott Axelrod, Gerasimos Potamianos, Chalapathy Neti, “Maximum Entropy And Mce Based Hmm Stream Weight Estimation For Audio-Visual Asr” IBM T. J. Watson Research Center-USA
85. Chi-Min Liu, Wen-Chieh Lee, and Yo-Hua Hsiao, “M/S Coding Based On Allocation Entropy”, Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003