International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 4, Issue 3 (March 2015), PP.01-07

PSoC BASED SPEECH RECOGNITION SYSTEM

[1]Shubhasini Sugumaran, [2]Mr.V.R.Prakash
[1]*Department of Electronics and Communication Hindustan University Chennai, India*
[2]*Department of Electronics and Communication Hindustan University Chennai,India*

Abstract:– Speech Recognition Systems(SRS) have been implemented by various processors including the digital signal processors(DSPs) and field programmable gate arrays(FPGAs) and their performance has been reported in literature. The fundamental purpose of speech is communication, i.e., the transmission of messages.In the case of speech, the fundamental analog form of the message is an acoustic waveform, which we call the speech signal. Speech signals can be converted to an electrical waveform by a microphone, further manipulated by both analog and digital signal processing, and then converted back to acoustic form by a loudspeaker, a telephone handset or headphone, as desired.The recognition of speech requires feature extraction and classification. The systems that use speech as input require a microcontroller to carry out the desired actions. In this paper, Cypress Programmable System on Chip (PSoC) has been studied and used for implementation of SRS. From all the available PSoCs, PSoC5 containing ARM Cortex-M3 as its CPU is used. The noise signals are firstly nullified from the speech signals using LogMMSE filtering. These signals are then sent to the PSoC5 wherein the speech is recognized and desired actions are performed.

Keywords: PSoC, LogMMSE, Speech Recognition

I. INTRODUCTION

The basic idea of speech is the transmission of messages. A message is represented as a sequence of discrete symbols that quantifies its information in bits and the rate at which information is transmitted as bits per second (bps). Speech recognition techniques have seemed to be more efficient and convenient for human-machine interaction. The speech recognition systems with fixed vocabulary were deployed in many applications[8] and [9]. Speech recognition systems for voice operated application have been implemented using various hardware platforms such as the DSPs [3], FPGAs [5] and microprocessors [10]. In speech production, the information to be transmitted is encoded in the form of a continuously varying analog waveform that can be transmitted, recorded, manipulated, and ultimately decoded by a human listener. This analog signal is the speech signal. These signals tend to be corrupted by noise in the real world. If the noise can be estimated from the noise source, this estimated noise can then be subtracted from the primary channel resulting in the desired signal. This task is usually done by linear filtering. In real time situations, the corrupting noise is a nonlinear distortion version of the source noise, so a nonlinear filter should be a better choice. To reduce the influence of noise in the speech, speech enhancement is done. The recognition algorithm without the use of enhancement algorithms proved to be less efficient.

Programmable System on Chip (PSoC) have and are being employed in a number of applications. They are cost effective due to which they have limited storage and computational power. In context to this, the recognition accuracy becomes important for PSoC and is addressed in the paper.

II. PROGRAMMABLE SYSTEM ON CHIP (PSOC)

Programmable System on Chip(PSoC) has been designed and implemented by Cypress semiconductors [2] and [4]. Every PSoC contains a microcontroller, programmable analog blocks such as ADC, DAC, I/O drivers and digital blocks such as Universal Digital Blocks (UDBs), CAN, I2C, PWM in a single chip. Embedded Development kits from Cypress contain one of the three PSoCs – PsoC1, PSoC3 and PSoC5. The processing performance, functionality, internal memories and configurability of the PSoC increases from PSoC1through PSoc5.

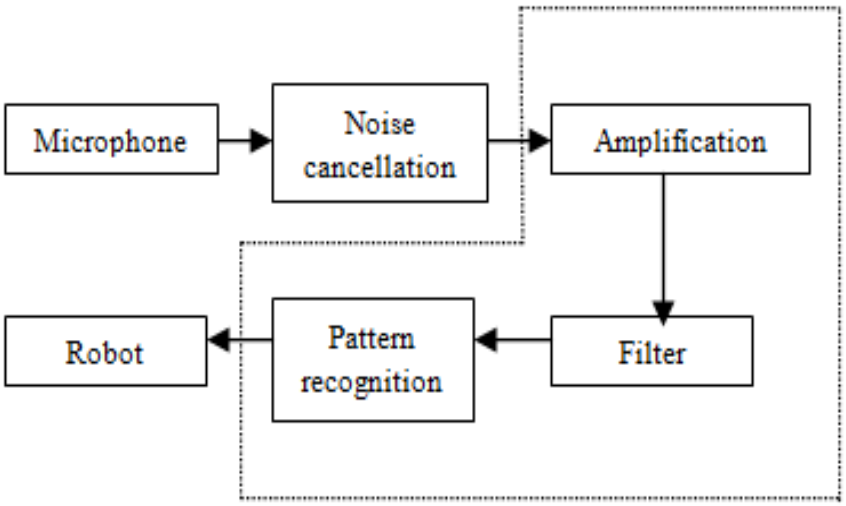
FEATURES	PSoC1	PSoC3	PSoC5
CPU	8-bit M8C core	8-bit 8051	32-bit ARM Cortex
Flash	4 to 32kB	8 to 64kB	32 to 256kB
Interface	I²C, SPI, UART, FS USB 2.0	I²C, SPI, UART, LIN, FS USB 2.0, I²S, CAN	I²C, SPI, UART, LIN, FS USB 2.0, I²S,CAN
ADCs	1 delta-sigma	1 Delta-Sigma	1 Delta-Sigma, 2SARs
DACs	2(6 bit)	4(8 bit)	4(8 bit)
I/Os	64	72	72

Table 1: Comparison of PSoCs

PSoC5 uses PSoC Creator for its development and implementation. PSoC Creator is a visual development tool and Integrated Development Environment for PSoC. It has a rich library of prebuilt components and a schematic design entry tool. It combines C based development flow with an automatically generated Application programmable interface (API). API rduces the errors in code and ensures proper interfacing with the peripheral devices which enables the software development to be faster, easier and less prone to errors. The PSoC Creator also has powerful, modern debugger, which is built in the IDE. It is used to display the values after execution at each point.

III. SPEECH RECOGNITION SYSTEM

The block diagram is as shown in Fig.1 below.



Speech Recognition System

The speech is given as input through a microphone. The given speech might contain external disturbances called noise. This noise is cancelled using LogMMSE algorithm. The speech signal after noise cancellation is amplified and filtered to remove further disturbances in the signal. The speech signal in analog form is then converted into digital signal using inbuilt analog to digital converter of PSoC. The feature extraction and pattern recognition is done using MFCC algorithm. The analog to digital conversion is done using delta sigma ADC present in PSoC5.

The sampling rate of ADC can be adjusted depending on the speech signal as required. The delta sigma ADC contains three blocks – an amplifier, a modulator and a decimator. The decimator is a four stage CIC decimation filter. It also contains a post processing unit. Continuous mode of the ADC is being used for conversion. The output of the speech recognition system is shown through the movement of a robot in different directions.

IV. TECHNIQUES USED

The performance of the SRS degrades when implemented in real world environment. This degradation is due to acoustic model mismatch. The acoustic model mismatch describes the difference between the environment in which the SRS is tested and the actual environment in which it is deployed. This can include echoes, background noise, speaker variability and transmission effects.

1. Log MMSE Filtering

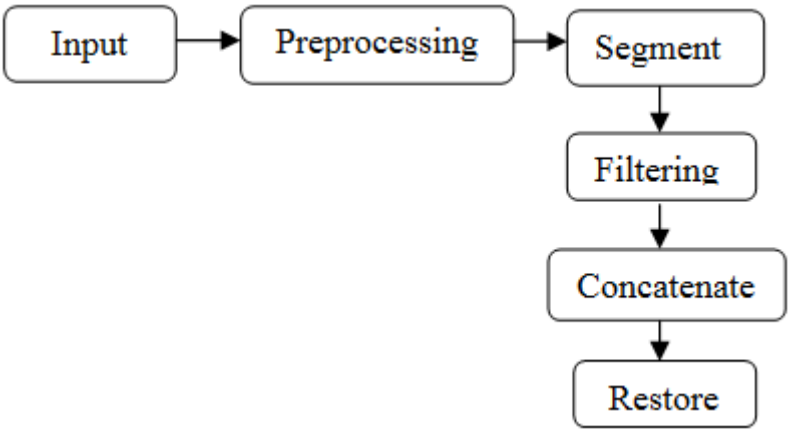


Fig.2 Noise reduction using Log MMSE

1.1 Preprocessing

In preprocessing of audio signals start with pre-emphasis refers to a system process designed to increase the magnitude of some frequencies with respect to the magnitude of other frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation distortion or saturation of recording media in subsequent parts of the system. The mirror operation is called de-emphasis, and the system as a whole is called emphasis.

Pre-emphasis is achieved with a pre-emphasis network which is essentially a calibrated filter. This network composed of two resistors and one capacitor. The frequency response is decided by special time constants. The cutoff frequency can be calculated from that value. Pre-emphasis is commonly used in telecommunications, digital audio recording, record cutting, in FM broadcasting transmissions, and in displaying the spectrograms of speech signals.

De-emphasis is the complement of pre-emphasis, in the anti noise system called emphasis. Emphasis is a system process designed to decrease, (within a band of frequencies), the magnitude of some (usually higher) frequencies with respect to the magnitude of other (usually lower) frequencies in order to improve the overall signal-to-noise ratio by minimizing the adverse effects of such phenomena as attenuation differences or saturation of recording media in subsequent parts of the system.

1.2 Segment

The signal samples are segmented into fixed number of frames and each frame samples are evaluated with hamming window coefficients.

The total frames are calculated by,

Fn = (Ls – Ns)/ (Ns*Sp) +1

Where,

Ls = length of signal

Ns = Length of each frame

Sp = Shift Percentage

Finally the samples of each frames are separated from input signal using Fn and Sp and its scaled by the hamming window coefficients.

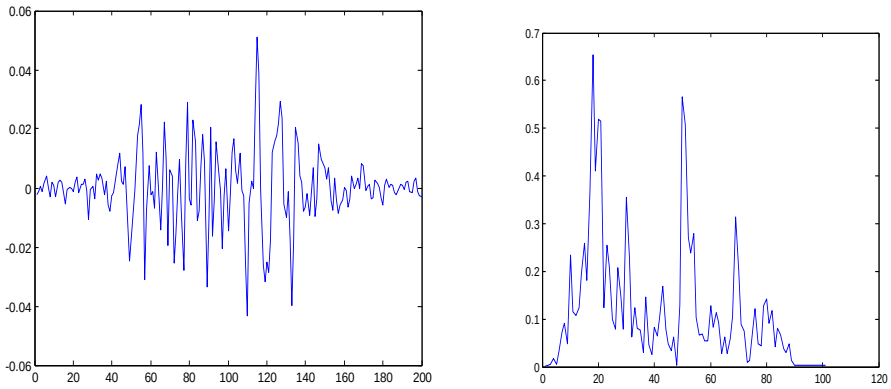


Fig.3 Segmented signal

1.3 Filtering

After the signal segmentation, the magnitude and phase spectrum from noisy signal are computed by applying fast fourier transform.The magnitude of noisy signal spectrum is further utilized for filtering process and signal phase kept same.

The restored signal magnitude spectra is obtained by,

Rs = G .* Y

Where, G – Log spectral amplitude Gain function

Y – Magnitude response of noisy signal

The log spectral gain function is defined by,

G = x./(1+x) exp(eint (v))

Where, **v = x./(1+x) * r**

x and r– Priori and posteriori signal to noise ratio

eint – Exponential integral

The posteriori snr is defined by,

r = (Y.^2)/ lamda

lamda = E[(Y).^2]

Where, lamda - Noise power spectrum variance

E – Mean value

Complex spectrogram obtained by Filtered magnitude spectrum is combined with noisy signal phase spectrum. The restored signal is reconstructed by applying inverse fast fourier transform to this complex spectrogram.The performance of filtering is measured with SNR evaluation and it is defined by,

SNR= 10log10 (Msig^2/(sum(in-out)^2/Ls))

Where, Msig = Maximum amplitude of signal

in, out= Noisy input signal and restored output.

1.4 Concatenate

The signals that were obtained on segmentation are now filtered and the noise content is reduced. These signals are then concatenated to reform the original signal.

1.5 Restore

The restored noiseless signal is restored to its original form.

2. MEL FREQUENCY CEPSTRUM COEFFICIENT(MFCC)

To increase the robustness in the frame selection process, a robust feature extraction with short time Fourier transform (STFT) domain uncertainty propagation (STFT-UP) was used. Our implementation followed using STFT-UP to compute a minimum mean square error (MMSE) estimate directly in the mel-frequency cepstral coefficient (MFCC) domain. For this particular implementation, amplitude based MFCCs with cepstral mean subtraction were used to attain improved performance.

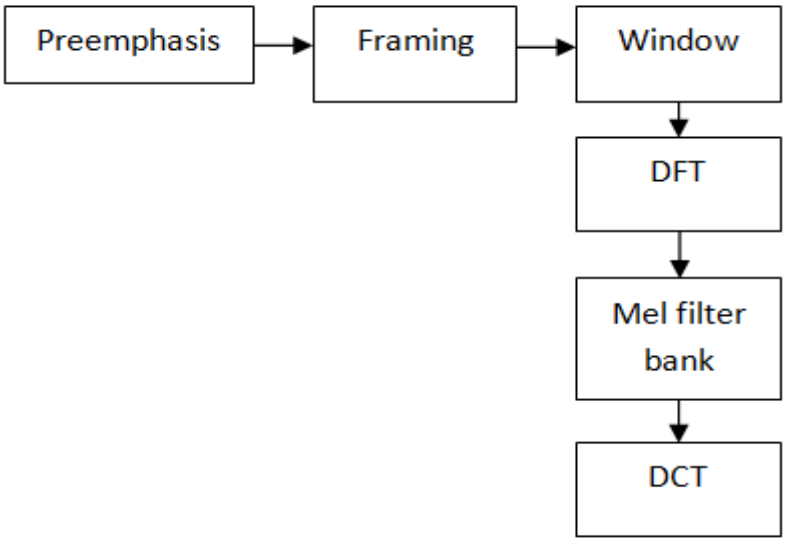


Fig. 4 MFCC

In speech recognition, the Mel-frequency cepstrum (MFC) is a representation of the shortterm power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency.

MFCC = DCT [LOG [ABS [FFT (SPEECH)]]]

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound.

V. TRAINING AND TESTING

The goal of the system training and inventory design stage is twofold: we need to divide the inventory into collections of phonetically similar segments with varying lengths and we need to arrive at a statistical description that tells us which set of collections is most likely to contain the inventory subsection that best matches the underlying clean frame of an incoming noisy frame . The division of the inventory into the collections is performed in a step-by-step fashion. First, is segmented and all silent segments are removed. The non-silent part of the inventory is then divided into sections that each belong to one of 40 phonetic classes.

We are applying the feature extraction to the entire segment stream of the inventory. Because the inventory signal is assumed to be virtually undistorted, it is sufficient to only retain the resulting short-time MFCC feature means and to discard the associated variance estimates. The feature means become, thereby, essentially feature vectors in their own right and we can develop a cluster model for them. We have decided to use the means and not the actual cepstral vectors at this stage to ensure that the impact of the mean-extraction-processing is captured in our feature representation.

VI. RESULTS

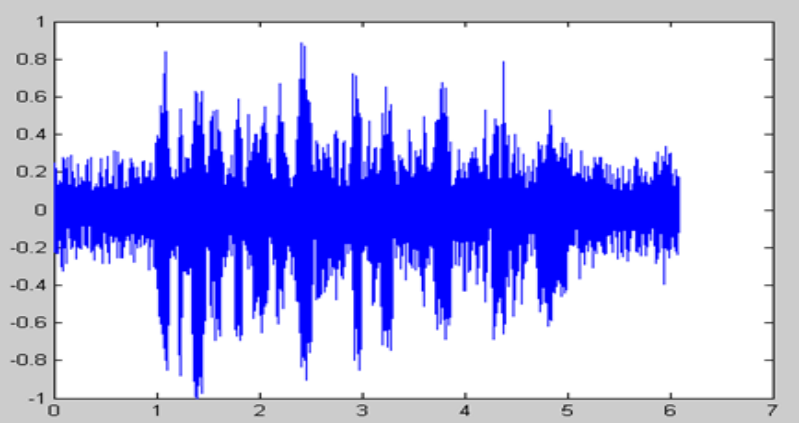


Fig. 5 Noisy Signal

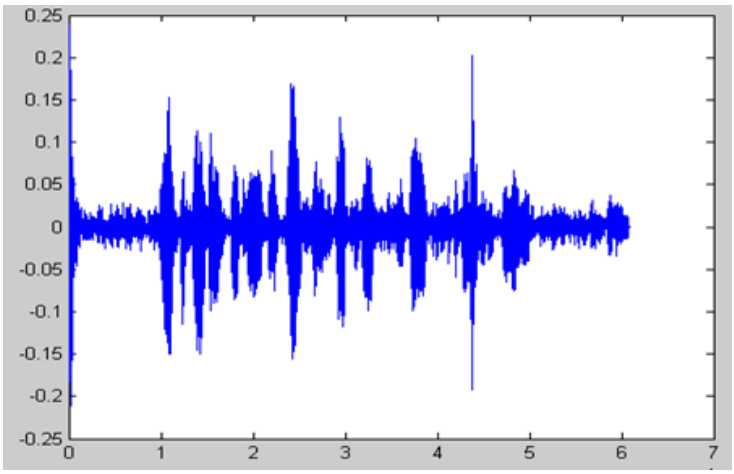


Fig. 6 Log MMSE Filtering

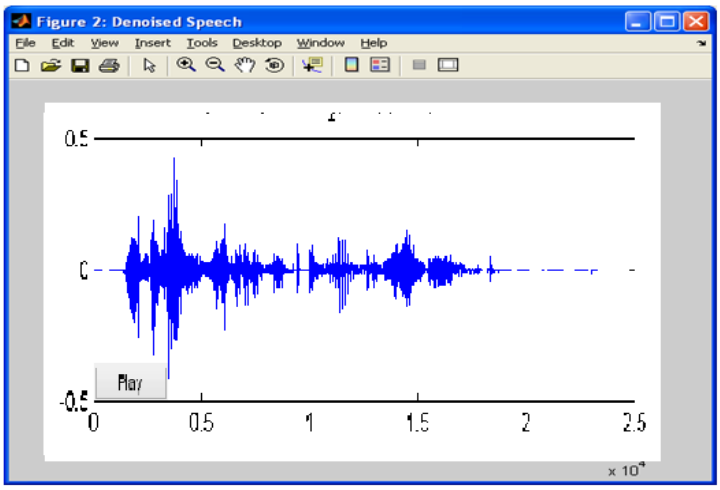


Fig. 7 MFCC Feature Extraction

For experimental basis, speech signals from noisy environments were taken. The result shown is fr a signal in the airport. The noise is cancelled using LogMMSE Filtering and feature extraction is done by MFCC.

VI. CONCLUSION

In this paper speech recognition is done and appliances are controlled using speech commands. The commands are given by the user in the form of speech. These signals are filtered using LogMMSE filtering due to which environmental noise is nullified. Further as a part of feature extraction, MFCC is used. A database is created where all the commands are saved. These commands are then given as input to the PSoC5 kit where the PSoC is programmed to give the desired result. According to the PSoC, the robot connected to the PSoC moves in different directions as specified.

REFERENCES

[1]. V. Naresh, B. Venkataramani, Abhishek Karan and J. Manikandan, " PSoCbased isolated speech recognition system, " International conference on Communication and Signal Processing, April 3-5, 2013.

[2]. R Namba, K Kobayashi, T Ohkubo and Y.Kurihara, "Development of PSoC microcontroller based solar energy storage system," Proceedings of SICE Annual Conference (SICE), 2011, pp.718-721, 2011.

[3]. J. Manikandan, B. Venkataramani, K. Girish, H. Karthic, V. Siddharth, "Hardware Implementation of Real-Time Speech Recognition System Using TMS320C6713 DSP",24th International Conference onVLSI Design (VLSI Design),pp.250-255, 2011

[4]. Jingchuan Wang and WeidongChen , "Integration of PSoC technology with educational robotics", International Conference on Field-Programmable Technology (FPT), 2010, pp.332-336, 2010.

[5]. Cheng-Yuan Chang , Ching-Fa Chen , Shing-Tai Pan , Xu-Yu Li, " The speech recognition chip implementation on FPGA ", Mechanical and Electronics Engineering (ICMEE), 2nd International Conference,2010.
[6]. V. Amudha, B. Venkataramani, J. Manikandan, "FPGA implementation of isolated digit recognition system using modified back propagation algorithm,"International Conference on Electronic Design ICED 2008, pp.1-6.
[7]. V.Amudha, B.Venkataramani, R.Vinoth Kumar and S. Ravishankar,"SOC Implementation of HMM Based Speaker Independent Isolated Digit Recognition System", in Proc. of IEEE Int. Conf. on VLSIDesign VLSI'07, 2007, pp.848-853.
[8]. Trihandyo, A. Belloum, A. Kun-Mean Hou, ″A real-time speech recognition architecture for a multi-channel interactive voice response system″, *Iternational Conference on Acoustics, Speech, and Signal Processing*ICASSP-97, vol.2, pp.1527-1530,1997
[9]. Mike Wald, Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality [A].In Proceedings of 34th ASEE/IEEE Frontiers in Education Conference S3G, Indianapolis, Indiana, 2005, pp 22-25.
[10]. N. Hataoka, H. Kokubo, Y. Obuchi, and A. Amano, "Compact and robust speech recognition for embedded use on microprocessors," IEEE Workshop on Multimedia Signal Processing, pp. 288-291, 9-11 Dec. 2002.