Speech Recognition using Processor DSP TMS320C6713 DSK

Wafaa. Mokrane,[a] Rachid Elkabil[b]
[a] University Hassan II Mohammedia-Casablanca, Casablanca, Morocoo
wafaa.mokran@gmail.com
[b] University Hassan II Mohammedia-Casablanca, Casablanca, Morocoo
relkabil@gmail.com

Abstract. Begin t. The speech signal transmits a lot of information on the speaker identity sex and emotion, numerous works have focused on the problem of automatic speaker recognition. The feature extraction and classification methods are the most important tasks in the speech recognition process.

As part of this work we are interested in the extraction of parameters by an algorithm that is inspired by the human auditory model, Mel Frequency Cepstral Coefficients (MFCC). The speech features extracted (MFCCs) are quantified using the algorithm of Vector Quantization. In this paper we used TMS32006713 Kit with softwares Code Composer Studio and Matlab for speech recognition in real time.

Keywords: Speech Recognition, Mel Frequency Cepstral Coefficients (MFCC), TMS32006713 Kit, Code Composer Studio.

1 INTRODUCTION

Speech recognition is an important area of digital signal processing , which is the subject of research of the automatic speaker recognition. The main objective of automatic speaker recognition is to extract the vocal characteristics of each individual, characterize and recognize the information about the identity of the speaker. The speaker identification using voice is an usualy stain commonly used, Identification can be opened, where it is possible that the speaker is not present in the system, or closed, when the speaker is necessarily in the system ,the task can be further divided into identification with dependent or independent text.

The voice recognition technical is used in many fields such as, verification for access control to services such as voice dialing, phone shopping, Access Services database, telephone banking, information service, Voicemail, security control for confidential information areas, and remote access to computers.

In this work we are interested in the extraction of parameters by an algorithm that is inspired by the human auditory model, this algorithm involves to analyze the speech signal in a frequency band. Mel Frequency Cepstral Coefficients (MFCC) is a feature vector most widely used for automatic speaker recognition. The extracted speech features (MFCCs) are quantified to a speaker on a number of centroids using the vector quantization algorithm (V. Tiwari *2010).* MFCC are calculated in the learning phase and again in the testing phase.

In this document, the TMS320C6713 DSP processor with Code Composer Studio (CCS) and Matlab was used for speech recognition in real time.

2 SPEECH RECOGNITION

The anatomical structure of the vocal tract is unique for each individual. So , the vocal information that characterizes each person available in the speech signal may be used to identify the speaker (S.Z.Boujelbene,D.B.A.Mezghani,N.Ellouze 2009).

2.1 Main stains in voice recognition

Speaker recognition called generally two tasks: the speaker verification and speaker identification.

The task of speaker verification is to recognize a person with a recording of his speech, the answer to this task is the binary type, either both recordings were pronounced by the same speaker, or two different speakers produced the recordings. In this case the decision is binary either accepted or refused. While the stain of speaker identification corresponds to the search of the identity of the person who is present in a database in the form a test sample of his voice, the model in the speaker is closest to the extracted model the test sample is declared as being the identity of the speaker under test. The latter task can be divided into two classes: Open identification system and closes identification system.

Opened system of identification: can be brought to supply an empty set, the decision must be made on the sample which the unknown word looks like most. If no correspondence is satisfactory with regard to a predefined threshold the system declares the speaker in test being unknown (S.Z.Boujelbene,D.B.A.Mezghani,N.Ellouze 2009).

System of identification closes: can be used in a very effective way who has to supply a set at least with a speaker, in this system every access of test is compared with all the models of the speakers referenced in the system. The identity of the speaker possessing the closest reference is emitted in exit of the system.

The fashions of voice recognition are given by the mode dependent of text and the mode independent of text.

Mode dependent of text: in mode dependent on the text, the text pronounced by the speaker is the same that the one that he pronounced during the learning of his voice. The levels of dependence to the text are classified according to the applications: systems free-text, systems text-prompted, systems vocabulary-dependant, or system user specific text dependent. The knowledge a priori of the vocal message returns the task of the systems of easier voice recognition and the performances more better (V. Tiwari *2010).*

Mode independent from the text: in mode independent from the text, the speaker can pronounce whatever words to be recognized. In this mode, there is no constraint on the message or on the language which the speaker can use.

2.2 Structure of a recognition system

The structure of automatic recognition of the speaker consists of three model main, modulates of extraction of parameter (acoustic analysis), models of modeling of the speakers and model of decision.

More, a system of speaker's automatic recognition possesses two modes of functioning.

Learning: where a model is estimated for every speaker of the system then will serve to decide for the tasks of recognition to come.

Test: where the recognition step (verification, identification ...) is performed. Output of this phase, the system sends a response: an identity for the identification task or decision access / rejection for verification (S.Z.Boujelbene,D.B.A.Mezghani,N.Ellouze 2009). Fig (1) represents a diagram illustrating the operation of an automatic system of identification of the speaker .

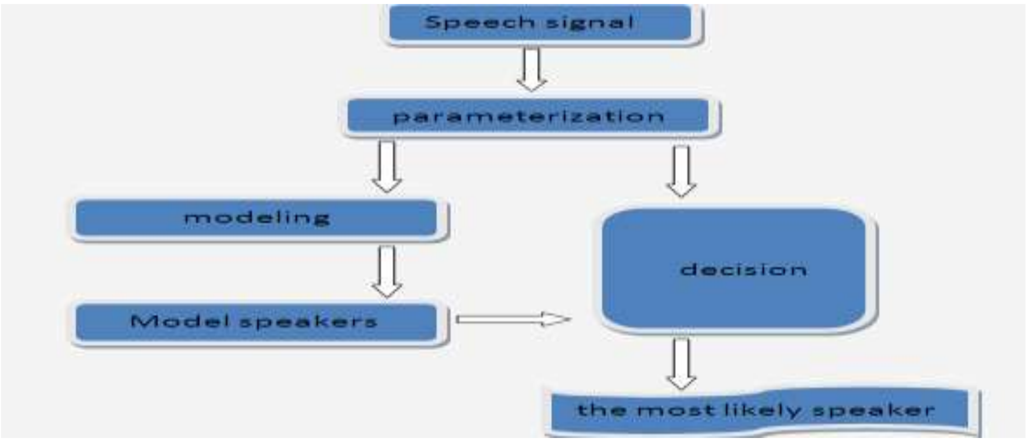


Fig. 1: Diagram a modular system automatic speaker recognition.

2.3 Methods of coding and Parameterization of the signal of word

The literature approaches numerous types of parameterization of the signal of the word the parameters of which must be frequent, easily measurable, not too sensitive to the variability intra-speaker, strong in front of imitators, etc (S.Z.Boujelbene,D.B.A.Mezghani,N.Ellouze 2009). It appears that the only types(chaps) of parameters usable effectively are the parameters of the spectrum analysis, the dynamic parameters and the prosodic parameters. Most of the techniques of parameterization consist in describing the envelope of the specter in the frequency domain. The phase of extraction of characteristics must be carefully made, because it contributes directly to the performances of the global system. The coders most usually used are the predictive linear coding (Linear Predictive Coding LPC), the cepstral coding (Mel Frequency Cepstrum Coding) MFCC or perceptual linear predictive coding (Perceptual Linear Predictive PLP).The coding MFCC and the coding PLP have the property to integrate knowledge of the human hearing model.

2.3.1 Linear Predictive Coding LPC

The coding LPC is a technique stemming from the analysis of the production of the word allowing to obtain coefficients of linear prediction From these coefficients we can calculate the cepstraux coefficients (LPCC) who will be used for the characterization of speaker [3].

2.3.2 Mel Frequency Cepstre Coding MFCC

MFCC bases himself on the analysis by a bench of filters with Mel's Scale this Scale gets closer to the frequency perception of the ear, Mel's scale contained a bench from 15 to 24 triangular filters spaced out linearly up to 1 kHz then spaced out logarithmically until the maximal frequencies It was conceived (designed) so that 1000 Hz corresponds to 1000 Mels. The conversion of the hertz to Mel is given by the formula (1).

$$\text{Mel} = 2595 \log (1+f(\text{Hz})/700) \qquad (1)$$

3 MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

Vector MFCC of characteristics most widely used for the automatic recognition of locuteur.il is used for the conception of a system of speaker's identification dependent on the text. MFCC based on the variation known for widths of critical band of the human ear. Mel-frequency spaced out linearly until 1 kHz after spaced out logarithmically until the maximal frequencies. The global process of calculation of MFCC is represented on the fig 2.

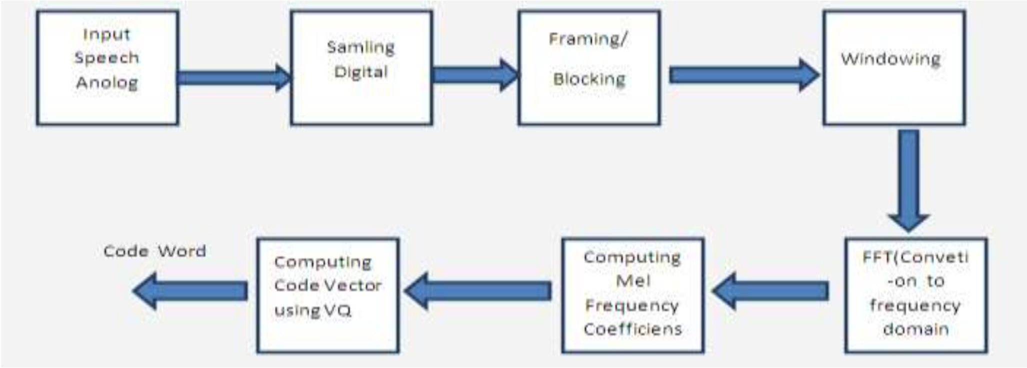


Fig. 2. Steps for speaker recognition implementation

We are going to make a detailed description of the implementation of this algorithm of parametric extraction.

Level detection: the beginning of a vocal signal is defined one base itself on a prerecorded threshold. The signal is captured just after its launch is passed on to the Framing stage.

Blocking frame: in this stage, the signal of continuous word is blocked in wefts of N samples, with the neighboring frames being separated by M (M <N). Typical values used are M = 100 and N = 256. The first weft is constituted by the first one N samples. This process continues until all the word is represented inside one or several. The process of segmentation of the samples of word obtained from an ADC in a frame small with length in the beach from 20 to 40 msec.

Windowing: after the framing the window of Hamming is used so as to minimize the discontinuities of signal at the beginning and at the end of every weft. The equation of the window of Hamming is given by W (n) avec 0 = n = N-1 the fig (3) represents the windowing of Hamming.

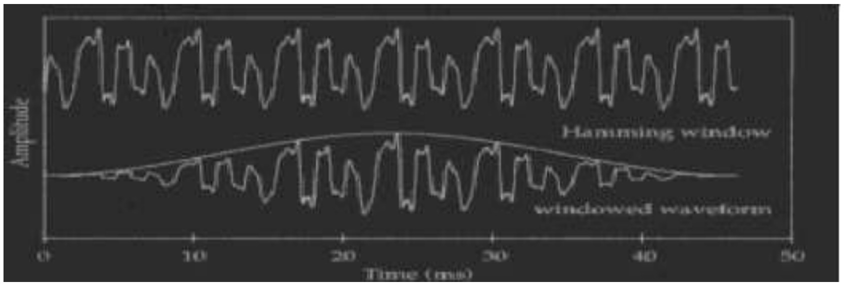


Fig 3: Hamming Window

Fast Fourier transform: to convert every weft of N samples of temporal domain into frequency domain .the word is a real signal, but FFT has two real and imaginary components.

Power spectrum calculation: the power of the frequency domain is calculated by summing the square of the real and imaginary components of the signal to yield a real signal. The second half of the samples in the frame are ignored since they are symmetric to the first half (the speech signal being real) (R.chassaing & D. Reay).

Mel-frequency wrapping: triangular Filters are conceived by using the scale of frequency Mel with a bank of filters to move closer to the human ear. The electric signal is then applied to this bank of filters to determine the contents of frequency through every filter. The scale Mel-Frequency between 0 et4 kHz. The specter Mel-Frequency is calculated by multiplying the specter of the signal with a set of triangular filters destined by using the scale Mel. For a given frequency, the Mel the frequency is given by the formula (2). And the bench of Mel's filter represented in the face (figure 4).

$$B(f) = \left[1125 \ln\left(\frac{1+f}{700} \right) \right] mels \qquad (2)$$

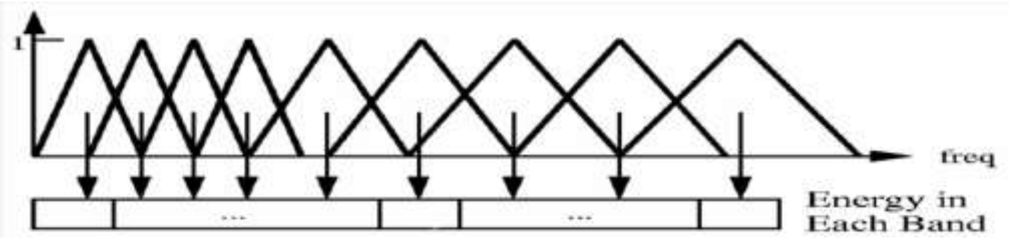


Figure 4: Mel scale filter bank, from (young et al,1997)

Mel-freequency cepstrum coefficients: after discovering the power spectrum, Mel's specter of log must be again converted to weather. The discrete cosine transform (DCT) gives the MFCCs. given by formula (3).

$$\mathrm{Ck} = \sum_{n=1}^{N} \mathrm{UnX}(n)\cos\left(\pi(2n-1)(k-1)\backslash 2\pi\right) \qquad (3)$$

3.1 VECTOR QUANTIZATION

The vectorial quantification is the classic technique of quantification of treatment of signal which allows the modeling of the functions of density of probability by the distribution of vectors prototypes. It works by dividing a large number of points into groups having approximately the same number of points which are close them. Every group is represented by its point centroïde (V. Tiwari *2010).* After the calculation of the MFCCs and for speaker's identification calculates the Euclidian distance between the MFCCs of every speaker in the phase of the learning for the center of gravity of every speaker individual and for the phase of test in measure the minimal Euclidian distance.

4 TMS320C6713 DSK

4.1 CPU description

The platform of TMS320C6000 of the processors of digital signals (DSP) is a part of the family of the DSP TMS320. Device Plan) TMS320C67x (' C67 x) is in floating decimal point DSP in the platform of TMS320C6000. The DSP TMS320C67x (including the device of TMS320C6713) composes the generation of DSP with floating point in the platform DSP (*TMS320C6713 DSK. Technical Reference).*

The device of C6713 is based on high-performance, advanced in very long word instruction (VLIW) structures developed by Texas Instruments (TI), making of this DSP an excellent choice for the multi-channel and multifunction applications (M.U.Nemade & S.K.Shah. 2014). Operating at 225 MHz,1800 million instructions per second (MIPS), two fixed multipliers / floating up to 450 million operations of multiplication- accumulation by second (MMACS). C6713 offers up to 1350 million operations in floating decimal point per second (MFLOPS), Operating at 300 mHz, 2400 million instructions per second (MIPS), C6713 offers up to 1800 million operations in floating decimal point per second (MFLOPS), and with two fixed multipliers / floating up to 600 million operations of multiplication - accumulation by second (MMACS). The device of TMS320C6713 has two modes of starting up. The diagram blocks describing the card is illustrated by the fig 5.

The kit of starting up DSK includes the following material elements:

- A Texas Instruments TMS320C6713 DSP operating at 225 MHz.
- An AIC23 stereo codec.
- 16 Mbytes of synchronous DRAM.
- 512 Kbytes of non-volatile Flash memory.
- 4 user accessible LEDs and DIP switches.
- Software board configuration through registers implemented in CPLD.
- Configurable boot options.
- Standard expansion connectors for daughter card use.
- JTAG emulation through on-board JTAG emulator with USB host interface or external emulator.
- Single voltage power supply (+5V).

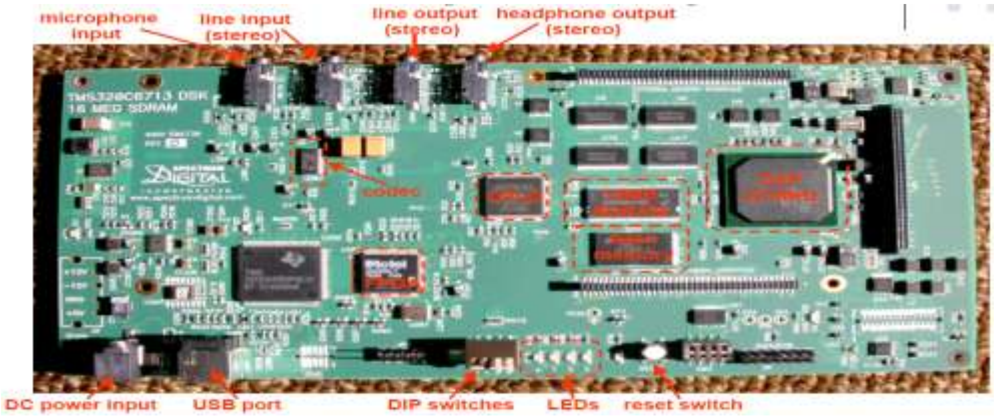
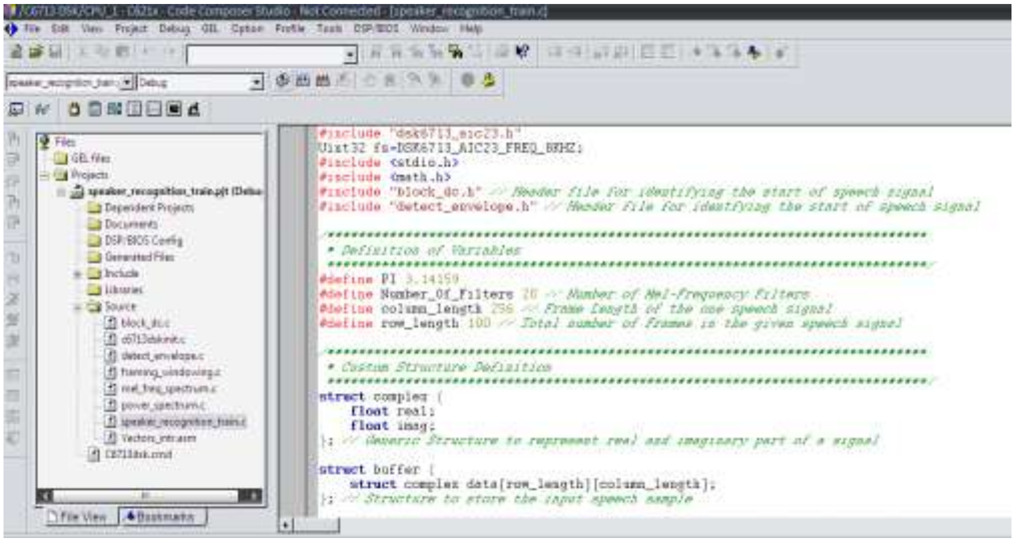


Fig 5: SYSTEM LAYOUT OF TMS320C6713

4.2 Code Composer Studio

To communicate with the DSK, we use the software code composer studio. CCS provides an IDE to integrate software tools. CCS includes tools for code generation, such as a compiler C / C + +, an assembly optimizer is the editor of links. He has graphic capacities and supports the real-time debugging. He supplies a software tool easy to use to build and debug programs. With DC, we can initialize various ports and registers of the DSK. Code Compose offers an environment of rich debugging which allows browsing the code, the stop points, and by examining the registers that the code is executed (*K.T.Talele. SPEAKER RECOGNITION USING TMS320C6713DSK*).

Target software Code Composer Studio includes DSP / BIOS ™ kernel for the TMS320 DSP TMS320 DSP standard algorithms to allow the re-use of the software, Chip Libraries of support to simplify the configuration of the device, and libraries of the DSP for optimal DSP feature. The figure 2 shows the interface of programming of the CSC. Figure 6 shows the interface of programming of DC.


C6713-DSK/CPU_1 - C621x - Code Composer Studio - Not Connected - [speaker_recognition_train.c]
File Edit View Project Debug GEL Option Profile Tools DSP/BIOS Window Help
speaker_recognition_train
Debug
Files
GEL files
Projects
speaker_recognition_train.pjt (Debug)
Dependent Projects
Documents
DSP/BIOS Config
Generated Files
Include
Libraries
Source
block_dc.c
c6713dskinit.c
detect_envelope.c
hamming_windowing.c
mel_freq_spectrum.c
power_spectrum.c
speaker_recognition_train.c
Vectors_intr.asm
C6713dsk.cmd
File View
Bookmarks

```
#include "dsk6713_aic23.h"
Uint32 fs=DSK6713_AIC23_FREQ_8KHZ;
#include <stdio.h>
#include <math.h>
#include "block_dc.h" // Header file for identifying the start of speech signal
#include "detect_envelope.h" // Header file for identfying the start of speech signal

/*****************************************************************************
 * Definition of Variables
 *****************************************************************************/
#define PI 3.14159
#define Number_Of_Filters 20 // Number of Mel-Frequency Filters
#define column_length 256 // Frame length of the one speech signal
#define row_length 100 // Total number of Frames in the given speech signal

/*****************************************************************************
 * Custom Structure Definition
 *****************************************************************************/
struct complex {
    float real;
    float imag;
}; // Generic Structure to represent real and imaginary part of a signal

struct buffer {
    struct complex data[row_length][column_length];
}; // Structure to store the input speech sample
```

Fig. 6 : The programming interface of CCS

5 ASPECT PRACTICE

The connection between host PC and target board is shown in fig.7. The audio input is appliedto Line-in of DSK TMS320C6713, The signal processing made in DSK by means of a code"C", which generates the required speech. The output is taken from the “Headphone out” Of the DSK and then it is applied to speaker.

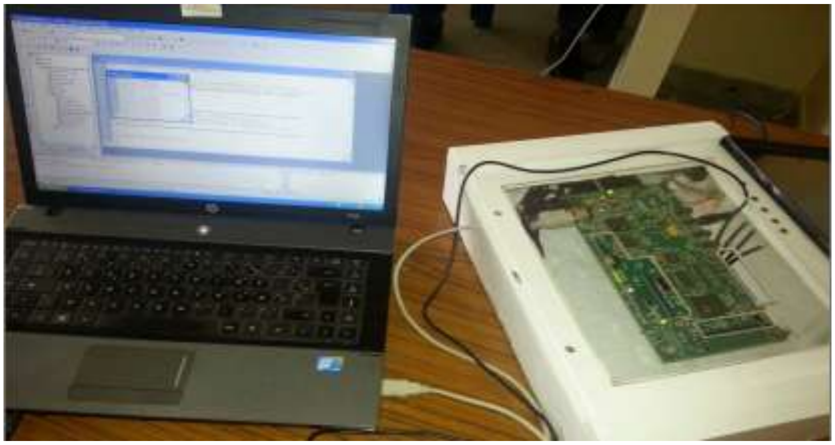


Fig. 7 : Connection between Host PC and Target board

For the analysis of the performance of the speech recognition, we considered here of eight speakers who pronounces the same zero number.

The purpose of this project is to determine the identity of the speaker from a vocal signal. All the sound productions stemming from various speakers were directly digitize in monophonic worlds with the format.wav with a sampling frequency Fe=15 kHz and a result of 16 bits the conception is examined at first with Matlab.

Eight speech samples from eight different people (eight speakers, labeled S1to S8) are used to test this project. Each speaker utters the same single digit, zero, once in a training session (then also in a testing session).A digit is often used for testing in speaker recognition systems because of its applicability to many security applications. MATLAB is used to convert files WAV registered in the file DAT. These files of data are then used as entrances for card DSK.

This project was implement by Sowmya Narayanan and Vasanthan Rangan.of the eight speakers, the system has correctly identified six (with an identification rate of 75%). Our work is to increase the recognition rate of 7/8 instead of 6/8.

6 CONCLUSIONS

In this paper we presented the real time hardware implementation of speech recognition using DSP processor software development kit, DSK-TMS320C6713 with Code Composer Studio (CCS). MFCC algorithm calculates cepstral coefficients of Mel frequency scale. After feature extraction from recorded speech, each Euclidean Distance (ED) from all training vectors is calculated using VQ.

The identification rate can be improved by adding more vectors to the training code words. The performance of the system may be improved by using two-dimensional or four-dimensional VQ or by changing the quantization method to dynamic time wrapping or hidden Marcov modeling.

References

V. Tiwari.*(2010) .MFCC and its applications in speaker recognition: International Journal on Emerging Technologies 19-22(2010).*

S.Z.Boujelbene,D.B.A.Mezghani,N.Ellouze.(*2009).Identification du Locuteur par Système Hybride GMM-SMO : Sciences of Electronic,Technologies of Information and Telecommunications March 22-26, 2009 – TUNISIA.*

R.chassaing & D. Reay. *Digital Signal Processing and Applications with TMS3206713 and TMS3206416 DSK.page 496-500.*

TMS320C6713 DSK. Technical Reference

M.U.Nemade & S.K.Shah.(2014). *Real Time Speech Recognition Using DSK TMS320C6713:International Journal of Advanced Research in Computer Science and Software Engineering.*

K.T.Talele. SPEAKER RECOGNITION USING TMS320C6713DSK towards the fulfillment of Bachelor of Engineering course in Electronics of the Mumbai University.