FINAL REPORT EE478

Voice-controlled Camera

Abstract: This report contains the process creating a voice-controlled camera. In this one, we will discuss about design specification, design implementation, and our testing for the product. At the end, readers will have an overview about our product about its shape, its characteristics, as well as its function also.

Group 13 - Class 09ECE

Nguyen Ngoc Minh	minh13055@gmail.com
Ngo Tran Duc Thang	tngotran@gmail.com
Nguyen Xuan Tien	tiennguyen148@gmail.com

Danang, 28-Apr-2013

Table of Contents

I. INTRODUCTION

In this project, we aim to create an embedded computer product to support people with their daily activities.

Supposing that you wanted to take snapshots during playing with your beloved children; so as not to miss that moment, an idea did across our mind with making a camera that people can control with their voice only. This camera will conduct immediately the task required by users afterwards. We call this product is ***"Voice-controlled Camera"***.

With that said – we would like to send our special thanks to authors of the articles in the attached links at the end of this report for their valuable knowledge that they represented in these articles and magazines.

The project is also using a music player for testing the Speech recognition.

II. BACKGROUND

- DSP knowledge in Speech processing.
- Another important background is experiment to working with the KIT DSKC6713. In the project, we are take too long time to familiar with the kit. The reason is that, the KIT is 10 years old from 2003, so almost tutorials or documents of this kit is presented on Window XP. For saving the classmate time for next subject on next semester, we did make some tutorial videos.
- MFCC, Vector quantization, HMM algorithms in analysis voice signal.
- Knowledge about the DAC, ADC, Timer, USCI, Interrupt, Memory of MSP430FG4618.
- Knowledge about I2C communication in Universal Serial Communication Interface (USCI) for communication between MSP430FG4618 and the Camera OV7670.

II. DISCUSSION OF THE LAB:

1. Design Specification

1.1. Requirements definition

From the specific demands of taking photos or recording videos immediately, or general demands of helping people; especially older ones and children in their daily activities because of their carelessness: we aims to design a voice-controlled camera, which receives voice command from users, do its tasks (take photos, record videos), save results to the memory.

1.2. System specification

In this product, we use:

- 1 KIT for DSP processing. It'll process the voice command of the user.
- 1 camera for taking pictures and recording videos.
- 1 MSP430FG4618 for testing the setting of Camera OV7670.
- 1 memory device for storing images and videos.
- 1 microphone for recording user's voice.
- A speaker for testing voice commands.

Use-case diagram:

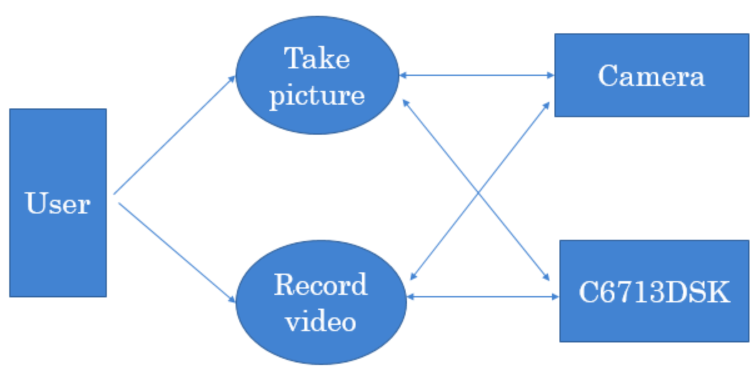


Figure 1: The Use-case diagram

1.3. Functional design

Hereunder is our functional diagram:

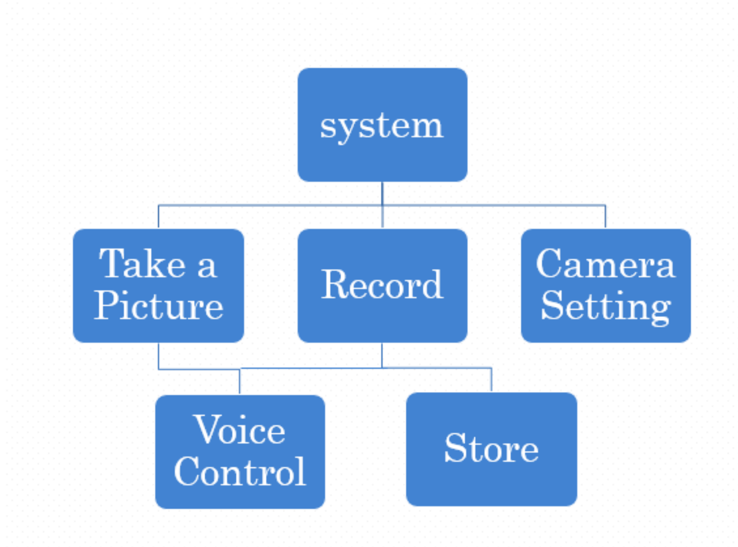


Figure 2: The functional diagram

1.4. Architectural design

Hereunder is our architectural diagram:

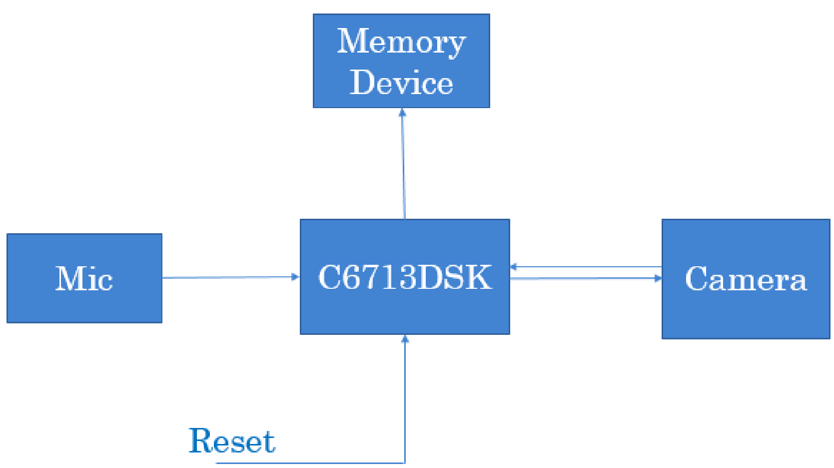


Figure 3: The architectural diagram

1.5. Prototyping

- Test speech recognition: plug one MIC and one speaker into the KIT; talk into the MIC; and test whether or not the commands were taken. The commands were: turn the music ON/OFF, next song, volume UP/DOWN. (You can see these things in the demo video).

- Test Camera's operation: programming for the connection between the testing KIT (MSP430) and the Camera; then checking the registers (OUPUT, TXBUFF_transmitting buffer, RXBUFF_receiving buffer) to know whether the Camera work well or not. (Values of these registers were shown in the figures in the Camera Module Part below).

2. Discussion_Design Implementation

2.1. Software implementation

Filter module

Description: The system uses DIP Switch to set output to DAC. When the input of DIP Switch (The Switch on Kit) equal 1 the output path will go form ADC through LMS (Least Mean Square) filter and to to DAC. The Least Mean Squares (LMS) Algorithm[3] can be used in a range of Digital Signal Processing applications such as echo cancellation and acoustic noise reduction.

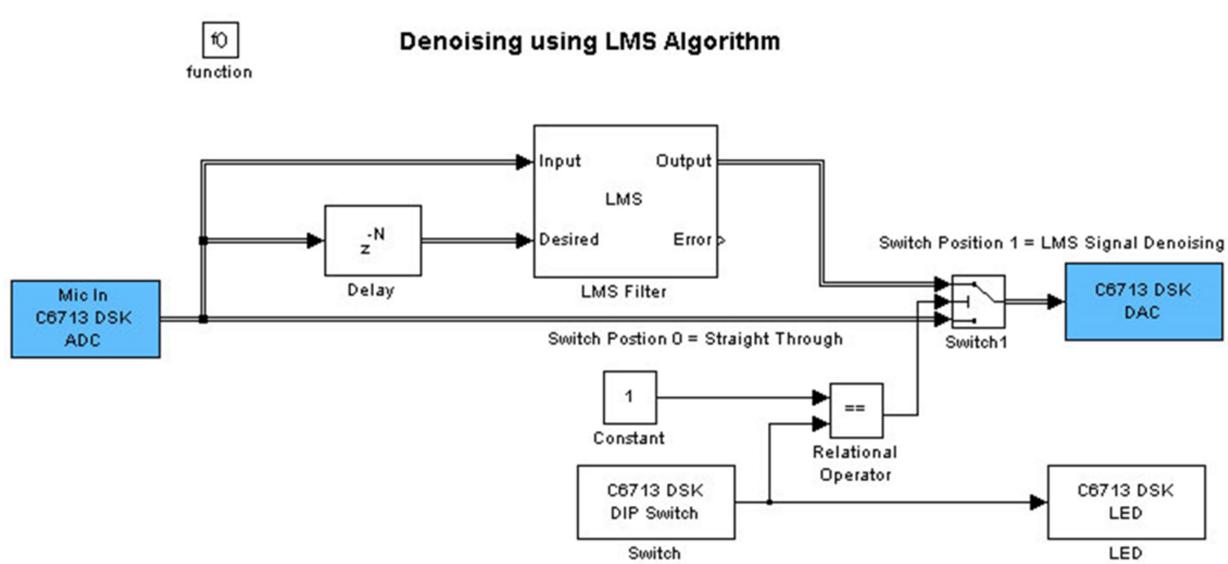


Figure 4: The input filter design scheme

Voice recognition module

In the project, we firstly create a model in Matlab to check all the module working well. After that we try to embed the simulation system on the kit through the Code Generation or Simulink tool of Matlab. This is a bad way because it is very passive for changing anything. The matlab, however, works with its main function of checking the whole system and some small project can be designed very fast with Simulink

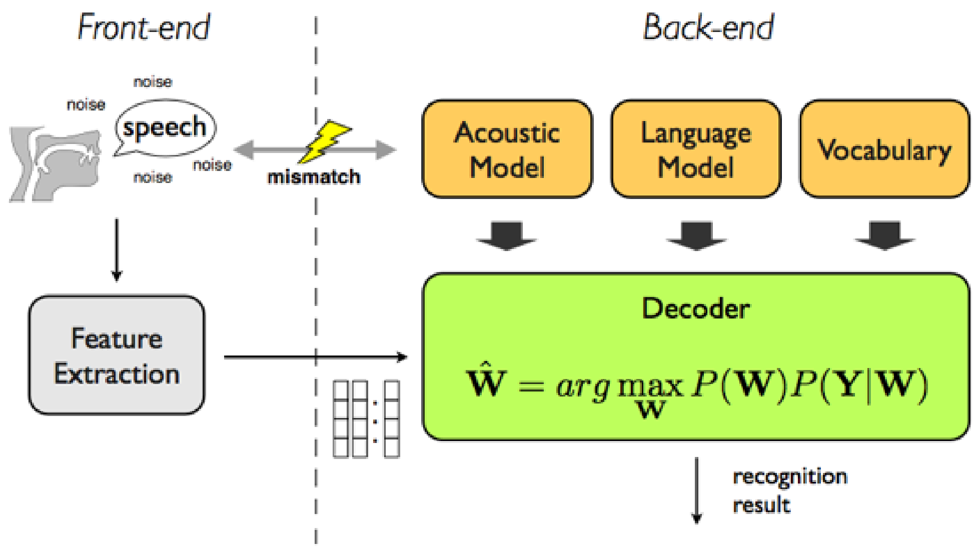


Figure 5: The Genernal Scheme for Speech Recognition

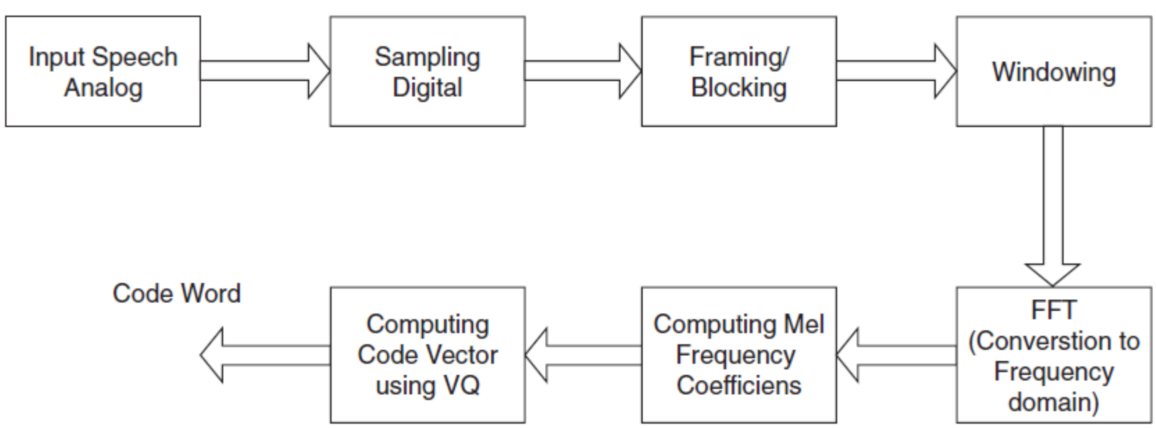


Figure 6: Step by step processing when someone say anything

Lecturer: Nathaniel McVicar Lab Assistants: Nguyen The Nghia
Pham Xuan Trung

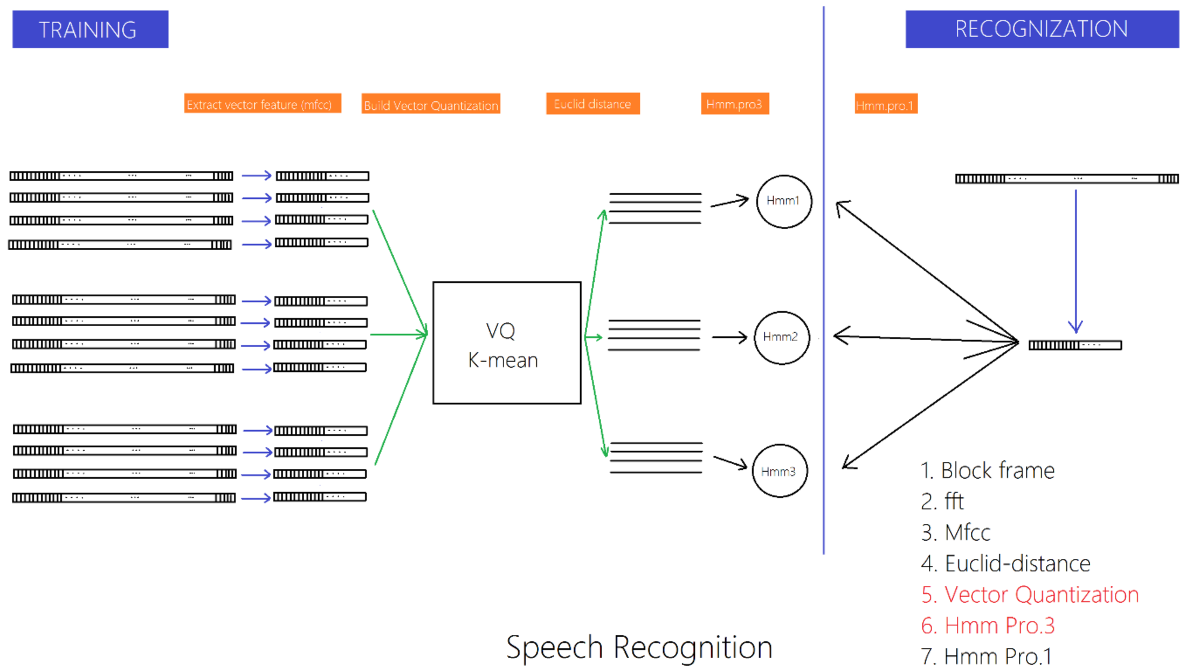


Figure7: Full scheme of speech recognition in vector form

Overview:

Speech recognition includes two processing

1. Training processing
   + Block Frame: Block the speech signal into frames, each consisting of a fixed number of samples. This step to make a motivated noise constant.
     - Hamming window: Window each frame to minimize the signal discontinuities at the beginning and end of the frame
     - Fourier Transform: Use FFT to convert each frame from time to frequency domain.

   + Starting and ending Point detection: this step detect where the voice begin and ending for reduce the number of calculation. This is the soul of the system that make it really fast.

   + DCT (Discrete Cosin Transform)
   + Mel - frequency scale: Convert the resulting spectrum into a
   + Vector quantization (Classification) : This step using K-mean method
   + Making Observation sequence: Using Euclid-distance to create observation sequence from VQ-Kmean and input signal

+ HMM problem 3: Build the system with these words by using Expectation Maximization method. Hmm problem 3 describes how to build the system.

2. Recognition processing
   + Block Frame:
     - Hamming window
     - Fourier Transform
     - Voice detection : used to decide whether the user say a command
   + Starting and ending Point detection
   + DCT (Discrete Cosin Transform)
   + Mel - frequency scale
   + Making Observation sequence
   + HMM problem 1: Build the system with these words using Forward algorithm.

   The HMM problem 1 describes this step.

Note: There are only two different parts in comparing Training and Recognition Processing – (VQ & Hmm problem3) versus (Voice detection & Hmm problem 1). The blue color above show this differences.

The detail explaination

Feature Extraction: This step is to extract the feature of each speech data. For example the feature of "Glass" and other word is different. We need to extract the feature of these words to build a model and to save the memory. We do not need to save whole word to memory, only need to save the model that built from its features.

There are many way to extract the feature of speech such as FBA, LPC, MFCC, and PLP. Because the time is limit so we only understand the LPC and MFCC method. In compare the LPC and MFCC, we choose the MFCC because it is used widly and efficience for speech. The efficient of LPC working on FPGA (other project) can be checked with the YouTube video []

(http://www.youtube.com/watch?v=KfzWQhKrTLM)

There are three smaller parts in this step

Hamming window : Prevent the lost in input data.[page 48- HaThucPhung thesis]

Fast Fourier Transform: transform the input sinal to frequency domain for effective calculations later.

Speech detection: this is the main part of the system that makes it really faster by reducing the number of need caculation. For example, if we do not use this part, the system will record 1s input speech and processing with this 1s input. In the view of the length of word or sentence is different such as "take" or "glass", we know that the take will be shorter than "glass". So the Speech detection will decide when the voice begin and ending. The "take" will only cover 10 frames as an example and glass takes 100 frames. The system in case of "take" will be processed with 10 frame data and it improve the speed. The page 44 in HaThucPhung-thesis presented this scheme very clearly.

Classification: consists of models for all words. This module classifies extracted features according to the individual word have been stored.

The recorded word patterns are used to derive a classification algorithm. Vector quantization (VQ) – K-means[page53 – HaThucPhungThesis] is used. This is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a *cluster* and can be represented by its center, called a *codeword*. The collection of all clusters is a *codebook*. In the training phase, a speaker - specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors. The distance from a vector to the closest codeword of a codebook is called a *VQ distortion*.

The matching method using the **Euclidean method** to calculate the distance between each MFCC and the trained vectors (from VQ) for speaker ID. The minimum distance is the decision for classify input signal.

This module combine with Extract feature vetor of each word to create the Obervation that is input to next step - HMM

$$d(x,y) = \sqrt{\sum_{i=1}^{k} |x_i - y_i|^2}$$

Training phase: The feature vectors are used to create a model for each speaker. During the *testing phase*, when the test feature vector is used, a number will be associated with each speaker model indicating the degree of match with that speaker's model. This is done for a set of feature vectors, and the derived numbers can be used to find a likelihood score for each speaker's model.

The VQ modle combine with Extract feature vector will be used to create an observation for each word. This observation will continue to process with Expactatoin Maximization, the HMM problem 3[page 75 - HaThucPhungThesis] to create a HMM modle. Each word will has its own modle.

In Recognization processing, we only need this result (all modles of all word) to calculate how many percentage input signal match with trained modles to decide what is that word. The problem 1 in HMM help us calculate this percentage[page 67-HaThucPhung thesis].

2.2. Hardware implementation

2.2.1. TMS320C6713 DSK

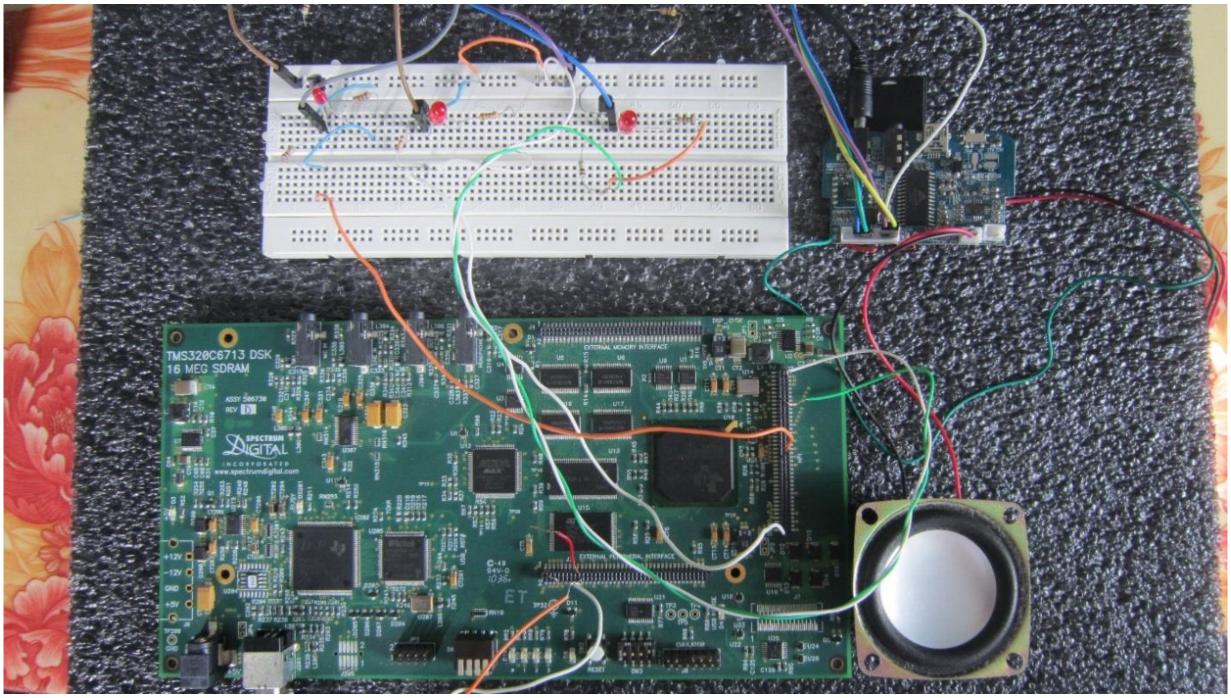
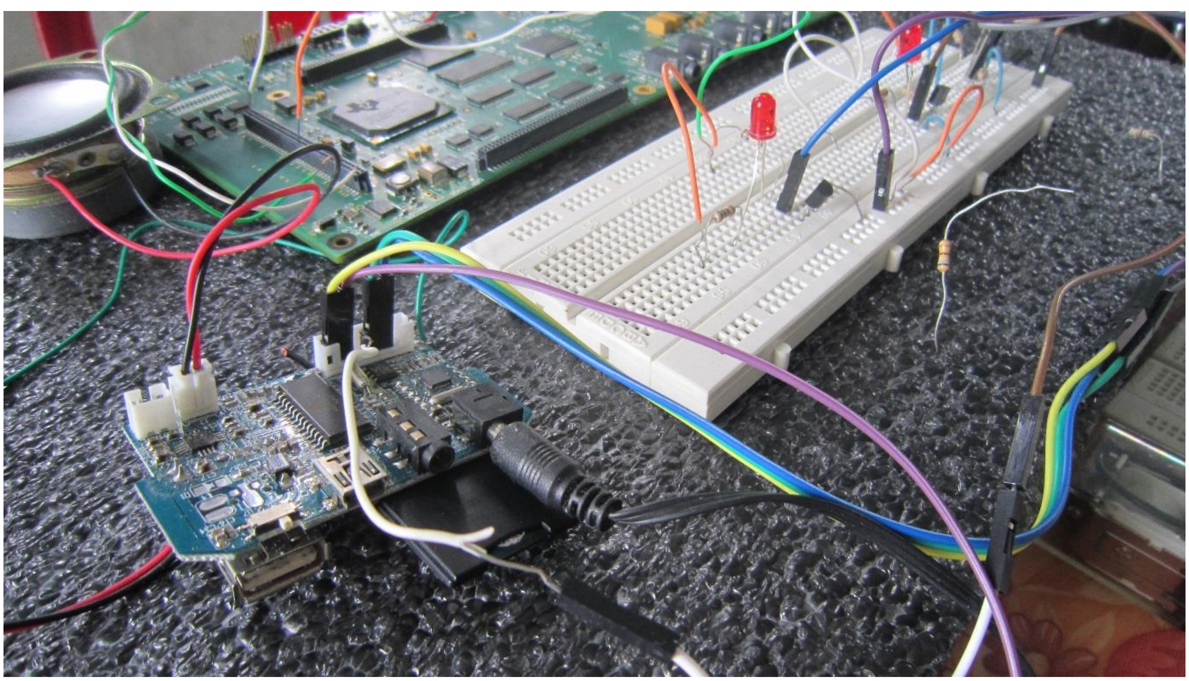


Figure8, 9: Voice control music player



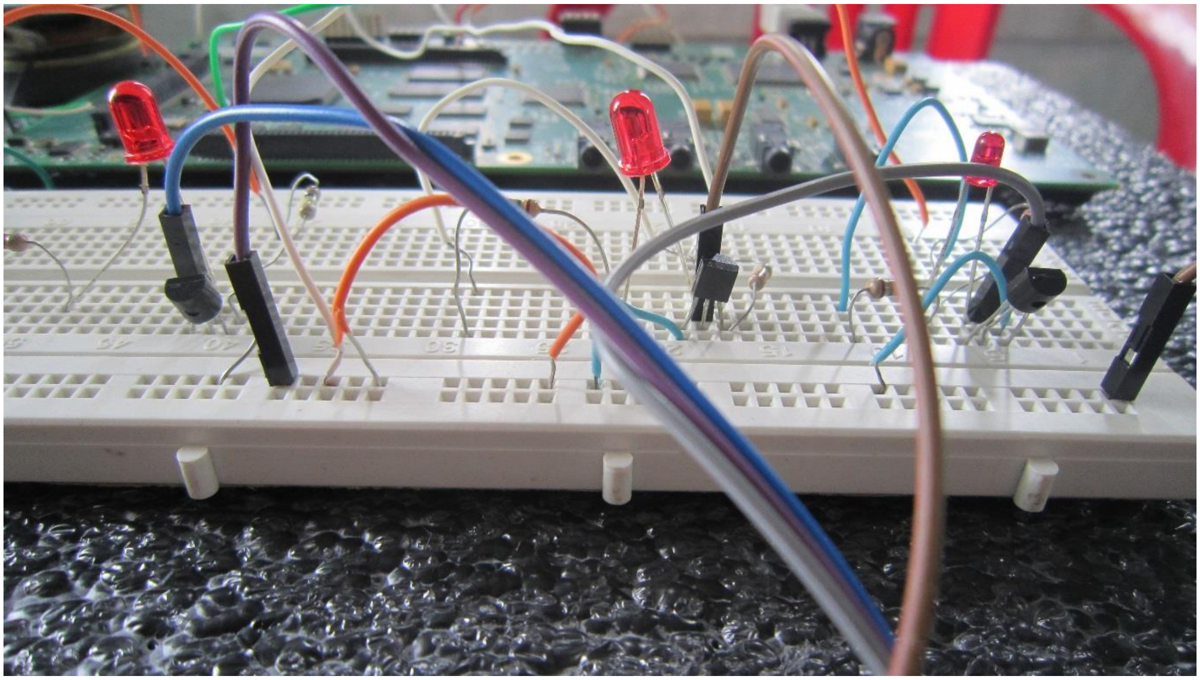


Figure: Control circuit using BJT
The circuit using BJTs (C9014 & two C1815) that are unpluged from old electronic (radio, remote)circuits

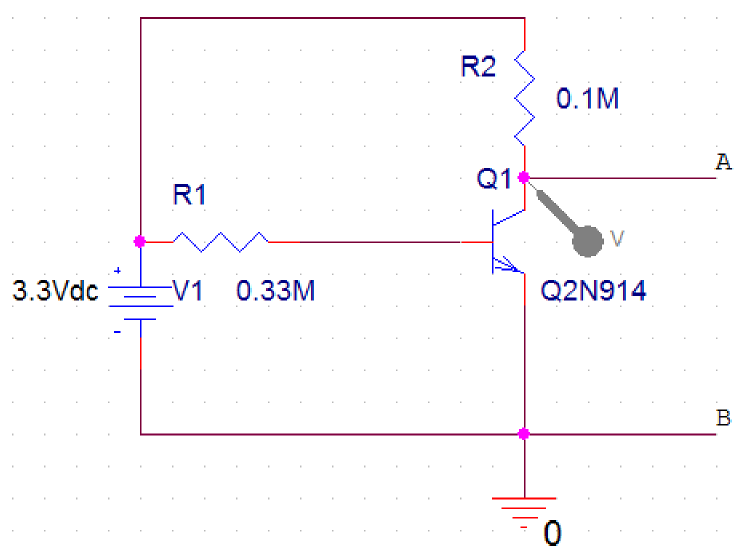


Figure 10: Control Player circuit, for only on function.

Description: The player control circuit receive control signal from DSKC6713. Node A and B is used to control function of player, when the current drain from A to B the according function will present. The music player have 4 functions that can be controlled by voice – next, back, stop, or on

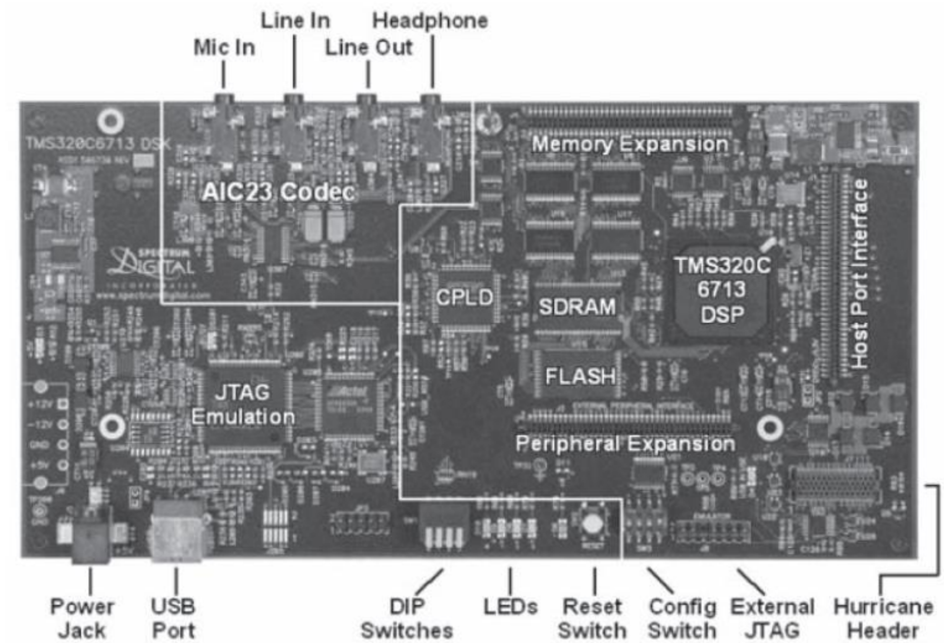


Figure 11: The real view of DSKC6713

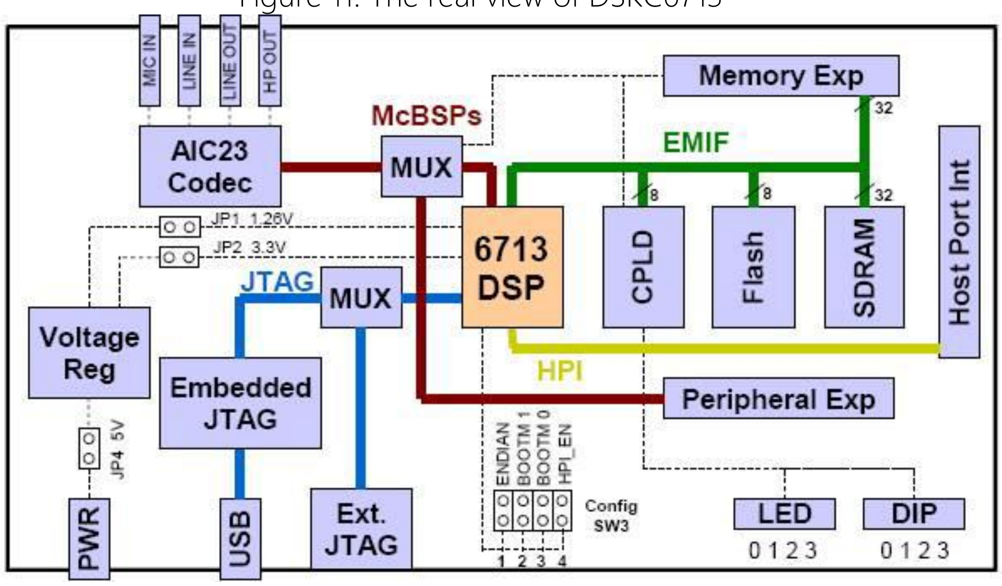


Figure 12: the C6713 DSK block diagram

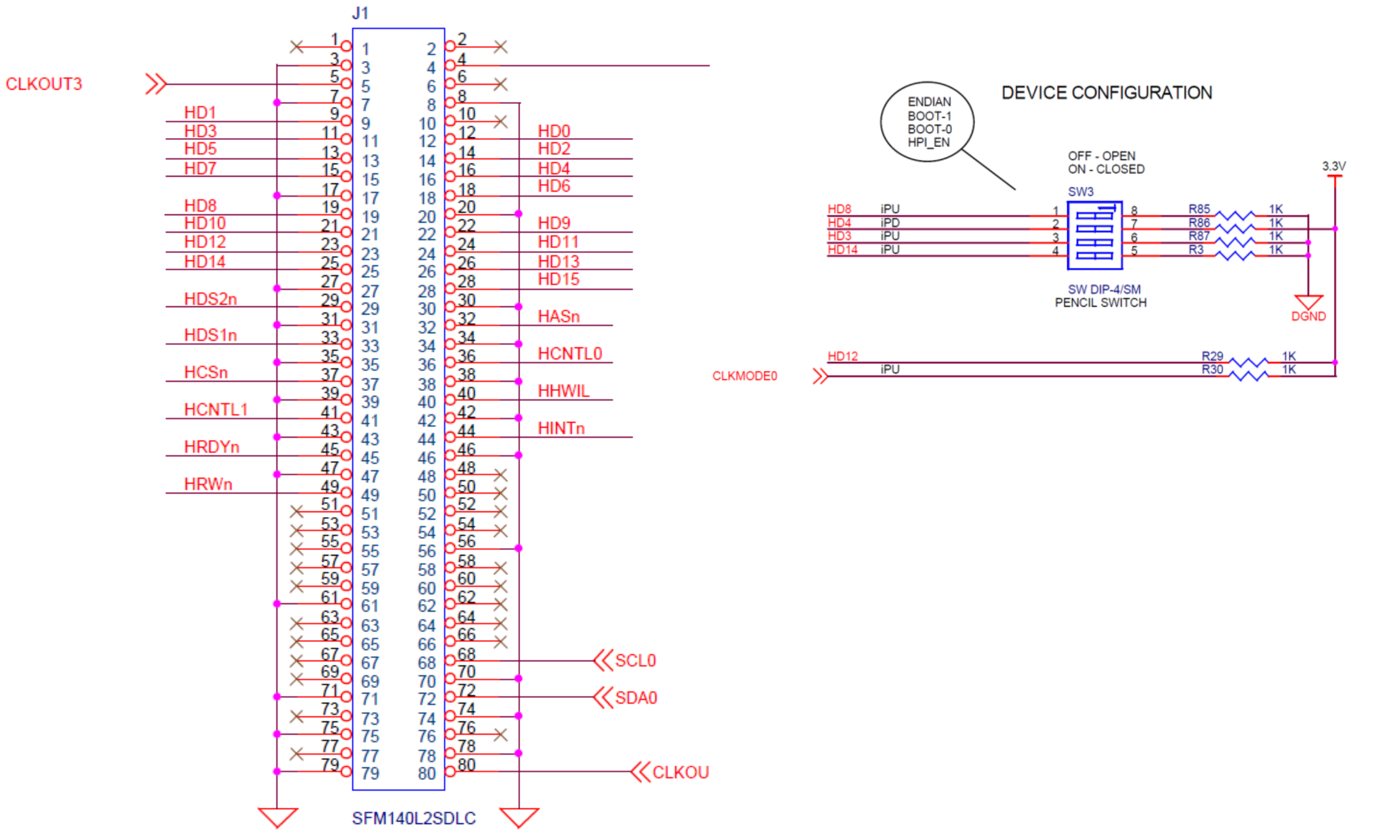


Figure 13: GPIO of DSK C6713

We only use 3 hardware parts on DSK C6713

No	Hardware	Purpose
1	Codec AIC23	Return the input speech
2	Ram (16Mb)	Store the input for processing
3	C6713	Main processor

Note: Turn the swith 3 in J3 on for using the GPIO

2.2.2. OV7670 Camera

OV7670's overview and functional diagram

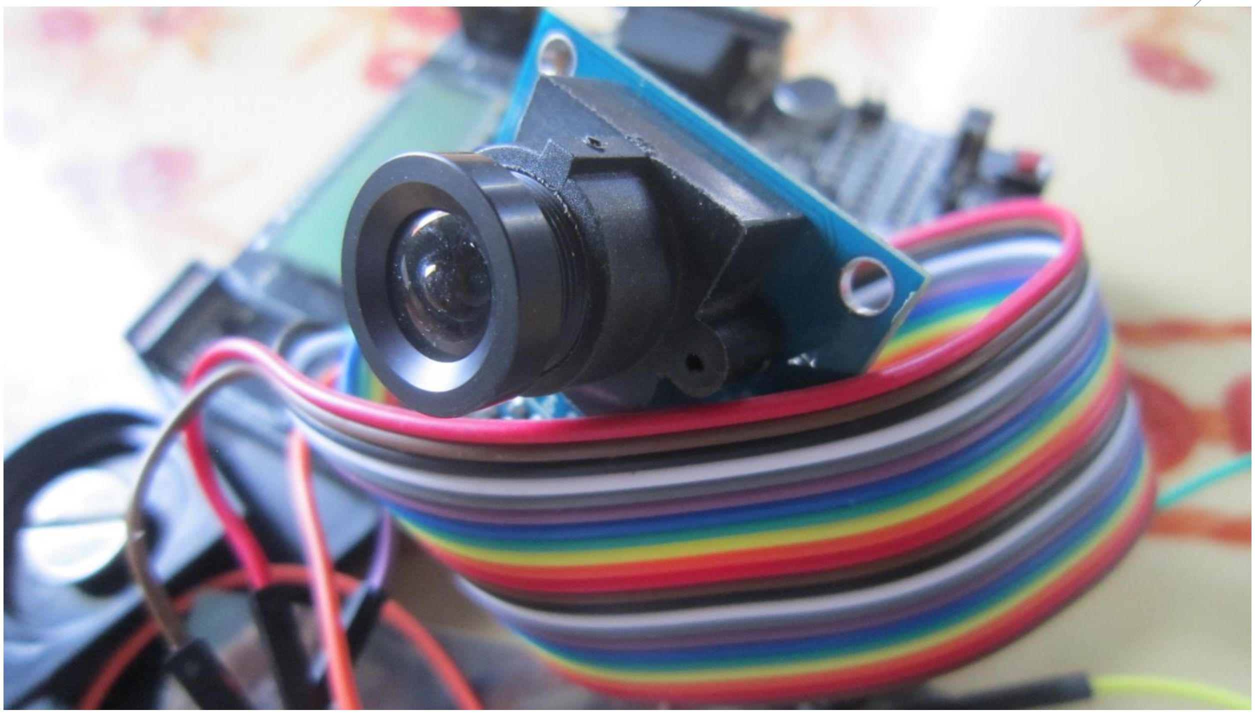


Figure14: OV7670 Camera

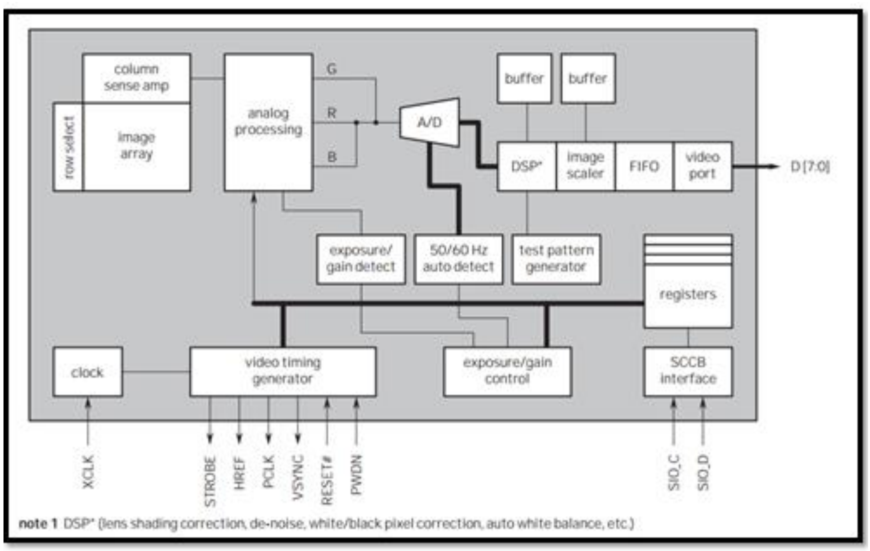


Figure 15: The scheme of Camera module OV7670 AL442 v2 with FIFO

This module adds an AL422 FIFO memory chip, and the module is designed so that the video data can be clocked directly to the FIFO. We can read the FIFO as slowly as we like, and don't need a big memory of our own for the data. This module also provides its own clock.

Hereunder is its schematic, which shows clearly pins inside it for connection purpose later:

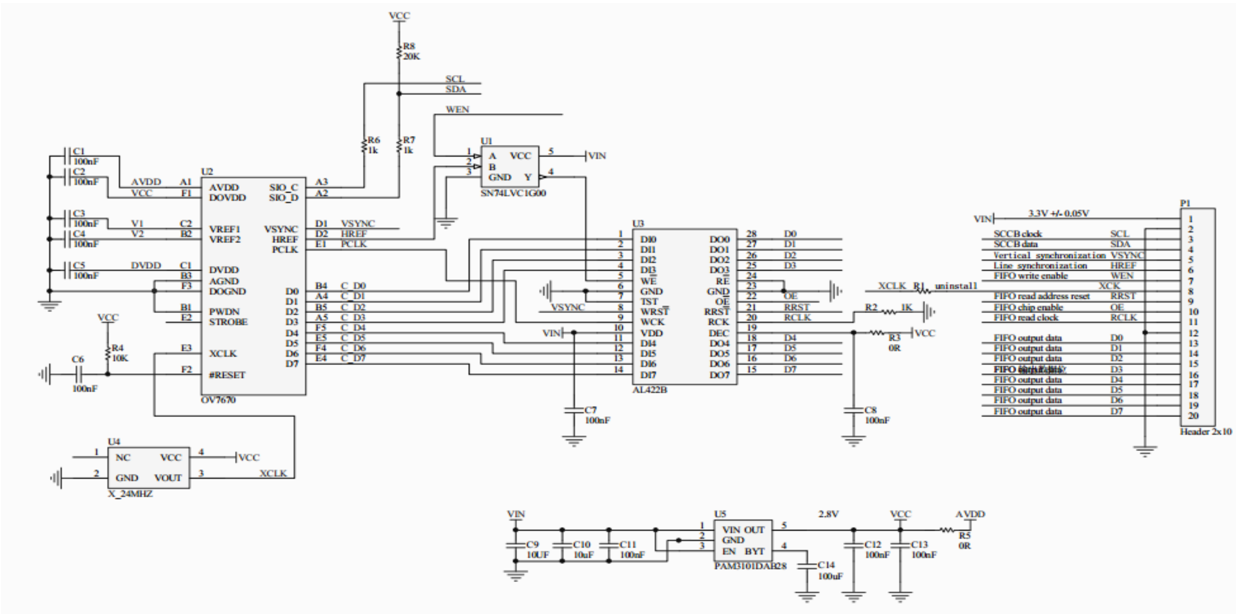


Figure 16: The internal pin of OC7670

At the beginning, we intended to connect directly the camera OV7670 to C6713DSK. After a time of working, we realized that we need to test whether or not the camera works well with the KIT; and it was convenient for team's members to work at home also; we decided to use another MCU.

For a proper application of the camera, my requirements from the MCU are as following:

- i2c support for setting the ov7670 registers
- 16 GPIO pins, 8 of which are for image data.

From the previous course (EE472), we've learnt so deeply about a kind of MCU; that is MSP430FG4618.

⇨ From the reasons above, we came up with using MSP430FG4618 for testing the operation of OV7670

Here-under is the schematic of MSP430. We just show a part of this with red circles for used PINs. (Detail of the full schematic is shown in the reference part at the end of the report).

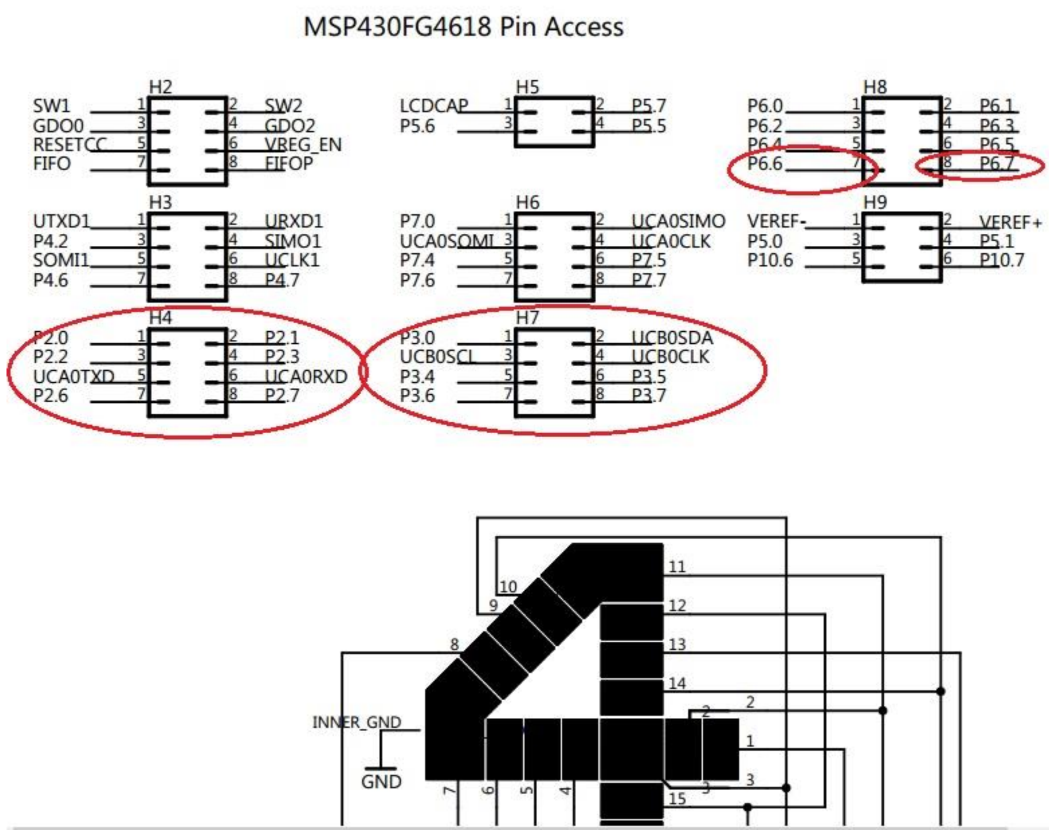


Figure 17: MSP430 scheme

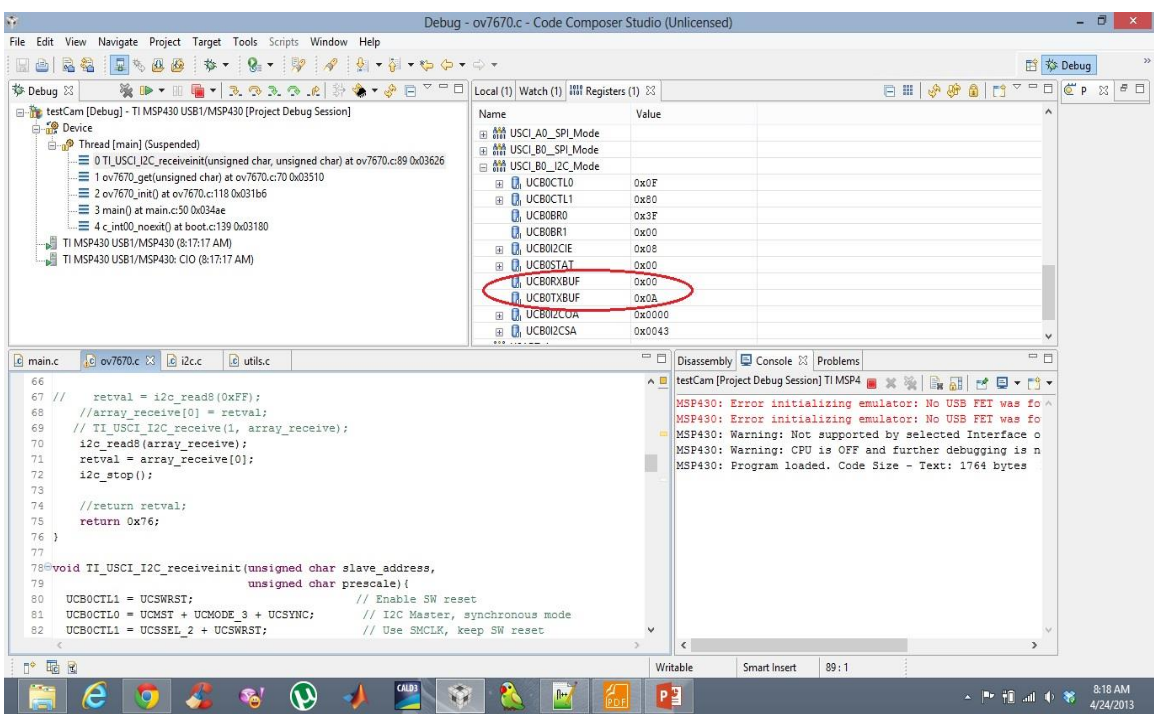
The connection pins between MSP430 and OV7670:

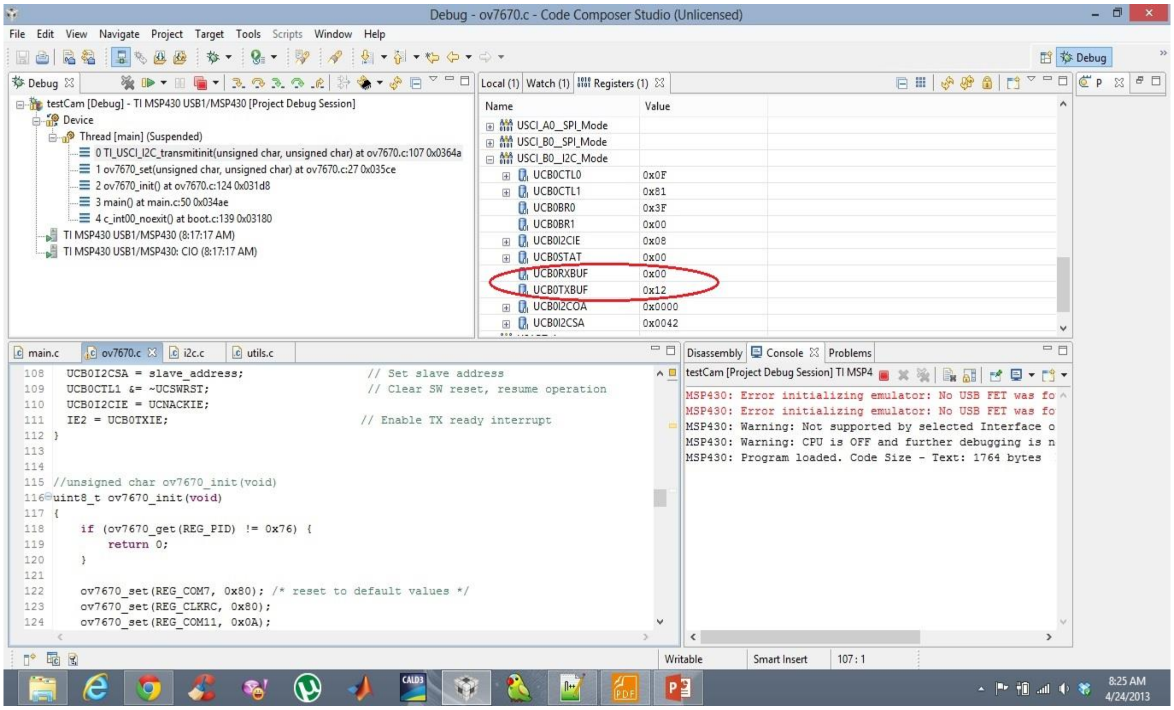
USB TTL adapter		C6713DSK		OV7670 FIFO
		P1.0 (VSYNC)	<-	5(VSYNC)
RXD	<-	P1.1(TXS)		
TXD	->	P1.2(RXD)		
		P1.3(RRST)	->	9(RRST)
		P1.4(RCLK)	->	11(RCLK)

		P1.5(WEN)	->	7(WEN)
		P1.6(SCK)		2(SCCB_C)
		P1.7(DSA)		3(SSCB_D)
		P2.0 – P2.7		13 – 20 (D0 – D07)
		VCC		1(3.3 V)
		GND		2 (GND)
		GND		10(OE)
GND		GND		

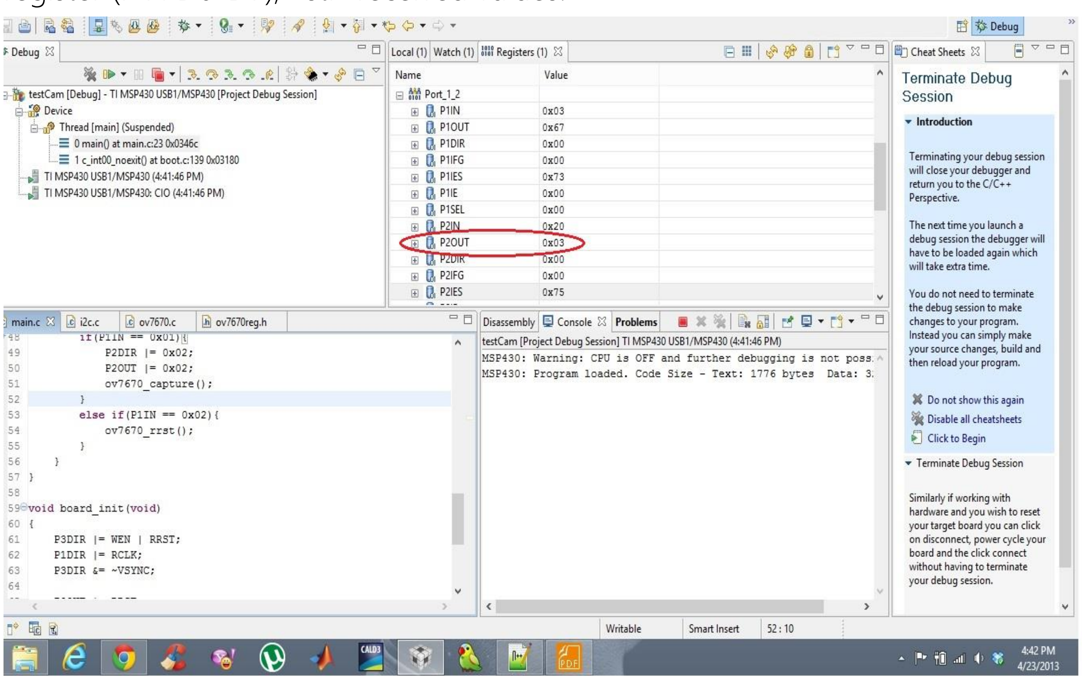
The result we got after programming for the connection between OV7670 and MSP430FG4618 is that:

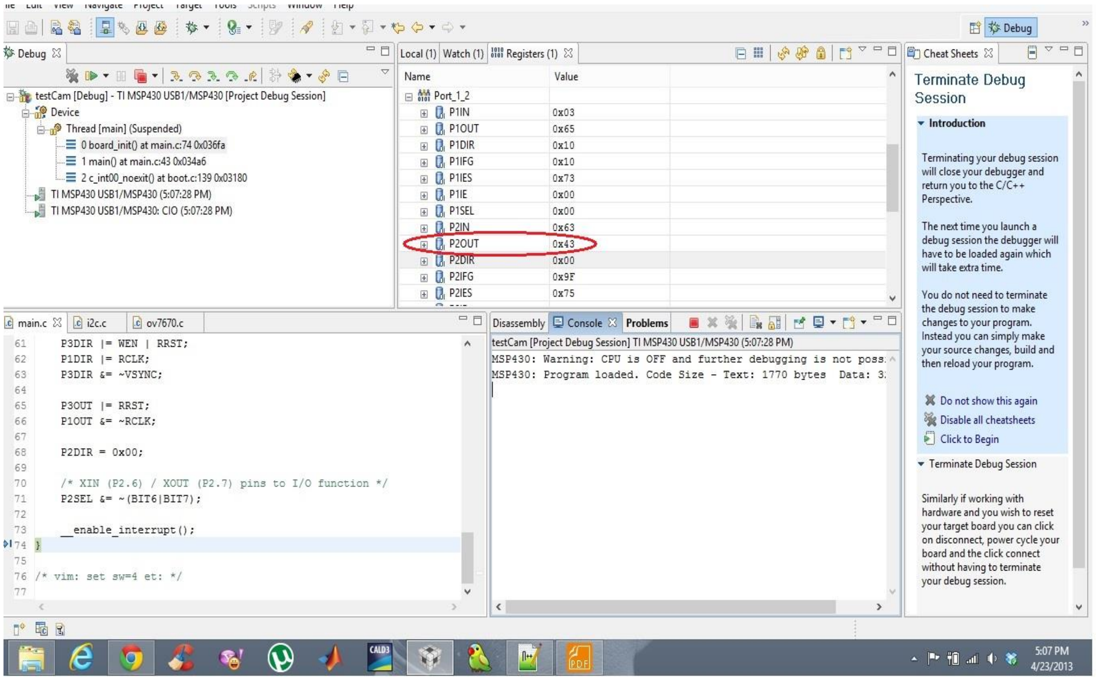
There are some errors in synchronizing between the camera and the KIT (problem in Clock). These errors may come from our code during programming process. It could not come from the setting since we did check the connection very clearly and carefully. And we are on the process to fix these errors. The reason why we know this error is: after running each line by line of code, we checked the registers of MSP430 and realized that there are just signals in transmitting buffer; unfortunately, it's not the case for receiving buffer. Pictures below show content of these registers clearly:

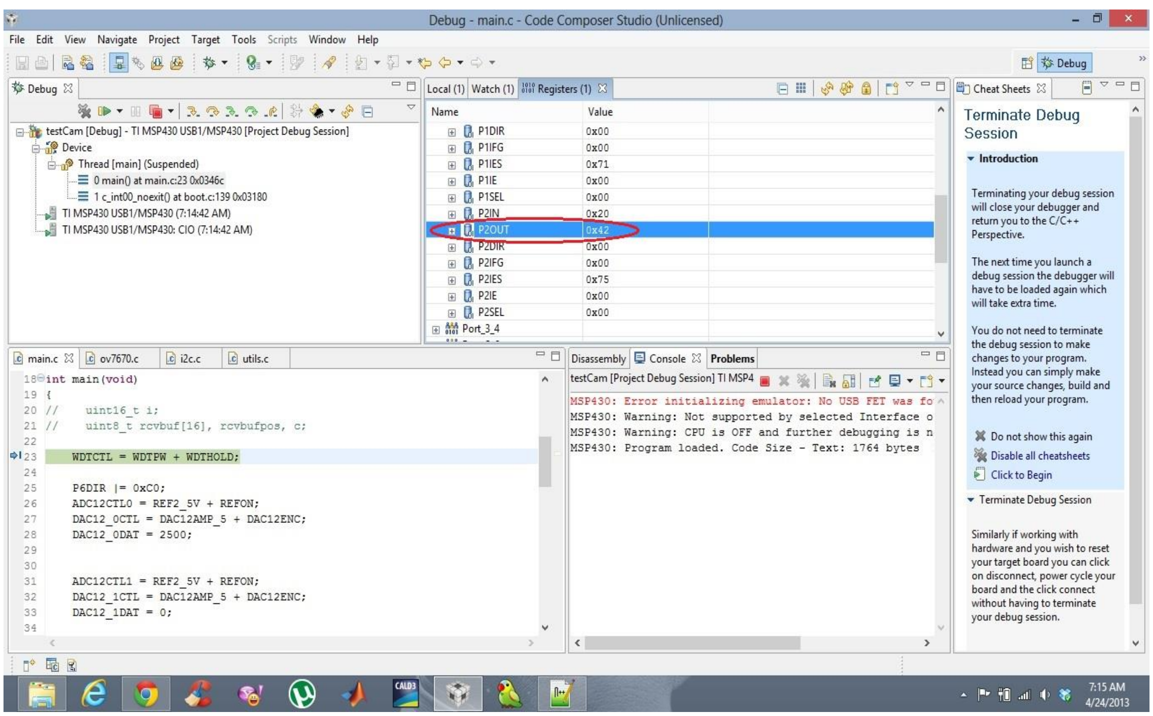




- There is something that is unreasonable in our result. Since with the error I mentioned above, the output, obviously, will be nothing (0x00), but whenever I check the output register (PIN D0-D7), I still received values.







- Obviously, we know that result is wrong, but we still want to show it in the report with the hope of having a discussion or comment to continue on developing our product; since we will not stop after this course.

3. Test Plan

3.1. Test plan

The test plan includes

<table>
<tr><th>No</th><th>Test cases</th><th>Action</th><th>Expected result</th><th>Status</th></tr>
<tr><td>1</td><td>Test the filter input speech design</td><td>You say anything to the microphone that connected to KIT DSKC6713 and listen the output.</td><td>You will listen a better output when press the button on kit. The button restart the filter block.</td><td>Done.</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">Test the Speech recognition</td><td>Say glass</td><td>The led 1 will turn on to wait the next command</td><td rowspan="3">This speech recognition is done but not for camera module.</td></tr>
<tr><td>Say “take a photo of this”</td><td>The camera will turn on and capture the picture and saving to the memory.</td></tr>
<tr><td colspan="2">Repeat the Speech test in the noisy environment for checking the efficiency of the system</td></tr>
<tr><td>3</td><td>Test the music player block</td><td>Say “stop” “next”, and “back”.</td><td>The music will response to user command</td><td>Done with average 85% true recognition. The circuit control is 100% true if SR give a</td></tr>
</table>

				control signal

3.2. Test Specification

Speak faster to check whether the system recognize your voice.

Checking with other people (different voice) to see how adaptive ability the system is.

4. Encountered Problem Discussion

4.1. Presentation and analysis of the result

- The system used matlab for training processing so we should make the DSK C6713 very compatible with the matlab. The problem is the gain of Codec and the gain of the laptop. While the Code return very big value, hunderd in silence and thousand in voice, then the Sound card and Matlab return value of average 0.002 in noiseor 0.01 in speech. So we create a gain for this and when changing the gain, it affect to quality of Speech Recognition(SR)
- The main different when implement the system on kit is that C-languge is constrained memory. The over-range problem is occurred often if we are not good ad control index of array. The system will turn on unstable at this points.
- The system on KIT when good organized will better than on Maltan. This is because in Malab, every thing is a complete module, so using it is very but slow because we cannot touch it inside. In C-language, we have to build (or download and modify) by hand so when it is good organized, it will faster than in Matlab.
- The system is very easy to build by Matlab because it's huge library. In contrast, the DSK6713 is ten year ago, the tutorial, application notes, or ebook for this kit working on Window 8 and new version CSSversion4 is rarely. So it's take a quite of time to study and make it working.

Camera module: only transmiter buffer has value, the camera does not return any value

The filter is working well

The speech recognition control music player is good

The simulation matlab code can detect 80% of words "glass" "take" "video" "photo" "record "and "picture"

The most difficult word is "picture"

The designed filter is good for reducing the noise

4.2. Analysis of any errors

Filter Block

<u>**Description**</u>**:** The filter working well with reduce the noise of environment but the user's voice is also compresses so much

<u>**Explain**</u>**:** The Least Mean Square filter is an adaptive filter that not

<u>**Solving method**</u>**:** - Try to design another filter and check

- We should only use the noise filter blo in the very noisy environment.

- Try the noise cancelation method.

Speech detection Block

Voice detection :The SR system will record noise signal of the first ten frames. The next ten frames 10 to 20 will be used in Speech detection block that decides whether the command is begin. For this reason, if you speech so fast after each command, your voice will be record as noise, the system will misunderstand the noise signal. Because your voice is large, so for this case the system will seem to die. You need to scream out over the normal to reset the system.

Sentence recognization: The sentence recognization is build base on word recognition. The SR system record whole command and block it in to word, so just recognize word by word. In some time, our system recognize very good but not all case.

Standalone System: The system now is not stand-alone system, we are not write the data to flash memory of DSKC6713. According to TI assistance, CCSv4 we are using is not compatible with flashburn. This project is quite difficult so we do not have enough time for study this part. We now rewrite data from laptop before using and after that we can bring it anywhere provided of power continuing.

Dynamic training. The system need to has ability of training a new command when user want it control something new.

Quality of SR: Some word has a high percentage than other, this can be explained below:

- The "glass" and "take" are two words that have a high recognition percentage. The reason is that "glass" and "take" are one vocal in comparing with two vocal such as "video" or "photo". So the processing would be easier and litter noise than other.
- The most difficult recognized word is "picture". The reason can be stay on record process because if we use the recorded data (that use to train the system) it is be able to recognize 100%.
- Another reason is that the database has only one person voice, hence its result can be accepted when recognizing 7 to 8 right times per 10 times.

Solving Methods:

- Training with more people to get the additive database.

Camera Block

Description: The camera module is not working for kit MSP430FG4618/F2013

Explain: When using oliscopter to check whether

Solving Methods:

- Try the camera module with another available kits – Launch pad, DSKC6713

4.3. Project Timeline

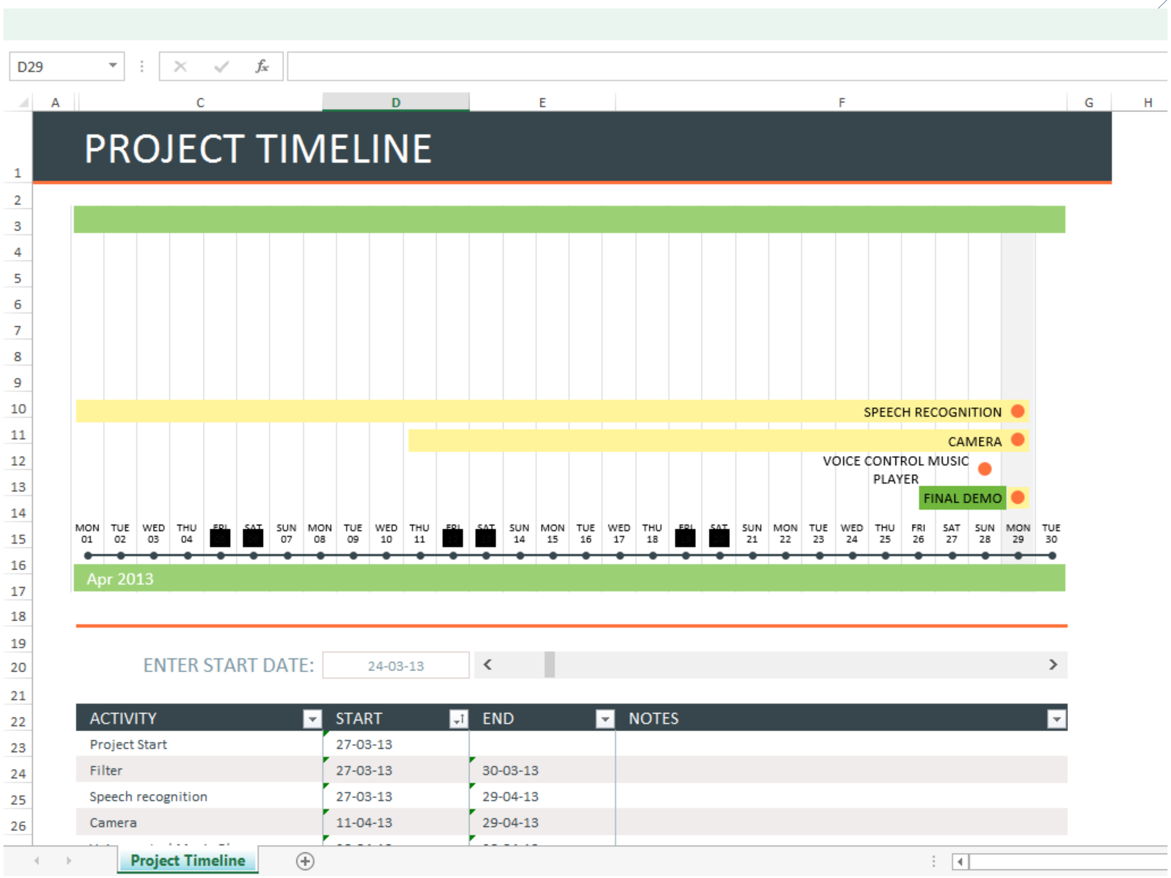


Figure 20: The project timeline

5. Future development

No	Description	Detail
1	Receive and show message	Imaging drive a car and message appear on the screen
2	Alarm setting	Use phone to search internet (3G)
3	Search information from internet	Using wifi or 3G of a cell phone
4	Translator	Record and translate real-time voice
5	Image processing	Take a picture and translate the language on that pictuer to familiar language

6	Reply message, email	

III. SUMMARY AND CONCLUSION

In summary, we've just represented to you detail of our voice-controlled camera, including design specification, design implementation, testing cases, as well as some discussion about this camera, our timeline for making this product, and the promising trend in the future that we aim to develop.

Although these short words and a few images cannot express fully our product, we hope that it somewhat conveys our ideal to you, helps you understand what we are doing in creating a product that is really helpful for people in daily life.

Through this project, we gained so many lessons about searching information, learning new knowledge about new kind of KIT and embedded product; especially practicing soft skills, team working in making a whole project.

IV. REFERENCES

The main knowledge to implement the project comes from

HaThucPhung Thesis

and the Kit tutorial comes from

[Rulph_Chassaing,_Donald_Reay]_Digital_Signal_Proc(Bookos.org)

The HaThucPhung present speech system recognization very complete with a lot of examples for understand.The Rulph_Chassaing give us a base to create and using Chip C6713 and Bord support library.

[1] NgoTranDucThang, "Tutorial how to create a project on CCSv4 to work with DSKC6713", http://www.youtube.com/watch?v=gPfbg_G39Xo

[3] *hayes1996_statisticaldigitalSignalProcessingandModeling.pdf*

[4] Morgan2008_TheTheoryofLinearPrediction.pdf

[5] http://www.youtube.com/watch?v=KfzWQhKrTLM

[6] Deepali Y Loni, "DSP Based Speech Operated Home Appliances Using Zero Crossing Features"

[7] http://mirlab.org/jang/matlab/toolbox/asr/ (matlab library tool)

[8] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html(matlab lybrary)

[9] http://www.mathworks.com/help/simulink/ug/creating-custom-block-libraries-with-matlab-function-blocks.html#bso5yl6(create your own block lybrary from function)

[10] http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm_download.html (HMMall)

Reference link for camera

1. "MSP430x4xx Family.pdf" http://www.ti.com/lit/ug/slau056l/slau056l.pdf
2. "MSP430FG4618/F2013 Experimenter's Board.pdf" http://www.ti.com/lit/ug/slau213a/slau213a.pdf
3. "OV7670 datasheet.pdf"

http://www.haoyuelectronics.com/Attachment/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29/OV7670%20datasheet%28V1.4%29.pdf

4. "OV7670 Implementation Guide.pdf"

http://www.haoyuelectronics.com/Attachment/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29/OV7670%20Implementation%20Guide%20%28V1.0%29.pdf

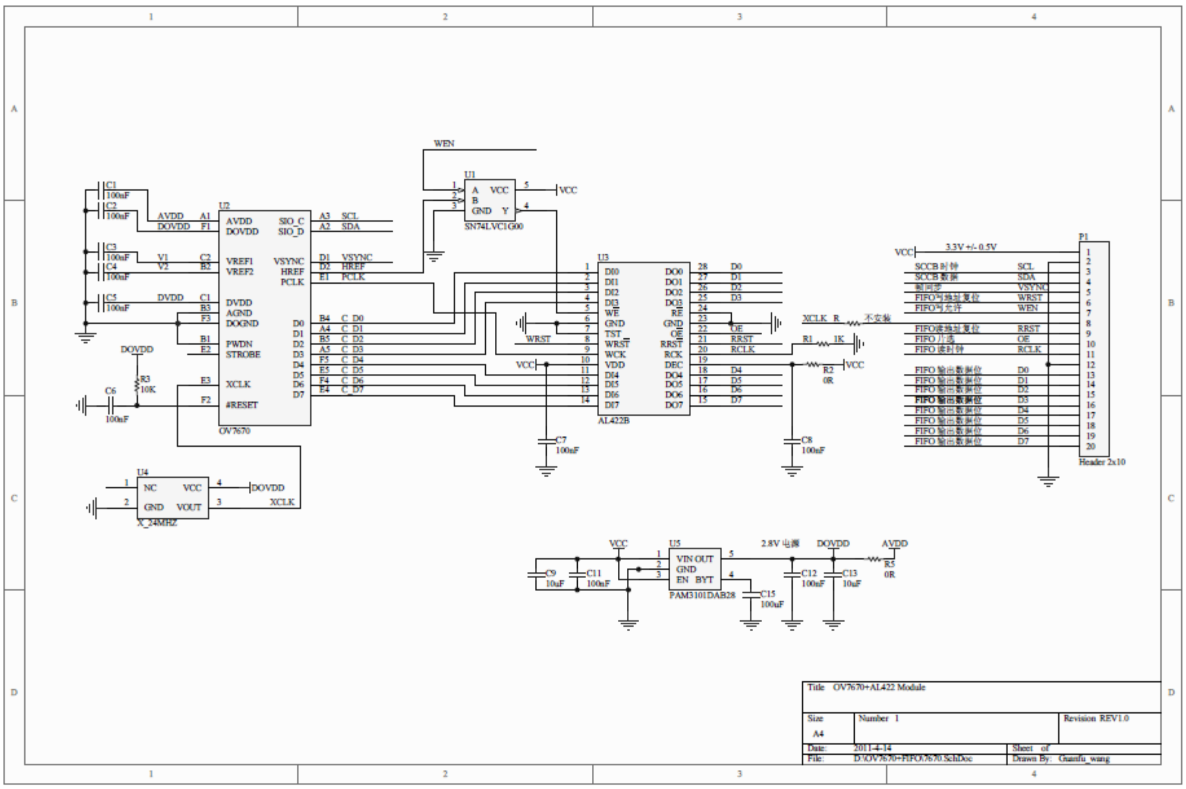
5. "OV7670 + AL422B(FIFO) Camera Module(V2.0) PCB PIN.pdf"

http://www.haoyuelectronics.com/Attachment/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29%20PCB%20PIN.pdf

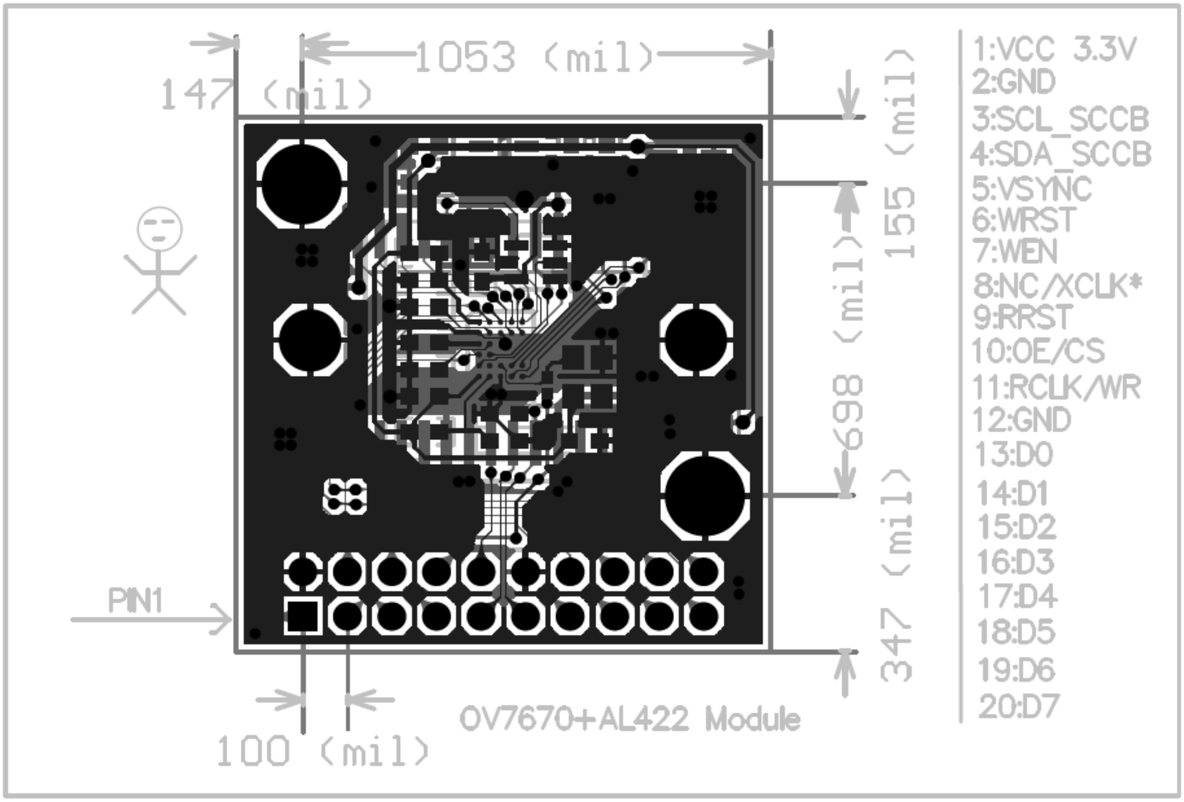
6. "OV7670 + AL422B(FIFO) Camera Module(V2.0) Sch6ematic.pdf"

http://www.haoyuelectronics.com/Attachment/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29/OV7670%20+%20AL422B%28FIFO%29%20Camera%20Module%28V2.0%29%20Schematic.pdf

IV. APPENDIX



THe camera scheematic


1053 (mil)
147 (mil)
155 (mil)
698 (mil)
347 (mil)
100 (mil)
PIN1
OV7670+AL422 Module
1:VCC 3.3V
2:GND
3:SCL_SCCB
4:SDA_SCCB
5:VSYNC
6:WRST
7:WEN
8:NC/XCLK*
9:RRST
10:OE/CS
11:RCLK/WR
12:GND
13:D0
14:D1
15:D2
16:D3
17:D4
18:D5
19:D6
20:D7

The PCB Size