A Real-Time Embedded Speech Recognition System

(Sang-Yep Nam[1], Chun-Woo Lee[2], Sang-Won Lee[3], In-Jung Park[3])

[1]Kyung Moon College [2] Wooding Information College, [3]Dankook Univ
TEL: +82-11-662-3556 TEL: +82-42-629-6252 TEL:+82-41-550-3544
r13337@unitel.co.kr, cwlee@ci.woosonginfo.ac.kr digitallab@kornet.net

Abstract

According to the growth of communication biz, embedded market rapidly developing in domestic and overseas. Embedded system can be used in various way such as wire and wireless communication equipment or information products. There are lots of developing performance applying speech recognition to embedded system, for instance, PDA, PCS, CDMA-2000 or IMT-2000. This study implement minimum memory of speech recognition engine and DB for apply real time embedded system. The implement measure of speech recognition equipment to fit on embedded system is like following. At first, DC element is removed from Input voice and then a compensation of high frequency was achieved by pre-emphasis with coefficients value, 0.97 and constitute division data as same size as 256 sample by lapped shift method. Through by Levinson - Durbin Algorithm, these data can get linear predictive coefficient and again, using Cepstrum - Transformer attain feature vectors. During HMM training, We used Baum-Welch reestimation Algorithm for each words training and can get the recognition result from executed likelihood method on each words. The used speech data is using 40 speech command data and 10 digits extracted form each 15 of male and female speaker spoken menu control command of Embedded system. Since, in many times, ARM CPU is adopted in embedded system, it's performed porting the speech recognition engine on ARM core evaluation board. And do the recognition test with select set 1 and set 3 parameter that has good recognition rate on commander and no digit after the several tests using by 5 proposal recognition parameter sets. The recognition engine of recognition rate shows 95%, speech commander recognizer shows 96% and digits recognizer shows 94%.

1. Introduction

After study early stage of speech recognition, the study of S/R up to far is mainly relate with well performed recognition in environment without any noise or large size S/R however, in recently, its getting turn out to an embedded system which is close to real life.

The menu dialing system of the embedded system is convenient when it is used during driving a car, working in the office or under the situation that we are all using our eye and hands both while moving by operating menu with only speech. The system memory limitation has never been considered in PC or any other electronic products. However, in the embedded system case, S/R or speech DB size is much influenced on system memory since system memory limitation. This paper intent to implement of compact speech recognition system with the most small size memory and using for well performing embedded system. Previous S/R is the speech recognition algorithm which is mainly involved in DTW(Dynamic Time Warping) that could be recognised by comparing reference pattern and ANN(Artificial Neural Network) based on brain neruo cell and HMM(Hidden Markov Model) applied probability process on speech recognition. Among of these, especially, HMM adaptable for large size S/R has been studied. Due to all of these facts, the S/R in personal computer is well controlled by user without any inconvenience and applied to using in various parts. However, the development of new technology is getting put more weigh on compact system and PCS , PDA and Embedded part made up primarily and the needs about S/R getting increase for the access hi-technology. According this, in this study, the points of S/R focused on small-size, compact system for satisfy these kinds of needs.

This paper is made up the order with, explanation on liner predictive analysis and cepstrum conversion which is mainly used in S/R completed and then HMM model, experiment and results.

2. Coefficient extraction of Speech feature and Recognition Algorithm

2.1 Speech feature extraction

Collected Audio data is completed with pre-emphasis for high-pass compensation using by filter shown in formula 2.1.

$$H(z) = 1 - az^{-1} \quad 0.9 \le a \le 1.0 \qquad \text{(F 2.1)}$$

Also, Audio data has been separated with standard size using Hamming window for the sake of Speech analysis. In this time, reiteration window using for decrease speech data loss among of frame and the half of frame size being used

$$w(n) = 0.54 - 0.46\cos\frac{2\pi n}{N-1} \qquad \text{(F 2.2)}$$

$$(0 \le n \le N-1) N\text{: frame size}$$

For pursuing the feature vector used in recognition, the Audio data separated with N size can get autocorrelation coefficient by using formula 2.3 above all.

$$\gamma_i(m) = \sum_{n=0}^{N-1-m} s_i(m)s_i(n+m) \qquad \text{(F 2.3)}$$

$$m = 0, 1, \cdot\cdot\cdot, p \quad (p\text{: order})$$

The autocorrelation coefficient searched from the above is used to find liner predictive coefficient, by using Levinson-Durbin Algorithm.

For using this study, LPC(liner predictive coefficient)can be converted to Cepstrum which is stronger and reliable feature than LPC by using formula 2.4.

$$c_0 = \ln\sigma^2 \qquad \text{(F 2.4)}$$

$$c_m = a_m + \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \qquad 1 \le m \le p$$

$$c_m = \sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \qquad m > p$$

2.2. HMM Model

1) Training

At first, training process is the one to control the model parameter in the behalf of getting the maximum probability of observation sequence of word. Consequently, by applying the repeated Baum-Welch algorithm at regular time like fig 1 possible to get optimized model parameter. Extracted from the above, Speech feature vector is using while the HMM training and recognition process

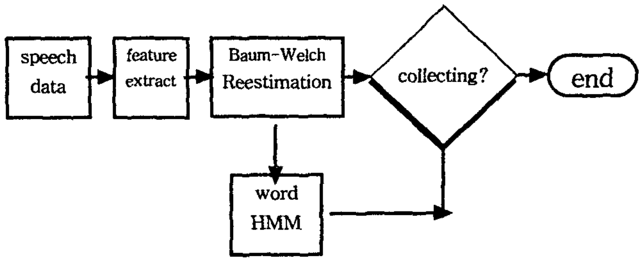

Figure 1. Block Diagram for Training

2) Recognition

The feature vectors of speech is using same feature getting from above, through out the likelihood calculation about the each word can choose the maximum word.

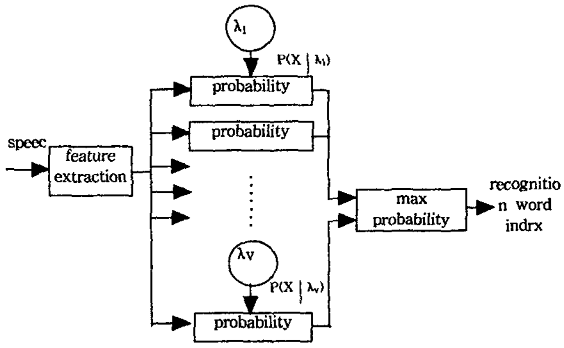

Figure 2. Block Diagram for Recognition

3. Implementation of Embedded system

The recognition system used in this study is given in Fig.3. The mic applied input tool using hand-free ear phone mic which is popular with people recently, for the process tool using KS32C50300 applied ARM7TDMI, For output tool use LCD display which shows speech recognition result.

The speech data obtained from the CODEC is stored in a memory on evaluation board. The above speech process method obtain the feature vector, the word index recognized by speech recognition program using HMM set that stored in flash transfer the word store map and finally output recognized word in LCD screen. The follow is the system overview used in here.

3.1 The feature extraction.

The speech data inputted with 8KHZ, 16 bit size by using the basic No like table 1 get a feature vector can be applied in HMM input.

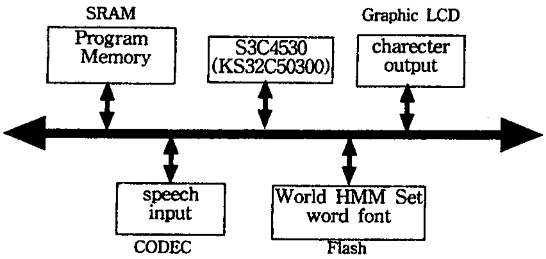

Figure 3. Embedded Speech Recognition System

Table 1. Analysis Conditions of Speech Data

Speech parameters	Contents
Sampling frequency	8 KHz
Resolution	16 Bit
Pre-emphasis	0.97
Hamming Window Frame rate (N)	256
Coefficient Predictive	LPC, Cepstrum
Order(p)	13 order
The number of Word	40 speech command/10digits

3.2. The construction of speech data base.

The speech data for recognition collected from the male and female 15 people each, mainly recorded in outer environment.

Word list used in recognition experiment can acquire speech data by using 40 speech command data and 10 digits for instance, main function demands for Embedded system operation and control word.

Recording is done for twice speak 40words during two days, its due to compensate changing each persons voice according to time change.

4. Experiment and Discussion

4.1 Speech recognition program.

In this study, speech recognition system implemented based on word unit speech recognition, also applied to evaluation board using ARM processor in direct way. Table 2 & 3 shows the detail total DB size and program.

In this study, the program implemented without any optimization which is not putting inline assemble code into necessary place. This result only shows the S/R parts, using C language in hardware program.

The recognition test with select set 1 and set 3 parameter that has good recognition rate on commander and no digit after the several tests using by Table 4 & 5 shows 5 proposal recognition parameter sets & results for optimized speech rates..

Table 2. Memory requirement for Speech Recognition program

Section	Memory Size(Kbyte)
code size	180
constant /initial value	4
wide variable	227
Heap / Stack	720
Code book	200
total	1,331

Table 3. DB size requirements from Standard Vocabulary Selection

System	DB size	Code Size	RAM Size
SI(40word)	150 Kbyte		
Digit(10word)	20 Kbyte		
SD(20word)	20 Kbyte		
Constant Data	10 Kbyte		
Total	200 Kbyte	180 Kbyte	Less than 50Kbyte

Table 4. Parameter Set

T-kit	SET1	SET2	SET3	SET4	SET5
BegDetEnable	1	1	1	1	1
BegDetAbsThreshold	20000	18000	18000	20000	22000
BegDetRelThreshold	1560	1560	1560	1560	1560
EndDetEnable	1	1	1	1	1
EndDetTmin	30	20	20	30	30
EndDetCounter	10	8	5	10	5
Tmax	90	60	60	90	60
Nutt	3	3	3	3	3
RejTolerance	0	1	0	1	0
ConfTolerance	0	1	0	1	0

4.2 Speech Recognition

The most words recognition rate shows more than 90%, The table 4 shows overall result.

The speech recognition result about word speech can earn as follows.

$$Recognition\ rate\ =\ \frac{recognized\ word}{total\ \ vmber\ of\ tests}\quad (F4.1)$$

Among the recognition result 4 words shows below 90%, 4 words has 100% and the rest shows the result above 90%..

5. Conclusion.

In this paper, a compact speech recognizer was designed and implementation for the embedded system like PCS phone & PDA.

Not concerning optimization and recognition speed at all, training necessary words and applied practically shows only 4 words recognition rate drop down remarkably.

ARM CPU is adopted in embedded system, it's performed porting the speech recognition engine on ARM core evaluation board. And do the recognition test with select set 1 and set 3 parameter that has good recognition rate on commander and no digit after the several tests using by 5 proposal recognition parameter sets.

Table 5. Parameter Sets Results for Optimized Speech Rates

Name	Version		Recognition Lab	Recognition Rate /BOS error	Parameter
SI Command	with MIC		LAB 6	96.79	Set 1
	OOV command		LAB 6	0.000	Set 1
SI Digits	with MIC		LAB 6	94.44	Set 3
	Alternative Digits		LAB 6	0.000	Set 3
SD Command	Environment	office	SD-Digit 4	98.65	Set 4
		outside	SD-Digit 8	93.66	Set 4
	OOV	office	SD-Digit 12	0.000	Set 4
		outside	SD-Digit 16	0.29	Set 4
SI Remote Command	remote speaker command		LAB 25	94.55	Set 1
	OOV		LAB 25	0.000	Set 1
SI Remote Digits	remote speaker digits		LAB 25	90.53	Set 3
	Alternative Digits		LAB 25	0.000	Set 3
SD Car Command	CAR speaker		SD-4	91.99	Set 4
	OOV Car speaker		SD-4	0.000	Set 4

The recognition engine of recognition rate shows 95%, speech commander recognizer shows 96% and digits recognizer shows 94%.

This is caused by outer noise while speech DB producing affect to recording speech. In addition, overall, no function of the speech signal processing in speech input, can't get the same result with the other normal PC since there was not compensation on optimized. Therefore, expect to made up stronger recognition if added contents not considered above.

Table 6. Word list & recognition rates.

word	R/rates	word	R/rates
Cu-so-lok	[redacted]	Nim-s$N-sa-s$-ham	95.0
nal-s$i	97.5	sI-K$i-CUl	90.0
Cin-toN	100.0	kEi-san-ki	[redacted]
mai-n$-mo-tD	95.0	Su-sin-ham	90.0
m$i-mo	95.0	Nin-T$-n$is	90.0
bi-sang-c$n-hoa	91.4	Nal-lam	97.5
Nap-Nt-lo	97.5	coN-lO	90.0
hal-Nil	95.0	hoak-Nin	100.0
Zaik-kai-Pi	95.0	s$n-Tak	95.0
cu-sik	94.7	Nilk-ki	95.0
Ni-m$i-Nil	95.0	nU-si	91.4
pel-so-ri	90.0	sa-C$n	100.0
ta-Nim	97.5	tIN-lok	90.0
kO-ToN-C$N-po	[redacted]	Zui-so	95.0
m$i-nU	95.0	C$n-hoa	91.4
Nim-s$N-sa-se-ham	95.0	Ci-um	97.5
NEI	100.0	Na-ni-No	95.0
young	94.0	kong	91.0
e-il	91.0	rhee	91.0
sam	97.0	sa	97.0
oh	88.57	rhok	90.0
chil	91.0	pal	94.0
gu	91.18	won-wi-chi	94.0

References

[1] K. Takagi, R. Oguro, K. Hashimoto, and K. Ozek:, "Performance Evaluation of word phrase and Noun Category Language Models for Broadcast News speech Recognition," *proc, ICSLP98*, pp 2507-2510, DEC, 1998.

[2] C.S.Ramalingam, Y.Gong, L.P.Netsch, W.W. Anderson, J.J.Godfrey, and Y.H.Kao. "Speaker Dependent Name Dialing in a Car Environment with Out f Vocabulary Rejection", *in Proceeding of ICASSP'99.* vol3 pp 1780-1783, May. 1999.

[3] D. van Compemolle, "Speech Recognition in the Car From Phone Dialing to Car Navigation," *in Proceedings of EUROSPEECH'97.* vol. 5, pp2431-2434, Sep.1997.

[4] Lawrence Rabiner, Biing-Hwang Juang, *FUNDAMENTALS OF SPEECH RECOGNITION*, Prentice-Hall International, Inc., 1993.

[5] Allen Gersho, Robert M.Gray, *"Vector Quantization and Signal Compression"*, Kluwer Academic Publisher, 1992

[6] BEN GOLD, NELSON MORGAN, *"SPEECH AND AUDIO SIGNAL PROCESSING"*, JOHN WILEY & SONS. INC, 2000.

[7] L. R. Rabiner and B. H. Juang, *Fundamentals of speech recognition*, Prentice-Hall Inc., 1993.

[8] J. Mariani, "Recent advances in speech processing," *Proc. of ICASSP*, pp. 429-440, 1989.

[9] Sang-Yep Nam, Chun-Woo Lee, In-Jung Park, Kang-Hyeon Rhee "Numeral Speech Recognition by Integration Both Acoustic and Visual Information". ICEIC 2000 Proceeding of The 2000, ICEIC , Liaoning University, China, pp125-128, 2000.

[10] Sang-Yep Nam, Sang-Won Lee, and In-Jung Park, "A Study on Embedded Speech Recognizer for Mobile Phone using HMM". Proceeding of ITC-CSCC '01, Tokushima, Japan, Vol.II, pp857-860, July10-12, 2001.