독립성분분석을 이용한 DSP 기반의 화자 독립 음성 인식 시스템의 구현

김창근* · 박진영* · 박정원* · 이광석** · 허강인*

Implementation of Speaker Independent Speech Recognition System Using Independent Component Analysis based on DSP

Chang-Keun Kim*, Jin-Young Park*, Jung-Won Park*, Kwang-Seok Lee**, Kang-In Hur*

본 논문은 2002년도 정보통신부 IT학과 장비지원사업의 동아대학교 대응자금에 의해 연구되었음.

요 약

본 논문에서는 범용 디지털 신호처리기를 이용한 잡음환경에 강인한 실시간 화자 독립 음성인식 시스템을 구현하였다. 구현된 시스템은 TI사의 범용 부동소수점 디지털 신호처리기인 TMS320C32를 이용하였고, 실시간 음성 입력을 위한 음성 CODEC과 외부 인터페이스를 확장하여 인식결과를 출력하도록 구성하였다. 실시간 음성 인식기에 사용한 음성특징 파라메터는 일반적으로 사용되어 지는 MFCC(Mel Frequency Cepstral Coefficient)대신 독립성분분석을 통해 MFCC의 특징 공간을 변화시킨 파라메터를 사용하여 외부잡음 환경에 강인한 특성을 지니도록 하였다. 두 가지 특징 파라메터에 대해 잡음 환경에서의 인식실험 결과, 독립성분분석에 의한 특징 파라메터의 인식 성능이 MFCC보다 우수함을 확인 할 수 있었다.

ABSTRACT

In this paper, we implemented real-time speaker independent speech recognizer that is robust in noise environment using DSP(Digital Signal Processor). Implemented system is composed of TMS320C32 that is floating-point DSP of Texas Instrument Inc. and CODEC for real-time speech input. Speech feature parameter of the speech recognizer used robust feature parameter in noise environment that is transformed feature space of MFCC(mel frequency cepstral coefficient) using ICA(Independent Component Analysis) on behalf of MFCC. In recognition result in noise environment, we knew that recognition performance of ICA feature parameter is superior than that of MFCC.

키워드

Speaker Independent Speech Recognition, MFCC, HMM, ICA, DSP, Parameter

Ⅰ. 서 론

음성은 인간의 가장 자연스럽고 편리한 의사소통 수단이며 이러한 음성 데이터를 이용한 Man Machine Interface에 대한 연구가 활발히 진행되고 있다. 그러나 음성 데이터는 화자간의 영향이

* 동아대학교 전자공학과
** 진주산업대학교 전자공학과
접수일자 : 2003. 11. 28

나 잡음환경에서의 특징변화가 상당히 심하기 때문에 이에 따른 특징 파라메터 추출에 대한 다양한 방법이 제시되고 있다.[1][2][3][4]

본 논문에서는 음성인식에 주로 사용되는 음성 특징 파라메터인 MFCC(Mel-Frequency Cepstral Coefficient) 외에 ICA(Independent Component Analysis)를 통해 MFCC의 특징 공간을 변화시킨 파라메터를 제안하고 음성인식 알고리즘으로 많이 사용되는 HMM(Hidden Markov Model)을 이용하여 각각에 대한 인식 성능을 비교 분석하였다.[2][3] 또한 실험결과를 바탕으로 범용 DSP를 이용한 음성 인식기를 구현하였다.

본 논문은 Ⅱ장에서는 ICA를 이용한 음성 특징 파라메터 추출에 관해 설명한다. 그리고, Ⅲ장에는 본 논문의 제안된 파라메터를 이용한 음성인식 시스템 구현에 사용된 범용 DSP에 대해 간략히 설명하고 Ⅳ장에서 기존의 특징 파라메터와 제안한 파라메터와의 성능비교 실험 및 결과에 대해 설명하고 Ⅴ장에서는 결론 및 향후 과제로 구성하였다.

Ⅱ. 독립성분분석에 의한 파라메터 재구성

독립성분분석은 식(1)에서 관측신호 $x(t)$를 확률적으로 독립인 신호 $y(t)$의 선형결합으로 생성된 것이라 가정하고 독립신호간의 통계적인 의존성이 최소가 되도록 가중치 W를 추정하여 식(2)을 통해 관측신호에 대한 독립신호를 구할 수 있다.

$$x(t) = Ay(t) \tag{1}$$

$$y(t) = Wx(t) \quad W = A^{-1} \tag{2}$$

독립신호간의 통계적 의존성은 상호정보량(Mutual Information)으로 정의되는데 이는 추정된 독립신호들의 결합엔트로피와 각각의 엔트로피의 차로 계산을 하거나 Kullback-leibler 발산정리에 의해 다음과 같이 정의된다.[2][3]

$$J(W) = \int p(y)\log\frac{p(y)}{\prod_{i=1}^{N} p_i(y_i)}dy$$

$$= -H(y(t)) + \sum_{i=0}^{N} H(y_i(t)) \tag{3}$$

상호정보량 $J(W)$는 항상 양수 값을 가지며 $y(t)$의 각 신호들이 독립적으로 분리되었을 때는 0의 값을 가진다. 따라서 독립신호간의 상호정보량 $J(W)$를 최소화하는 W를 학습에 의해 추정하여 식(2)에 의해 독립신호 $y(t)$를 얻을 수 있다.

학습방법은 식(4)과 같은 자연 감소법(natural gradient)을 사용하여 가중벡터 W를 추정하였다.

$$\Delta W = -\eta\frac{\partial J(W)}{\partial W}[W^T W]$$

$$= \eta[I - \varphi(y(t))y(t)^T]W \tag{4}$$

여기서, $\varphi(y(t)) = [\varphi_1(y_1(t)) \cdots \varphi_N(y_N(t))]$, I는 단위행렬, 그리고 $\varphi_i(y_i(t))$는 식(5)와 같이 나타낸다.

$$\varphi_i(y_i(t)) = -\frac{\partial \log(y_i(t))}{\partial y_i(t)} \tag{5}$$

본 연구에서는 위의 자연 감소법과 상호정보량의 최소화법을 이용하여 추정된 W에 의해 얻어진 독립신호 $y(t)$를 ICA에 의한 음성 특징 파라메터로 사용하였다.

Ⅲ. 범용 DSP를 이용한 시스템 구현

(1) 시스템 구성

본 논문에서 구현한 음성인식 시스템은 그림 1과 같이 세 부분으로 구분할 수 있다.

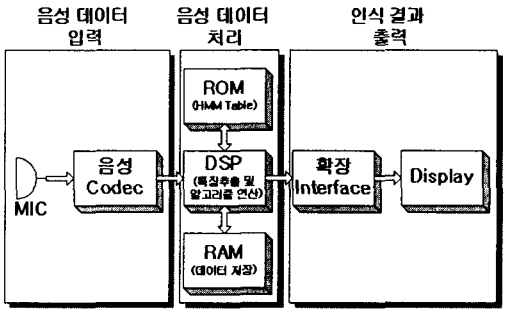


그림 1. 시스템 구성

화자의 발성을 실시간으로 입력 받기 위한 음성 데이터 입력 부분과 받아들인 음성 데이터에 대한 특징을 추출하고 음성인식 알고리즘을 수행하는 부분, 인식결과를 출력하는 부분으로 나뉘어진다.

표 1. Training Kit Spec

	TI TMS320C32 50Mhz 144 PQFP Package
	32KByte
	사용가능
	사용가능
	2채널사용가능, 우선순위지정가능
	1
	2
	Level, edge/level triggered Interrupt Vector Table 이동지정 가능
	5V

구현된 시스템은 TI사의 범용 DSP인 TMS-320C32를 사용한 ㈜ADITEC의 SITK-C32를 이용하였다. SITK-C32 Kit의 구성은 표 1과 같다.[10]

(2) 음성 데이터의 입력

실시간 음성 데이터 입력을 위해 PHILIPS사의 Dynamic Microphone SBC MD680을 사용하고 마이크의 출력은 음성 CODEC의 입력으로 사용된다. 본 시스템에서 사용하는 음성 CODEC은 Crystal사의 단일칩으로 구성된 Stereo Audio CODEC으로 8bit의 해상도를 갖도록 설정하고, 표본 주파수는 8㎑로 고정했다. 입력된 음성 신호를 A/D 변환하여 음성 데이터를 구성하고 시리얼 모드 4의 전송 방식을 이용해서 TMS320C32의 시리얼 포트에 전송한다.

(3) 음성 데이터의 처리

CODEC으로부터 입력되는 음성 데이터를 이용한 음성인식 알고리즘은 그림 2의 흐름으로 수행된다.

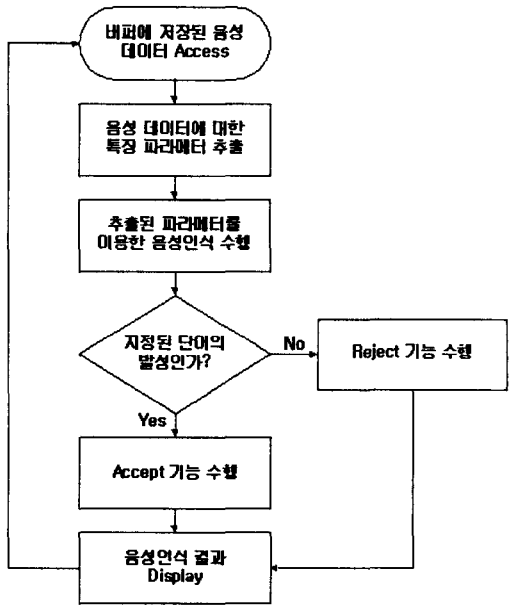


그림 2. 음성 데이터의 처리

입력된 음성 데이터는 TMS320C32의 DMA를 이용하여 내부 메모리의 순환 버퍼에 저장된다. DSP에서는 버퍼에 저장된 음성 데이터를 읽어 들여 음성에 대한 특징 파라메터로 MFCC를 추출한다. 추출된 MFCC를 앞서 언급한 ICA를 이용하여 파라메터를 재추정한 다음 ROM에 저장된 HMM Table을 참조하여 음성인식 알고리즘을 수행한다.

(4) 음성인식 결과 출력

음성인식의 결과는 SITK-C32에서 외부 인터페이스 확장을 위해 사용된 8055를 이용하여 7-Segment를 통해 출력된다. 확장된 인터페이스

에는 기본적으로 7-Segment가 연결되어 있고, 다른 종류의 표시 장치와 범용 컨트롤러의 확장 연결이 가능하다. 본 시스템에서는 7-Segment에 화자의 발성이 미리 정의한 단어와 일치하는 결과라면 해당 단어를 표시하고, 그렇지 않다면 거절기능을 수행한 다음 오류메시지를 표시한다. 해당 단어의 인식결과에 대한 출력 데이터는 ROM에 저장되어 있다.

(5) 음성인식 시스템 적용 예

본 논문에서 구현한 음성인식 시스템은 독립적인 음성인식 시스템으로의 활용가능성이 다양하다. 그림 3은 위의 음성인식 시스템을 모형자동차에 응용한 것이다.

그림 3. RC Car의 구성

모형자동차 구성을 간략히 설명하면 SITK-C32 Kit의 확장 인터페이스에 Atmel사의 마이크로 컨트롤러인 AVR 90S2313을 확장연결하고 인식 결과 데이터를 BIM418 무선 모듈을 통해 송신한다. 모형자동차는 BIM418을 통해 수신 받은 데이터를 AVR 90S8515을 이용해 분석하고 해당명령의 동작을 수행한다.

IV. 실험 및 결과

(1) 음성 데이터의 분석조건

본 논문에서 사용되는 HMM의 학습과정을 위한 실험 데이터로는 일상생활 잡음환경에서 남성화자 50명을 대상으로 PC를 사용하여 녹음하였다. 화자 당 5개의 단어(전진, 후진, 좌회전, 우회전, 정지)를 10번씩 발성한 데이터를 표 2의 분석조건을 이용하여 음성 특징을 추출하였다.

표 2. 분석조건

	8kHz, 16bit
	Hamming
	32ms (256 samples)
	7.5ms (60 samples)
	10 order MFCC
	ICA (Convert MFCC)

(2) HMM을 이용한 음성 데이터 학습

수집된 음성 데이터에 대한 특징 추출 과정을 통해서 우선적으로 음성인식에 주로 사용되는 MFCC를 구하고 이 MFCC를 제안된 알고리즘인 ICA를 이용하여 파라메터를 추출하였다. 또한, 두 가지의 특징을 각각 HMM을 이용해서 학습을 수행하였다. 학습과정은 PC를 이용하였고, 이 학습과정을 통하여 DSP에 사용하기 위한 MFCC를 이용한 HMM Table과 ICA를 이용한 HMM Table을 생성하였다.

(3) DSP를 이용한 음성인식 실험

4.2절의 학습과정을 통해서 구한 HMM학습 Table과 HMM 인식 알고리즘을 본 논문에서 사용한 DSP Kit의 프로그램 메모리(64KByte)에 저장하고 음성 인식 실험을 수행하였다.

표 3. 인식결과

	인 식 률
	89.33%
	96.25%

인식 실험은 화자 30명을 대상으로 5개의 단어를 5번씩 발성하였고 화자 독립의 음성 인식 시스템 구현이 목적이므로 인식 대상은 모두 학습 미 참여자로 구성하였다. 인식실험 결과는 표3과 같다.

V. 결론 및 향후과제

본 논문에서는 MFCC 파라메터를 ICA를 통해 변환한 파라메터를 제안하고 실생활의 잡음환경에서 음성데이터를 수집하여 음성 인식을 수행하였다. 화자독립 음성인식 시스템 구현을 위해 범용 DSP의 프로그램 메모리에 HMM 알고리즘을 저장하고 실시간 음성 인식 실험을 수행했다. 표 3의 인식결과와 같이 제안된 파라메터의 인식 성능이 기존의 MFCC 파라메터보다 약 7%의 성능 향상을 확인할 수 있었다. 비록 현장에서의 실험적인 결과이기는 하지만 ICA 알고리즘의 잡음에 강인한 특징을 확인하는 작업으로는 충분한 결과라고 사료된다.

향후 과제는 더 많은 음성 데이터와 다양한 환경에서의 인식 실험을 통해 신뢰성 있는 시스템을 구현할 계획이다. 또한, 본 논문의 시스템 구현에 사용된 알고리즘의 고찰을 통해 범용 DSP에 최적화 할 수 있도록 수정 보완하고, 이를 통해 시스템의 인식 속도 향상 및 대 어휘 음성인식 시스템을 구현할 계획이다. 또한 신호처리를 전용으로 하는 DSP가 아닌 범용 컨트롤러를 이용한 음성인식 시스템을 구현하고 이를 현장에 적용 가능한 범용적인 시스템의 개발이 이루어져할 것이다.

참고 문헌

[1] L.R.Rabinner, R.W.Schafer, "Digital Processing of Speech Signals", Prentice-Hall, 1978

[2] A.Hyvarinen, J.Karhunen, E.Oja, "Independent Component Analysis", John Wiley & Sons, 2001

[3] S.Choi, A.Cichocki, S.Amari, "Flexible independent component analysis", IEICE Trans. Fundamentals, Vol.E83-A, No.12, pp.2715 -2722, 2000

[4] C.K.Kim, S.B.Kim, S.H.Kim, K.I.Hur, "Performance Improvement of Speech Recognition Based on Independent Component Analysis", ICSP2001, Vol.2 of 2, pp.663-666.

[5] 박정원, 김평환, 김창근, 허강인, "음성인식기 구현을 위한 SVM과 독립성분분석 기법의 적용", 2003년도 대한전자공학회 하계종합학술대회 논문집, 제26권 제1호, pp. 2164, 2003년 7월.

[6] Texas Instruments, "TMS320C3X User's Guide", 1994

[7] Texas Instruments, "TMS320C3x/4x Optimizing C Compiler User's Guide", 1997

[8] Texas Instruments, "TMS320C3x General Purpose Applications user's Guide", 1998

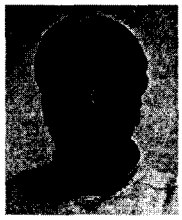
[9] 윤덕용, "TMS320C32 마스터", Ohm사, 1999

[10] 이지홍, 서일DSP㈜, "DSP Chip의 활용 SITK-C3x Training Kit", 서일DSP주식회사, 2000

저자 소개

김창근(Chang-Keun Kim)

1994년 동아대학교 공과대학 전자공학과(공학사)
1998년 동아대학교 대학원 전자공학과(공학석사)
2002년 동아대학교 대학원 전자공학과(공학박사 수료)
2002년~현재 동명정보대학교 겸임교수
※ 관심분야 : 신호처리, 음성인식, DSP



박진영(Jin-Young Park)

2002년 동아대학교 공과대학 전자공학과 (공학사)
2002년 3월 ~ 현재 동아대학교 전자공학과 공학석사 예정
※ 관심분야 : 신호처리, 음성인식, DSP

박정원(Jung-Won Park)

2002년 동아대학교 공과대학 전자공학과 (공학사)
2002년 3월 ~ 현재 동아대학교 전자공학과 공학석사 예정
※ 관심분야 : 신호처리, 음성인식, DSP

이광석(Kwang-Seok Lee)

International Journal of KIMICS, Vol. 1, No. 4, pp.209-212, December 2003 참조

허강인(Kang-In Hur)

International Journal of KIMICS, Vol. 1, No. 4, pp.209-212, December 2003 참조