論文93-30B-3-1

화자 확인 시스템의 설계 제작 및 성능 분석

(Implementation and Performance Analysis of a Speaker Verification System)

權錫奎*, 李秉基*

(Seok Kyu Kweon and Byeong Gi Lee)

要 約

본 논문에서는 실시간 처리가 가능한 자동 화자 확인 시스템의 설계 제작 및 성능에 관하여 기술한다. 이 시스템은 TI(Texas Instrument)사의 디지탈 신호처리 전용칩인 TMS320C25와 고속의 SRAM을 채용하여, 기다림 상태가 없는(no wait state) 고속 연산을 수행할 수 있도록 설계된 것이다. 본 시스템은 또한 독립적 운용 방식과 IBM PC를 사용자 PC 접면으로 하는 PC 연결 운용 방식으로 동작이 가능하다. 화자인식을 위한 음성요소(음성 특징 파라미터)로서, 우수한 성능을 나타내는 것으로 알려져 있는 PARCOR 계수와 LPC-cepstrum 계수를 채용하였다. 판단 논리에 있어서는 일반화된 가중치 거리 개념을 적용하였다. 제작된 화자 인식 시스템으로 화자 인식 실험을 수행한 결과, PARCOR 계수에 대해 5.3%, LPC-cepstrum 계수에 대해 4.7%의 오차율을 얻었다.

Abstract

This paper discusses issues on the design and implementation of a real-time automatic speaker verification system, as well as the performance analysis of the implemented system. The system employs TI 's TMS320C25 digital signal processor TMS320C25 and high speed SRAMs. The system is designed to be used stand-alone as well as via hand-shaking with IBM-PC. The speech parameters used for speaker verification are PARCOR and LPC-cepstrum coefficients, and the employed decision logics are those based on the generalized weighted distance concept. The implemented system showed the performance of 5.3% error rate for the PARCOR coefficient, and 4.7% error rate for the LPC-cepstrum coefficient.

Ⅰ. 서 론

인간의 음성신호를 다루는 음성신호처리 기법은 최근 들어 디지탈 컴퓨터의 발달과 통신의 고급화에 힘입어 많은 발전을 보아왔다. 특히 근래에 들어 디지탈 신호처리 전용의 마이크로프로세서가 많이 개발됨에따라 실시간처리가 가능한 하드웨어구현과 알고리듬에 관한 연구가 활발히 진행되고 있다.[1,2,3] 화자인식(speaker recognition)은, 음성 신호 처리의 한 분야로서, 음성에서 추출한 화자의 개인성 정보를 이용하여 화자의 신원을 파악하는 기법이다.[3,4,5] 화자인식은, 기준 화자와 시험 화자의 동일인 여부를 판정하는 화자 확인(speaker verification)과 시험 화자의 음성 신호가 누구의 것인지 판별하는 화자 식별(speaker identification)로 구분할 수 있다.

현재까지의 화자 인식 분야의 연구 방향은 크게 두

*正會員, 서울大學校 電子工學科
(Dept. of Elec. Eng., SeoulNat'lUniv.)

接受日字:1992年 10月 29日

가지로 나눌 수 있다. 즉, 화자의 개인성 정보를 나타내는 음성 요소에 관한 연구와, 같은 화자를 판별해내기 위한 음성 요소들간의 유사도 측정에 관한 연구가 그것이다. 화자 인식에 유용한 것으로 알려진 음성 요소로는 선형 예측 계수(linear prediction coefficeint : LPC), PARCOR(Partial correlation) 계수, 선형 예측 계수를 변환한 cepstrum 계수(LPC-cepstrum) 등이 있다. 유사도를 측정하는 방법은 두 패턴간의 거리를 측정하는 것으로서, 이것은 거리의 개념을 어떻게 정의하는가 하는데 따라서 달라진다. 널리 이용되는 거리에는 유클리드 거리, 공분산 행렬을 이용한 가중치 거리[4, 6], 기준 화자와 외부 화자의 통계적 성질을 고려한 일반화된 거리[5] 등이 있다. 또 화자 인식 기법은 패턴정합(pattern matching) 방법과 통계적 성질을 이용한 분류 방법 등으로 구분할 수도 있다.[6, 7]

본 논문에서는 이러한 기존의 화자 인식 연구를 바탕으로 하여 실시간 처리 화자 확인 시스템을 설계, 구현한 결과를 기술하고자 한다. 실시간 처리가 가능한 화자 확인 시스템을 구현하기 위해서는, 빠른 속도로 동작하는 디지탈 신호처리 전용 마이크로프로세서와 고속의 기억 장치가 필요한데, 이를 위해서는 TMS320C25와 고속의 SRAM을 사용하였다. 또한 화자 인식에 사용되는 음성 요소로는 PARCOR과 LPC-cepstrum 계수를 선택했고, 유사도 측정 방법은 일반화된 거리를 채용했다. PARCOR 계수를 추출하는 알고리듬은 고정 소수점 연산에서도 계산 결과가 정확한 LeRoux-Gueguen 알고리듬을 사용했다.

본 논문의 구성은 다음과 같다. Ⅱ절에서는 화자 인식 시스템을 제작하는데 필요한 전반적인 사항을 검토하도록 하겠다. 이어서 Ⅲ절과 Ⅳ절에서는 각각 본 시스템의 하드웨어와 소프트웨어 구조를 설명하도록 하겠다. 끝으로, Ⅴ절에서 본 시스템의 성능에 관하여 검토하도록 하겠다.

Ⅱ. 화자 확인 시스템 구현을 위한 기본 고려 사항

화자 확인 시스템은 이미 등록된 화자의 특징(기준 패턴)과 확인하려고 하는 화자의 특징(시험 패턴)의 유사도를 측정하여 동일 화자 여부를 결정하는 장치이다. 본 시스템에서 수행되는 화자 확인 프로그램의 전체구조는 그림 1과 같다. 그림에서 알 수 있듯이, 화자 확인 과정은 시스템 초기화, 음성 데이터 채취, 음성 요소 추출, 패턴 작성, 시간 보정 과정을 거쳐, 기준 패턴과의 유사도 측정 결과를 판단 논리에 적용시켜 화자의 신원을 확인하는 것으로 구성된다.

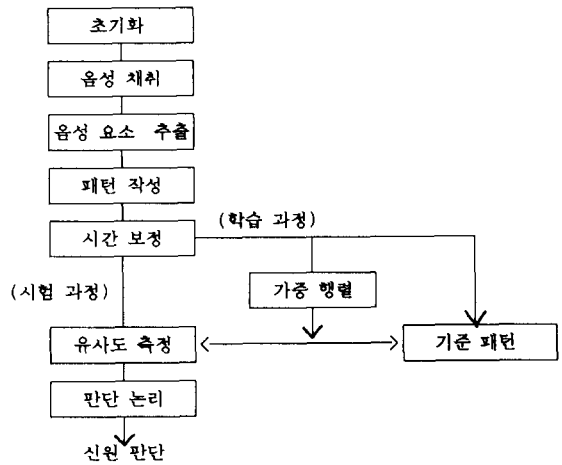


그림 1. 화자 확인 처리 과정의 흐름도
Fig. 1. Flowchart of speaker verification procedure.

음성 요소 추출 과정은 화자의 음성 표본으로부터 LPC, PARCOR, LPC-cepstrum 계수를 계산하는 과정을 의미한다. 이때 빠른 속도의 연산을 위하여 고정 소수점 연산용 하드웨어를 사용해야 하기때문에, 그에 알맞는 알고리듬을 개발하여야 한다. 패턴 작성 과정은 추출된 음성 요소들을 이용하여, 기준 패턴과 시험 패턴을 작성하는 것을 의미한다. 여기서 패턴이란 음성 요소들을 시간 순서로 나열한 것을 의미하며, 기준 패턴은 화자 확인의 대상이 되는 원래의 화자의 음성 요소를 미리 등록해 놓은 것을 가리킨다. 기준 패턴은 여러번의 발음을 통해 추출한 음성 요소들을 평균하여 구한다. 시험 패턴은 시스템을 설치한 후, 임의의 화자가 등록된 화자와 동일인인지를 확인하기 위하여, 입력 화자의 음성 요소를 나열한 것을 의미한다. 시간 보정은 발음할 때마다 달라지는 음성 요소들의 시간적인 위치를 기준 패턴에 맞도록 조정하는 것을 의미하며, 이를 위해 널리 이용되고 있는 방법은 DTW(dynamic time warping)이다.[8] 가중 행렬[5]은 학습과정에서 구하며, 시험 과정에서 기준 패턴과 시험 패턴 사이의 유사도 측정에 이용된다. 이 가중 행렬은 기준 화자와 외부 화자의 특징들의 통계적인 상관관계를 고려해 주기 위한 것이다. 각 단계에 대한 자세한 내용은 Ⅳ절에서 다루기로 하겠다.

그림 1의 화자 확인 프로그램을 수행할 수 있는 실시간 화자 확인 시스템을 설계하기 위해서는 매우 빠

른 속도와 방대한 계산 능력을 갖춘 마이크로프로세서가 요구된다. 위의 프로그램에서 음성 표본 한 개당 음성 요소를 계산하는데 약 1000 개 정도의 어셈블리 명령어가 필요하다. 화자의 음성 주파수 대역을 5kHz 정도까지 잡아, 10kHz의 표본화를 한다면, 표본 간의 시간차는 0.1msec가 된다. 이를 표본당 명령어 수와 비교해 보면 대략 10 MIPS 정도의 처리 속도가 요구된다. 이 정도의 처리 속도는 범용 마이크로프로세서로의 구현은 어렵다. 따라서, 신호처리 전용의 DSP 칩을 사용하는 것이 필요하다.

본 연구에서는 DSP 칩으로서 미국 TI(Texas Instrument)사의 TMS320C25를 사용하였다. 화자의 음성을 채취하기 위해서는 오디오 증폭기와 A/D 변환기가 필요한데, 이것 역시 TI사의 TLC32040을 채택했다. 이것은 소프트웨어 명령에 의해, 표본화 주파수를 변경시킬 수 있으며, 최고 19.2kHz까지 표본화가 가능하다. 또한 TMS320C25와 직렬 포트를 이용한 직접적인 데이터 전송이 가능하여, 전체 구조를 간단히 할 수 있다. TMS320C25는 544 워드의 내부 램(on-chip RAM)과 4K 워드의 내부 롬(on-chip ROM)을 가지고 있으며 총 64K 워드의 프로그램 기억 장치와 64K 워드의 데이터 기억 장치를 접근(access)할 수 있다. 한편 본 시스템은 48K 워드의 프로그램 기억 장치와 32K 워드의 데이터 기억 장치를 가지고 있다.

Ⅲ. 하드웨어의 구조

화자 인식 시스템의 전체 구조를 간략히 표시하면 그림 2의 블록도와 같다. 본 시스템은 크게, CPU, 기억 장치, PC 접면(PC-interface), 그리고 입출력 장치, 음성 입력부의 다섯가지 부분으로 구성된다. 각각의 블럭들은 서로 번지, 데이터, 제어 신호를 주고 받는다. CPU와 PC 접면간의 통신은 버스 개폐기를 조절하여 기억 장치를 점유하기 위한 목적으로 이용된다.

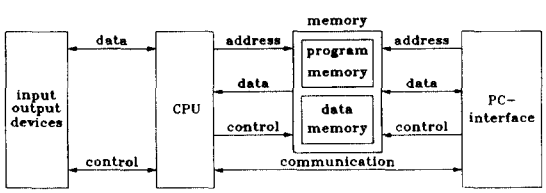


그림 2. 화자 확인 시스템의 하드웨어 블록도
Fig. 2. Block diagram of speaker verification system hardware.

본 화자 확인 시스템은 IBM-PC/AT를 호스트 컴퓨터로 한 PC 연결 운용 방식과, 호스트 컴퓨터 없이 시스템 단독으로 동작하는 독립적 운용 방식 두가지를 모두 수용한다. 핸드 쉐이킹(hand shaking) 방식으로 동작하는 경우에는, TMS320C25의 모든 버스가 하이 임피던스(high impedance)인 상태에서, PC와 시스템의 기억 장치 간에 데이터가 전송된다. 신호처리 칩이 동작할 때는, 기억 장치와 PC 접면 부분이 차단되어, TMS320C25가 기억 장치를 단독으로 점유하게 된다. 독립적 운용 방식으로 동작하는 경우에는 PC 접면과 PC간의 연결이 단절되므로 PC 접면으로부터 나오는 모든 신호는 하이 임피던스 상태, 혹은 비가동(inactive) 상태에 있게 된다. 전원은 PC 연결 운용 방식으로 동작할 때는 PC로부터 공급받으며, 독립적 운용 방식일 때는 외부 스위칭 전원 장치로부터 받게 되어 있다.

실제 화자 확인 시스템의 하드웨어는 3장의 PCB 기판과 음성 입력을 위한 마이크로폰으로 이루어져 있다. 그림 3은 전체 시스템의 실물 사진이다. 그림 3의 좌측 부분은 주 기판으로서 그림 2의 CPU 부분과 PC 접면 부분과, 음성 입력을 위한 오디오 증폭기, A/D 변환기 등을 담고 있다. 그림 3의 중앙 부분은 메모리 기판으로서, 시스템 전체의 기억장치 부분을 담당하며, 고속의 SRAM과 ROM으로 이루어져 있다. 또한 입출력 장치를 제어하기 위한 입출력 제어기를 포함하고 있다. 그림 3의 우측 부분은 입출력 기판으로서 화자의 코드번호 입력을 위한 키보드와 화자확인 결과를 출력하는 LED로 이루어져 있다. 주기판과 메모리 기판, 메모리 기판과 입출력 기판은 커넥터로 서로 연결되어 있으며, 주기판에 마이크로폰을 연결하기위한 커넥터가 장착되어 있다.

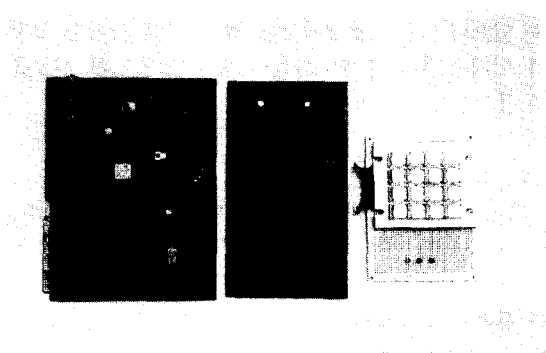


그림 3. 전체 시스템의 실물 사진
Fig. 3. Photograph of the speaker verification system

그림 2 전체 블록도의 각 부분을 좀더 구체적으로 설명하면 다음과 같다.

1. PC 접면 부분 [9]

호스트 컴퓨터가 시스템의 기억 장치에 직접 접근하기 위해서는 신호처리 칩에서 나오는 신호와 동일한 신호를 확장용 슬롯에서 생성시킬 수 있어야 한다. 이때, 신호처리 칩으로부터 나오는 신호들과 버스충돌이 일어나지 않도록 하여야 한다. 이러한 기능을 담당하는 것이 PC 접면 부분이다. 이 부분은 PC 연결 스위치를 통하여 호스트 PC와 연결되어 있으며, PC로부터 입력을 받아 CPU, 기억 장치, 버스 개폐기를 위한 제어 신호 발생기와 호스트 PC와 시스템의 버스를 개폐시키는 버스 개폐기로 구성되어 있다. 제어 신호 발생기는 여러개의 복호기로 구성되어 있는데, 본 시스템에서는 회로를 간단히 하기위하여, 그것들을 한 개의 EPLD 칩으로 구현하였다.

2. CPU 부분 [10]

CPU는 기억 장치, 입출력 장치 및 음성 입력부, PC 접면 부분을 제어하고, 화자 확인에 필요한 연산을 수행한다.

데이터 버스 D0부터 D15는 16 비트의 데이터를 전송하고, 번지 버스 A0부터 A15는 64K 워드의 기억 장치를 접근할 수 있다. CPU는 PC 접면 부분과 마찬가지로 기억 장치 선택을 위한 제어 신호를 발생한다. 또한 호스트 PC와의 통신을 통해 PC 접면 부분의 버스개폐기를 제어하여 CPU와 호스트 PC에서 나오는 번지와 데이터 버스의 충돌이 일어나지 않도록 조정한다. CPU를 초기화시키는 리셋(reset) 신호는 입출력 장치로부터 입력된다.

TMS320C25는 최대 40MHz의 속도로 동작시킬 수 있으나, 본 시스템에서는 빠른 기억 장치를 구하기 어려운 관계로 20MHz의 외부 발진기의 신호를 주 클럭으로 사용하였다.

3. 음성 입력부

음성 입력부는 마이크로폰, 오디오 증폭기(audio amp), A/D 변환기로 구성되어 있으며, 마이크로폰을 통하여 아날로그 음성신호를 받아 오디오 증폭기를 통하여 증폭시킨후 A/D 변환기를 통해 디지탈 신호로 바꾼후 CPU에 전송한다. 오디오 앰프에서는 입력 신호의 잡음을 제거하고 원하는 음성신호만이 A/D 변환기에 입력되도록 한다.

A/D 변환기로는 TI사에서 나온 TLC32040을 사용하는데, 이 칩은 A/D, D/A 겸용으로서, 14 비트의 생동폭(dynamic range)과 10 비트의 선형성(linearity)을 제공하고 있다. 19.2kHz까지의 프로그램 가능한 표본화 주파수를 가지고 있으며, TMS320C25와 직렬 포트를 통하여 데이터를 교환하고, TMS320C25로부터 주 클럭을 전달 받는다. TMS320C25와 A/D 변환기 사이를 왕래하는 제어 신호들은 직렬 포트를 통한 데이터 전송에서 데이터를 동기시키는데 사용 된다.

+12V, -12V 및 아날로그 접지가 필요한데이는 호스트 컴퓨터(독립적 운용일 때는 외부 전원)로부터 공급받는다. TLC32040은 매우 안정된 전원을 필요로 하므로, 전압 조절기(voltage regulator)인 7805와 7905를 사용하여 +12V와 -12V로부터 정류된 +5V와 -5V를 인가한다.

4. 기억 장치와 기억 장치 복호기

기억 장치는 4개의 32K 바이트 SRAM과 2개의 32K 바이트 EPROM으로 구성되어, 총 96K 바이트의 용량을 가진다. 이 중 2개의 SRAM과 2개의 EPROM은 프로그램 기억 장치에 사용되고, 나머지 두개의 SRAM은 데이터 기억 장치에 사용된다.

TMS320C25는 대단히 빠른 처리 속도(10 MIPS)를 가지기 때문에 그에 부속되는 기억 소자도 접근 속도가 극히 빨라야 한다. 그러나, 독립적 운용 방식으로 동작할 때에는 자체의 기억 장치에 프로그램과 데이터를 저장하기 위하여 롬이 필요하다. 이런 점들을 고려하여 기억 장치를 다음과 같이 구성한다. 2개의 EPROM은 기억 장치의 하위 번지를 구성하며, 2개의 SRAM은 상위 번지를 구성하도록 한다. 즉, TMS320C25는 프로그램 기억 장치, 데이터 기억 장치 자체 내에서, 혹은 상호 간에 데이터 전송이 가능하기 때문에, EPROM에 전체 프로그램과 데이터를 저장해 놓고, 시스템을 리셋시키면 프로그램과 데이터가 각각 프로그램 기억 장치의 상위 번지와 데이터 기억 장치로 옮겨지도록 한다. 그리고, 실제 시스템 동작은 SRAM에 들어 있는 프로그램에 의해 수행되도록 한다. TMS320C25를 최대 속도로 동작시키 위하여, SRAM을 접근할 때는 기다림(wait) 상태가 없도록 하고, EPROM을 접근할 때는 10 번의 기다림 상태를 두도록 설계하였다. 이때, 사용된 SRAM의 접근 시간은 35nsec이다.

기억 장치 복호기는 CPU와 PC 접면부분이 시스템의 기억 장치를 접근하는데 이용된다. 즉, PC 접면부분의 제어신호 발생기와 CPU로부터 생성된 제

어 신호들을 입력으로 하여 각 기억 장치를 가동(enable)시키는 신호를 발생시킴으로써, 기억 장치를 접근하는 기능을 담당한다. 이때 가동시키는 신호로는 프로그램 기억 장치 접근 신호, 데이터 기억 장치 접근 신호, 읽기/쓰기 가동 신호 등이 있다. 기억 장치 복호기 역시 EPLD 칩을 사용하여 설계함으로써 회로 구조가 간단해지도록 하였다.

5. 입출력 장치

입출력 장치는 외부로부터 필요한 데이터를 받아들이고, 시스템의 화자 인식 결과 및 기타의 결과를 외부로 출력하는 역할을 한다. 입출력 장치는 크게 입출력 제어 장치, LED 디스플레이, 키보드의 세 부분으로 구성되며, 이중 키보드는 입력 장치를, LED 디스플레이는 출력 장치를 구성한다.

입출력 제어 장치는 입출력 제어 신호를 발생시키는 논리 회로 부분과 입출력 버퍼로 구성되며, CPU로부터 입출력 선택 신호, 읽기/쓰기 가동 신호 등의 제어 신호를 입력으로 받아 입출력 제어 신호를 발생시키고, 필요한 신호를 버퍼를 통하여 전달하는 역할을 수행한다.

Ⅳ. 소프트웨어의 구조

화자 확인 시스템을 제대로 동작시키기 위해서는 그림 1에 나타난 바와 같이 화자 확인을 위한 음성 요소 추출, 거리 계산 등을 수행하는 소프트웨어가 필요하다. 본 화자 확인 시스템의 소프트웨어는 크게 두가지로 구분된다. 첫째는 호스트 컴퓨터를 통해 화자 확인 시스템을 작동시킬 때, 시스템을 제어하기위한 프로그램으로서, C 언어로 작성되어 있다. 둘째는, 실제 화자 확인을수행하는 프로그램으로서 C 언어와 TMS320C25의 어셈블리어로 작성되어 있다.

1. 제어 프로그램

제어 프로그램은, 시스템을 PC 연결 운용 방식으로 동작시킬 때 호스트 PC가 화자 확인 시스템을 제어시키기 위한 프로그램이다. 이 프로그램들은 주로 PC에 저장되어 있는 화자 확인 프로그램과 데이터를 시스템의 기억 장치에 옮겨 싣고, 시스템이 계산한 결과를 기억 장치에 저장시키면, 그 데이터를 PC로 옮기는 역할을 하게 된다. PC에서 시스템 기억 장치로 옮겨지는 데이터는, 기준 패턴과 거리 계산에 필요한 가중치 및 문턱값 등이다. 독립적 운용방식으로 동작할 때는, 이러한 데이터는 모두 프로그램 롬 상에 저장되어 있다. 또한, 제어 프로그램은 시스템을 리셋시키고, 동작 상태로 만드는 기능도 수행한다.

2. 화자 확인 프로그램

화자 확인 프로그램은 실제 시스템 상에서 수행되는 프로그램으로, 전체 프로그램의 흐름도는 그림 1과 동일하다. 그림의 각 단계를 좀더 자세히 살펴 보면 다음과 같다.

먼저 시스템 초기화 과정은 그림 4의 흐름도와 같다.

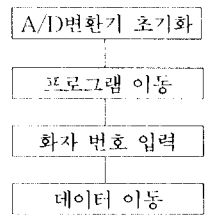


그림 4. 시스템 초기화 프로그램의 흐름도
Fig. 4. Flowchart of system initialization program.

시스템 초기화 과정은, TMS320C25가 기억 장치와 A/D 변환기, 입출력 버퍼를 초기 상태로 만드는 것을 의미한다. 먼저, 시스템을 리셋 상태에서 풀어주면 TMS320C25는 A/D 변환기와의 이차적 통신(secondary communication)을 통해 표본화 주파수를 9.6kHz로 초기화시키고, 입력 신호 레벨을 -3V에서 +3V로 맞춘다. 또한, 음성 데이터를 주고 받는 일차적 통신(primary communication)에 필요한 여러가지 데이터를 초기화한다. 이 과정이 끝나면, TMS320C25는 프로그램 롬에서 프로그램 램으로 화자 확인에 필요한 프로그램들을 이동시킨다. 이것은 독립적 운용 방식으로 동작하는 경우를 염두에 둔 것으로, 호스트 PC가 없는 경우 화자 확인에 필요한 프로그램과 데이터들은 롬에 저장되어 있다가, 시스템 동작시에는 램으로 옮겨져 빠른 속도로 동작하게 되는 것이다. 프로그램 이동이 끝난 후, 시스템은 LED 디스플레이를 통해 키보드 입력을 기다리고 있음을 알린다. 입력이 들어 올 때까지 기다림 상태에 있다가, 키보드 입력이 들어 오면, 키보드 번호를 화자의 번호로 인지하고, 그 번호에 해당하는 화자의 데이터를 프로그램 롬으로부터 데이터의 램으로 이동시킨다.

시스템 초기화가 끝나면, 음성 채취 단계에 들어간다. 마이크로폰, 오디오 앰프를 거친 음성 신호는

A/D 변환기를 거치면, 디지탈 신호로 변환되고, 직렬 포트를 통해 직접 TMS320C25에 전달된다. CPU는 전달 받은 음성 데이터를 데이터 기억 장치에 저장하고, A/D 변환기로부터 올, 다음 표본을 기다리게 된다. 표본 수가 20480개(256 표본 × 80 프레임)가 되면, 음성 채취를 끝마치게 된다.

다음 단계로 화자 확인에 필요한 여러가지 음성 요소들을 추출하게 되는데, 추출되는 음성 요소는 LPC, PARCOR, LPC-cepstrum 등이다. 그 과정을 나타내는 흐름도가 그림 5이다.

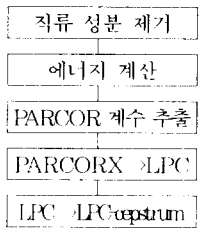


그림 5. 음성 요소 추출 프로그램의 흐름도
Fig. 5. Flowchart of speech parameter extraction program.

먼저 데이터 기억 장치에 저장된 음성 신호를 프레임 단위(256 표본)로 취하여, 직류 성분을 제거하고, 프레임의 에너지를 계산한다. 그 값이 일정치 이하가 되면, 묵음 구간으로 처리하고, 다음 프레임을 취하게 된다. 만약, 그 프레임이 발음 구간이면, 다음 단계로 PARCOR 계수를 계산한다. PARCOR 계수의 계산을 위해서는 고정 소수점 연산에서도 계산 결과가 정확한 LeRoux-Gueguen 알고리듬[11] 을 사용한다. PARCOR계수만으로 화자를 판단할 때는, 이 단계에서 곧장 패턴 작성 과정으로 넘어간다. LPC-cepstrum을구하기 위해서는 먼저, 주어진 PARCOR 계수를 LPC 계수로 바꾸고, 이 LPC 계수를 다시 LPC-cepstrum 계수로 변환한다. 이 과정에서, 버림 오차(truncation error)에 의한 정보의 손실을 막기 위하여, 32비트 곱셈과 64비트 덧셈, 뺄셈을 채용하였다.

음성 요소 추출이 끝나면, 음성 요소를 시간에 따라 나열하여 패턴을 작성 한다. 이렇게 만들어진 패턴은 기준 패턴을 기준으로 삼아 시간 보정을 시켜 새로운 패턴으로 만든다. 시간 보정 방법으로서는 DTW(Dynamic Time Warping)를 이용하였다. 기준 패턴은 학습 과정을 통해서 만들어지는데, 시간 보정이 된 5개의 패턴의 평균을 취하여 기준 패턴으로 삼는다. 시험 과정에서는 확인할 화자의 시험 패턴을 기준 패턴에 시간 정합시킨다. 이 과정이 끝나면, 최종적으로 학습과정에서 구한 가중치, 문턱값들을 이용하여, 시험 패턴과 기준 패턴 사이의 유사도를 측정하여 화자 여부를 판단하게 된다.

학습과정은 결정함수를 계산할 때 필요한 가중행렬과 문턱값을 구하는 것으로서[5], C언어로 작성되어 있으며, 외부 컴퓨터 상에서 수행된다.

Ⅴ. 화자 확인 시스템의 성능 검토 분석

Ⅲ, Ⅳ절에서 설명한 과정대로 제작한 시스템의 성능을 분석하기 위하여, 실제 화자를 대상으로 하여 화자 확인 실험을 수행하였다. 실험에 사용된 계수는 모두 10차의 PARCOR과 LPC-cepstrum 계수이다. LPC 계수에 Itakura-Saito 방법을 적용한 것은 LPC-cepstrum 계수에 Euclid 거리를 적용한 것과 같은 물리적 의미를 가지므로, 본 실험에서는 LPC 계수 대신 LPC-cepstrum 계수만으로 대신하였다. 실험 데이터는, 10명의 화자가 각각 50번씩 발음한 음성 데이터에 대하여 동시에 PARCOR 계수와 LPC-cepstrum 계수를 구하여 사용했다. 이중 25개는 학습과정에 사용하고 나머지 25개는 시험과정에 사용하였다.

이때, 사용된 발음은 '통신 연구실'이다.

학습 과정에서는 우선 기준화자의 25개의 데이터와, 외부화자 (기준화자가 아닌 나머지 9명의 데이터) 각각에서 5개씩을 취한 45개의 외부화자 데이터를 실험용 데이터로 취하였다. 이렇게 취한 데이터를 가지고 결정함수에 필요한 가중행렬과 문턱값을 결정한다. 기준화자와 외부화자의 데이터가 완전히 분리되지 않은 경우에는 알고리듬이 수렴하지 않으므로[5], 변화의 정도가 일정치 이하가 되면 계산을 끝내도록 프로그램을 작성하였다. 이 학습과정은 본 시스템 상에서 수행하지 않고, 수행속도가 매우 빠른 외부 컴퓨터(MIPS사의 RS2030)를 사용하였다.

시험과정에서는 학습과정에서 사용되지 않은 나머지 데이터(기준화자 25개, 외부화자 180개)를 입력 데이터로 하여 오척(false reject:FR)율과 오인(false accept:FA)율을 계산하여 평균을 취한 것을 확인 오차율로 정하였다.[5]

위에서 설명한 방법으로 PARCOR과 LPC-cepstrum 계수에 대해 화자인식 실험을 수행한 결과는 표 1과 같다.

표 1. 화자 확인의 오차율

Table 1. Error rate of speaker verification.

계수	화자1	화자2	화자3	화자4	화자5	화자6
PARCOR	1.0	2.4	8.6	2.6	3.8	8.3
LPC-cepstrum	1.0	2.0	4.3	6.6	2.7	6.3
계수	화자7	화자8	화자9	화자10	평균	분산
PARCOR	13.2	4.5	4.3	3.8	5.3	12.3
LPC-cepstrum	8.1	4.6	5.0	6.6	4.7	5.2

표 1에 나타나 있는 바와 같이, 본 시스템의 화자 확인 오차율은 LPC-cepstrum에 대해 4.7%로 나타났으며, 이것은 기존의 연구 결과[5]인 1.4% 보다 조금 높은 편이다. 이것은 본 시스템은 기존 연구와는 달리 실시간 처리가 가능하도록 고정 소수점 연산을 이용하여 음성 요소를 추출하였기 때문에, 계산오차가 발생한 것으로 간주된다

또한, 이 표에서 알 수 있듯이, 기존의 연구결과[5]대로, LPC-cepstrum이 PARCOR 계수보다 대체로 우수한 성능을 나타냄을 확인할 수 있다. 즉, LPC-cepstrum이 PACOR 계수보다 오차율도 낮고, 각 화자 간의 분산도 낮다. 그러나 화자에 따라서는 그 반대의 결과를 나타내는 것도 있었다. 이는 계수들을 추출하는데 있어 발생하는 버림 오차(truncation error)가 누적되어 최종적으로 구해지는 LPC-cepstrum에 가장 큰 오차가 발생하기 때문인 것으로 여겨진다. 이것을 해결하기 위해서는 알고리듬 상의 문제점을 개선하거나 계산 방법에 있어 부동 소수점 연산을 도입할 필요가 있겠다.

또한 화자에 따라서는 오차율이 비교적 높게 나타나는 경우도 있는데 그것은 실험 과정이 이상적으로 이루어지지 못했기 때문인 것으로 간주된다. 즉, 본 시스템에서는 한 프레임 내의 에너지를 계산하여, 일정치를 기준으로 발음 구간과 묵음 구간을 구분하고 있다. 그런데, 묵음 구간으로 판정된 구간은 버려지게되므로 기준 패턴을 구성하는데 기여를 하지 못하는 것이다. 화자 7의 경우는 발음하는 동안 너무 소리가 작아, 발음의 상당 부분이 묵음 구간으로 처리된 것 같다. 이러한 관계는 그림 6에 보인 LPC-cepstrum 기준패턴들을 비교해 보면 잘 알 수 있다. 이 그림은 10차 LPC-cepstrum 계수들 중에서 하나의 계수값을 프레임에 따라 나타낸 것이다. 화자 1의 기준 패턴은 발음구간이 길고, 전체적으로 특징적인 부분이 많이 나타나지만, 화자 7의 경우는 발음 구간도 짧고, 특징적인 부분이 거의 나타나지 않는 것을 볼 수 있다. 이러한 문제는 실험과정에서, 좀더 세심한 주의를 기울이면 해결될 수 있으리라 생각된다.

하드웨어 측면의 성능분석을 위하여 화자확인 시간, 편리성 등을 검토해 보았다. 화자 확인 평균 시간은 발음 후 3초 이내로서, 실시간 처리 기준을 충분히 만족시킨다고 볼 수 있다. 본 시스템은 회로를 간단히 하기위하여 TMS320C25 만을 CPU로 사용하고 있지만, 범용 프로세서와 DSP 전용 프로세서를 사용하여 시스템 제어 프로세서와 화자 확인 전용 프로세서로 구분하면 화자 확인 시간을 더욱 단축시킬 수 있다.

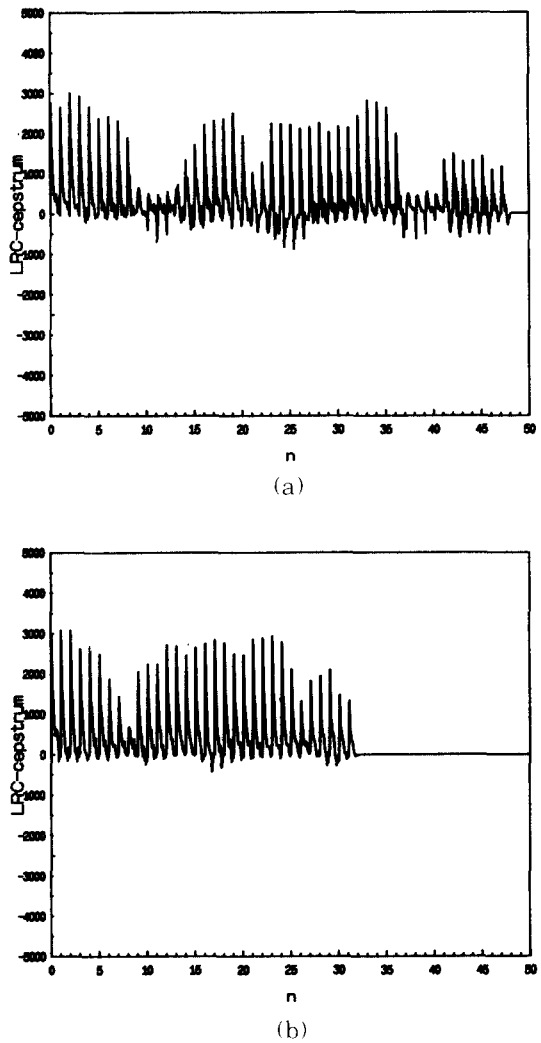


그림 6. LPC-cepstrum 기준 패턴
(a) 화자 1 (b) 화자 7

Fig. 6. LPC-cepstrum reference pattern
(a) Speaker 1. (b) Speaker 7.

편리성 측면에서 살펴보면, 본 시스템은 독립적 운용 방식과 PC 연결 운용 두가지를 모두 수용하고 있

어, 독립적으로 완전한 화자 인식기로서 동작할 뿐만 아니라, PC와 연결하여 범용 음성 신호처리에 이용하면, 음성신호 처리에 필요한 소프트웨어 개발에 이용할 수 있으리라 간주된다.

Ⅴ. 결론

본 논문에서는 DSP 칩을 사용하여 실시간 처리가 가능한 화자 확인 시스템을 설계제작하고 그 성능을 검토 분석하였다. 이 시스템은 발음후 3초 이내에 화자 확인 결과를 출력하며, 화자 확인 오차율은 평균 5% 정도이다. 본 시스템은 IBM-PC 상에서 확장카드로 사용할 수 있으며, 독립적 운용으로도 사용 가능하다. 시스템의 구조를 간단히 하기 위하여, PC 접면 부분과 기억 장치 복호기를 EPLD 칩을 이용하여 구현하였다. 시스템에서 수행되는 프로그램은 계수 추출 프로그램, DTW 프로그램, 판단 프로그램이다.

제작된 화자 확인 시스템으로 화자 확인 실험을 수행함으로써, 이미 발표된 바 있는 음성요소(음성 특징 파라미터)들 간의 성능 분석 결과를 비교 확인하였다. 이로부터, LPC-cepstrum이 PARCOR 계수에 비해 좋은 성능을 나타냄을 확인할 수 있었다. 또한 문헌 [5]에서 제안한 가중치 방법을 적용함으로써 기준 화자 뿐만 아니라 외부 화자의 통계적 성질까지도 고려하여 주었다. 그 결과, 기준화자 데이터의 공분산 행렬을 가중치 행렬로 사용한 방법보다 인식율이 우수한 화자 확인 시스템을 구현할 수 있었다.

본 논문에서 제작된 화자 확인 시스템을 실제 사용하기에는 아직 미흡한 상태라 하겠다. 그것은 실험결과에서 보여지듯이 오차율이 5% 정도로서 비교적 높은 편이기 때문이다. 좀 더 낮은 오차율을 얻기위해서는, 여러가지 음성 요소들을 결합하여화자 확인에 적용하는 것을 검토해 볼 수 있겠다. 이러한 설계 제작 연구를 계속하면 실제 사용이 가능한 화자 확인 시스템을 구현하는 것이 가능하리라 간주된다

본 시스템은 IBM-PC 상에서 제어하여 실시간 처리 프로그램을 수행시킬 수 있으므로, 화자 확인과 같은 음성신호 처리의, 새로운 알고리듬 개발을 위해 이용할 수 있겠다. 또, 국내에서 이루어진 대부분의 음성인식 연구들에 있어서 음성신호를 A/D 변환시킨 후에는 모든 처리 과정을 순전히 소프트웨어 상에서 처리해온 점을 감안할 때, 본 연구는 화자 확인 시스템을 하드웨어 상에서 실시간처리가 가능하도록 구현하였다는데 의의가 있다 하겠다.

參 考 文 獻

[1] Texas Instrument, *Digital Signal Processing Applications with the TMS320 Family*, Texas Instrument Incorporated, 1986.

[2] 전자 통신 연구소, 숫자음 인식을 위한 실시간 H/W 설계연구,.1986.

[3] J.B. Attili, M. Savic and J. P. Campbell,Jr., "A TMS32020-based real time, text-independent, automatic speaker verification system", *Proceedings. ICASSP 1988*, vol.1, pp.599-602.

[4] M.R. Sambur, "Speaker recognition using orthogonal linear prediction", *IEEE Trans. on ASSP*., vol. ASSP-24, No. 4, pp. 403-409, Aug. 1976.

[5] 이 혁재, 이 병기 "화자인식을 위한 음성요소들의 성능 분석 및 새로운 판단 논리", 전자공학회논문지 제 26권 제 7호 pp.146-156, 1989.

[6] L.R. Labiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prenice-Hall, New Jersey, 1978.

[7] J.T. Buck, D.K. Burton and J.E. Shore, "Text-dependent speaker recognition using vector quantization", *Proceedings. ICASSP*, pp. 381-384, 1985.

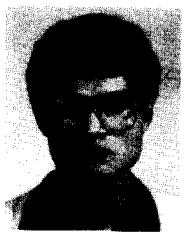
[8] C. Myers, L.R. Rabiner and A.E. Rosenberg, "Performance tradeoffs in dynamic time warping algorithm for isolated word recognition", *IEEE Trans. on ASSP*., vol. ASSP-28, No. 6, pp. 623-635, Dec.1980.

[9] IBM, *Technical Reference Personal Computer AT, International Business Machines* Corporation, 1986.

[10] Texas Instrument, *TMS320C25 User's Guide*, Texas Instrument Incorporated, 1986.No. 4, pp. 4-17, 1986.

[11] P.E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, New-Jersey, 1987.

著者紹介



權 錫 圭 (正會員)

1966年 9月 20日生

1989年 2月 서울대학교 전자공학과 (학사). 1991年 2月 서울대학교 전자공학과 (석사). 1992年 ~ 현재 미국 미시간 대학 재학중. 주관심 분야는 신호처리

李 秉 基 (正會員)

1951年 5月 12日生. 1974年 서울대학교 전자공학과 (학사). 1978年 경북대학교 대학원 전자공학과 (공학석사). 1982年 University of California, Los Angeles (공학박사). 1974年~1979年 해군사관학교 전자공학과 교관. 1982年~1984年 미국 Granger Associates 연구원. 1984年~1986年 미국 AT&T Bell Laboratory 연구원. 1986年 ~ 현재 서울대학교 전자공학과 조교수. 주관심분야는 디지탈 신호처리, 광대역통신 및 광통신, 회로이론 등임.