

음성인식 기반 콘텐츠 네비게이션 시스템

김기백*, 최종호*

Contents Navigation System using Speech Recognition

Kee-Beak Kim*, Jong-Ho Choi*

요 약

최근 들어 인간의 의지를 각종의 전자시스템에 전달하기 위한 수단으로 음성인식 기술을 이용하고자 하는 연구가 널리 진행되고 있다. 음성인식 인터페이스에서 가장 중요한 이슈는 처리시간의 감소 및 범용 인터페이스의 개발이다. 이러한 문제점을 해결하기 위하여 본 연구에서는 하드웨어 기반의 상용 IC로 생산되고 있는 음성인식 프로세서인 RSC-4128이 내장된 음성인식 모듈 VR-STAMP를 사용하였다. 본 연구에서 새롭게 개발한 시스템은 T2SI(Text To Speaker Independent) 기반의 화자(話者)독립 방식으로 음성인식 신호를 콘텐츠 네비게이션 시스템의 제어신호로 활용하여 임베디드 시스템 및 PC 등에 설치된 윈도우즈 기반의 응용 소프트웨어를 제어할 수 있는 시스템이다. 필드 테스트를 통해 그 유용성을 확인한 결과, 본 연구에서 개발한 시스템은 콘텐츠 네비게이션은 물론 가전기기 제어 및 홈 네트워크 등에 널리 응용될 수 있을 것으로 판단된다.

▶ Keyword : 화자독립, 음성인식, 임베디드 시스템, PC, 콘텐츠 네비게이션

1. 서 론

최근 들어 인간의 의지를 직감적으로 기계에 전달하기 위한 수단으로 음성을 이용하고자 하는 연구가 널리 진행되고 있다. 일상의 대화에서 자연스럽게 사용되고 있는 음성을 이해하는 인터페이스를 구현할 수만 있다면 음성은 의사전달에 대응하는 자연스러운 입력수단으로 널리 이용될 수 있을 것이다.

음성인식(Voice Recognition)은 사람의 음성신호를 전자장치에 입력하여 이를 처리하는 방식을 총괄하

여 표현하는 것으로 그 구현 방법(1,2,3)으로는 크게 화자종속 방식(Speaker Dependent)과 화자독립 방식(Speaker Independent)이 있다. 화자종속 방법으로는 화자(話者)의 음성신호 파형을 메모리와 같은 저장장치에 저장 후 이를 인식하는 방법으로 대표적인 예로 기존의 핸드폰 전화 저장 장치를 들 수 있다. 이 방식은 사용자의 음성파형을 사전에 입력하는 방식이기 때문에 인식오류가 적고, 특정 사용자만을 위한 시스템의 구축이 가능하다는 장점이 있으나 외부잡음이 크게 영향 받는다는 치명적인 단점이 있다.

• 제1저자 : 김기백
* 강남대학교 전자공학전공

이와는 반대로 화자독립 방식은 화자(話者)의 음성신호와는 별도로 텍스트 형태로 저장되는 발음기호를 입력되는 음성신호와 비교하여 인식하는 방식이다. 대표적인 예로 ARS 서비스에서 제공하는 음성서비스를 들 수 있다.

본 연구에서는 범용의 시스템을 구현하기 위하여 T2SI(Text To Speaker Independent) 기반의 화자독립 방식을 이용하여 콘텐츠 네비게이션 시스템을 개발하였다. 음성인식 인터페이스에서 가장 중요한 이슈는 처리시간의 감소 및 범용 인터페이스의 개발이다. 이러한 문제점을 해결하기 위하여 본 연구에서는 하드웨어 기반의 상용 IC로 생산되고 있는 음성인식 프로세서인 RSC-4128이 내장된 음성인식 모듈 VR-STAMP[4,5,6]를 사용하였다. 본 연구에서 새롭게 개발한 시스템은 T2SI(Text To Speaker Independent) 기반의 화자(話者)독립 방식으로 음성인식 신호를 콘텐츠 네비게이션 시스템의 제어신호로 활용하여 임베디드 시스템 및 PC 등에 설치된 윈도우즈 기반의 응용 소프트웨어를 제어할 수 있는 시스템이다. 실험을 통하여 그 유용성을 확인한 결과, 본 연구에서 개발한 시스템은 콘텐츠 네비게이션은 물론 가전기기 제어 및 홈 네트워크 등에 널리 응용될 수 있을 것으로 판단된다.

2. 음성입력 시스템

2.1 음성인식 주요 단어

본 연구에서는 음성인식 인터페이스를 자동차용 네비게이션 시스템에 적용하는 것을 목표로 발음하기 편하고, 발음의 구별이 비교적 명확한 단어를 선정하여 음성인식 단어로 활용하였다. 본 연구에서는 음성인식을 이용하여 자동차용 네비게이션 시스템에 내장된 응용소프트웨어를 제어하는 음성인식 인터페이스 시스템의 개발을 목표로 발음이 자연스러우면서 각 발음이 서로 뚜렷하게 구별되는 것에 주안점을 두고, <표 1>의 의미를 갖는 단어를 입력인식 주요 단어로 선정하였다.

<표 1> 음성인식 대상 주요 단어

단 어	기 능	비 고
나 비	네비게이션 실행	주요 목적지 선택가능 연계
DMB	DMB 실행	처음, 이전, 다음으로 채널설정
DVD	DVD 실행	처음, 이전, 다음으로 콘텐츠 선택
MP3	MP3 실행	처음, 이전, 다음으로 콘텐츠 선택
처 음	목록지정	콘텐츠 네비게이션

이 전	목록이동	콘텐츠 네비게이션
다 음	목록이동	콘텐츠 네비게이션
메신저	메신저 실행	콘텐츠 선택 / TTS 연계
이메일	이메일 실행	콘텐츠 선택 / TTS 연계
뉴 스	사이트 접속	콘텐츠 선택 / TTS 연계
노래방	노래방 실행	처음, 이전, 다음으로 콘텐츠 선택

2.2 음성입력 마이크 시스템

마이크 시스템에서 가장 중요하게 고려해야 할 사항은 외부에서 입력된 신호에 포함되는 외부 노이즈를 최소화하는 것이다. 마이크를 회로에 마운트 할 경우 공기와 접촉할 수 있는 공간을 제거해야만 한다. 먼저 마이크 소자를 회로 표면에 부착하여 일차적인 외부 환경으로부터의 노이즈를 제거하는 것이 필요하다.

또한 회로 표면에 장착된 마이크 앞부분에는 인식을 방해할 수 있는 모든 상황에서 자유로워야 한다는 것이다. 0.7mm 이상의 플라스틱 케이스 등이 이 부분에 부착될 경우, 인식률의 저하를 가져올 수 있다.

그리고 마지막으로, 하우징(Housing)된 마이크 주변에 고무와 같은 피막을 둘러서 외부 노이즈를 흡수할 수 있도록 제작하는 것도 매우 중요하다. 이를 통해 시스템의 부조화로 인한 환경 노이즈들을 제거할 수 있기 때문이다.

한편, 마이크 제조회사에서 제공하는 민감도 값 (Sensitivity)을 이용하면 식 (1)을 이용하여 외부 저항값을 얻을 수 있다.[4]

$$R_s = R \times 10^{\frac{((-46dB - (Sensitivity))}{20})} \dots\dots\dots (1)$$

본 연구에서 사용하고 있는 VR-STAMP의 마이크 입력 시스템의 부분회로를 <그림 1>에 나타냈다.[5]

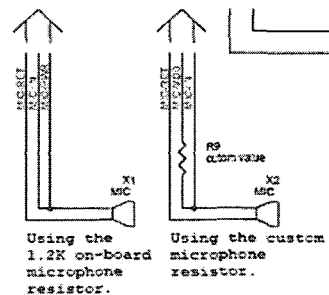


그림 1. 마이크 입력 부분 회로

3. 음성인식 및 콘텐츠 네비게이션 시스템

T2SI(Text To Speaker Independent)는 음성 DB를 만드는 프로그램으로 텍스트를 입력하여 발음기호로 변환하여 SV(Speaker Voice) 템플릿을 만드는 것이다. 먼저 트리거(Trigger) 워드와 실제 음성인식에 사용될 단어를 작성한 후, 단어들의 인식률을 테스트한다. 인식률이 낮은 단어는 실제 발음기호를 수정하여 튜닝하는 과정을 거친다. 마지막으로 완성된 음성신호를 빌드하여 실제 데이터 파일과 이를 사용할 수 있도록 하는 라이브러리를 생성한다. 음성인식 DB 제작절차를 <그림 2>에 나타냈다.

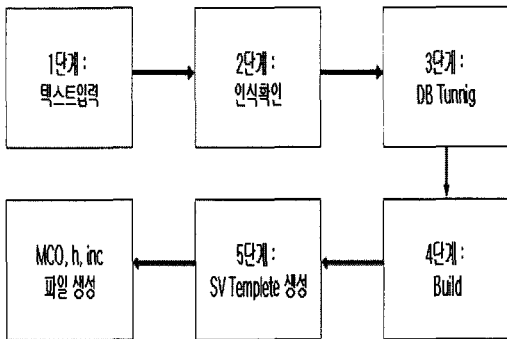


그림 2. 음성인식 DB

음성인식 DB를 기반으로 음성을 인식하는 과정은 다음과 같다. 트리거 신호가 입력될 때까지 입력을 기다린다. 트리거 신호 입력 후, 입력되는 음성신호와 일치하는 단어가 DB에 존재하면 음성인식을 수행하여 저장된 단어의 인덱스 값(Index)을 10진수의 형태로 출력한다. 입력된 음성신호와 일치하는 단어가 DB에 존재하지 않는 경우에는 트리거 입력상태로 다시 진입한다.

음성인식 결과를 윈도우즈 기반의 시스템으로 전송하는 인터페이스 시스템은 Atmega128을 사용하여 구성하였다. 외부 인터럽트 신호의 에지에서 인터럽트 서비스루틴이 실행되면 1-10까지의 인덱스를 RS-232C를 통해 콘텐츠 네비게이션 대상 시스템으로 전송된다. 이 과정이 완료되면 다시 피드백 되어 신호를 대기한다. <그림 3>에 통신 인터페이스 시스템의 절차를 나타냈다.

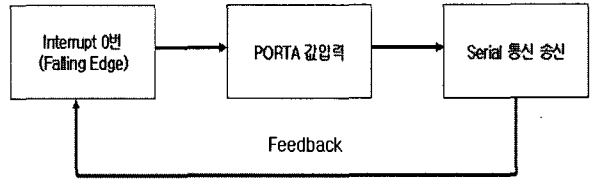


그림 3. 인터페이스 절차

콘텐츠 네비게이션은 음성인식 모듈의 출력포트를 실시간으로 감지하여 수신되는 인덱스 값에 따라 매크로를 실행시켜 응용 프로그램을 동작시키는 절차이다. 매크로를 사용하여 윈도우즈 어플리케이션(Windows Application)을 실행하는 절차를 <그림 4>에 나타냈다.

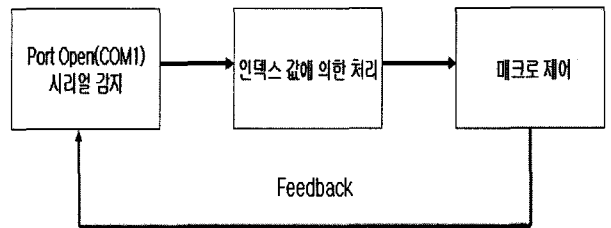


그림 4. 매크로 실행절차

매크로 제어부분은 크게 두가지로 분류될 수 있다. 모든 윈도우즈 어플리케이션(Windows Application)이 독립적으로 동작한다는 점을 고려하여, 먼저 프로그램을 실행시키기 위한 매크로 루틴과 이를 내부에서 동작할 수 있게 하는 매크로 루틴으로 구분하여 구현한다.

실행을 위한 매크로는 API 함수를 이용하여 실제 타겟 프로그램을 실행시킬 수 있도록 제어한다. 이를 통해 음성인식 시스템으로부터 출발하여 최종 시스템까지 넘어온 10진수 형태의 인덱스 데이터에 따라 각 1~4까지, 그리고 9~11까지의 문자를 사용하여 각 독립적인 프로그램들을 실행한다.

위 과정을 먼저 수행한 후, 두 번째로 5~8까지의 단어를 통해 내부 프로그램을 동작시킨다. 이는 키보드 값을 각 매크로 값으로 변환하여, 마치 키보드 입력값이 입력되는 것처럼 각 독립적인 어플리케이션을 제어하는 것이다. 예를 들어, 만약 종료루틴의 실행을 원할 경우, 종료라는 단어가 입력되었다고 먼저 가정하면, 이 후 콘텐츠 네비게이션 내부에서 동작하는 제어 어플리케이션(Application)은 윈도우즈의 종료 Function 키와 같은 Alt + F4 키가 입력되어 독립적인 프로그램을 제어할 수 있도록 한다.

또한 DVD, DMB, MP3 Player 등과 같은 다른 응용 프로그램에서도 이를 직접 제어할 수 있도록 각 상황에 맞는 매크로 키 값을 입력하여 이를 처리한다.

4. 실험 및 결과

본 연구에서 개발한 음성인식 기반 콘텐츠 네비게이션 시스템의 시제품을 (그림 5)에 나타냈다.

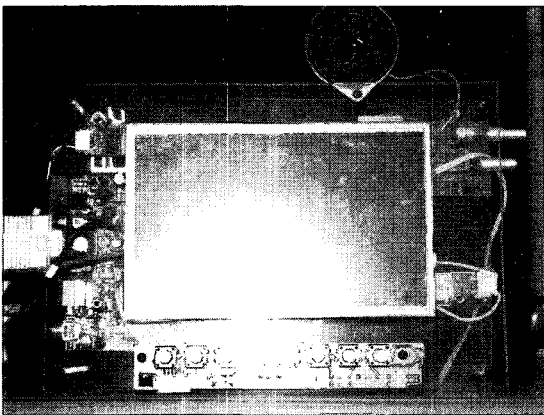
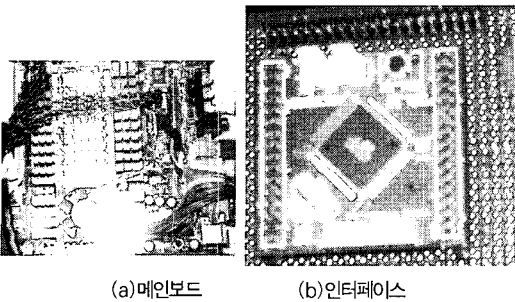


그림 5. 음성인식 기반 콘텐츠 네비게이션 시스템

본 연구에서 개발한 시제품의 성능검증을 위한 필드 테스트에서 정상적인 동작을 확인하였다. 본 연구에서 사용한 음성인식 모듈에는 약 20개의 입력 데이터를 저장할 수 있는 메모리가 내장되어 있다. 그러나 콘텐츠 네비게이션 시스템의 기능을 다양하게 구현할 목적으로 입력대상 단어를 15개 이상으로 증가한 필드 테스트에서는 인식률이 저하되는 것을 확인하였다.

5. 결론

본 연구에서는 T2SI(Text To Speaker Independent) 기반의 화자(話者)독립 방식으로 음성인식 신호를 콘텐츠 네비게이션 시스템의 제어신호로 활용하여 임베디드 시스템 및 PC 등에 설치된 윈도우즈 기반의 응용 소프트웨어를 제어할 수 있는 시스템을 개발하였다.

12개 내외의 음성인식 대상 단어를 사용하는 시제품의 성능검증을 위한 필드 테스트에서 정상적인 동작을 확인하였다. 한편, 본 연구에서는 외부환경으로부터 발생하는 각종의 소음을 제거하는 문제는 고려하지 않았으나, 외부 소음에 강건한 실시간 콘텐츠 네비게이션 시스템을 개발하기 위해서는 필수적으로 고려하여야 할 사항이다. 향후의 연구과제이다.

필드 테스트를 통하여 본 연구에서 개발한 시제품의 유용성을 확인한 결과, 본 시스템은 콘텐츠 네비게이션은 물론 가전기기 제어 및 홈 네트워크 등에 널리 응용될 수 있을 것으로 판단된다.

참고문헌

- [1] 류준형, 박찬경, 이의택, 김경태, "음성 Dialing 시스템을 위한 실시간 음성 인식 소프트웨어의 한구현", 대한전자공학회 하계종합학술대회 논문지 Vol.9 ,No 1, pp. 211~214, 1986.
- [2] 이경록, "News Data Analysis Using Acoustic Model Output of Continuous Speech Recognition", 한국콘텐츠학회 논문지 제6권 제10호, pp. 9~16, 2006.
- [3] 김영주, 김은주, 김명원, "Performance improvement of Command Speech recognition using Context Information", 한국컴퓨터종합학술대회 논문집(B), pp. 718~720, 2005.
- [4] *Microphone Housing*, TMSensory, 2005.
- [5] *VR STAMP Programmable Guide*, TM Sensory, 2005.
- [6] *VR STAMP Quick Guide*, TMSensory, 2005.