

Windows 환경에서 음성인식을 이용한 멀티미디어 게임의 구현

윤재선^o, 최광표, 홍광석
성균관대학교 전기전자컴퓨터공학부 HCI연구소

An Implementation of Multimedia Game using Speech Recognition for Windows

Jeh-Seon Youn, Kwang-Pyo Choi, Kwang-Seok Hong
HCI Lab. Electronic Engineering, Sung Kyun Kwan University
sunhci@chollian.net, kshong@yurim.skku.ac.kr

요 약

본 논문에서는 음성인식 알고리즘인 HMM을 사용하여 Windows 환경에서 온라인으로 사용할 수 있는 음성인식 게임 "Voice Illust Magic" 개발에 관하여 소개한다.

사용자와 컴퓨터가 상호작용(Interaction)할 수 있는 매체를 마우스와 키보드뿐만 아니라 게임에 필요한 명령어를 음성인식으로 실행함으로써 정보전달이 매우 효과적으로 이루어져 사용자가 접근하기 쉽고 편리하게 되었으며 의사전달 효율을 높일 수 있었다.

음성인식 과정은 온라인으로 마이크를 통해 들어온 음성을 자동으로 설정을 검출한 후, Mel-Cepstrum을 추출하여 Word 단위의 reference HMM과 비교하여 최적의 model이 선택되면, 윈도우즈에게 메시지를 보내어 마우스나 키보드가 동작하는 것과 마찬가지로 실행되도록 하였다. 또한, 입력 음성을 모든 reference pattern과 비교하는 것이 아니라 그 상황에 적용될 수 있는 표준 패턴을 한정함으로써 탐색시간을 줄일 수 있었으며 높은 인식률을 나타내었다.

I. 서 론

현재 인간의 편의를 위하여 인간과 컴퓨터 사이의 정보 전달을 원활하게 해주는 Man-Machine 인터페이스 기술에 대한 연구가 활발히 진행되고 있다.[1] 지금은 컴퓨터의 입력 장치로 키보드와 마우스가 주로 사용되고 있지만, 앞으로는 인간에게 보다 자연스럽고 편리한 음성용어를 이용해서 컴퓨터와 인간이 통신할 수 있는 기술이 적극적으로 개발되고 활용되고 있다. 요즘은 많은 컴퓨터 게임에 있어서 3D 용량에 사용!!와 그래픽이 뛰어난 많은 기능도 많아져 마우스를 바빠 움직여야 하고 여러 가지 단축키들을 이용함으로써 인해 양손을 모두

사용해도 의도한대로 게임을 하기 어려운 실정이다. 그러나 여러 가지 단축키들 대신에 음성을 이용한다면 좀 더 쉽고 편리하게 게임을 즐길 수 있을 뿐만 아니라 컴퓨터와 서로 의사소통을 함으로써 훨씬 더 현장감 있고 흥미진진하며 쉽게 게임에 몰두할 수 있을 것이다. 따라서 본 논문에서는 음성인식 기술을 이용하여 윈도우즈환경에서 게임 진행 명령어를 인식하여 그림을 맞추는 퍼즐게임 "Voice Illust Magic"을 구현하였다.

II. 음성인식게임의 구조

음성인식게임은 펜티엄 CPU 200 이상, Sound Blaster 호환카드, Windows 95의 PC환경에서 음성을 온라인으로 받아 들어 음성 인식을 통해 게임을 구현하였다.[2] 음성인식부분에서는 마이크를 통해 들어온 음성신호의 시작점과 끝점을 판별하는 기능을 수행하며 온라인으로 게임을 수행하기 위해서는 연속적인 제어를 필요로 한다. 구현된 음성게임에서는 오디오 디바이스의 자세한 부분을 제어할 수 있는 Low Level 오디오 디바이스를 이용하여 입력 음성신호의 녹음을 위한 버퍼를 설정하였다. 음성의 시작점과 끝점을 판별하기 위해 버퍼에 입력된 데이터의 에너지값을 계산하며, 계산된 에너지와 설정된 임계값을 비교하여 음성의 시작점과 끝점을 결정하였다.

PC에서의 사용자 인터페이스는 Visual C++ 5.0의 MFC를 이용하여 구현하였으며, 마이크로 입력된 신호는 사운드 카드에서 11.025 kHz, 16 bit로 양자화하고, 음성특징벡터는 Mel-Cepstrum 16차를 사용하였고, 훈련데이터는 게임에 필요한 데이터 28개를 남자 40명, 여자 20

명이 각각 3번씩 발생한 데이터를 자동으로 끝점을 검출하여 각 단어를 DIHMM으로 모넨링하였으며 VQ 레벨은 128로 사용하였다[1].

입력된 음성 구간에 대하여 인식 후보 단어와의 비교를 통해 인식된 결과는 윈도우즈에서 제공되는 함수를 사용하여 메시지를 전달하였다. 게임의 그림피율을 맞추는 상태에서 인식되는 방향키 명령어는 OnKeyDown 함수를 사용하여 키보드의 방향키입력 메시지를 보내 작동하도록 하였다. 또한 모든 메뉴의 명령어는 단축키를 설정하여 메뉴 명령어가 인식되면 PostMessage 함수를 사용하여 WM_SYSKEYDOWN, WM_KEYDOWN의 메시지를 전달하여 키보드의 "Alt + 키입력"의 효과를 게임에 수행하였다.

PC에서 구현된 음성인식 커준 게임의 구조는 그림 1과 같다.

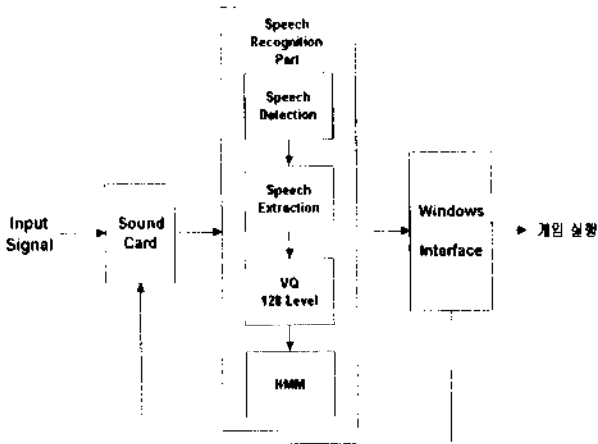


그림 1. 음성 인식 게임의 구조

III. Voice Illust Magic의 구현

음성으로 게임구현의 편의성을 위하여 메뉴에 나타나는 명령어와 메뉴에 선택된 후 풀다운 메뉴에 나타나는 명령어를 모두 다 인식하도록 게임을 구성하였으며, 대화상자에서 인식되는 명령어는 볼륨 단추에 적혀 있는 명령어를 인식하도록 하였다. 그리고, 메뉴 명령어를 잘못 인식하였을 경우에는 "취소"를 말하면 초기상태로 돌아가도록 하였다. 그림 2와 같은 "게임"메뉴가 선택되었을 때는 메뉴와 서브메뉴에서의 "시작, 환경설정, 종료, 보기, 도움말, 취소" 6개의 명령어가 인식된다. 이 상태에서 "취소"를 말하면, "게임, 보기, 도움말"을 인식할 수 있는 상태로 이동하게 된다. 그림 3과 같은 대화상자인 경우에는 볼륨 단추에 쓰여있는 "남자, 여자, 쉬운 게임, 어려운 게임, 확인, 취소" 6개의 명령어를 인식하도록 하였다.



그림 2. 메뉴에서 인식된 명령어의 예
인식 명령어 : 시작, 환경설정, 종료, 보기, 도움말, 취소

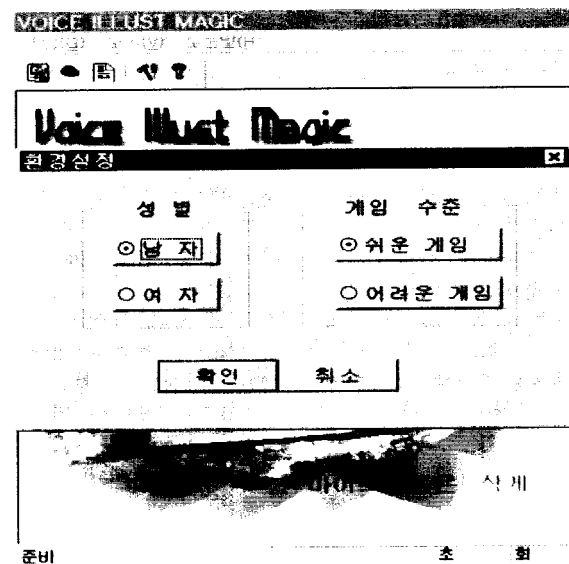


그림 3. 대화상자에서 인식된 명령어의 예
인식 명령어 : 남자, 여자, 쉬운 게임, 어려운 게임, 확인, 취소

3.1 상태에 따른 인식 명령어

피즐게임을 구현하기 위한 사용자 인터페이스를 완벽하게 구현하기 위해 인식되고자 하는 명령어의 참조패턴을 9개의 상태로 구성하였다. 상태에 따른 참조패턴을 구성함으로써 음성인식을 위한 탐색시간을 줄일 수 있었으며 상태변에 따른 Reject기능을 추가함으로써 더 높은 인식율을 나타내었다. 표 1은 9개의 상태에 따른 참조패턴을 나타내었다.

표 1. 상태에 따른 참조 패턴

상태	참조 패턴
상태 0	게임, 보기, 도움말
상태 1	시작, 환경설정, 종료, 보기, 도움말, 취소
상태 2	전체그림, 잠수관, 게임, 도움말, 취소
상태 3	사용방법, 만능사람들, 게임, 보기, 취소
상태 4	그림설정, 삼, 사, 오, 육, 확인, 취소
상태 5	남자, 여자, 쉬운 게임, 어려운 게임, 확인, 취소
상태 6	확인, 취소
상태 7	확인
상태 8	위, 아래, 선택
상태 9	위, 아래, 좌, 우, 왼쪽, 오른쪽, 게임, 보기, 도움말

3.2 음성인식 게임의 진행방법

그림 퍼즐 게임을 음성으로 제어하기 위해 표 1의 참조 패턴을 이용하여 그림퍼즐 맞추는 모드가 시작되는 상태 9까지의 오토마타를 구성하여 그림 4에 나타내었다[4]

그림 4에서 화살표는 시작점위치 위에 단어가 인식되면 다음의 상태로 이동되도록 표시하였다 예를 들어 상태 1에서 "환경설정"이 인식되면, 상태 5로 천이되며, 상태 4인 "시작"대화상자에서 "확인"이 인식되면, 퍼즐그림을 맞추는 모드가 시작되는 상태 9로 넘어가게 되며, 참조 패턴으로는 방향키 (위, 아래, 좌, 우, 왼쪽, 오른쪽)6개의 명령어와 메뉴명령어(게임, 보기, 도움말)3개를 들었다. 상태 9에서 방향키 명령어가 인식되면 다시 상태 9로 돌아오며, 메뉴명령어가 인식되면 상태 1, 2, 또는 상태 3으로 천이하게 된다. 그리고 대화상자에서 인식된 명령어가 "확인" 또는 "취소"가 인식되면 다시 그림 퍼즐을 맞추기 위한 상태 9로 이동한 후 다음의 음성구간을 인식하게 된다.

게임의 난이도를 조절하기 위하여 "시작"대화상자에 난이도를 조절하기 위해 매트릭스와 크기를 결정할 수 있는 "삼, 사, 오, 육"의 세 가지 모드를 두었다 예를 들어 "오"를 선택하게 되면 게임의 매트릭스형태는 5×5의 구조로 시작하게 된다. 또한, 음성으로 퍼즐을 맞추기가 어렵기 때문에 "환경설정"대화상자에서 쉬운 게임과 어려운 게임으로 구분하여 두었다 "쉬운 게임"은 그림 퍼즐클릭이 반대편으로 화살이 가능한 모드이며, "어려운 게임"은 화살이 불가능하도록 게임이 진행되는 모드이다. 그림 5에 "쉬운 게임"일 경우의 퍼즐 움직임에 대해 설명하였다. 설명에 따라 "남자"와 "여자"를 환경설정에서 설정하도록 되어있고, "환경설정"모드에

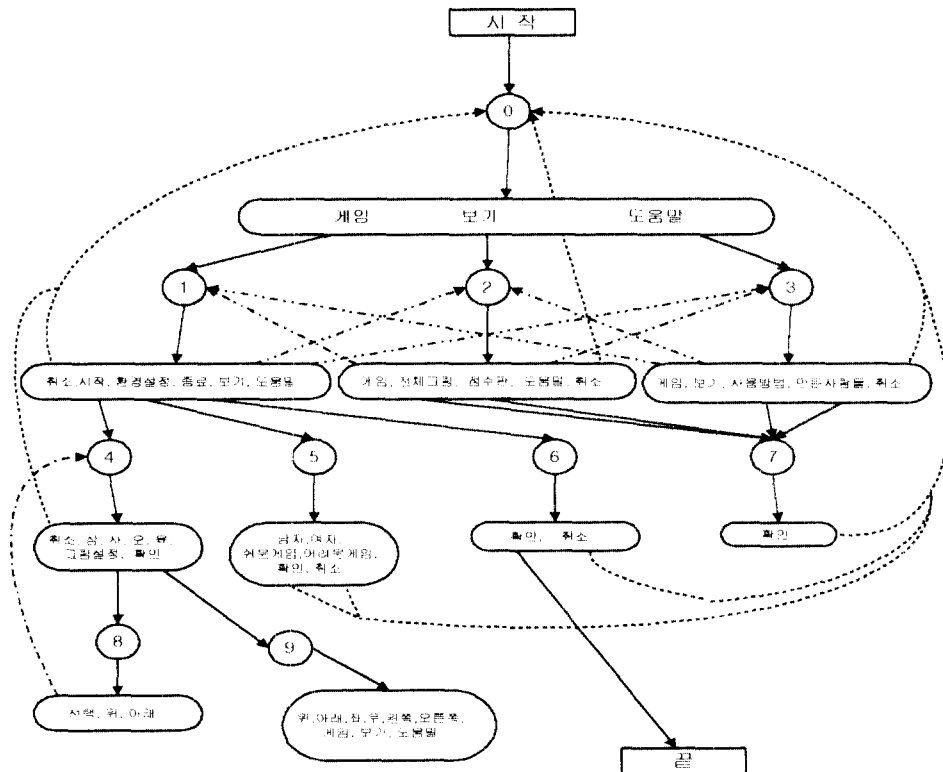


그림 4. 유한상태 오토마타

서 “확인”이 선택되면 VQ의 데이터와 표준패턴을 선택된 성별의 데이터로 변환하여 인식하도록 하였다. 퍼즐을 다 맞추게 되면 게임진행시간과 동작회수를 이용하여 점수로 환산하여 보여주도록 되어있다.

그럼 퍼즐 게임은 음성뿐만 아니라 마우스나 키보드로도 게임을 진행하도록 하였다. 따라서 메뉴모드에서의

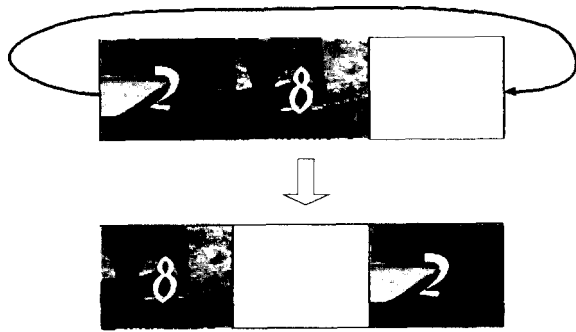


그림 5. 쉬운게임인 경우의 퍼즐 움직임

“왼쪽”이라고 인식하면 2번 그림은 회전율하여 오른쪽으로 이동한다

마우스 동작과 키보드 입력은 OnInitMenuPopup함수, 퍼즐 맞추기 모드에서는 OnLButtonDown함수를 사용하여 마우스의 동작을 인지하였고, 키보드의 경우에는 OnKcyDown함수를 이용하여 키보드 입력을 인지하였다

IV. 결론

멀티미디어의 대표적인 환경인 PC의 윈도우즈상에서 명령어를 음성으로 실행시킬 수 있는 음성인식 퍼즐게임 “Voice Illust Magic을 구현하였다.

상태에 따른 참조패턴을 구성함으로써 실시간으로 게임을 진행시킬 수 있었으며, 높은 인식률을 나타내었다. 그러나, 음성인식 기술을 게임에 적용하는데는 커다란 제약조건은 게임의 사운드가 마이크의 입력으로 들어가 음성을 제대로 인식하기가 부차 어렵다는 것이다. 앞으로 좀 더 완벽한 음성인식 게임을 구현하기 위해서는 입력으로 들어오는 게임의 사운드를 제거하는 방법에 대한 연구가 진행되어야 한다.

※ 이 게임은 성균관대학교 HCI연구실의 홈페이지 <http://hci.skku.ac.kr>에 올려 놓았습니다.

참고 문헌

- [1] J.C. Junqua, J.P. Haton, *Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, Boston, 1996.
- [2] Microsoft[®] Speech API Developer's Guide, Microsoft Press, 1998.
- [3] J.G. Wilpon, L.R. Rabiner, "A Modified K-Means Clustering Algorithms for use in Isolated Word Recognition," *IEEE Trans on Acoust. Speech and Signal Proc.*, Vol. 33, No. 3, pp. 587-594, 1985.
- [4] 심상엽, 이영재, "HMM에 의한 연속음성 인식시스템의 구현," 제 13회 음성통신 및 상호처리 워크샵 논문집, 1996.