

음성인식 끝말 이어가기 게임의 구현

김동환, 윤재선, 홍광석
성균관대학교 전기전자 및 컴퓨터공학부 휴먼컴퓨터 연구실

An Implementation of Word Relay Game using Speech Recognition

Dong-Hwan Kim, Jeh-Seon Youn, Kwang-Seok Hong
HCI Lab, Electrical & Computer Engineering, SungKyunKwan University

kdh2k@popsmail.com, sunhci@ece.skku.ac.kr, kshong@yurim.skku.ac.kr

요약

최근에 음성인식의 상용화가 급격히 추진되고 있다. 그러나 음성인식 응용제품의 부족과 음성인식 시스템의 성능문제로 인하여 일반인의 이용은 그다지 많지 않다. 본 논문에서는 연구실에서 만든 가변 어휘 음성인식기를 이용하여 음성인식 끝말 이어가기 게임을 구현하였다. 가변어휘 음성인식기는 VCCV(Vowel+Consonant+Consonant+Vowel) 기반의 화자독립으로 구현하였다. 끝말 이어가기 게임을 위해서 약 500만 어절이 포함된 문장에서 추출한 단어의 일부를 이용하여 사전을 구축하였고, 같은 음절로 시작하는 단어가 많은 경우에는 그 수를 제안하였다.

본 연구에서 구현한 음성인식 끝말 이어가기 게임은 제한된 단어사전을 이용하도록 하였으나 음성인식기의 성능향상과 완전한 사전구축이 이루어지면 음성인식을 이용한 언어 학습기나 게임 등의 개발과 이용의 활성화에 크게 기여할 것이라 생각된다.

I. 서론

본 음성인식 끝말 이어가기 게임은 음성인식인 실생활의 어느곳에 응용될 수 있는가라는 점에 초점을 맞춰 구현된 것이다. 현재 음성인식을 상용화하여 실생활의 여러 서비스측면에 사용하고

있는 부분은 많다. 하지만 일반인들에게 음성인식이라는 측면을 부각시키기에는 아직 많은 개선점과 홍보 부족등이 있다. 본 논문에서는 일차적 방법으로 일반인들이 널리 즐길 수 있는 게임을 선택하여 음성인식의 응용 측면을 좀 더 부각시키고자 한다. 프로그램 구성은 요즈음 게임과 비교해 볼때 부족한 면이 많지만, 게임적인 측면을 고려하여 음성인식 뿐만 아니라 음성합성을 추가시켜 컴퓨터가 끝말을 이어갈때 그 단어를 소리로 출력하도록 하였다. 또한 승패를 판가름하는 처리 루틴도 포함시켰다.

II. 끝말 이어가기의 구성

끝말 이어가기 프로그램에서는 사용되어질 단어 목록을 미리 구성한다. 우선 이전에 구축되어진 모음열 사전[1]으로부터 2음절에서 10음절에 해당하는 명사만을 추출한 결과 표 2.1과 같은 분포를 얻었다. 이 가운데서 실제로 끝말 이어가기에서 사용되어질 수 있는 단어(2-3음절)들을 프로그램을 통하여 수작업으로 선별한 결과 합성 명사를 제외시켜 총 9,687개의 단어들을 얻었다.

끝말 이어가기 프로그램의 구성은 일반적으로 게임 규칙에 적용되어지는 방법들을 그대로 적용하였고 그 규칙들은 다음과 같다.

(1) 한글에 적합한지를 검사하여 부적합한 단어

- 들이 입력되었을 경우 재입력하도록 한다.
- (2) User와 Computer 모두 상대방의 끝음절로 시작되는 단어만을 사용하도록 한다.
- (3) User와 Computer 모두 이전에 사용한 단어는 사용하지 못한다.
- (4) 합성명사는 사용하지 못한다.

표 2.1 추출된 명사 분포

계열	단어수	계열	단어수
ㄱ	9115	ㅅ	167
ㄴ	387	ㅇ	10484
ㄷ	2197	ㅈ	7616
ㄹ	3866	ㅊ	93
ㄺ	222	ㅌ	3005
ㄻ	296	ㅋ	551
ㄼ	4147	ㅍ	1329
ㄽ	5296	ㅑ	1847
ㅁ	97	ㅓ	4637
ㅂ	8589	총합	63941

끝말 이어가기 프로그램의 모든 처리 과정은 그림 2.1과 같은 순서로 이루어진다.

- (1) User가 사용할 단어 목록을 랜덤하게 생성해 낸다. 또한 후에 음성인식에서의 모델 생성을 위하여 단어 목록수를 제한하였다. User는 이 목록중에서 하나를 택하여 입력을 하게 된다.
- (2) 적합성 판단 과정에서는 위에서 언급한 끝말 이어가기 규칙들이 적용된다. 이때 User가 더 이상 사용할 단어가 없을 시에는 Computer가 승리하게 된다. 반대로 적합성에 합당한 단어가 선택되었을 시에는 User단어를 "사용된 단어 리스트"에 추가시키고, 제어권을 Computer로 넘긴다. Computer는 User 단어의 끝음절을 검사하여 그 음절로부터 시작되는 단어 목록을 생성한다. 이 때는 단어 목록수를 제한하지 않으며, Computer는 단지 단어 목록 중에서 랜덤하게 하나를 택하고 후에 음성합성으로 이 단어를 처리해 주면 된다. 이 때도 User와 마찬가지로 "사용된 단어 리스트"에 Computer 단어를 추가시킨다.
- (3) 승패 판단 과정은 User 또는 Computer 어느 한쪽에서 더 이상 끝말 이을 단어를 선택하지 못했을 때 처리되게 된다. 즉, 사용하고자 하는 단어가 이미 "사용된 단어 리스트"에 모두 들어 있

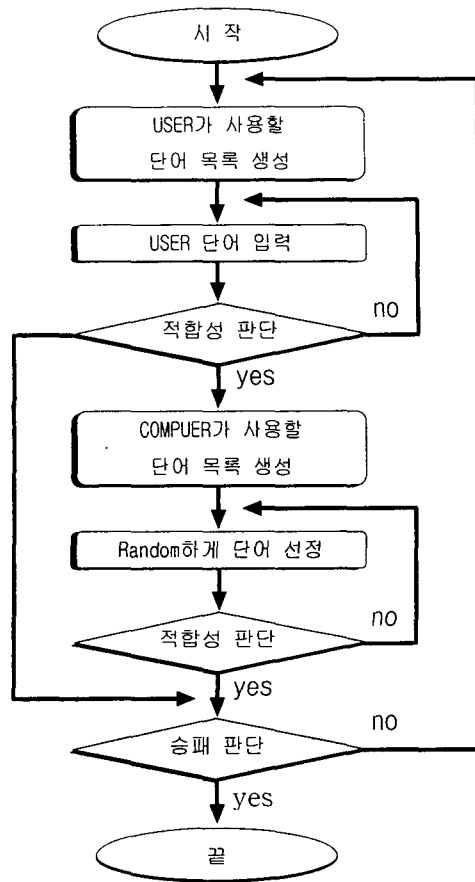


그림 2.1 프로그램 처리 과정

어 더 이상 사용할 단어를 생각하지 못했을 경우가 이에 해당한다. 반대로 끝말 이을 단어가 계속 존재한다면 (1)번 과정부터 다시 반복한다.

III. 음성 인식 및 합성과의 결합

본 끝말 이어가기 프로그램에서는 음성인식 시스템과 더불어 보다 게임적인 측면을 고려하여 음성합성 시스템도 함께 사용하였다.

3.1 음성인식 시스템

본 논문에서의 무제한 음성인식 시스템은 어떠한 단어라도 인식가능한 가변 어휘 인식 시스템을 말한다. 가변 어휘 인식 시스템의 구성은 먼저 사전에 남녀 화자로부터 입력받은 데이터를 가지고 Reference Model을 구성한다. 훈련데이터로는 PBW데이터 1,001개, 성명데이터 1145개,

단음절 데이터 521개를 사용하였고 남녀 각각 약 100여명 정도로부터 모델을 구성하였다. 인식 단계에서는 입력 음성으로부터 Mel Cepstrum 특징 파라미터를 추출해 내고, VQ 코드북을 통과시켜 Index Sequence를 얻는다. 이렇게 생성된 Index Sequence와 CV·VCCV·VC Reference Model로부터 생성되어지는 단어 Reference Model과의 비교가 이루어진다. 이때는 HMM 알고리즘이 적용되어지며 확률적으로 가장 큰 값을 갖는 단어가 선택되어 진다. [2] 인식 과정은 그림 3.1과 같다.

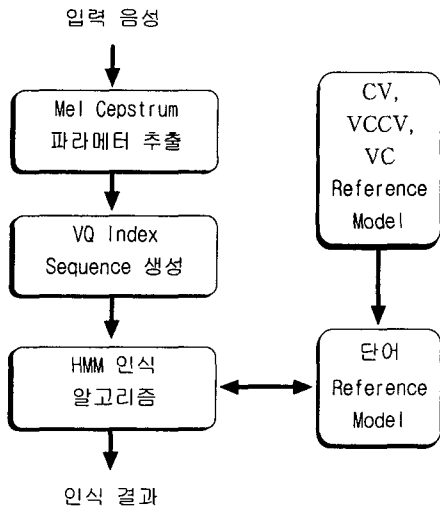
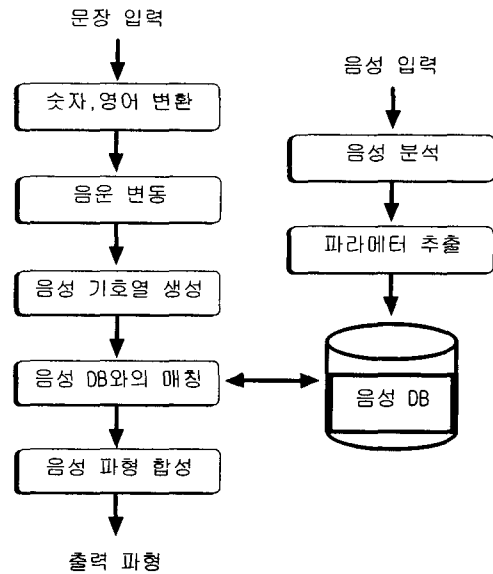


그림 3.1 음성인식 처리 과정

실제로 끝말 이어가기 프로그램에서 위의 인식 처리 부분은 User의 음성 입력을 받을 때 사용되어 진다. 그림 2.1에서 User 단어 목록을 생성하는 경우가 그림 3.1에서의 단어 Reference Model을 생성하는 경우에 해당한다. 따라서 매번 Computer가 사용한 단어의 끝음절을 이용하여 이로부터 시작되는 단어들에 대한 모델을 생성하는 것이다. 이때 본 프로그램에서는 User가 사용할 수 있는 단어 수를 제한하도록 하였다. 이 단어 수는 User가 임의로 조정할 수 있지만, 그 수가 증가할수록 그에 대한 모델 생성 시간이 길어지게 된다. 또한 모델 수가 증가함에 따라 인식률도 그만큼 떨어지는 결과를 낼 수 있다. 이렇게 생성된 모델과 User로부터 입력된 음성과의 비교를 수행하여, 그로부터 인식된 결과를 User 단어로 선정한다.



3.2 음성합성 처리 과정

3.2 음성합성 시스템

본 논문에서의 무제한 음성합성 시스템은 어떠한 문장이라도 음성으로 변환하여 출력해 내는 시스템(TTS)을 말한다. 크게 두 부분으로 나눌 수 있는데 음성 데이터 베이스를 구축하는 부분과 실제로 음성합성을 수행하는 부분이 이에 해당한다. 음성 데이터 베이스는 남녀 화자 각각 1명씩 약1,000여 단어를 발성시켜 구축하였다. 합성 방식은 TD-PSOLA 방법을 적용시켰으며, 합성 단위는 반음절을 사용하였다. 합성 과정은 그림 3.2와 같다. 우선 문장이 입력되면 우선 숫자나 영어에 대한 변환 과정을 거친 후, 한국어로 소리나는 대로 변환시키는 음운 변동 과정을 거친다. 이로부터 음성 기호열을 생성한 후 사전에 구축되어진 음성 DB로부터 해당하는 소리를 가져와 음성 합성을 하게 된다. [3] 실제로 프로그램에서 위의 음성합성 부분은 Computer가 랜덤하게 선택한 단어의 합성음을 출력시키는데 사용된다.

음성 인식과 음성 합성이 결합된 끝말 이어가기 프로그램의 실행 화면은 그림 3.3과 같다. 그림 3.3에서 User가 "정류장"이라는 단어를 발성하여 인식되었고, Computer는 "장소"라는 단어로 끝말을 이어 음성합성으로 소리를 출력해 낸다. 다음 User 단어 창에는 "장소"의 끝음절인 "소"

로 시작되는 단어를 랜덤하게 생성하게 되며, 이때 생성된 단어들에 대한 모델들이 만들어 진다. 여기서 User 단어수를 MAX=20으로 제한하였으므로 20개 이내의 단어 목록이 생성된다.

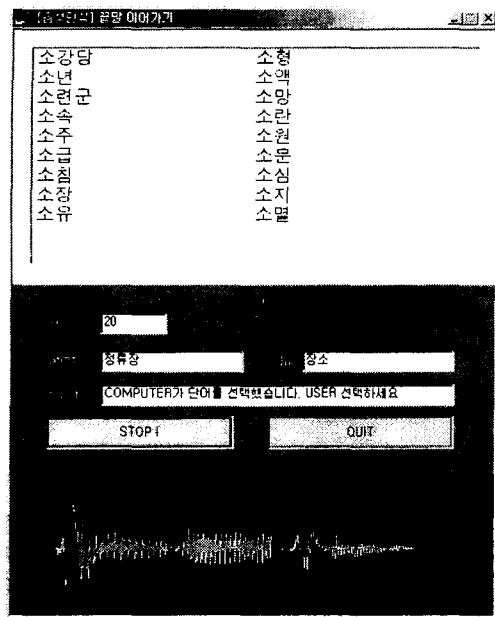


그림 3.3 프로그램 실행 화면

IV. 실험 및 결과

본 실험에서는 총 5명의 남성화자가 User 단어수를 5개, 10개, 20개로 정하여 각각 20번씩 단어 인식실험을 하여 인식률을 측정하였다. 그 결과는 표 4.1과 같았다.

User 단어장에는 매번 새로운 단어를 생성할

표 4.1 끝말 이어가기 게임의 인식실험

	5개	10개	20개
화자 1	20	19	17
화자 2	18	19	18
화자 3	19	18	18
화자 4	19	18	18
화자 5	20	20	19
평균인식률	96(%)	94(%)	90(%)

때마다 2음절, 3음절이 섞여 나오게 되어 있다. User 단어의 MAX값을 5, 10, 20으로 조정하여 실험한 결과 전체 평균 약 93.3%의 인식률을 나타내었다. 음절길이를 다르게 발생하거나 연음을 발생할 경우 인식률 저하를 보였다.

V. 결론

본 논문에서는 음성인식의 응용 분야 중 일부인 게임 프로그램에 대해 간략히 다루어보았다. 게임적인 측면만을 생각한다면 상당히 단순한 게임에 불과하겠지만, 음성인식과 음성합성을 게임에 접목시켜 응용 분야의 한 측면을 부각시켰다는데 그 의의를 둘 수 있다. 음성인식 프로그램 제작에 있어 가장 큰 문제로 대두되는 인식률 문제는 기존에 구축해 놓은 데이터 베이스 양이 음성인식을 하는데 있어 충분히 컸기 때문에 인식면에서는 별 문제가 되지 않았다. 따라서 향후 연구 과제로서 음절길이나 연음 문제가 있을 것이고, 본 논문의 실험 장소가 조용한 연구실 내였던 만큼 인식률도 좋게 평가되었지만 실외에서와 같이 잡음 현상이 두드러진 곳에서도 이와 같은 높은 인식률을 보장하지는 못할 것이다. 이러한 몇가지 문제들이 해결되었을 때 음성인식의 응용분야는 더욱 발전하리라 본다.

※ 본 연구는 “스피치인 테크놀로지”의 연구비 지원으로 이루어졌습니다.

참고문헌

- [1] 김동환, 윤재선, 홍광석, “무제한 단어 음성인식을 위한 모음열 사전의 구축”, 한국 신호처리·시스템학회 논문집 제1권 제1호, pp.177-180, 2000.
- [2] 윤재선, 홍광석, “무제한 어휘 독립 단어 인식 시스템의 구현”, 음성통신 및 신호처리 학술대회 논문집 제 17권 제1호, pp.145-148, 2000.
- [3] 양원렬, 윤재선, 홍광석, “영한 음차 변환을 이용한 무제한 음성인식 및 합성기의 구현”, 한국 신호처리·시스템학회 논문집 제1권 제1호, pp.181-184, 2000.