

비인식 대상 문장 거부 기능을 위한 음소 기반 인식 네트워크의 구성에 관한 연구

이병혁^o 하진영

강원대학교 컴퓨터정보통신공학과

ssamtwo^o@kwnu.kangwon.ac.kr, jyha@kangwon.ac.kr

Research on Recognition Network Structures for Non-recognition Sentence Rejection

Byeong-Hyeok Lee^o, Jin-Young Ha

Department of Electrical and Computer Engineering, Kangwon National University

요 약

음성인식 시스템에서 입력된 음성 데이터에 대해 비인식 대상에 대한 거부기능은 신뢰도 보장 측면에서 상당히 중요하다. 비인식 대상의 단어 거부 는 지금까지 여러 연구가 이루어져 왔으나, 문장 거부에 대한 연구는 사실상 부족한 실정이다. 본 논문에서는 비인식 대상 문장 거부기능의 신뢰도를 한층 높일 수 있도록, 음소 기반 네트워크에 유성자음(VC), 무성자음(C), 모음(V) 단위의 필러 음향 모델을 생성하여 다양한 음소기반 인식 네트워크의 구성방법을 적용하여 비인식 대상 문장에 대해 거부 기능을 구현하고, 그에 따라 인식률과 거부율이 달라질 수 있음을 보인다. 구현된 시스템에서 제안한 3가지 음소단위 인식 네트워크 중 문장의 각 단어별 필러 모델을 구성했을 때가 가장 좋은 구성임을 알 수 있었다.

1. 서 론

음성인식 기술이 발전함에 따라 음성인식을 이용한 다양한 음성대화 시스템이 등장하고 있다. 이 시스템들은 사용자를 시스템과 좀 더 편리하고, 자연스럽게 한다는 장점을 갖고 있다. 이런 대표적인 음성대화 시스템 중 영어 발음 교정 시스템에서는 인식 대상 어휘 외에 다른 어휘들이 입력되었을 때는 이를 처리하기 힘든 단점을 가지고 있다[1].

영어 발음 교정 시스템에 있어서, 사용자가 인식 단어 외 비인식 단어를 발성하였을 경우 신뢰도 측면에서 거부기능이 이루어져야 원활한 영어발음 교정을 할 수 있다. 즉, 인식 대상 단어 외의 말이 입력되었을 때, 이를 다른 단어로 오인식 하지 않고, 입력이 잘못되었음을 판단하는 음성인식 거절 기능에 대한 연구가 최근 들어 활발하게 진행되고 있다. 하지만 단어에 비해 문장에 대해서는 연구가 부족한 실정으로 본 논문에서는 거절기능 구현에 문장을 대상으로 한다[2-8].

그리고 음성 인식 시스템에서 입력된 음성 데이터에 대해 비인식 대상을 거부하는 기능은 신뢰도 보장 측면에서 상당히 중요하므로, 본 논문에서는 비인식 대상, 본 논문에서는 비인식 대상 문장에 대하여 거부기능을 구현하고, 그에 따라서 인식률과 거부율이 달라질 수 있음을 보인다. 음향모델에 사용된 음소 단위의 문장별 음소 단위 네트워크를 구축하여 인식기를 구성하였으며, 이 구축된 음소 기반 네트워크를 다양하게 구성하여 비인식 대상 문장에 대한 거부기능을 구현하였다. 인식 네트워크는 유성자음(VC), 무성자음(C) 그리고 모음(V) 단위의 필러 음향 모델을 생성하여 다양하게 변형하여 병렬 연결하여

적용시켰다.

본 논문은 2절에서 전반적인 시스템 구현과 필러 모델 구성을 통한 문장 거부 네트워크의 구성을 설명하고, 문장별 음소 단위 인식 네트워크의 구성을 보인 후 3절에서 실험 및 결과분석을 제시하고 4절에서는 결론을 맺는다.

2. 시스템의 구성 및 구현

2.1 전반적인 시스템 구현

비인식 대상 문장 거부 기능을 수행하기 위하여 HTK (Hidden Markov Model Toolkit) 3.2.1을 사용하였으며, 비인식 대상 문장 거부기능은 아래 그림 1과 같은 흐름도와 같이 진행된다[9].

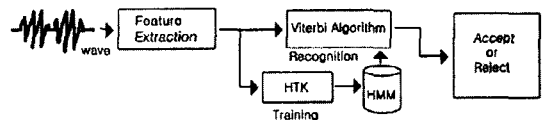


그림 1. 비인식 대상 문장 거부기능 수행 흐름도

2.2 필러 모델 구성을 통한 문장 거부 네트워크의 구성

비인식 대상 문장 거절 기능을 구현하는 방법으로는 대표적

으로 거절을 위한 별도의 모델을 사용하는 방법[2,6]과 적절한 후처리 과정을 통하여 인식결과를 확인 하는 방법이 있다.[2,4-6] 별도의 모델을 사용하는 방법에는 반단어 (Anti-keyword)를 모델링 하는 방법[5]과 필러 모델(Filler model)을 이용하는 방법[7]이 있는데, 본 논문에서는 유성자음(VC), 무성자음(C), 모음(V)단위의 가비지(Garbage)를 표현할 수 있는 필러 모델을 구축하고, 이 가비지 음향모델 단위를 연결하여 인식 네트워크에 병렬연결로 적용함으로써 거절기능을 구현하는 접근 방법을 택하였다[4].

2.3문장별음소단위인식네트워크의구성

필러 음향모델 단위를 연결하면 비인식 대상 문장에 대한 거부기능을 수행하는 인식 네트워크를 구성 할 수 있다. 언어 모델은 각 문장의 단어별 또는 문장별로 표준 발음 음소와 병렬로 연결 시켰으며, 그림 3, 그림 4, 그림 5에서와 같이 본 논문에서는 “Don’t miss the bus” 문장의 필러 인식 네트워크를 구성하였다. IPA(International Phonetic Alphabet: 국제 음성 기호) 발음 표기에 의하면 /d/ /oʊ/ /n/ /t/ , /m/ /ɪ/ /s/ , /ð/ /ə/ , /b/ /ɪ/ /s/ 로 나타낼 수 있으며, 이때 각 음소는 그림 2와 같은 HMM (Hidden Markov Model)모델의 구조를 갖는다.

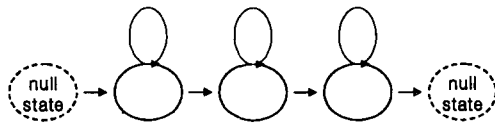


그림 2. 음소모델의 구조

첫 번째 음소기반 인식 네트워크의 구성으로, 그림3은 문장의 각 단어별로 가비지를 사용한 필러 모델의 음소기반의 인식 네트워크의 구조이며, Don't_g, miss_g, the_g, bus_g는 Don't, miss, the, bus의 가비지 형태이다. 이 가비지는 표 1에서와 같이 문장의 각 단어의 음소에서 유성자음(VC)과 모음(V)을 무성자음(C)로, 무성자음(C)는 유성자음(VC)로 변환한 것이다. 두 번째는 그림 4는 문장의 단어마다 유성자음(VC), 무성자음(C), 모음(V)을 반복 인식하는 네트워크의 구성이다. 세 번째 구성은 문장 전체에 대하여 유성자음(VC), 무성자음(C), 모음(V)을 반복하는 필러 모델을 구성하여 만든 음소 기반의 인식 네트워크 구조이다.

표 1. 문장의 음소 단위의 Garbage 변환

모델 타입	단어별	음소 단위 변환
General	don't	/d/ /oʊ/ /n/ /t/
Garbage	don't_g	C C VC VC
General	miss	/m/ /ɪ/ /s/
Garbage	miss_g	C C VC
General	the	/ð/ /ə/
Garbage	the_g	C C
General	bus	/b/ /ɪ/ /s/
Garbage	bus_g	C C VC

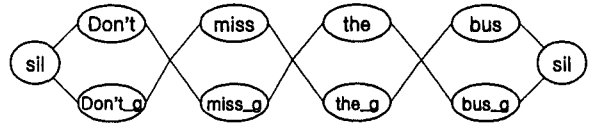


그림 3. 문장의 단어별 Garbage 모델 인식 네트워크

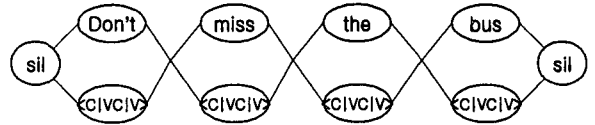


그림 4. 문장의 단어별 CIVCIV 반복 인식 네트워크

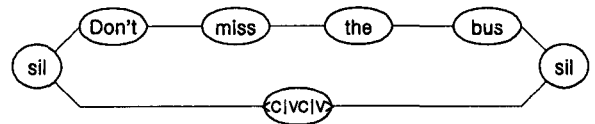


그림 5. 문장전체에 대한 CIVCIV 반복 인식 네트워크

3. 실험 및 결과 분석

3.1 실험 환경 및 DB

비인식 대상 문장 거부 기능을 수행하기 위하여 HTK를 사용하여 전반적인 인식 실험을 수행하였다. 본 실험에서 사용한 영어 음성 데이터베이스는 언어교육을 위한 영어 발음 교정용 음향 모델 생성을 목적으로 PC 환경에서 영어를 모국어로 사용하는 성인 400명이 문장을 발음한 영어 음성 DB를 사용하였다. 음성 데이터는 16Khz, 16bit, Mono, linear PCM 으로 녹음 되었으며, 남자 200명, 여자 200명이 각각 12~13문장을 발음한, 총 9920개의 영어 문장을 인식 실험을 위해 기존의 필러 학습 모델을 사용하여 실험에 사용하였다.

사용된 사전은 4.58MB의 크기를 갖는 표준 발음사전이며, HTK 3.2.1을 사용하여 비인식 대상 문장 거부기능을 위한 3가지 음소 기반 인식 네트워크의 구성에 따라 실험 하였다.

3.2 비인식 대상 문장에 대한 인식률과 거부율

문장에 대해서는 음소 기반 인식 네트워크 구성에 따라 인식률과 거부율이 달라질 수 있다. 아래 그림 6, 그림 7, 그림 8 은 실험에 사용된 3가지 음소기반 네트워크의 구성에 따라 가지율에 따른 Error rate의 결과를 FAR (False Acceptance Rate)과 FRR (False Rejection Rate)을 통하여 그래프로 표현 하였다. FAR은 엉뚱한 문장이 들어왔을 때 잘못 Accept한 경우이며, FRR은 제대로 된 완전한 문장이 들어왔을 때 잘못 Reject하는 경우를 말한다.

그림 6은 첫 번째로 구성했던 문장의 각단어별 가비지 모델 인식 네트워크를 사용한 결과이며, 그림 7은 문장의 단어별 CIVCIV를 반복시킨 인식 네트워크이다. 그림 8은 문장 전체에 대한 CIVCIV를 반복시킨 인식 네트워크의 결과이다. 각각의

결과 그래프에서 FAR과 FRR의 평균을 최소화하는 문턱 값 (Threshold)은 첫 번째 실험에서 Garbage rate이 27.5%, Error rate가 21% 일 때였으며, 두 번째 실험에서는 Garbage rate 55%, Error rate가 22%, 세 번째 실험에서는 Garbage rate 77%, Error rate가 36% 일 때 최소가 됨을 알 수 있었다.

따라서, 본 논문에서 제안한 3가지 인식 네트워크의 구성 중 문턱 값이 제일 작은 문장에서의 각 단어별 Garbage 인식 네트워크의 구성을 적용했을 때 가장 좋은 결과를 가져왔다.

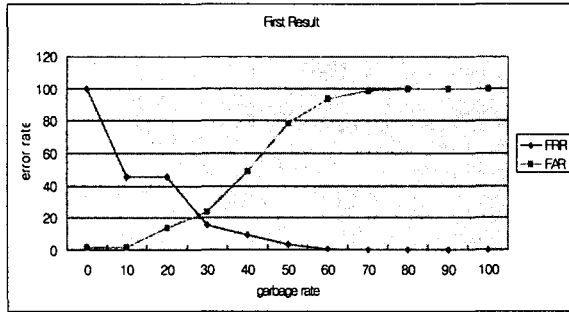


그림 6. 문장의 단어별 Garbage 모델 인식 네트워크에 따른 결과

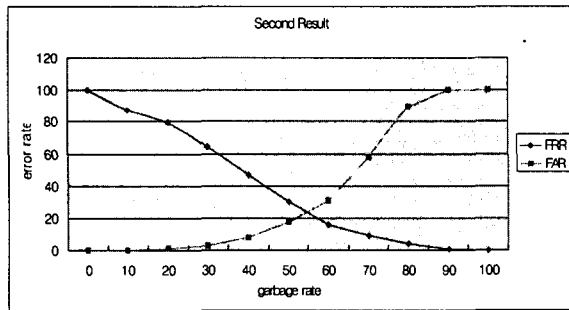


그림 7. 문장의 단어별 CIVCIV 반복 인식 네트워크의 결과

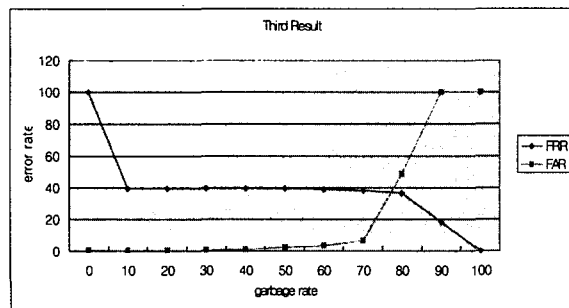


그림 8. 문장전체에 대한 CIVCIV 반복 인식 네트워크의 결과

4. 결론 및 향후 과제

본 논문에서는 음성인식 시스템에서 사용자의 잘못된 발성 패턴이 입력되었을 때 인식대상 문장은 검출하되 비인식 대상 문장은 거절하여 정확도를 높이는 음성 시스템에 있어서 인식 네트워크를 다양하게 구성함으로써 인식률과 거부율이 달라질 수 있음을 보였다. 본 실험은 인식 대상 문장을 거절하기 위해서 인식 대상 문장과 비인식 대상 문장간에 음소기반의 필터 모델을 구성하였으며, 필터 모델은 유성자음(VC), 무성자음(C), 모음(V)를 사용하여 구성하였다.

향후 연구에서는 본 논문에 사용된 3가지 네트워크의 구성 외에 좀더 다양한 음소 기반의 인식네트워크 구성의 실험이 필요하며, 문장 내 단어 수에 따른 적응적 Threshold를 적용하는 방법에 대한 연구가 필요하다[10].

감사의 글

본 연구는 강원대학교 BK21 사업단의 지원을 받았음을 밝힙니다.

참고문헌

- [1] 김무중, 김효숙, 김선주, 김병기, 하진영, 권철홍, "한국인을 위한 영어 발음 교정 시스템의 개발 및 성능 평가," 말소리, 제46호, 대한음성학회, pp.87-102, 2003.
- [2] 김동화, 김형순, 김영호 "고립단어 인식 시스템에서의 거절 기능 구현", 한국 음향 학회지, 제 16권 제 6호, pp.106-109, 1997.
- [3] RC.Rose "Discriminant wordspotting techniques for rejection nonvocabulary utterances in unconstrained speech," Proc. IEEE Conf, Acoustics, Speech, and Signal Processing, Vol.2, pp. 105-108, Mar. 1992.
- [4] 김무중, 김병기, 하진영 "음소기반 인식 네트워크에서의 비인식 대상 단어 거부 기능 성능 분석," 한국음향학회 하계 학술발표, 제22권, 1(s) pp.85-88, 한국음향학회, 2003.
- [5] 김우성, 구명완 "반음소 모델링을 이용한 거절 기능에 대한 연구," 한국 음향학회지, 제 18권, 제 3호, pp.3-9, 1999.
- [6] N.Moreau, D.Charlet and D.Jouvet,"Confidence Measure and Incremental Adaptation for The Rejection of Incorrect Data", ICASSP'2000, pp. 1807-1810, 2000.
- [7] Richard C. Rose and Douglas B. Paul, "A Hidden Markov Model Based Keyword Recognition System", ICASSP'90, pp. 129-132, 1990.
- [8] Hoi-Rin Kim, SiongHun Yi and Hang-Seop Lee, "Out-of-Variable Vocabulary Word Recognition, ICSP, Vol, 1, pp.337-339, 1999
- [9] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason Valtcho Valtcher, Phil Woodland, The HTK Book (for HTK 3.2.1), Entropic Cambridge Research Laboratory, 1999.
- [10] Sunil K. Gupta and Frank K. Soong, "Improved Utterance Rejection Using Length Dependent Thresholds", ICSLP, 1998.