음소기반 인식 네트워크에서의 단어 검출률을 이용한 문장거부

김형태* 하진영**

* 강원대학교 컴퓨터정보통신공학과
** 강원대학교 전기전자정보통신공학부

Sentence Rejection using Word Spotting Ratio in the Phoneme-based Recognition Network

Hyung-Tai Kim*, Jin-Young Ha**

* Department of Computer and Information Communication Engineering, Kangwon National Univ.
** Department of Electrical and Computer Engineering, Kangwon National Univ.

ds2swd@kangwon.ac.kr , jyha@kangwon.ac.kr

Abstract

Research efforts have been made for out-of-vocabulary word rejection to improve the confidence of speech recognition systems However, little attention has been paid to non-recognition sentence rejection. According to the appearance of pronunciation correction systems using speech recognition technology, it is needed to reject non-recognition sentences to provide users with more accurate and robust results. In this paper, we introduce standard phoneme based sentence rejection system with no need of special filler models. Instead we used word spotting ratio to determine whether input sentences would be accepted or rejected. Experimental results show that we can achieve comparable performance using only standard phoneme based recognition network in terms of the average of FRR and FAR.

Ⅰ. 서론

최근 들어 음성 인식 기술이 발전함에 따라 사용자들로부터 하여금 좀 더 자연스럽고, 편리한 인터페이스 방식을 갖는 음성인식 시스템이 등장하고 있으나, 음성인식 시스템의 신뢰도(confidence measure)를 높이기 위해서는 단순 인식기능 외에 부적절한 입력 패턴으로 인한 시스템의 오작동을 미리 방지 할 수 있는 거부 기능의 필요하다.[1][2] 여기서 신뢰도란 인식된 결과인 음소나 단어에 대해서, 그 외의 다른 음소나 단어로부터 그 말이 발화되었을 확률에 대한 상대 값을 말하며 특히 연속 음성인식 시스템의 경우에는 발화한 문장속의 한 단어만 다른 패턴의 비 인식 대상 단어 일지라도 오인식으로 판단됨으로써 엉뚱한 말로 인식을 해 버리는 문제점을 지니게 된다. 따라서 사용자가 고의나 실수로 인식 단어 외의 다른 단어를 발성 하였을 경우 이를 무조건 인식하려 하지 않고 거부하여 사용자에게 제대로 된 문장 음성을 재입력하게 하는 것이 중요하다 즉, 인식 대상 단어에 대해서만 인식을 하고, 그 외는 인식 결과를 내지 않고 거부함으로서 시스템의 성능을 향상시키고자 하는 것이 목적이다.[3]

본 논문에서는 이러한 거부기능을 구현하기 위하여 필러 모델(filler model)을 사용하지 않고 입력된 문장

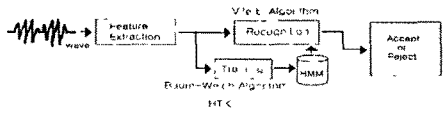
※ 이 논문은 2005년도 강원대학교 두뇌 한국 21 사업에 의하여 지원되었음.

내 단어를 선택하여 인식 할 수 있도록 음성인식 거부 네트워크를 구성하였고, 문장의 단어수별 인식된 단어와 인식에서 누락된 단어의 비율에 의해 문장에 대한 거부를 판단하도록 최적 문턱값을 구하였다. 그리고 단어수에 따른 문장의 거부 성능을 FAR(False Acceptance Rate : 제시된 문장 이외의 다른 문장이 입력되었을 때 거부하지 못하는 오류)와 FRR(False Rejection Rate : 제시된 문장이 입력되었음에도 이를 거부하는 오류)의 평균을 최소화 하는 것으로 평가를 하였다.[4][5]

Ⅱ. 문장 거부 네트워크의 구성

1. 시스템의 구성 및 구현

시스템은 음성인식 분야에서 우수한 성능을 보여 많이 사용하고 있는 HMM(Hidden Markov Model)을 채택하였고, 음소 모델의 훈련 및 인식 실험을 위해 연구용으로 공개된 HTK(Hidden Markov Model Toolkit)를 사용하여 <그림 1>과 같이 전반적인 비인식 대상 문장 거부 기능을 수행하였다.

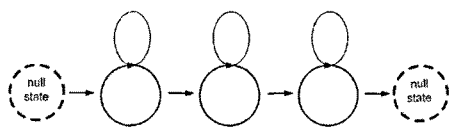


<그림 1> 비인식 대상 문장 거부 기능 수행 흐름도

비인식 대상 문장거부 기능을 구현하는 방법으로는 대표적으로 거부를 위한 별도의 모델을 사용하는 방법[6]과 적절한 후처리 과정을 통하여 인식 결과를 확인하는 방법[2][5]이 있으나 본 논문에서는 별도의 필러 모델이나 후처리 과정 없이 문장내의 단어 인식을 선택(option)처리하여 인식 네트워크에 연결 적용함으로써 거부기능을 구현하는 방법을 택하였다.

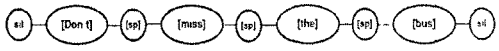
2. 단어 선택 검출 모델과 음소단위 인식네트워크 구성

IPA(International Phonetic Alphabet: 국제 음성 기호) 발음 표기에 의하면 /d/ /óu/ /n/ /t/, /m/ /i/ /s/, /ð/ /ə/, /b/ /ʌ́/ /s/ 로 나타 낼 수 있으며, 이때 각 음소는 <그림 2>와 같은 HMM 모델의 구조를 갖는다.



<그림 2> 음소모델의 구조

단어 검출률을 이용한 문장 거부모델은 각 문장내의 모든 단어에 대하여 표준 발음의 인식 처리를 선택적으로 하게끔 이들 음소를 직렬 연결함으로서 비인식 대상 문장에 대한 거부기능을 수행하는 인식 네트워크를 구성 할 수 있다. 본 논문에서는 "Don't miss the bus" 문장의 단어 선택 인식 네트워크를 <그림 3>과 같은 구조로 하여 전체 문장을 구성하고 있는 단어를 선택적으로 인식하여 인식에서 누락된 단어의 비율을 구하였다.



<그림 3> 문장의 단어 선택 검출 모델

Ⅲ. 실험 및 결과 분석

1. 실험 환경 및 데이터베이스

비인식 대상 문장 거부 기능을 수행하기 위하여 HTK V.3.2.1을 사용하여 음향 모델 훈련과 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 PCM으로 녹음되었으며, 남자 200명, 여자 200명이 각각 발음한, 총 4120개의 영어 문장을 사용하였다.

사용된 사전은 4.58MB의 크기를 갖는 표준 발음사전이며 CMU 사전을 근간으로 하여 만들었으며, sp, sil 외 108개의 음소 모델을 사용하였다. 또한 가우시안 믹스처(Gaussian Mixture) 7개의 Continuous density HMM을 사용하였다.

2. 실험 및 결과 분석

본 논문에서의 비인식 대상 문장 거부 기능을 실현하는 기본적인 방법으로 인식 네트워크에서 최적 경로를 Viterbi 알고리즘으로 구한 결과에 표준 음향 모델의 인식된 모델의 비율과 인식이 누락된 모델이 어떤 비율로 포함 되어 있는가를 조사하는 것이다. 입력 음성에 대해 네트워크에서 주어진 문장의 모든 단어가 인식 된다면 입력 음성은 주어진 문장을 발화한 것이라고 판단할 수 있고, 반대로 모든 단어가 인식이 되지 않았다면 주어진 문장 대신 다른 문장을 발화하거나 소음이 들어간 것이어서 거부해야 한다고 판단 할 수 있다. 하지만, 이러한 이상적인 결과는 대부분의 경우 어렵고, 주어진 문장의 발화 데이터에 대해서도 일부는 인식이 누락된 것으로 나올 수도 있고, 주어진 문장과 전혀 다른 발화 데이터에 대해서도 정상 인식된 결과가 나올 수 있다

따라서 전체 인식 결과 중 인식 누락된 모델이 차지하는 비율을 문턱값으로 설정하여 입력 발화 데이터에 대해서 이 문턱값을 넘는 누락비율을 보이면 거부하고, 그렇지 않으면 받아들이는 방법을 사용한다. 문턱값을 낮추면 FRR(False Rejection Rate)가 커지고, 문턱값을 높이면 FAR(False Acceptance Rate)이 커지기 때문에 FAR과 FRR의 평균치를 최소화 할 수 있는 문턱값을 선택하는 방법을 택하였다. 또한 이를 목표 문장의 단어수(2~6단이)에 따라 가변적인 길이 의존 문턱값을 사용하는 방법으로 실험을 하였다.

본 실험의 결과를 얻기 위하여 먼저 실험에 사용될 각 문장의 단어수 별로 2~6단어의 문장으로 분류 하였다. 그 후, 분류된 문장들의 각각의 인식누락 단어수를 계산하여 <표 1>, <표 2>에서와 같이 백분율로 나타내고 이것을 누적하여 계산된 결과를 <표 3>에 나타내었다. <표 1>, <표 2>에 나타낸 Result_S는 인식 모델과 같은 음성 문장의 데이터로 인식 실험을 한 결과로서 FRR로 나타내었으며, Result_O는 전혀 다른 음성 문장의 데이터로 인식 실험한 결과로서 FAR로 나타내었다.

단어검출 누락비율(%) / 단어수	0	1~10	10~20	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	백분율(%)	문장수
2개	78.75	0	0	0	0	20	0	0	0	0	1.25	100	160
3개	62.03	0	0	0	29.84	0	0	7.65	0	0	0.46	100	1280
4개	44.31	0	0	34.08	0	15.3	0	0	7.25	0	0.06	100	1640
5개	29.17	0	39.17	0	22.09	0	6.77	0	2.7	0	0.1	100	960
6개	17.5	0	38.75	0	30	10	0	3.75	0	0	0	100	80

<표 1> 단어수별 단어검출 누락수의 백분율값 Result_S

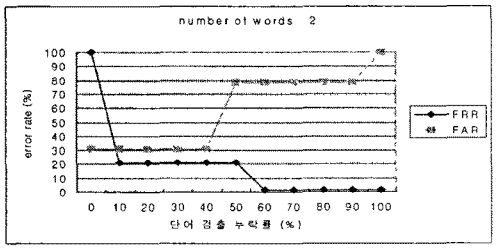
단어검출 누락비율(%) / 단어수	0	1~10	10~20	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	백분율(%)	문장수
2개	30.62	0	0	0	0	47.5	0	0	0	0	21.88	100	160
3개	4.15	0	0	0	33.04	0	0	5.5	0	0	7.81	100	1280
4개	1.46	0	0	9.75	0	36.6	0	0	44.63	0	7.56	100	1640
5개	0	0	1.66	0	13.22	0	33.54	0	46.9	0	4.68	100	960
6개	0	0	0	0	1.25	7.5	0	56.25	0	35	0	100	80

<표 2> 단어수별 단어검출 누락수의 백분율값 : Result_O

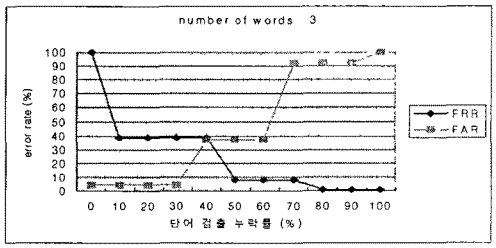
단어검출 누락비율(%) / 단어수	0	1~10	10~20	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	비고	문장수
2개	100	21.25	21.25	21.25	21.25	21.25	1.25	1.25	1.25	1.25	1.25	FRR	160
	30.62	30.62	30.62	30.62	30.62	78.12	78.12	78.12	78.12	78.12	100	FAR	
3개	100	37.95	37.95	37.95	37.95	8.11	8.11	8.11	0.46	0.46	0.46	FRR	1280
	4.15	4.15	4.15	4.15	37.19	37.19	37.19	92.19	92.19	92.19	100	FAR	
4개	100	56.69	56.69	56.69	22.61	22.61	7.31	7.31	7.31	0.06	0.06	FRR	1640
	1.46	1.46	1.46	11.21	11.21	47.81	47.81	47.81	92.44	92.44	100	FAR	
5개	100	70.83	70.83	31.66	31.66	9.57	9.57	2.8	2.8	0.1	0.1	FRR	960
	0	0	1.66	1.66	14.88	14.88	48.42	48.42	95.32	95.32	100	FAR	
6개	100	82.5	82.5	43.75	43.75	13.75	3.75	3.75	0	0	0	FRR	80
	0	0	0	0	1.25	8.75	8.75	65	65	100	100	FAR	

<표 3> 단어수별 단어검출 누락수의 백분율 누적 값

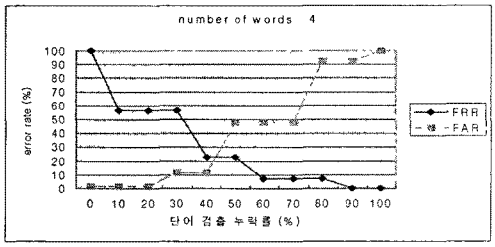
FRR과 FAR이 평균이 최소가 되는 가비지 비율의 문턱값을 찾기 위해 <표 1>~<표 3>에서 계산 되어진 결과를 사용하여 <그림 4>~<그림 8>과 같이 FRR과 FAR의 그래프로 나타냈다. 그리고 각각의 단어별로 오류가 최소가 되는 최적의 단어검출 누락비율을 구할 수 있었다.



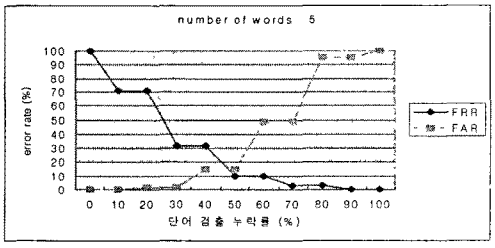
<그림 4> 단어수가 2개일 때 FRR과 FAR의 그래프



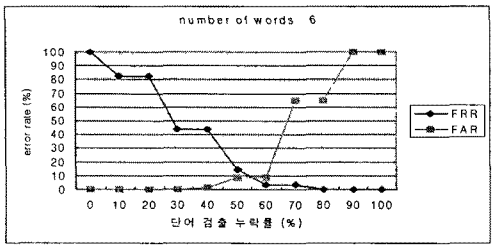
<그림 5> 단어수가 3개일 때 FRR과 FAR의 그래프



<그림 6> 단어수가 4개일 때 FRR과 FAR의 그래프



<그림 7> 단어수가 5개일 때 FRR과 FAR의 그래프



<그림 8> 단어수가 6개일 때 FRR과 FAR의 그래프

실험결과 단어수가 2개일때와 3개일때만 FRR과 FAR의 평균값이 특정구간에서 계속 같은 값을 유지하여 시작되는 값을 오류가 최소가 되는 최적의 평균 단어 검출 누락률 기준으로 잡았으며, <표 4>에서와 같이 나머지 4단어 이상부터는 단어의 수가 증가하면 할수록 이 비율도 증가 하는 것을 볼 수 있었다. 단어수를 고려하여 최적의 적응 문턱값을 가변적으로 적용하였을 때 FRR과 FAR의 평균값은 17.25%이고 2~6개까지의 각 단어별 FRR과 FAR의 평균값을 구하였다.

단어수	문 턱 값	FRR과 FAR의 평균(%)	문 장 수
2개	10	25 94	160
3개	10	21 05	1280
4개	40	16 91	1640
5개	50	12 23	960
6개	60	6 25	80
평균		17 25	계 · 4120

<표 4> 단어수별 오류가 최소가 되는 최적의 평균 단어 검출 누락률

Ⅳ. 결론 및 향후 과제

본 논문에서는 음성 인식을 이용한 발음 교정 시스템에서의 거부 성능을 표준 음소 모델만을 이용한 단어 검출률을 활용하여 구현하는 방법을 보여준다.

음성인식 거부 네트워크를 필러모델이 아닌 인식 단어 검출률에 의존함으로써 음성인식시스템의 신뢰도를 높임을 보였다. 인식 대상 문장을 거부하기 위하여 문장 내 비인식 단어들을 선택적으로 검출하여 인식된 단어와 인식에서 누락된 단어의 비율에 의해 문장에 대한 거부를 판단하고, 이를 최적화 시킬 수 있는 문턱값을 구할 수 있었고, 문장을 구성하고 있는 단어수에 따른 문장의 거부 성능을 FAR과 FRR의 평균을 최소화 하는 값을 구했다.

향후 연구에서는 보다 다양한 음성인식 네트워크 구성의 실험이 필요하며, 문장의 거부 방법에 있어서 필러모델을 사용하여 문장 거부 성능을 높이는 방법과 단어 검출률 이용한 방법, 필러모델 방법을 통합한 새로운 방법을 연구하는 것도 필요하다.

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] 김무중, 김병기, 하진영 "음소기반 인식 네트워크에서의 비인식 대상 단어 거부 기능 성능 분석," 한국음향학회 하계학술발표, 제22권, 1(s) pp.85-88, 한국음향학회, 2003.

[3] 김우성, 구명완 "반음소 모델링을 이용한 거절 기능에 대한 연구," *한국 음향학회지*, 제 18권, 제 3호, pp.3-9, 1999.

[4] Sunil K. Gupta and Frank K. Soong, "Improved Utterance Rejection Using Length Dependent Thresholds", *ICSLP*, 1998.

[4] 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.

[5] B.-H. Lee and J.-Y. Ha, "Length Dependent Threshold for Non-Recognition Sentence Rejection," The 4th Asia Pacific International Symposium on Information Technology, pp.462-465 Gold Coast, Austrailia, 2005.

[6] 김동화, 김형순, 김영호 "고립단어 인식 시스템에서의 거절기능 구현", *한국 음향 학회지*, 제 16권 제 6호, pp.106-109, 1997.