Aurora 특징파라미터 추출기법에 따른 한국어 연속숫자음 전화음성의 인식 성능 비교

김 민 성 *, 정 성 윤 *, 손 종 목 *, 배 건 성 *, 김 상 훈 * *

경북대학교 전자공학과 *, 한국전자통신연구원 * *

Performance Comparison of Korean Connected Digit Telephone Speech Recognition According to Aurora Feature Extraction

Min Sung Kim *, Sung Yun Jung *, Jong Mok Son *, Keun Sung Bae *, Sang Hun Kim * *

School of the Electronic & Electrical Engineering, Kyungpook National University *,

Electronics Telecommunications Research Institute * *

kmslove@mir.knu.ac.kr

Abstract

To improve the recognition performance of Korean connected digit telephone speech, in this paper, both Aurora feature extraction method that employs noise reduction 2-state Wiener filter and DWFBA method are investigated and used. CMN and MRTCN are applied to static features for channel compensation. Telephone digit speech database released by SITEC is used for recognition experiments with HTK system. Experimental results has shown that Aurora feature is slightly better than MFCC and DWFBA without channel compensation. And when channel compensation is included, Aurora feature is slightly better than DWFBA with MRTCN.

I. 서론

음성인식 기술의 발달과 휴대전화 사용의 증가로 전화망을 이용한 음성인식 기술이 자동 음성다이얼링이나 증권안내, 자동응답시스템 등의 다양한 곳에 적용되고 부분적으로 실용화 단계까지 이르렀다. 하지만 전화망 환경에서의 음성인식은 채널왜곡과 부가잡음으로 인해 원래의 음성이 훼손되어 인식률이 저하되기 때문에, 이를 개선하려는 연구가 여러 방면에서 계속되어 왔다[1]. 이러한 연구들에는 특징추출 단계에서 잡음제거 기법이나 전화채널보상 기법을 이용한 것과 HMM(Hidden Markov Model) 기반의 적응 방법을 적용한 것이 있다. 그 중에서 잡음제거 기법을 이용한 것으로, ETSI(European Telecommunications Standards Institute)에서 표준으로 정한 Aurora 특징추출 방법은 단말기의 전처리 단에서 2단계 Wiener 필터를 이용한 잡음제거 기법이 포함된 특징추출 방법으로 휴대 단말기를 이용한 음성인식에서 좋은 성능을 나타내고 있다[2]. 또한 켑스트럼 정규화 방법을 통해 전화채널의 영향을 줄이고자 하는 연구도 이루어져 왔으며 대표적인 채널보상기법으로는 CMN (Cepstral Mean Normalization)과 MRTCN(Modified Real Time Cepstral Normalization) 방법 등이 있다[3, 4]. 이외에도 전화채널과 부가잡음에 대한 강인성을 높여주는 DWFBA(Direct Weighted Filter Bank Analysis)와 같은 특징파라미터 추출 방법이 연구되기도 했다[5].

본 논문에서는 한국어 연속숫자음 전화음성의 인식성능을 개선하기 위한 연구로 각 특징파라미터 추출 기법에 따른 연속숫자음 전화음성의 인식실험을 수행하고 인식률을 비교하였다. 이를 위해, 기존 특징파라미터인 MFCC(Mel Frequency Cepstrum Coefficient) 및 DWFBA 특징추출 기반의 특징파라미터를 기본으로

하고, ETSI에서 표준으로 정한 Aurora 프로젝트의 DSR(Distributed Speech Recognition)에서 휴대 단말기의 전처리단에 적용한 특징파라미터를 사용하여 인식성능을 비교하였다. 또한 Aurora 특징추출 기법에 채널 보상기법을 적용하여 그에 따른 인식실험을 수행하였다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 Aurora 특징추출 방법과 DWFBA 특징추출 방법 및 채널보상기법에 대해 설명하고 3장에서는 본 연구에서의 인식 실험과정에 대해 기술한다. 4장에서 인식 실험 결과를 제시한 후, 5장에서 결론을 맺는다.

II. 특징추출 기법 및 채널 보상기법

본 논문에서 전화음성 인식성능 향상을 위해 특징파라미터 추출단계에서 적용한 Aurora 특징추출 방법과 DWFBA 기법 및 채널 보상기법에 대해 설명한다.

1. AURORA 특징추출

ETSI에서 표준으로 정한 휴대단말기에서의 음성인식 전처리단은 크게 3 부분으로 이루어져 있다[2]. 첫 번째는 특징추출 단계이며 두 번째는 특징파라미터의 전송률을 높이기 위한 압축단계, 마지막으로 전송 시 발생하는 에러를 최소화하기 위한 채널코딩 부분으로 이루어져있다. 본 논문에서는 전화음성 인식을 위해 특징추출 단계만을 사용하였으며 이를 Aurora 특징추출 방법이라고 정의했다. 특징추출 단계에서는 2단계 Wiener 필터를 이용한 잡음제거 단을 포함하고 있다. 이 2단계 Wiener 필터는 전화음성에 내재한 잡음을 제거하여 인식성능을 향상시키는 역할을 할 수 있을 것으로 생각된다. 첫 번째 단계에서는 전체적인 잡음을 제거하면서 유색 잡음을 백색잡음화 시켜준다. 두 번째 Wiener 필터에서는 백색잡음의 correlation 특성을 이용하여 잡음을 한번 더 제거하게 된다[2]. 잡음제거시 입력신호의 VAD(Voice Activity Detection) 정보를 이용하여 먼저 잡음신호의 스펙트럼을 추정하고 두 번째 Wiener 필터에서 잡음신호와 음성신호에 따른 이득을 조절하여 잡음제거 정도를 다르게 적용하고 있다[2].

특징추출 단계에서 WP(Waveform Processing)는 잡음이 제거된 신호의 SNR(Signal Noise Ratio)에 따른 가중치를 주어 전체적인 SNR을 증가시켜 주는 역할을 한다. 즉, 상대적으로 에너지가 큰 부분을 강조시켜주는 효과를 가져온다. 그리고 특징추출 단계의 마지막은 BE(Blind Equalization)으로 추출된 특징파라미터를 주파수 상에서 평탄한 특성을 보이는 신호의 캡스트럼으로 등화시키는 역할을 한다. 일종의 켑스트럼 정규화 기능이므로 Aurora 특징파라미터에 채널 보상기법을 적용할 때에는 이를 제거하고 실험하였다. 그림 1에 Aurora 특징추출 기법의 과정을 나타내었다. Aurora 특징추출에서 음성분석은 프레임 길이는 25 ms, 프레임 이동은 10 ms, 필터뱅크의 수는 23개, 켑스트럼 차수는 13차(0차 켑스트럼), 전처리계수는 0.9로 표준에서 정하고 있다. 전화음성 인식실험에서도 위와 같은 조건으로 특징파라미터를 추출하여 결과를 비교하였다.

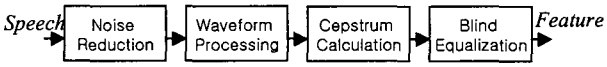

그림 1. Aurora 특징추출 과정

2. DWFBA 특징 추출

DWFBA는 켑스트럼이 주변 잡음의 간섭에 강인하도록 하기위해 log 필터 뱅크 에너지의 높은 에너지 부분을 강조해 주는 것인데, MFCC 추출과정 중 DCT(Discrete Cosine Transform) 전에 critical band의 log 에너지에 비례하도록 하는 가중함수를 곱하여 특징파라미터를 추출하는 방법이다[5]. 관련 수식은 식 (1), (2)와 같다. 여기서 L과 Q는 특징파라미터의 차수, critical band의 수를 나타낸다. $e(i)$는 i번째 critical band의 energy이다.

$$C_m = \sum_{i=1}^{Q} w(i)\log[e(i)+1.0]\cos[m(\frac{2i-1}{2})\frac{\pi}{2}] \quad 1 \le m \le L \quad (1)$$

$$w(i) = \frac{\log[e(i)+1.0]}{\sum_{j=1}^{Q}\log[e(j)+1.0]} \quad (2)$$

3.채널보상기법

전화채널 특성 변화가 켑스트럼 영역에서는 합의 형태로 나타나고 그 특성이 단시간에 큰 변화가 생기지 않는다고 볼 수 있으므로 전체 켑스트럼의 평균값을 빼줌으로써 전화채널 특성의 변화에 의한 영향을 줄여주는 방법이 CMN 기법이다[3]. 본 논문에서는 CMN과 MRTCN을 적용하였는데, CMN은 4연숫자음의 켑스트럼 평균으로 켑스트럼을 정규화 하는 방식이며, MRTCN은 4연숫자음 마다 켑스트럼을 구해 전체 켑스트럼의 평균을 추정하고 전체 켑스트럼의 분산도 같은 방식으로 추정하여 켑스트럼 영역에서 정규화해 줌으로써 채널에 의한 영향을 줄여주는 보상기법이다.

III. 인식 실험

4연숫자음 인식을 위해 음향모델의 생성 및 훈련과 인식은 HTK(Hidden Markov ToolKit)를, 특징파라미터 추출은 ETSI에서 공개한 프로그램과 직접 구현한 프로그램을 사용하였다. 인식시스템에서 정의한 기본 유사음소는 4연숫자음의 개별숫자음과 묵음을 표현하기 위해 초성 4개, 종성 5개, 모음 6개 그리고 묵음 2개로 총 17개의 유사음소를 정의하였다. 음향모델로는 CHMM(Continuous Hidden Markov Model)을 사용했으며 연속숫자음 인식시 앞뒤 음소의 영향을 고려하기 위해 트라이폰 단위로 인식실험을 하였다. 트라이폰 모델을 사용함에 따라 훈련시켜야 할 모델의 수가 증가하므로 모델별 훈련데이터의 수를 확보하기 위해 TBC(Tree Based Clustering) 기법을 이용하였다[7]. HMM의 상태수는 5개, 상태당 mixture의 수는 9개로 정하여 실험하였으며, 언어모델로는 연속숫자음은 개별 숫자음들의 나열이므로 비교적 간단한 언어모델인 FSN(Finite State Network)을 적용하였다.

4연숫자음에서 특징파라미터 추출시 음성분석은 Aurora 특징파라미터의 경우 표준 문서에 정의된 조건으로 수행하였다[2]. 직접 추출한 특징파라미터의 경우는 전화음성 분석프레임의 길이는 20 ms, 프레임의 이동은 10 ms로 하였다. 전처리 계수를 0.97로 하였으며 프레임별로 해밍윈도우 처리하였다. 인식 실험에 이용된 모든 특징파라미터에 보상기법의 적용은 13차의 static 파라미터(로그 에너지 포함)에서 이루어졌으며, 보상기법을 적용한 후 delta 및 delta-delta 특징파라미터를 구해 총 38차의 특징파라미터로 인식 실험하였다. 채널보상기법 CMN에서는 4연숫자음 단위로 켑스트럼 평균을 구해 정규화를 적용하였으며, MRTCN 기법에서는 4연숫자음 단위로 전체 켑스트럼 평균과 분산을 추정하였는데 추정계수로 0.125를 사용하였다.

실험에서 전화음성 연속숫자음 인식실험을 위해 SITEC에서 제작한 전화음성 4연숫자음 DB를 사용하였다. 전화음성은 전체 2000명 화자의 음성으로 구성되었으며 유선/무선 전화 및 cellular, PCS 전화망 환경에서 녹음되어 linear PCM(Pulse Code Modulation)으로 저장되어 있다[6].

음향모델의 훈련데이터로는 SITEC 전화음성 DB의 구성에 따라 훈련데이터로는 58292개, 인식테스트 데이터로는 6468개의 4연숫자음을 사용하였다.

전체 인식 실험은 크게 2부분으로 나누어 수행하였다. 첫 번째 실험은 잡음제거 기법이 들어간 Aurora 특징파라미터와 그것에 보상기법을 적용한 특징파라미터를 이용한 실험이고 두 번째 실험은 DWFBA 특징파라미터 추출 기법과 Aurora 기법을 모두 적용하여 얻은 특징파라미터를 이용한 실험이다. 또한 Aurora 특징파라미터 추출에서 잡음제거 기법이 전화음성 인식에 미치는 영향을 고려하기 위해 BE 대신 보상기법을 적용한 경우에 대해서도 인식 결과를 확인하였다.

IV. 실험 결과

SITEC 전화음성 4연숫자음을 가지고 Aurora 특징파라미터와 DWFBA 특징파라미터에 채널 보상기법을 적용한 후 인식 실험한 결과를 표 1과 표 2에 나타내었다. 표 1에서 Aurora 특징파라미터에 채널 보상기법을 적용할 때, Aurora 특징추출 기법에서 Blind Equlization이 켑스트럼을 정규화 하는 역할을 하므로 이 과정을 제거한 후 특징파라미터를 얻었다. Aurora 특징추출 기법을 적용한 경우 전화음성에 존재하는 잡음 성분이 제거되어 기존의 MFCC보다 더 나은 성능을 보여주고 있으며 채널 보상기법을 적용한 경우에도 Aurora 특징파라미터에서 다소 나은 성능을 보였다.

표 2 에서는 MFCC 보다 더 나은 인식성능을 보였던[5] DWFBA 특징추출 방법을 Aurora 특징추출 방법에 적용 시켜 얻은 특징파라미터로 인식실험 한 결과이다. 즉, Aurora 특징추출 기법에서 잡음이 제거된 음성을 가지고 DWFBA 방식으로 특징파라미터를 추출하여 인식 실험한 결과이다.

인식 결과에서 Aurora 특징파라미터에 DWFBA를 적용한 경우 인식 성능은 향상되어 90.37%의 4연숫자음 인식률을 얻었으며 채널보상기법을 적용한 경우에는 MRTCN 채널 보상기법을 적용한 경우가 91.05 %의 4연숫자음 인식률로 가장 좋은 성능을 나타내었다. 즉, Aurora 특징추출에서 잡음이 제거된 신호를 DWFBA 방식을 적용하고 MRTCN 채널보상기법을 적용한 경우 DWFBA에 MRTCN을 적용한 파라미터보다 0.2 %정도의 인식률 향상을 얻을 수 있었다.

V. 결론

전화음성 인식에서는 채널과 부가잡음의 영향으로 인식성능 저하를 가져온다. 본 논문에서는 연속숫자음 전화음성의 인식 성능을 향상시키고자 DSR의 전처리단으로 이용되는 Aurora 특징추출 기법과 MFCC 및 DWFBA 특징추출 기법을 적용하여 인식실험을 하였다. 전화 채널의 영향을 줄이기 위해 채널보상기법으

표 1. Aurora 특징파라미터와 채널보상기법에 따른 4연숫자음 인식률

특징파라미터 종류/인식률	4연숫자음 인식률(%)	개별숫자음 인식률(%)
MFCC	87.06	96.17
Aurora	89.59	96.97
Aurora+CMN	90.09	97.16
Aurora+MRTCN	91.02	97.38

로는 CMN과 MRTCN을 적용하였다. 실험결과 2단계 Wiener 필터를 이용하여 잡음을 제거한 Aurora 추출 기법이 기존의 MFCC보다 2.5%, DWFBA보다는 1%의 성능 향상을 보였으며 채널 보상기법으로 MRTCN을 적용한 경우에도 91.02%의 4연숫자음 인식률을 보여 MFCC보다 더 나은 성능을 얻을 수 있었다. 그리고 Aurora 특징추출 기법에 DWFBA 기법을 적용하여 특징파라미터를 추출한 후 채널 보상기법으로

표 2. Aurora 기법과 DWFBA 기법을 적용한 실험결과

특징파라미터 종류/인식률	4연숫자음 인식률(%)	개별숫자음 인식률(%)
Aurora	89.59	96.97
DWFBA	88.54	96.71
DWFBA+MRTCN	90.85	97.37
Aurora+DWFBA (BE 제외)	89.52	96.93
Aurora+DWFBA	90.37	97.20
Aurora+DWFBA+CMN	91.02	97.41
Aurora+DWFBA+MRTCN	91.05	97.41

MRTCN을 적용한 경우에도 약간의 성능 향상을 얻을 수 있었다. 향후 인식률을 높이기 위해서는 인식이 어려운 한국어 숫자음 쌍들에 대한 근본적인 변별력을 높여줄 수 있는 특징추출 기법에 대한 연구가 지속되어야 한다.

본 논문은 한국전자통신연구원 네트워크기술연구소 음성정보연구센터의 연구비 지원으로 수행되었습니다.

참고문헌

[1] P.J. Moreno, *Speech Recognition in Telephone Environment,* Master Thesis, Carnegie Mellon University, 1992.

[2] ETSI ES 202 050 V1.1.1, "Speech Processing, Transmission and Quality aspects; Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms", 2002

[3] J.D. Veth and L. Boves, "Comparision of Channel Normalization Technique for Automatic Speech Recognition Over the Phone," *Proc. ICSLP*, pp. 2332-2335, 1996.

[4] 정성윤, 김민성, 손종목, 배건성, 김상훈, "한국어 연속숫자음 전화음성의 인식성능개선," *대한전자공학회, 추계학술발표대회 논문집* 25권 2호, pp. 582-585, 2002.

[5] 최종연구보고서, 전화망 환경에서의 연속숫자음 신호왜곡 연구, 한국전자통신연구원, 2002.

[6] http://www.sitec.or.kr/index.asp.

[7] Steve Young, Gunnar Evermann and D. Kershaw, The HTK Book (for HTK Version 3.1), *Cambridge University Engineering Department.*