2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

자동 입력레벨 조절기의 구현 및 인식 성능 향상

김 상 진, 한 민 수

한국정보통신대학원대학교

Implementation of Automatic Microphone Volume Controller and Recognition Rate Improvement

SangJin Kim, Minsoo Hahn

Information and Communications University

E-mail : sangjin@icu.ac.kr / mshahn@icu.ac.kr

요 약

본 논문에서는 마이크 입력레벨 조절기의 구현과 이를 이용한 인식률의 향상을 다룬다. 마이크를 통한 음성 입력이 너무 작거나 너무 크면 인식률에 직접 영향을 미치므로 인식에 적합한 입력레벨로 조절할 필요가 있다. 자동 입력레벨 조절기의 구현을 위해 고려할 사항을 연구했으며, 이를 통해 PC환경의 입력레벨 조절기를 구현했다. 수집된 음성 데이터베이스는 켑스트럼 평균 차감법(CMS)을 이용하여 채널왜곡을 보상했으며, 구현된 조절기를 이용하여 실험한 결과, 이용하지 않은 경우에 비해 약 50%의 오인식율을 줄일 수 있었다.

ABSTRACT

In this paper, we describe the implementation of a microphone input level control algorithm and the speech improvement with this level controller in personal computer environment. The volume of speech obtained through a microphone affects the speech recognition rate directly. Therefore, proper input volume level control is desired for better recognition. We considered some conditions for the successful volume controller implementation firstly, then checked its usefulness on our speech recognition system with common office environment speech database. Cepstral mean subtraction is also utilized for the channel-effect compensation of the database. Our implemented controller achieved approximately 50% reduction, i.e., improvement in speech recognition error rate.

Index Terms : speech recognition, microphone level control

1. 서 론

최근 음성인식을 이용한 상품이나 서비스가 일상 생활에서 급증하고 있다. 예를 들어, 일반 PC에서 음성을 이용하여 프로그램을 실행시키는가 하면, 인터넷 웹 서핑을 즐기며, 또한 받아쓰기를 이용한 문서작업도 어느 정도 가능하게 되었다.

이 모든 작업은 마이크로폰을 이용한 음성의 입력이 필수적이다. 고성능의 마이크로폰이나 A/D 변환기 등을 이용한다면 보다 좋은 인식결과를 얻을 수 있겠지만, 일반 사용자들의 환경은 그렇지 못하다. 이 경우, 잡음제거나 채널 왜곡 보

상 등 여러 전처리 작업이 인식율 향상에 도움을 줄 수 있다[1][2][3][4].

만일 동일한 환경의 조건이라면 같은 입력에 대해 마이크로폰의 입력레벨 조절이 인식율에 영향을 미칠 수 있다. 따라서 보다 높은 음성인식율을 위해서는 적절한 마이크로폰 입력레벨 조절이 요구된다.

2절에서 자동 마이크 입력레벨 조절기의 구현에 대해 설명하고, 3절에서는 실험에 사용된 음성 데이터베이스에 대해 먼저 설명한 뒤, 구현된 입력레벨 조절기와 음성 데이터베이스를 이용한 인식 실험 및 결과를 설명하며, 4절에서 결론을 맺었다.

2. 자동 마이크로폰 입력레벨 조절기

음성의 입력은 마이크로폰을 통하게 된다. 이 때 마이크로폰의 입력레벨에 따라 양자화되는 음성샘플이 차이가 나게 된다. 그림 1에서처럼 입력의 음성 레벨이 낮은 경우와 높은 경우를 비교해 보면 A/D 변환을 통해 양자화 될 때 주어진 양자화 레벨이 같다면 후자의 경우가 더 높은 인식율을 얻게 해 준다.

2.1 구현 알고리즘

입력 음성과 마이크로폰, 음성입력단자 및 조절기 등의 관계를 그림 2에 블록도로 보였다.

먼저 마이크로폰의 입력레벨과 입력 음성의 최대값의 상관 관계를 구한 뒤, 이를 마이크로폰의 입력레벨 범위에 대하여 적절한 변환 테이블을 완성한다.

입력된 음성으로부터 최대값을 찾고, 미리 완성한 변환 테이블을 통해 적절한 입력레벨 값을 얻는다. 이 입력레벨로 고정을 하면, 조절 후의 입력에 대해서는 효율적으로 음성 입력을 양자화 할 수 있다.

2.2 자동 마이크로폰 입력레벨 조절기의 구현

마이크로폰의 입력레벨은 음성의 샘플 값을 이용하여 조절할 수 있는데, 입력레벨 조절기의 구현을 위해서는 먼저 다음과 같이 음성의 특성 및 입력 환경을 고려해야 한다:

1) 입력 음성 샘플의 최대값은 음성의 특성상 +/- 최대값을 모두 고려해야 한다.
2) 음성 샘플에 존재할 수 있는 돌출잡음을 고려한다. 즉, median 필터를 적용하거나 몇 개의 입력 샘플의 평균을 이용하여 음성의 최대값을 구한다.
3) 한국어의 발음상 강하게 발성되는 말과 약하게 발성되는 말을 모두 고려한다. 즉, 이렇게 선정된 단어를 레벨 조절 시 발성하도록 유도한다.
4) 입력 음성의 크기가 일정해야 한다. 즉, 조절에 사용될 음성을 입력할 때마다 음성의 크기가 계속 변한다면 적절한 레벨로 수렴하기가 곤란하다. 따라서, 3~4번으로 한정된 횟수의 반복 입력을 받고 이들의 평균을 이용한다.
5) 최대 입력레벨로 조절되어도 너무 작은 음성

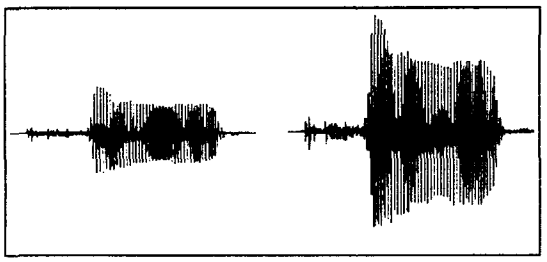

그림 1. 마이크 입력레벨의 차이에 따른 입력 음성파형의 비교
Figure 1. The variation of input speech waveform depending on microphone input level

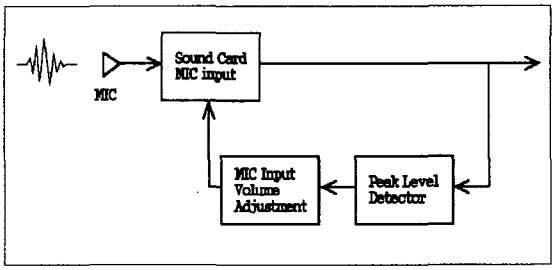

그림 2. 자동 입력레벨 조절기 블록도
Figure 2. The block diagram of the automatic microphone input level controller

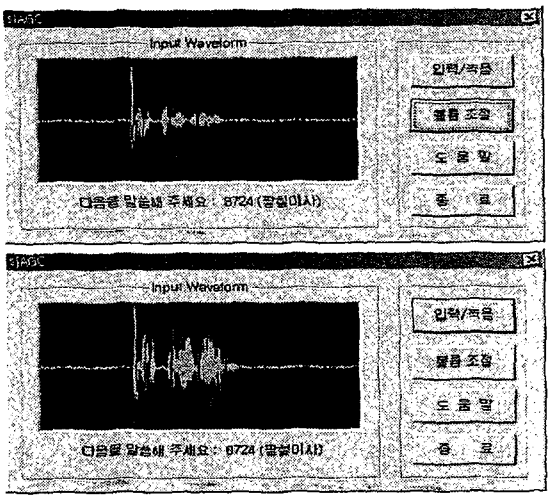

그림 3. 자동 입력조절기의 적용 전과 후 입력 음성 파형 비교
Figure 3. Comparison of the speech waveform before and after automatic microphone input level control

을 입력하는 경우와, 너무 크게 발성하여 오버플로우가 발생하는 경우도 고려한다.

구현된 자동 마이크로폰 입력레벨 조절기를 통해 레벨을 조절하기 전과 후의 입력음성 파형을 그림 3에 보였다.

위와 같은 사항을 고려하여 자동 입력레벨 조절기를 구현할 수 있지만 입력 레벨이 적절히 조절되어도 사용자의 입력이 레벨 조절 시의 음성 크기와 항상 같을 수는 없다. 일반적으로 사용자의 음성은 시간이 지날수록 점점 작아진다. 이를 극복하기 위해 다음과 같은 두 가지 방법을 제시할 수 있다:

1) 마이크로폰의 입력에 대하여 매번 입력레벨을 조절하며 다음 음성을 입력받고 인식을 실행한다.
2) 한 번 조절된 입력레벨은 고정되지만, 인식기에서 인식결과에 대한 적절한 통계를 바탕으로, 인식률이 특정 레벨 이하로 저하되면 입력레벨을 다시 조절한다.

위 방법 중 두 번째 방법은 인식기와 연동되어야 한다.

3. 실험 및 결과

3.1 사용된 음성 데이터베이스

훈련용 데이터베이스는 국어공학센터의 주관으로 원광대에서 제작된 한국어 4연 숫자음성 데이터베이스로 공, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구의 10개 숫자로 이루어졌으며, 남녀 40명이 발성한 연속 숫자 음성이다. 조용한 환경에서 녹음되었고 16 kHz로 샘플링되고 16 bit로 양자화 되었다.

테스트용 데이터베이스는 8인의 남녀 화자들이 훈련용 데이터베이스와 동일한 35개의 4연속 숫자를 사무실 환경에서 두 번씩 발음한 음성을 이용했다. 처음의 35개 연속 숫자 음성은 입력레벨 조절을 하지 않고 화자는 일정하지 않은 크기의 음성으로 발성했으며, 두 번째의 35개 음성은 입력레벨 조절 후 일정한 크기의 음성으로 발성했다. 발성된 음성은, 보편적으로 사용되는 저가의 마이크를 통해 PC 환경으로 입력되었으며, 16 kHz, 16 bit로 A/D변환되어 저장되었다.

테스트용 데이터베이스는 일반 사무실 환경에서 일반 마이크로폰을 통해 녹음되었으므로 잡음 제거 및 채널왜곡 보상이 필요하다. 이를 위해 가장 일반적인 전처리 방법인 켑스트럼 평균 차감법(CMS)을 훈련용과 테스트용 데이터베이스 모두에 적용했다.

3.2 인식 시스템

HMM을 이용한 인식기를 사용했다. 음성 특징 파라미터는 총 24차로써, 13개의 Mel frequency band를 가지는 12차의 MFCC (Mel Frequency Cepstral Coefficients)와 이 MFCC들의 delta coefficients를 기본으로 사용하고, 여기에 에너지와 에너지의 delta coefficients를 포함한 총 26차의 특징파라미터도 사용하였으며, 에너지를 포함한 경우와 포함하지 않은 경우도 비교해 보았다. 먼저 0.97로 pre-emphasis를 한 뒤, 20 msec의 해밍윈도우를 이용하여 10 msec씩 이동하며 특징 파라미터를 추출하였다. Monophone과 triphone에 대해 실험하였으며 음소 당 상태 수는 3개를 사용하였다[5].

표 1. 자동 입력레벨 조절기 적용 전/후 인식 결과 비교

Table 1. Comparison of the speech recognition with and without automatic microphone input volume level controller

12차 MFCC & Delta with LT-CMS		Monophone		Tied-Triphone			
		Mix-1		Mix-1		Mix-5	
		숫자열	숫자	숫자열	숫자	숫자열	숫자
w/o E (24차)	Before AVC	32.14	75.45	75.71	92.23	77.14	92.77
	After AVC	30.00	74.64	82.50	95.09	83.57	95.54
	오차율 향상	-3.2%	-3.3%	28.0%	36.8%	28.1%	38.3%
with E (26차)	Before AVC	19.29	66.34	75.71	91.79	78.57	93.48
	After AVC	32.14	75.62	86.43	96.16	88.57	96.79
	오차율 향상	15.9%	27.6%	44.1%	53.2%	46.7%	50.8%
오차율향상 (After AVC) w/o vs. with		3.1%	3.9%	22.5%	21.8%	30.4%	28.0%

※ AVC : Automatic MIC input Volume-level Control
MFCC : Mel Frequency Cepstral Coefficients
LT-CMS : Long-Term Cepstral Mean Subtraction
Delta : Delta Coefficients
E : Energy
w/o : without

3.3 인식 실험 및 결과

마이크로폰 입력볼륨 조절 없이 일정하지 않은 크기로 입력받은 음성에 대한 인식 실험 결과를 baseline으로 정했으며, 반면에 입력볼륨 조절을 통해 일정한 크기로 입력받은 음성에 대한 인식 결과를 baseline에 대해 비교하고 오차율 향상을 구했다. 또한, 특징 파라미터에 에너지를 포함한 경우와 포함하지 않은 경우로 나누어 실험한 뒤 오인식율이 얼마나 개선되었는지 알아보았다.

표 1에서 보면, triphone 기반의 숫자인식 경우, 마이크 입력 볼륨 조절을 통해 일정한 크기로 입력을 받은 음성에 대한 인식 실험 결과가 그렇지 않은 경우에 대해서, 에너지를 포함하지 않은 경우 약 38%, 포함하는 경우는 약 50%의 오인식율 감소를 보였다. Mixture 개수를 1개, 5개로 실험한 결과 향상의 폭이 크지 않음을 알 수 있었다. 또한, 특징 파라미터에 에너지를 포함한 경우와 포함하지 않은 경우를 서로 비교해 보면, 포함한 경우 포함하지 않은 경우에 대해 약 28%의 오인식율 감소를 보였다.

4. 결 론

자동 마이크로폰 입력 레벨 조절기를 구현할 때에는 상기한 바와 같이 음성의 특성상 여러 가지 고려해야 할 사항들이 있다.

구현된 자동 마이크로폰 입력 레벨 조절기를 통해, 볼륨조절 전과 후의 음성 입력에 대해 인식 실험을 수행한 결과, 약 50%의 오인식율을 감소시킬 수 있었다.

현재 본 논문에서 제안한 알고리즘의 보편적 유용성을 확인하기 위해 숫자음이 아닌 고립 단어 및 문장 인식 등으로 확장 실험을 진행 중이다.

참고문헌

[1] Richard J. Mammone, Xiaoyu Zhang, Ravi P. Ramachandran, "Robust Speaker Recognition, A Feature-based Approach," *IEEE signal processing mag.*, Sep. 1996.

[2] 김원구, 임용훈, 차일환, 윤대희, "잡음 환경에서 음성 인식을 위한 신호 처리," *한국음향학회지 11권 2호*, 1992.

[3] Aaron E. Rosenberg, Chin-Hui Lee, Frank K. Soong, "Cepstral Channel Normalization Techniques for HMM-Based Speaker Verification," *proc. of ICSLP*, 1994.

[4] 오영환, *음성언어정보처리*, 홍릉과학출판사, 1997.

[5] Steve Young, et al., *The HTK Book (for HTK version 2.2)*, Entropic Ltd., 1995-1999.