말소리와 음성과학 제6권 제3호 (2014.9.30)
ISSN 2005-8063, pp. 165~169
www.speechsciences.or.kr
http://dx.doi.org/10.13064/KSSS.2014.6.3.165

소음 환경에서의 명료한 청취를 위한 음절형태 기반 음소 가중 기술

Syllable-Type-Based Phoneme Weighting Techniques for Listening Intelligibility in Noisy Environments

이 영 호[1] · 주 종 한[2] · 최 승 호[3]
Lee, Young Ho · Joo, Jong Han · Choi, Seung Ho

ABSTRACT

Intelligibility of speech transmitted to listeners can significantly be degraded in noisy environments such as in auditorium and in train station due to ambient noises. Noise-masked speech signal is hard to be recognized by listeners. Among the conventional methods to improve speech intelligibility, consonant-vowel intensity ratio (CVR) approach reinforces the powers of overall consonants. However, excessively reinforced consonant is not helpful in recognition. Furthermore, only some of consonants are improved by the CVR approach. In this paper, we propose the *corrective weighting (CW)* approach that reinforces the powers of consonants according to syllable-type such as consonant-vowel-consonant (CVC), consonant-vowel (CV) and vowel-consonant (VC) in Korean differently, considering the level of listeners' recognition. The proposed CW approach was evaluated by the subjective test, Comparison Category Rating (CCR) test of ITU-T P.800, showed better performance, that is, 0.18 and 0.24 higher than the unprocessed CVR approach, respectively.

Keywords: noisy environment, speech intelligibility, syllable types, consonant-vowel intensity ratio (CVR), corrective weighting (CW)

1. 서론

강연회장이나 기차역 등에서 청취자에게 전달되는 강연 음성이나 안내방송 음성은 주변 소음에 의해 음성의 명료도(intelligibility)가 크게 저하될 수 있다. 이러한 상황은 TV 시청, 휴대폰 음성 통화 등의 경우에도 발생한다. 주변 소음이 존재하는 청취 환경에서는 음성신호가 주변 소음에 의해 마스킹되어 청취자가 일부 음성신호를 인지하기 어려우며 특히, 청각장애자와 노인에게 영향이 더 크다. 마스킹된 음성신호는 청취자가 일부 음성신호를 인지하기 어려운 상황이므로 음성 출력 단에서 음성신호를 강화시켜 음성의 명료도를 향상시키는 기법이 필요하다.

소음 환경에서 음성의 명료도 향상을 위해서 음성개선(speech enhancement) 즉, 잡음감쇄 (noise suppression)에 대해 연구 [1, 2]가 이루어져 왔지만, 소음을 직접적으로 통제할 수 없는 즉, 청취자가 소음감쇄 기기를 착용하지 않은 일반적인 상황에서는 적용할 수 없다.

일반적으로, 음성의 명료도를 향상시키기 위해 신호대잡음비 (signal-to-noise ratio, SNR)나 스피커 볼륨을 높이는 방법은 명료도와 관련이 없는 출력 음성의 모든 성분이 일정하게 증가되는 문제점이 있으며, 음질이 저하되거나 왜곡되는 현상이 발생할 수 있고, 청취자의 피로감과 거부감을 유발시키므로 좋은 방법이 아니다.

또 다른 방법으로는 자음의 전력 (power)을 강화시키는 것이다. 자음은 음성의 명료도 관점에서 모음보다 더 중요한 정보를 전달하며 [3], 자음은 모음에 비해 전력이 작아 소음 환

1) 서울과학기술대학교, cap97029@gmail.com
2) 서울과학기술대학교, whgks36@naver.com
3) 서울과학기술대학교, shchoi@seoultech.ac.kr
본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업 [2014-044-055-002, 라우드니스 기반의 방송음량 기술 및 실내 환경 소음의 스트레스 평가 기술 개발]과 한국연구재단의 기초연구사업 [No. 2013R1A1A2007971]의 일환으로 수행하였음.

접수일자: 2014년 8월 14일
수정일자: 2014년 9월 14일
게재결정: 2014년 9월 22일

경에서 청취하는데 어려움을 겪는다. 자음의 전력을 강화시키기 위해서는 자모음비 (consonant-vowel intensity ratio, CVR) 기법을 이용하여 해결할 수 있다. CVR은 모음에 대한 자음의 전력비로서 자음의 음압레벨 (sound pressure level, SPL)을 증가시키며, CVR 레벨이 증가할수록 자음 청취에 유리하다고 알려져 있다 [4, 5]. 그러나 CVR 기법을 통한 단순한 자음의 강화는 음성의 명료도를 향상시키는데 크게 도움을 주지 않을 수 있다 [6]. 대부분 CVR 레벨이 증가하면 자음의 청취 능력이 향상한다는 의견이 지배적이지만 [5, 7], CVR 기법에 따른 단어 인지도 (word recognition score)의 효과에 대한 의견은 현재까지도 일치를 보지 못하고 있다. 국내에서는 CVR 기법에 따른 단어 인지도의 효과를 확인하기 위한 관련된 연구 [8, 9]에서 한국어에서 청취가 어려운 자음을 몇몇 모음과 결합한 자음-모음 (consonant-vowel, CV) 음절을 CVR 기법을 통해 인지도의 변화를 확인해본 결과, CVR 레벨이 증가해도 인지도 차이는 크게 개선되지 않는다. 이는 자음 중 몇몇 자음만 인지도 개선에 효과가 있으며, 자음의 전력을 과도하게 강화시킨 CVR 기법의 적용은 모든 청취자에게 인지도 개선에 반드시 도움을 주는 것이 아님을 의미한다 [4]. 자음만을 강화하는 CVR 기법이 인지도 개선에 대한 효과가 여러 측면에서 관찰하는데 어려움이 있으므로 근본적으로 인간의 청각 지각 특성을 이용하여 인지도가 어떻게 변하는지 고려할 필요가 있다.

한국어에서 모음은 주로 저주파수 영역에 분포하며 음성에너지와 관련성이 높고, 자음은 고주파수 대역에 주로 분포하며 명료도에 큰 영향을 미친다 [10]. 그러나 자음의 주파수는 후행하는 단모음에 따라서 주파수 범위가 다르게 나타난다. 예를 들어, 자음 /ㅅ/의 평균 중심 주파수는 약 6,200 Hz, /ㅆ/는 약 6,600 Hz에 나타나지만, 단모음 /오/, /우/, /어/, /아/, /으/, /애/, /이/ 이 후행되었을 때 /ㅅ/의 주파수 범위는 4,044~6,461 Hz 이고 /ㅆ/은 4,357~6,767 Hz이다 [11]. 이러한 음향학적인 단서에서 인지하기 어려운 고주파수 영역에 해당하는 자음일지라도 모음에 따라서 청취자의 인지도가 변할 수 있다는 의미이므로 모음도 음성의 명료도에 고려해야 할 음소이다. 따라서 한국어의 음절형태에서 음소가 2개 이상으로 구성되는 자음-모음-자음 (consonant-vowel-consonant, CVC), CV, 모음-자음 (vowel-consonant, VC) 그리고 모음 (vowel, V) 등 음절형태별 주파수 영역이 다르므로 인지도가 변할 수 있다. 본 논문에서는 음절형태별 청취자의 인지도를 고려하여 음성의 명료도 향상을 위한 새로운 방법을 제안하였다.

2. 음성의 명료도 향상 기법

2.1 기존의 CVR 기법

2.1.1 소음 환경에서 자음 강화 방법

소음 환경에서 명료한 청취를 위한 기존의 CVR 기법은 <그림 1>과 같이 소음을 추정하여 음성출력 단에서 자음을 검출해 모든 자음의 전력을 소음의 크기에 따라 비례하여 일률적으로 강화시킨다. 예를 들어, 추정한 소음 정도에 따라 소

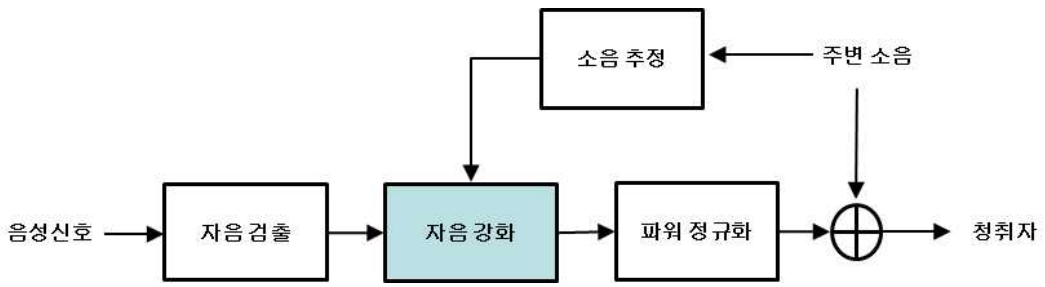

그림 1. 음성의 명료도 향상을 위한 CVR 기법의 블록도

Figure 1. Block diagram of the CVR approach for improving intelligibility of speech

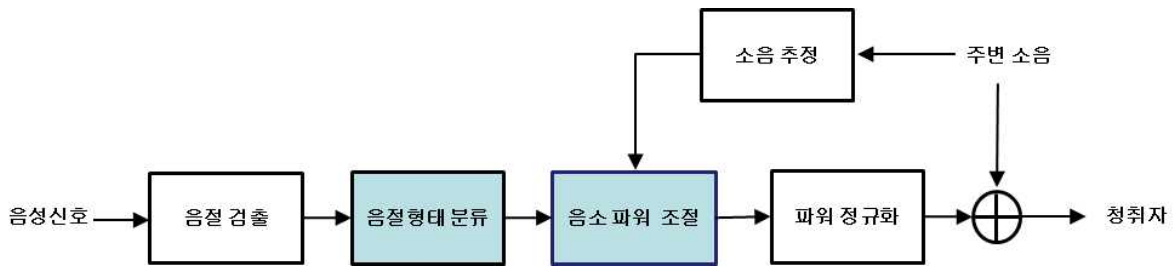

그림 2. 음성의 명료도 향상을 위한 제안된 CW 기법의 블록도

Figure 2. Block diagram of the proposed CW approach for improving intelligibility of speech

음의 크기가 작으면 +3 dB, 소음의 크기가 크면 +12 dB를 모든 자음의 전력을 일률적으로 증가시킨다. 그 후, 전력 정규화를 통해 자음의 전력을 강화한 음성 신호와 아무런 처리하지 않은 음성신호의 전력을 같게 함으로써, 신호의 전체 전력을 같게 하여 형평성을 유지한다.

2.1.2 자음 강화를 위한 가중치 결정

자음 강화는 자음 신호의 전력 P_C에서 자음 신호 x_C에 가중치 (weight)를 적용하여 강화된 자음 신호의 전력 $P_C^{'}$가 도출될 수 있도록 그 가중치를 계산하도록 한다. 원래 자음 신호에 곱할 가중치를 α라고 한다면 x_C 대신 αx_C를 대입하면 식 (1)과 같이 표현된다.

$$P_c^{'} = 10\log_{10}\frac{1}{N_c}\sum_{n=0}^{N_c-1}[\alpha \cdot x_c(n)]^2$$

$$= 10\log_{10}\alpha^2 + P_c \quad (1)$$

가중치 α는 아래와 같이 구할 수 있다.

$$\alpha = 10^{\frac{\triangle P_C}{20}} \quad (2)$$

여기에서 $\triangle P_C$는 자음의 SPL 증가치 즉, $P_c^{'} - P_c$이다.

2.2 제안된 CW 기법

2.2.1 소음 환경에서 음절형태별 자음 강화 방법

소음 환경에서 명료한 청취를 위한 모든 자음의 전력을 일률적으로 강화시키는 기존의 CVR 기법과 달리 본 연구에서는 음절형태별로 자음의 전력을 각각 다르게 강화시키는 *교정 가중 (corrective weighting, CW)* 기법을 제안하며, <그림 2>와 같다. 실제 시스템에서는 자동으로 음절 검출 및 분류를 할 필요가 있으나, 본 연구에서는 제안 알고리즘의 타당성 검증을 위하여 직접 청취 후 수작업으로 검출 및 분류하였다.

2.2.2 음절형태별 자음 강화를 가중치 결정

한국어 형태소 및 어휘 사용 빈도의 분석 [12]에서 빈도수를 고려하여 남녀 각각 250개의 검사용 어음을 선정하여 실험한 결과, 남녀 각각 250개의 각 음절형태에 대한 단음절 어음 인지도 평가를 11개의 등급으로 분류하였다 [13]. 등급이 1에 근접할수록 인지도가 높은 상위권 등급의 음절이며 11로 갈수록 인지도가 낮은 음절이다. 상위권 등급에서는 각 음절형태마다 균일하게 분포되어 있지만 하위권 등급에서는 CVC와 CV이 보다 많이 분포되어 있다. 따라서, CVC와 CV는 명료도가 좋지 않은 음절형태이므로, 다른 음절형태에 비해서 자음의 전력을 더 강화시킬 필요가 있다.

본 연구에서는 단음절 어음 인지도 평가결과를 기반으로 각 음절형태의 등급 기대치를 구했으며, 결과는 <표 1>에서 보여준다.

음절형태별 음소 가중을 위한 SPL 조절치 (dB) y는 선형적으로 결정하였으며, 다음 식 (3)과 같이 나타낼 수 있다.

$$y = aE[X] + b \quad (3)$$

여기에서 X는 각각의 음절형태 V, VC, CVC 그리고 CV의 등급을 나타내는 랜덤변수이고 $E[X]$는 등급 기대치이다.

그리고 음절형태별 다르게 적용될 가중치 범위는 y_{CV}와 y_V간의 동적 범위 (dynamic range) $\triangle_{CW}$에 의해 결정된다.

$$\triangle_{CW} = y_{CV} - y_V = a(E[X_{CV}] - E[X_V]) \quad (4)$$

여기에서 $\triangle_{CW}$는 실험적으로 구했으며, 12 dB이다. 그리고 음절형태별 등급 기대치에 곱할 상수 a는 식 (5)과 같다.

$$a = \frac{\triangle_{CW}}{E[X_{CV}] - E[X_V]} \quad (5)$$

상수 b는 식 (2)의 $\triangle P_C$와 평균적으로 같아지도록 다음과 같이 구한다.

$$\triangle P_C = \frac{y_{VC} + y_{CVC} + y_{CV}}{3} \quad (6)$$

$$b = \triangle P_C - \frac{a}{3}(E[X_{VC}] + E[X_{CVC}] + E[X_{CV}]) \quad (7)$$

음절형태별 SPL 증가치를 계산한 결과는 <표 2>와 같이 정리되어 있으며, VC, CVC, CV 음절형태별 SPL 증가치가 서로 다르다는 것을 알 수 있다. 음절형태 V의 SPL 증가치도 조절해서 식 (8)과 같이 $\triangle_{CW}$를 12 dB가 되도록 하면, y_V 는 -3.78 dB가 된다.

$$\triangle_{CW} = y_{CV} - y_V = 8.22 - y_V \quad (8)$$

표 1. 음절형태별 등급 기대치
Table 1. The expected value of group rank based on syllable type

음절형태	V	VC	CVC	CV
등급 기대치	4.83	5.81	6.59	6.71

표 2. 음절형태별 음소의 음압 레벨 증가치
Table 2. The increases in SPL of phoneme in accordance with syllable type

음절형태	V	VC	CVC	CV
음소 대상	V	C	C	C
SPL 증가치 (dB)	-3.78	+2.36	+7.42	+8.22

3. 실험 및 토의

3.1 실험 대상

정상 청력을 가진 20 세와 29 세 사이의 12명의 피시험자들 (7 명의 남성과 5 명의 여성)이 실험에 참여하였으며, 이들 중 5명은 음성신호처리를 전공하는 학생이며, 나머지 7 명은 아니다.

3.2 실험 자료

제안된 CW 기법의 성능을 평가하기 위해서 ITU-T P.800 의 표준 음질 측정 방법인 CCR (Comparison Category Rating) 테스트를 실시하였다 [14]. 음성신호는 8초 길이의 4명의 남성과 4명의 여성이 발음한 8개의 음성파일들로 구성되었다 [15]. 각 파일은 2 개의 문장으로 구성되어 있으며, 8 kHz 로 샘플링되었다 [15]. 배경 잡음은 NOISEX-92 데이터베이스 [16]의 다화자 잡음 (babble noise)과 백색 잡음 (white noise)을 사용하였으며, 잡음 레벨이 SNR 0 dB, 5 dB, 10 dB, 15 dB 로 혼합되도록 하였고, 헤드폰을 통한 청취 실험을 진행하였다.

표 3. 음성의 명료도의 주관적 평가 기준
Table 3. Performance scale for intelligibility of speech

Rating	Description
3	Much better
2	Better
1	Slightly better
0	About the same
-1	Slightly worse
-2	Worse
-3	Much worse

3.3 실험 절차 및 평가 방법

12명의 피시험자들은 4가지 종류의 음성 파일을 청취하게 된다. 첫 번째 파일은 잡음을 혼합하지 않은 파일로서 참고로 청취한다. 두 번째 파일은 아무 처리를 하지 않은 음성신호에 잡음을 부가한 신호, 세 번째 파일은 기존의 CVR 기법을 적용한 음성신호에 잡음을 부가한 신호, 그리고 네 번째 파일은 제안된 CW 기법을 적용한 음성신호에 잡음을 부가한 신호이며, 이 파일은 두 번째 파일과 세 번째 파일의 비교 대상이 된다. 이때 참고로 청취하는 첫 번째 파일을 제외한 3 가지 종류의 파일은 순서 없이 재생된다. 제안된 CW 기법으로 적용된 음성신호의 파일이 비교대상이 되는 파일들보다 얼마나 명료하게 잘 인지하는지를 평가기준으로 하여 <표 3>과 같이 -3점부터 +3점까지 주관적인 점수로 매긴다.

피시험자들로부터 얻은 점수를 SNR에 따라 평균값으로 나타내었는데, 결과에 나타난 점수가 0 보다 큰 값 일수록 성능이 더 좋다는 것을 의미하고 반대로 0 보다 작은 값을 갖게 되면 성능이 나쁘다는 것을 의미한다.

3.4 실험 결과

<표 4>와 <표 5>는 제안된 CW 기법에 의해 자음이 강화된 음성신호가 각각 아무런 처리되지 않은 것이나 CVR 기법보다 얼마나 더 명료하게 들리는지 평가한 결과이다. 각각의 SNR에서 평균값이 0보다 큰 값을 가짐을 볼 수 있는데, 이를 통해서 제안된 CW 기법이 성능이 더 우수함을 알 수 있다. 또한 백색 잡음뿐만 아니라 실생활 소음에 가까운 다화자 잡음을 혼합한 상황에서도 CCR 성능 평가 결과, 명료도가 향상된 것을 볼 수 있다. 즉, 일률적으로 자음을 강화 시키는 CVR 방식보다 음절형태별로 가중치를 다르게 하여 강화시키는 CW 방법이 더 우수함을 알 수 있다.

표 4. 주관적 명료도 선호 테스트 결과: 제안된 CW 기법을 적용한 음성과 아무 처리를 하지 않은 음성의 비교
Table 4. Subjective preference test result: the proposed CW approach vs. the unprocessed

Test set	CW - unprocessed	
noise	babble	white
0 dB	0.10	0.25
5 dB	0.15	0.14
10 dB	0.17	0.24
15 dB	0.29	0.10
average	0.18	0.18

표 5. 주관적 명료도 선호 테스트 결과: 제안된 CW 기법을 적용한 음성과 CVR 기법을 적용한 음성의 비교
Table 5. Subjective preference test result: the proposed CW approach vs. the CVR approach

Test set	CW - CVR	
noise	babble	white
0 dB	0.03	0.43
5 dB	0.21	0.17
10 dB	0.33	0.15
15 dB	0.23	0.19
average	0.20	0.24

4. 결 론

본 연구에서는 소음 환경에서의 음성의 명료도 향상을 위해 실제 청취자의 단음절 어음 인지도를 기반으로 한국어의 음절형태별로 자음을 강화시키는 CW 기법을 제안하였다. 구체적으로, 명료도가 좋지 않은 음절형태는 자음의 전력을 많이 강화시키고 명료도가 좋은 음절형태는 자음의 전력을 조금 강화시키는 것이다. 음성명료도를 묻는 주관적 평가 실험 결과로부터 아무 처리를 하지 않은 음성과 기존의 CVR 기법에 의해 처리된 음성에 비해 제안한 CW 기법의 성능이 우수함을 알 수 있었다. 음성의 명료도를 저하시키는 요인은 주변 소음뿐만 아니라 잔향음 (reverberation sound)이 있다. 잔향 환경에서는 청취자의 인지도가 저하되므로 주변소음과 잔향음 두 가지 요인을 모두 고려하여 음성의 명료도 향상을 위한 연구가 계속되어야 한다.

5. 참고문헌

[1] Y. Ephraim and D. Malah. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109-1121.

[2] 3GPP2 Document C.S0014-0 v1.0. (1999). *Enhanced Variable Rate Codec (EVRC)*.

[3] G. A. Miller. (1963). *Language and Communication*, McGraw-Hill

[4] L. Hickson and D. Byrne. (1997). Consonant perception in quiet: effect of increasing the consonant-vowel ratio with compression amplification. *Journal of the American Academy Audiology*, vol. 8, no. 5 , pp. 322-332.

[5] E. Kennedy, H. Levitt, A. C. Neuman, and M. Weiss. (1998). Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners, *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 1098-1114.

[6] C. A. Sammeth, M. F. Dorman, and C. J. Stearns. (1999). The role of consonant-vowel amplitude ratio in the recognition of voiceless stop consonants by listeners with hearing impairment, *Journal of Speech Language and Hearing Research*, vol. 42, no. 1, pp. 42-55.

[7] D. A. Preves, T. W. Fortune, B. Woodruff, and J. Newton. (1991). Strategies for enhancing the consonant to vowel intensity ratio with in the ear hearing aids, *Ear Hear*, 12(6 supply):139s-153s.

[8] 신유정, 이경원. (2011). 한국어의 자모음비에 따른 인공와우 착용아동의 무의미 음절의 인지도 변화, *청능재활*, 제7권, 제2호, pp. 200-205.

[9] 이소예, 이경원. (2010). 한국어의 자모음비(CVR)에 따른 무의미음절 단어인지도 변화, *청능재활*, 제6권, 제1호, pp. 25-29.

[10] 주연미, 장현숙. (2009). 노인성 난청의 청력손실 정도에 따른 어음인지능력, *청능재활*, 제5권, 제1호, pp. 36-41.

[11] 이주현, 장현숙, 정한진. (2005). 한국어 음소의 주파수 특성에 관한 연구, *청능재활*, 제1권, 제1호, pp. 59-66.

[12] 김홍규, 강범모 (2000). *한국어 형태소 및 어휘 사용 빈도의 분석*, 고려대학교 민족문화연구소.

[13] 김은옥, 임덕환. (2006). 어음 자극 난이도 및 화자 변수가 어음평가에 미치는 영향. *청능재활*, 제2권, 제2호, pp. 102-106.

[14] ITU-T P.800. (1996). *Methods for Subjective Determination of Transmission Quality*.

[15] J. W. Shin, N. S. Kim (2007). Perceptual reinforcement of speech signal based on partial specific loudness, *IEEE Signal Processing Letters,* vol. 14, no. 11, pp. 887-890.

[16] A. Varga and H. J. M. Steeneken. (1993). Assessment for automatic speech recognition, Ⅱ－NOISEX-92; A database and an experiment to study the effect of additive noise on speech recognition systems, *Speech Communication*, vol. 12, pp. 247－251.

• **이영호 (Lee, Young Ho)**
서울과학기술대학교 전자공학과
서울시 노원구 공릉로 232
Email: cap97029@gmail.com
관심 분야: 음성신호처리, 입체음향

• **주종한 (Joo, Jong Han)**
서울과학기술대학교 전자공학과
서울시 노원구 공릉로 232
Email: whgks36@naver.com
관심 분야: 음성신호처리, 음향반향제거, 음성인식

• **최승호 (Choi, Seung Ho)** 교신저자
서울과학기술대학교 전자공학과
서울시 노원구 공릉로 232
Tel: 02-970-6461
Email: shchoi@seoultech.ac.kr
관심 분야: Speech/audio reinforcement, 잡음처리, 음성인식