자연성 평가를 위한 객관적 음질 평가 방법

장 경 아, 이 희 원, 송 종 희, 배 명 진

숭실대학교 정보통신공학과

Objective Speech Quality Measurement for Naturalness Assessment

KyungA Jang, MyungJin Bae

Dept. of Telecomm. Engr., Soongsil Univ.
kajang74@hotmail.com

Abstract

Speech quality measurement is sorted subjective and objective speech quality measurements based on the mathematics representation. Between two, subjective speech quality measurement is able to be evaluated more accurate speech quality then objective one but it has a demerit such as taking much more time and cost to performance it. However, using objective speech quality measurement being able to predict the result of evaluating the subjective speech quality compliments the demerit of subjective standards which is mentioned former. In this paper, we propose the objective speech quality measurement in order to evaluate naturalness assessment approximately like subjective speech quality measurement. We measured naturalness of speech with estimating speaking rate, variation of pitch and energy.

1.서 론

코딩 알고리즘을 공정하게 비교하고 코딩 기법을 적용한 시스템의 성능을 평가하고 음성 코딩 기능을 갖춘 통신 네트워크를 최적으로 디자인하기 위해서 적당한 음질 평가 방법이 요구된다.

무선 통신의 중용한 수단인 이동 전화기와 같은 통신 시스템에서는 채널을 통해 전송되는 대분분의 음성이 채널 잡음과 배경잡음에 의해 왜곡된다. 따라서 서비스 업자들이 고려해야 할 중요한 사항 중의 하나가 이동 통신 채널 상의에서 음질을 좋게 유지하는 것이다. 일반적으로 이동 전화기의 통화 품질을 측정하기 위해서는 먼저 서비스 통화 지역 내의 다양한 장소에서 반복 청취 실험에 의한 주관적인 평가를 실시 해야 한다. 그러나 이 방법은 매우 수고스럽고 비용이 많이 들기 "때문에 실제적으로 불가능하다. 따라서 이동 통신 시장의 활성화와 더불어 다양한 잡음 환경과 채널 손상 하에서 녹음된 음성의 주관적 평가 결과를 정확히 추정할 수 있는 객관척도 알고리즘의 개발에 대한 연구가 활발히 진행되고 있다[5][6][7].

본 논문에서는 주관관 음질 평가에 근사한 자연성을 평가하는 객관적 음질 평가 방법을 제안한다.

2. 음질 평가 방법

음질 평가 방법은 크게 수학적인 표현식에 근거한 객관적 방법과 청취자들의 주관적 평가 결과에 근거한 주관적 방법으로 구분할 수 있다. 그 중에서 주관적인 평가 방법이 더 정확한 음질을 나타내지만, 이 방법은 시간과 비용이 많이 소모되고 일관성이 없다는 단점이

있다. 따라서 주관적 음질 평가 결과를 예측할 수 있는 객관적 평가 방법을 사용함으로써 이러한 주관척도의 단점을 보완할 수 있다. 본 장에서는 일반적으로 널리 사용되고 있는 객관적인 평가 방법 및 주관적인 평가 방법에 대하여 간단히 설명한다.

2.1 객관적 음질 평가 방법

2.1.1 SegSNR(Segmental SNR)

객관적 음질 평가에 주로 사용되는 SNR은 원 음성과 왜곡된 음성 파형 간의 자승 오차 평균이다. 또한 SNR은 음성 신호 전체에 대한 계산이므로, 음성 신호 중 에너지가 큰 부분에 크게 영향을 받는다. 그러나, 음성 신호는 에너지가 큰 부분과 작은 구간이 반복되는 비정적 신호(*nonstationary*)이므로 SNR을 사용하는 것보다 구간별 SNR을 구하여 이것의 통계적 특성을 이용하는 것이 바람직하며 이러한 방법을 SegSNR이라 한다[5].

2.1.2 LPC_CD(LPC-Cepstral Distance)

LPC(Linear Prediction Coefficient)는 음성의 주파수 스펙트럼 포락선 모양을 나타낸다. 그리고 LPC_CD 방법은 이 포락선 성분의 대수적 차이를 계산한 것이다[1].

2.1.3. BSD(Bark Spectral Distance)

바크 스펙트랄 거리(BSD)[2][6]는 인간의 청각적 임계 대역을 기본으로 두고, 인간 청력을 모델링하는 필터 뱅크를 사용하는 주관적인 접근 방법이다. 인간의 심리 음향을 고려하므로 음질의 주관적인 평가와 상관관계가 매우 높다고 알려져 있다[8]. 인간의 청각 특성은 800Hz 이상의 주파수에 대해서는 주파수가 증가함에 따라 청각의 분해능이 감소하고, 중간 주파수 영역에서 보다 민감하다. 이러한 특성을 반영한 것이 바크 스펙트럼이며, BSD는 원 신호와 왜곡된 신호의 바크 스펙트럼의 차를 구하는 방법이다.

2.1.4 PSQM

PSQM(Perceptual Speech Quality Measure)은 인간이 어떤 소리를 들었을 때 그 소리를 어떻게 인지하는지를 모델링한 것이다. 즉, 신호의 물리적 표현을 인간의 심리 음향학적 표현 방식으로 바꾸는 방법이다.

2.2 주관적 음질 평가 방법

주관척도[5][6]는 크게 명료도 테스트와 자연도 테스트로 나뉜다. 명료도 테스트 방법에는 DRT(diagnostic Rhyme Test), MRT(Modified Rhyme Test) 등이 있고 자연도 테스트에는 DAM(Diagnostic Acceptability Measure), DCR(Degradation Category Rating), ACR(Absolute Category Rating), CCR(Comparison Category Rating)[3][4] 등이 있다.

2.3 객관척도로부터 주관적 음질 예측

객관적 음질 평가 척도로부터 주관적 음질을 예측하기 위해서는 이차함수가 널리 사용되고 있으며 비교적 우수한 성능을 나타낸다[7]. 또한 음질평가 시스템의 성능을 평가하기 위해서 예측한 주관척도의 결과와 실제 주관척도 사이의 상관 계수를 구한다[9]. 이 때 상관 계수가 1인 것은 MOS 예측 오차가 없음을 의미한다.

3. 제안한 음질 평가 방법

3.1 피치 변화도

현재 프레임의 피치를 측정하는 방법으로는 다음과 같이 NAMDF를 정의하여 사용할 수 있다[10].

$$NAMDF(d)=\frac{\sum_{n=1}^{N}|s(n)-s(n-d)|}{\sum_{n=1}^{N}|s(n)|+|s(n-d)|} \qquad (3.1)$$

여기서 s(n)은 음성신호이고 N은 NAMDF를 구하려는 윈도우 구간이다. 지연인자 d를 점차 증가시키면서 NAMDF를 구해보면, 지연인자가 프레임내 음성피치에 정수배가 될 때마다 NAMDF는 거의 영이 된다. 그림 3-1에서 보면 $Y=X^2$ 그래프와 $Y=|X|$ 그래프는 자기상관함수와 AMDF를 취했을 경우 피크점에 서의 그래프를 나타내고 있다[10]. 영점 위치를 살펴보면 자기상관함수의 정확한 피크 값을 찾는 것이 AMDF의 피크 값을 찾는 것 보다 더 어렵다는 것을 볼 수 있다.

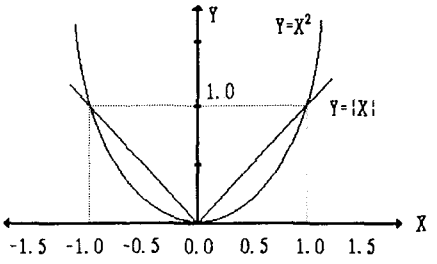


그림 3-1. 1차와 2차 함수의 에러 함수 비교

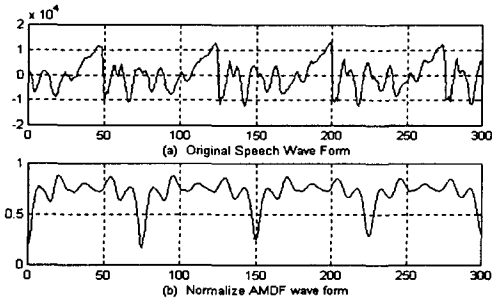


그림 3-2. (a)음성파형 (b) NAMDF 파형

이러한 이유 때문에 피치검색시에 잘못된 피크 값을 얻게 됨으로써 피치검색시 오차를 발생시킬 수 있는 문제를 내포하고 있어 AMDF가 자기상관함수 대신에 주기성을 강조하는데 오랫동안 적용되어 왔다[10]. 또한 AMDF는 곱셈을 사용하지 않는 장점이 있다. 단 규준화시 한 번의 나눗셈은 전체 계산량에 커다란 영향을 주지 않기 때문에 NAMDF의 장점을 유지할 수 있다.

구해진 피치를 이용하여 아래와 같이 피치 변화도를 구한다.

$$V_{P(n)} = |Pitch_n(i) - Pitch_{n+1}(i)|^2 \qquad (3.2)$$

3.2 에너지 변화도

다음 식과 같이 각 프레임 단위로 에너지를 구한다.

$$E = \frac{\sum_{n=1}^{N} |s(n)|^2}{N} \qquad (3.3)$$

구해진 에너지를 이용하여 에너지가 갑자기 커지거나 작아지는지 판정하기 위해 에너지 변화도를 구한다.

$$V_{E(n)} = |E_n - E_{n+1}|^2 \qquad (3.4)$$

3.3 발성 속도

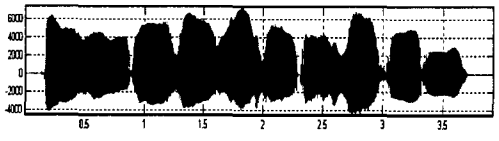
본 논문에서 고려하는 발성속도는 묵음 부분이 제거된 음성신호에서의 발성속도이다. 발성속도 측정에 있어 묵음 구간이 고려된다면 실제의 빠른 발성에 대해서도 다른 결과를 나타내게 된다. 따라서 유효한 발성속도를 측정하기 위해서는 음성구간의 검출이 먼저 선행되어야 한다. 본 논문에서는 먼저 묵음구간의 에너지와 LSP 파라미터를 정보를 이용하여 음성 검출을 수행하였다. 음성 검출을 위한 파라미터를 추출하는 묵음구간은 발성시료의 처음 부분을 이용하였다. 이 구간에서 음성 검출에 필요한 묵음 데이터를 추출하여 이후에 나타나는 음성구간과 비교하였다.

8KHz의 샘플링신호에서 처음 150msec의 신호를 묵음으로 간주하여 에너지와 LSP 데이터를 추출하고, 구해진 에너지의 200%를 에너지 문턱값으로 결정하였다.

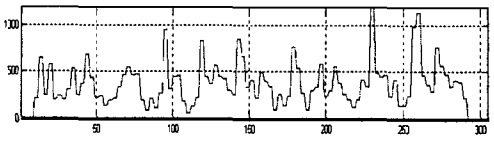
본 논문에 LSP 거리를 구하기 전에 먼저 음성신호의 발성모델에 의해 음성신호의 스펙트럼은 무성음을 제외하고는 짧은 시간동안 변화하지는 않는다. 따라서 60msec동안의 평균 LSP 값을 사용하여 거리를 측정하였다. 거리 계산은 유클리디안 거리측정법을 사용하였다.

$$D(n) = \frac{1}{P} \sum_{i=0}^{P} |LSP_n(i) - LSP_{n+1}(i)|^2 \qquad (3.5)$$

D(n)는 n번째 분석구간과 n+1번째 분석구간과의 LSP 거리를 나타내고, P는 LSP 분석 차수이다. LSP_n은 n번째 분석구간의 평균 LSP이고 LSP_{n+1}은 n+1번째 분석구간의 평균 LSP이다. 그림 3-2은 음성신호에 대한 LSP거리의 변화를 보여주고 있다. (a)는 음성신호의 시간 영역 파형이고 (b)는 식 3.5과 같이 계산한 LSP 거리를 나타낸다. 발성한 음성시료는 /아야어여오요우유으이/로 유성음인 모음이다. 그림(b)와 비교하면 거리가 크게 나타나는 영역이 각 모음의 경계 영역임을 알 수 있다. 따라서 음소의 변화를 추정할 때 큰 거리의 차이를 보이는 영역을 검출하게 된다.



(a) 음성신호



(b) LSP 거리 변화

그림 3-2. LSP 거리 측정

입력음성의 발성속도를 계산하기 위해서는 먼저 현재 처리되는 분석구간이 묵음인지 판정해야 된다. 묵음의 판정은 미리 구한 에너지 문턱값과 LSP 파라미터를 이용한다. 처리되는 분석구간의 에너지가 문턱값 보다 낮고 묵음구간의 LSP와의 거리가 문턱값 보다 낮은 경우 묵음으로 간주한다. 묵음으로 간주된 구간은 발성속도 계산에서 제외된다.

묵음 판정이 끝난 후 인접 분석구간과의 LSP 거리를 측정한다. 측정된 거리 값이 문턱값을 넘는 경우는 음소의 변화가 일어난 것으로 판정하고 이전에 음소가 변화된 구간에서 진행된 시간을 계산한다.

$$SPR = \frac{Fs}{VST(n) - VST(n-1)} \tag{3.6}$$

SPR은 1초당 변화하는 음소의 수를 나타내는 발성속도이고 VST(n)은 현재 음소의 변화가 일어난 시간이고, VST(n)은 이전에 음소의 변화가 일어난 시간이다.

4. 결론

코딩 알고리즘을 공정하게 비교하고 코딩 기법을 적용한 시스템의 성능을 평가하고 음성 코딩 기능을 갖춘 통신 네트워크를 최적으로 디자인하기 위해서 적당한 음질 평가 방법이 요구된다.음질 평가 방법은 크게 수학적인 표현식에 근거한 객관적 방법과 청취자들의 주관적 평가 결과에 근거한 주관적 방법으로 구분할 수 있다. 그 중에서 주관적인 평가 방법이 더 정확한 음질을 나타내지만, 이 방법은 시간과 비용이 많이 소모되고 일관성이 없다는 단점이 있다. 따라서 주관적 음질 평가 결과를 예측할 수 있는 객관적 평가 방법을 사용함으로써 이러한 주관척도의 단점을 보완할 수 있다. 본 논문에서는 주관관 음질 평가에 근사한 자연성을 평가하는 객관적 음질 평가 방법을 제안한다. 피치 변화도, 에너지 변화도, 발성 속도를 측정하여 자연성을 측정하였다. 향후에는 여러 가지 환경에서 녹음된 데이터들을 이용하여 객관척도 결과를 구한 다음 이미 평가한 주관 척도와 객관척도 결과 사이의 매핑 함수식을 구해야 한다. 즉 이 왜곡된 음성과 원 음성과의 차이를 구하여 객관적 음질을 평가한 후 이렇게 계산된 객관적 음질 척도로부터 주관적 음질 척도인 MOS를 예측하는 과정이 필요하다.

5. 참고문헌

[1] Sadsoki Furui, M. Mohan Sondhi. Advance in Speech Signal Processing, Dekker

[2] Nynek Hermanky, Percepture Linear Prediction(PLP) analysis of speech, J. Acoust. Soc. AM., vol. 87, pp. 1734-1752, April, 1990.

[3] W. B. Kleijn, K.K. Paliwal, Speech Coding and Synthesis, Elseevier, 1995.

[4] ITU-T, Method for subjective determination of transmission quality, Rec. pp. 800, Aug, 1996.

[5] S.Quackenbush, T. Barnwell and N.Clements, Objective Measures of Speech Quality, Englewood Cliffs, NJ: Prentice Hall, 1988.

[6] Shihua Wang, et al, An Objective Neawsures for Predicting Subjective Quality of Speech, IEEE J. Select. Areas Commun., vol 10, No. 5, pp.819-829, June 1992.

[7] ITU-T, "Model for Predicity Transmission Quality from Objective Measurements", Supplement2(Series p), Narch 1993.

[8] K.H. Lam, O.C. Au, et al, Objective Speech Measures for Chinese in Wireless Environment, in Proc. IEEE Int,\. Conf. Acoust., Speech Signal Pricess., vol 1, pp.277-280, May 1995.

[9] N.R. Draper, H. Smith, Applied Regressing Anylysis, John Willey & Sons, New York, 1981

[10] L.R. Rabiner, and R.W. Schafer *"Digital Processing of Speech Signals"*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.