論文93-30B-9-5

규칙합성음의 객관적 품질평가에 관한 연구

(A Study on Objective Quality Assessment for Synthesized Speech by Rule)

洪鎭祐*, 金淳協**

(Jin Woo Hong and Soon Hyob Kim

要約

본 논문은 규칙합성음의 품질을 LPC 켑스트럼 거리 (LPC CD) 를 이용하여 객관적으로 평가하고 그 결과를 주관적 평가와 비교한 것이다. 본 논문의 실험에 사용된 음성은 한국어의 속성 및 빈도분포를 고려한 단어 선택법에 의해 선정된 108개 단어를 반음절 규칙합성 방식에 의해 생성된 합성음이다. 규칙 합성음의 품질을 객관적으로 평가함으로써 주관적 평가에 의한 문제점 (평가 시간 및 규모의 확대, 평가 결과의 가변성등) 을 해결하였다. 그리고, 규칙 합성음의 주관적 품질 평가 척도인 이해도와 MOS 를 이용한 평가 결과와 객관적 평가를 비교함으로써 객관적 평가의 타당성을 입증하였으며, 이 결과는 규칙합성 방식의 연구, 개발 및 상용 시제품에 유효한 품질평가 지침을 제공해 줄 수 있다.

Abstract

In this paper, we evaluate the quality of synthesized speech by rule using the LPC CD as a objective measure, and then compare the test result with the subjective one. Speech used for the test consists of 108 words which are selected by word construction method using Korean attribute and frequency distribution, synthesized by demi-syllable rule. By evaluating the quality of synthesized speech by rule objectively, we have tried to resolve the problems such as lots of evaluation time, expansion of test scale, and variables of analysis result arised by subjective measure. We have, also, proved the validity of the objective test using the LPC CD, by comparing intelligibility which is the index for the subjective quality evaluation of synthesized speech by rule with MOS. From this results, we can provide a guide for quality assessment that would be useful in the R&D of synthesis method and the commercial products using synthesized speech.

I. 서론

최근 맨-머신 인터페이스 (Man-Machine Interface) 의 중요성이 강조되고 있으며 특히, 기계에 의한 음성의 합성에 대한 연구가 활발히 진행되고 있다. 또한 음성 합성기술의 발전과 적용 범위가 넓어짐에 따라 다양한 방법에 의한 합성방식 (녹음 편집 방식, 분석 합성 방식, 규칙 합성 방식)이 대두되고 있어 각 방식에 의한 합성 음성의 품질을 평가하고자 하는 연구의 필요성이 점차 확산되고 있다.

이러한 상황에서 규칙 합성음의 품질을 정량화할 품질 평가법을 확립할 필요가 있으며, 이에 기초한

* 正會員, 韓國電子通信研究所 音響情報處理研究室
(ETRI, Acoustics Information Processing Section.)
**正會員, 光云大學校 電子計算機工學科
(Dept. of Computer Eng. Kwang Woon Univ.)
接受日字 : 1992年 7月 2日

평가 분석에 의해 양호한 품질의 합성음성을 서비스함은 물론 합성방식의 연구, 개발에 유효한 지침을 제공해줄 수 있다.

일반적으로 음성의 품질 평가는 주관적인 평가 방법과 객관적인 평가 방법으로 크게 분류할 수 있으며 각각 장.단점을 가지고 있다. 주관적인 평가 방법은 품질을 평가하는 객체가 사람임을 중시하여 품질의 평가를 사람이 직접 결정하는 방법으로써 명료도, 만족도, 이해도등의 평가 척도가 사용된다.[2] 그러나, 주관적인 평가 방법은 품질을 평가하기 위한 실험 규모가 커지고 실험에 소비되는 시간이 길며, 품질 평가 결과가 평가하는 사람의 심리적 환경에 매우 민감하게 작용하는 등의 문제점이 있다.[6]

본 논문에서는 이러한 문제점을 감안하여 규칙 합성음의 품질을 객관적으로 평가하는 연구를 수행하였다. 본 논문에 적용한 객관적인 평가 방법의 적용 알고리즘은 LPC CD (Linear Prediction Coefficient Cepstrum Distance) 이며 평가 음성은 한국어의 속성 및 빈도 분포를 고려한 단어 선택법에 의해 선정한 108개 단어를 반음절 규칙 합성 방식에 의해 생성된 합성음이다.

Ⅱ. 음성 품질 평가법

인간은 주위의 환경으로 부터 다양한 형태의 자극을 받고 그것에 적절한 반응을 하면서 생활하고 있다. 자극에는 시각적 자극과 청각적 자극이 있으나 청각적 자극인 음성에 대해 그 품질을 평가하는 수행 방법과 평가 대상은 다양하게 적용될 수 있다. 반응을 평가하는 데는 그 대상과 목적에 가장 적절한 평가 체계를 찾는 것이 중요하다. 물체의 칫수를 잴때에 자를 사용하듯이 평가 체계에는 그 목적으로 부터 유도된 평가척도가 필요하다. 그러나, 하나의 평가 척도만으로 목적에 맞는 평가를 하는 경우는 거의 없고, 몇개의 평가 척도의 집합체가 필요하게 된다.[8]

다수의 일반 상대자를 대상으로 하는 음성응답시스템 (ARS) 에서 음성 품질의 좋고 나쁨은 가장 중요하게 고려되야 할 점의 하나이다. 많은 연구자들이 각종 음성품질 특성을 규정하고 품질기준을 설정하려는 것은 가장 경제적인 최상의 음성 품질 실현에 그 목적이 있다. 그러나 이러한 음성 품질의 좋은 정도 (goodness) 를 시스템 설계자가 독단적 판단으로 결정한다든지, 다른 나라에서 사용하고 있는 품질기준을 그대로 모방한다든지 할 경우 사회 문화적인 환경, 성별, 연령별 등 개인적 요인에 따라 많은 차이가 있기때문에 국내 실정에 맞는 음성품질의 평가를 기대 하기가 어렵다.

일반 사용자를 대상으로 하는 음성 품질 주관 평가치는 시간과 사람에 의해 상당히 가변적이기 때문에 시변적이고, 개인 가변적인 어떤 정량적인 표현을 나타나게 된다. 일단 어떤 평가치가 구해지면 그 음성 품질의 최상치를 실현하는 데 필요한 가장 경제적인 방법을 고안해야 한다. 이때 음성 품질 평가치는 연산 가능한 것이 바람직하며 결과적으로 사용한 기기의 구성요소 설계에 명확히 귀환될 수 있다. 그러나, 이러한 주관 평가량은 항상 시간 가변적, 개인 가변적인 성질에서 얻어지는 것은 아니며, 이 성질들을 매개 변수로 한 시간 불변적, 개인 불변적인 주관적 반응과 객관적 반응에 연관지어 얻어야 한다.

품질 평가법은 크게 주관적인 방법과 객관적인 방법으로 나눌수 있다. 주관적 방법에는 자연성, 이해성, 음량감과 만족도 등이 있으며 객관적 평가에는 LPC CD 법과 FFT 분석에 의한 Coherence함수법 등이 제시 되고 있다. 주관 평가법은 두개의 커다란 범주인 실용적 방법과 분석적 방법으로 나눌수 있다.[8]

그러나, 주관적인 평가 방법은 평가 실험을 위해 대규모의 모델 시스템이 필요하고 다수의 평가 인원이 요구되고, 평가 실험 시간이 많이 소요될 뿐만아

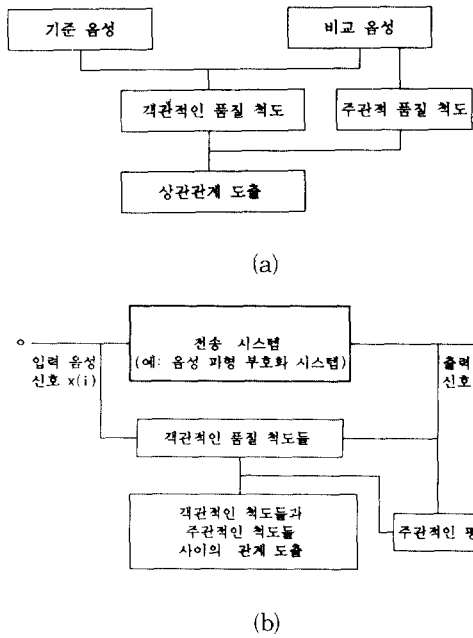


그림 1. 객관적인 품질 척도를로 부터 주관적인 값 평가 방법

Fig. 1. Subjective values estimation method from objective quality measures.

니라 평가 결과가 평가자의 심리적 요인이나 평가 환경에 많은 영향을 받는 문제점이 지적되고 있다. 본 논문에서는 이러한 문제점을 해결하기 위하여 객관적인 평가 방법을 도입하였다. 객관적인 품질측정은 기준 신호와 비교 신호와의 유사 특성의 거리차, 또는 시스템을 경유한 입력과 출력 신호 사이의 왜곡으로 정의되어 진다. 그림 1은 객관적인 평가 값을 결정하기 위한 측정방법의 블록도이다.

객관적인 품질 평가를 위해 시간 영역과 주파수 영역에서 다양한 객관적인 품질 측정 방법이 제안되고 있으며 이것들은 다음과 같다. [6] [8]

1) 파형 왜곡 척도 (Waveform Distortion Measures)

파형 왜곡 척도들은 입력과 출력 음성 파형 사이의 왜곡에 의해서 정의되며 다음과 같이 2가지로 분류된다.

- 파형 왜곡 척도 (Waveform Distortion Measures)
- 구분화 신호 대 잡음비 (Segmental SNR; SNR_{seg})

2) 스펙트럼의 왜곡 척도 (Spectral Distortion Measures)

주파수 영역에서 객관적인 품질측정은 입력과 출력 음성 스펙트럼 (speech spectra) 사이에 왜곡으로 정의되며 다음과 같이 분류된다.

- 스펙트럼 왜곡 (Spectral Distortion; SD)
- 주파수 가중치된 스펙트럼 왜곡(Frequency Weighted Spectral Distortion; WSD)

3) 스펙트럼 Envelope 왜곡 척도 (Spectral Envelope Distortion Measures)

스펙트럼의 envelope 왜곡에 기초한 객관적인 측정은 FFT 기술과는 다른 선형 예측 부호화에 의해서 계산되어 진다. 음성 신호 envelope 왜곡은 가장 중요한 음성신호 특성들을 육안으로 볼 수 있는 스펙트럼의 정보를 가진다는 것이다. 스펙트럼 envelope 왜곡은 부호화 시스템의 입력과 출력 신호사이의 스펙트럼의 왜곡에서 가중치된 함수 (weighting function) 의 차이에 의해서 다음과 같이 분류되어 진다.

- 유사도 비율(Likelihood Ratio; LR)
- 가중치된 유사도 비율 (Weighted Likelihood Ratio; WLR)
- LPC Cepstrum Distance Measure (CD)
- COSH Measure (COSH)

본 논문에서는 객관적인 품질 평가의 측정 방법중 음성 신호내의 가장 중요한 특징만을 추출하여 특징들간의 차를 거리값으로 이용하는 LPC 켑스트럼 거리 측정 방법을 사용하였다.

Ⅲ. 객관적 품질평가를 위한 LPC CD 계산

음성의 특징을 추출하는 데에 다양한 방법이 있으나 LPC 를 이용해서 특징 파라메터를 추출하는 것이 매우 효과적이고, 합리적인 방식이라고 널리 알려져 있다. 음성 신호의 파형을 관찰하여 보면 음성 파형의 이웃한 샘플들은 상관 관계가 높음을 알 수 있다. 이와 같은 관계를 간단한 선형예측 형태로 표시하면 다음과 같다.

$$\overline{x_n} = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} \tag{1}$$

위 관계식은 음성파형의 한 샘플을 과거의 p개의 샘플들의 선형 결합으로 예측할 수 있다는 것을 가정하고 있는데 이때 각 샘플들에 곱하여 지는 가중치 $\{a_i,\ i = 1, 2, \cdots, p\}$ 를 선형예측 계수라 한다. [4] 선형예측 분석에서는, 예측오차 신호의 평균 자승치가 최소가 되도록 선형예측 계수들을 정한다.

LPC 에 의한 스펙트럼 추정이 화자간의 고유한 성질을 완전히 제거하지 못하여 화자의 독립성을 고려하는 데는 장애 요소가 되어 왔다. 따라서 LP 계수에 대해 변형이 요구되는데 LP 계수를 각각 독립적으로 다룰 수가 없으므로 각 계수가 독립적이 되도록 변화시킨다. 본 논문에서는 이러한 특성을 갖는 LPC 켑스트럼 계수를 특징 파라메터로 그림 3과 같은 절차에 의해 추출하였으며, 적용된 수식은 식 (2)와 같다. LPC 켑스트럼 계수는 LPC 분석에 의해 유도된 스펙트럼 envelope의 켑스트럼 계수를 나타내며 이 계수들은 실제 음성파형 스펙트럼에 대응되는 켑스트럼 계수와 같지 않다.

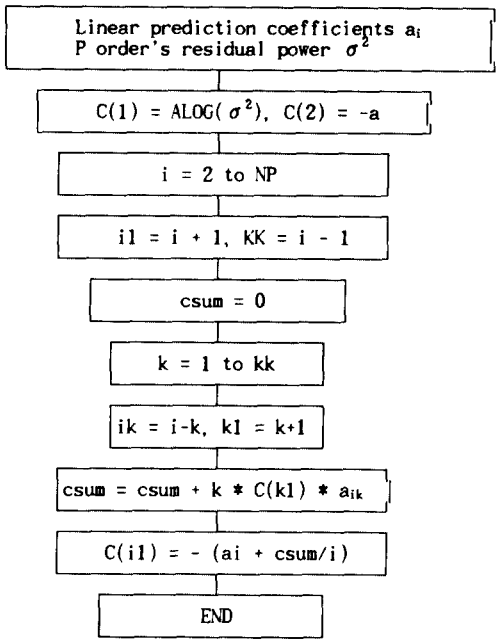
$$C_0 = \ln \sigma^2, C_i = -a_i\, 1/i \sum_{k=1}^{i-1} K C_k a_{i-k},\ 1 \le i \le p \tag{2}$$

선형예측 계수가 켑스트럼 계수에 입각해서 거리척도를 계산하기 위해서 사용되어 진다. 음성 파형으로 부터 직접 계산된 켑스트럼과 다르게, 예측 계수로 부터 계산된 켑스트럼은 평활화된 음성 스펙트럼의 평가를 제공한다. 이것은 다음 수식 (3)과 같이 표현되어 진다. [3] [4]

$$\log\left(\frac{1}{A(z)}\right) = \sum_{k=1}^{\infty} c(k) z^{-k} \tag{3}$$

여기서, A(z)은 LPC 모델 다항식이고, c(k)는 켑

스트럼 계수이고, z은 e^{jw}와 같게 설정되어 진다. 계수들은 다음에 따르는 수식을 사용해서 예측계수로부터 재귀적으로 계산되어 진다.



NP:분석 차수, ai:LPC 계수, c:cepstrum계수

그림 2. LPC 캡스트럼 계수 계산
Fig. 2. A calculation of LPC cepstrum coefficients

$$nc(n)-nc(n)=\sum_{k=1}^{n-1}(n-k)c(n-k)a(k) \quad for\ n=1,2,3,\cdots \qquad (4)$$

여기서, a(0) = 1 이고, k 〉 p 이면 a(k) = 0 이다. 이 수식에서 a(k)는 LPC 예측계수이고, p는 LPC 분석의 차수이다. 캡스트럼 계수에 입각한 척도는 다음과 같이 계산되어 진다.

$$d(c\varphi,\ c_d,\ 2,\ m) = \left\{\left[c\varphi(0)-c_d(0)\right]^2+2\sum_{k=1}^{L}\left[c\varphi\, c(k)-c_d(k)\right]^2\right\}^{1/2} \qquad (5)$$

여기서, d는 프레임 m에 대한 L_2 거리이고, cφ가(k)는 자연 음성의 캡스트럼 계수이고, c_d(k)는 규칙합성음성의 캡스트럼 계수이다.

Ⅳ. 실험 및 결과 고찰

1. 실험 환경

규칙합성 음성의 품질을 객관적으로 평가하기 위해서는 기준 음성 (reference speech) 이 있어야 한다. 그러나, 국내에 기준이 될만한 음성이 규정되어 있지 않기 때문에 본 논문에서는 규칙 합성음의 품질이 실생활 환경에 사용되는 자연음성 (natural speech)에 가까울수록 좋은 품질이라는 관점에서 자연음성을 기준 음성으로 정하였다.

자연 음성과 규칙합성 음성의 LPC 캡스트럼 거리를 측정하기 위한 대상어는 임의로 설정하지 않고 한국심리학회에서 조사한 "한국어 어휘 빈도 조사" 자료에 수록된 17,883 단어에 대해 속성 및 빈도 분포를 고려한 단어 선택법에 의해 선택된 108개 단어로 선정하였다. 선정된 대상어는 한국어 음성의 중요한 속성인 품사, 단어의 길이, 중요도, 변칙발음현상, 음운의 종류등이 고려된 단어들로써 실험 결과에서 이것들에 대한 속성별 결과도 기술하였다. [1] [9] 측정을 위해 사용된 음성중 자연 음성은 일반 성인 남성이 방음 환경에서 발성하여 테이프에 녹음한 것이다.

본 논문에서 LPC 캡스트럼을 추출하고 캡스트럼 거리를 구하기 위해 사용된 실험 시스템은 그림 3과 같다.

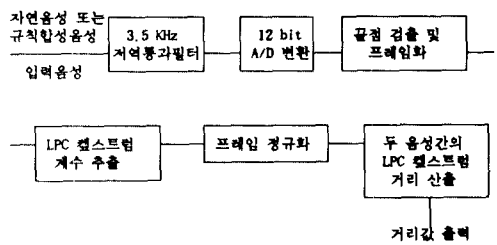


그림 3. LPC CD 계산 시스템의 구성도
Fig. 3. An organization diagram of LPC CD calculation system.

입력 신호를 받아 3.5 KHz의 저역통과 필터를 통과한후 8 KHz의 표본화와 12 bit의 부호화를 시켜 디지탈 데이타로 만들었다. 그리고, 시작점과 끝점 검출을 한후 128 샘플을 1 프레임으로 설정하여 각 프레임마다 10차 LPC 캡스트럼 계수를 추출하였다. 같은 단어에 대해 자연 음성과 규칙 합성음성의 LPC 캡스트럼 계수가 추출되면 서로 다른 프레임의 크기를 정규화시키기 위해 프레임 압축/팽창에 의한 정규화 기법을 적용한후 전체 프레임에 대해 LPC 캡스트럼 거리를 산출하는 방법에 의해 두 음성간의 거리값을 산출하여 출력시켰다. 한편, 규칙 합성음의 객관적인 평가와 주관적인 평가와의 상관관계 및 객관적 평가의 타당성을 입증하기 위해 주관적 평가를 수행

하였다. 주관적 평가의 척도로는 이해도 (intelligibility)와 MOS (Mean Opinion Score) 에 의한 만족도를 선정하였다. 이해도는 언어로서 의미를 갖는 단어 또는 문장을 송화하여 정청율을 구한 것으로서 참고문헌 [9] 의 실험 결과를 이용하였다. MOS에 의한 만족도 시험은 규칙 합성음을 청취하고 난후 자연 음성과의 품질을 비교하여 1) 매우 좋다(4점) 2) 좋다(3점) 3)보통이다(2점) 4) 나쁘다(1점) 5) 매우 나쁘다(0점) 로 판정한 결과를 합산하여 평균한 값으로 결과를 유도하였다. 본 논문의 주관적 평가인 MOS 시험에 참가한 피험자는 훈련이 되지않은 성인 남성 12명이다.

2. 결과 고찰

1) LPC CD와 이해도 및 MOS와의 관계

본 논문에서 수행한 종합적인 실험결과는 그림 4와 그림 5와 같다. 그림 4는 자연 음성과 규칙합성 음성간에 LPC 켑스트럼 거리 계산에 의해 산출된 거리값 (객관적 평가 결과) 과 참고문헌 [9] 에서 수행한 이해도 결과 (주관적 평가 결과) 와의 상관관계를 나타낸 것이고, 그림 5는 계산된 LPC 켑스트럼 거리값과 본 논문에서 주관적 평가로 실시한 MOS 값과의 관계를 나타낸 것이다.

그림 4에서 가로축은 자연 음성과 규칙합성 음성간의 거리 값의 관계를 dB로 표시한 것이며 세로축은 이해도 평가에 의한 percentage(%)를 나타낸 것이다. 그림 5에서 가로축은 자연 음성과 규칙합성 음성간의 거리값 관계를 dB로 표시한 것이며 세로축은

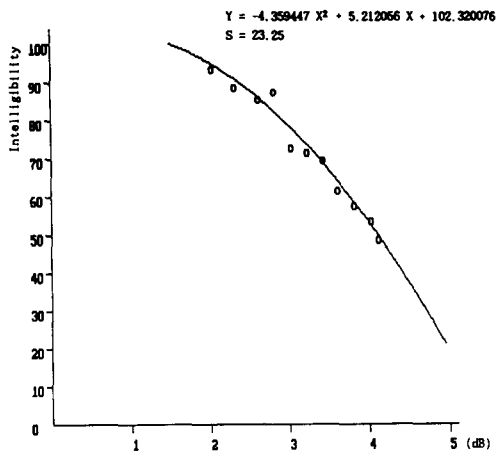


그림 4. 이해도와 LPC CD와의 관계

Fig. 4. Relation between intelligibility and LPC CD.

만족도 평가에 의한 MOS 값을 나타낸 것이다. 그림에 표시된 점들은 108개 단어들중 같은 LPC CD 값을 갖는 단어들의 평균 LPC CD 값이며, 2차 곡선은 이 값들에 의해 구성되는 곡선 회귀 모형 (curvilinear regression model) 을 표현하고 있으며 각각 다음과 같은 식으로 구성된다. 만일 가로축을 X, 세로축을 Y라 하면

$Y = -4.359447\ X^2 + 5.212056\ X + 102.320076$ (그림 4의 경우)

$Y = 0.141032\ X^2 - 1.555205\ X + 5.634372$ (그림 5의 경우)

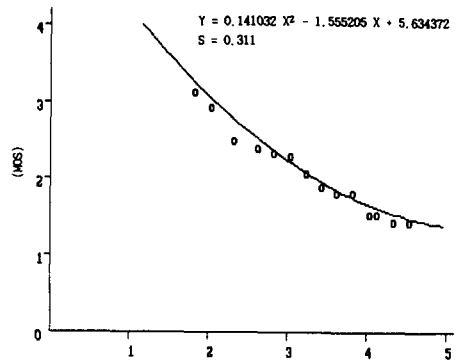


그림 5. MOS와 LPC CD와의 관계

그림 5. MOS와 LPC CD와의 관계

Fig. 5. Relation between MOS and LPC CD.

그림으로 부터 LPC CD 값이 커지면 커질수록 즉, 규칙합성 음성의 품질이 자연 음성의 품질에서 멀어지면 질수록 이해도와 MOS 값이 하락되는 것을 관측할 수 있어 객관적 평가 결과와 주관적 평가 결과가 일치하고 있음을 알 수 있다. 이해도에 의한 주관적 평가 (그림 4)에서 보통 80% 이상의 이해도가 확보되면 양호한 상태라고 할 수 있으며 이때 객관적 평가치인 LPC CD는 2.9dB가 된다. 반면에 MOS에 의한 주관적 평가 (그림 5)에서 50% 이상의 피험자가 보통 이상이라고 응답하는 MOS 값이 2.5 이상이기 때문에 이 이상의 품질이 확보되야만 실제의 응용에 사용할 수 있을 것으로 판단되며 이때의 객관적 평가치인 LPC CD는 2.6 ~ 2.7 dB가 된다. 따라서, 규칙합성 음성의 객관적 품질은 주관적 평가 (이해도 및 MOS 평가)에 의해 양호한 품질이라고 판단되는 값의 범주에 대응할 수 있어야 하므로 자연 음성 (기준 음성)과의 LPC CD 값이 약간의 차를 보이고 있는데 이것은 이해도가 음성 통신의 주 목적인 의미 전달을 평가하는데 반해 MOS는 품질의 만족도

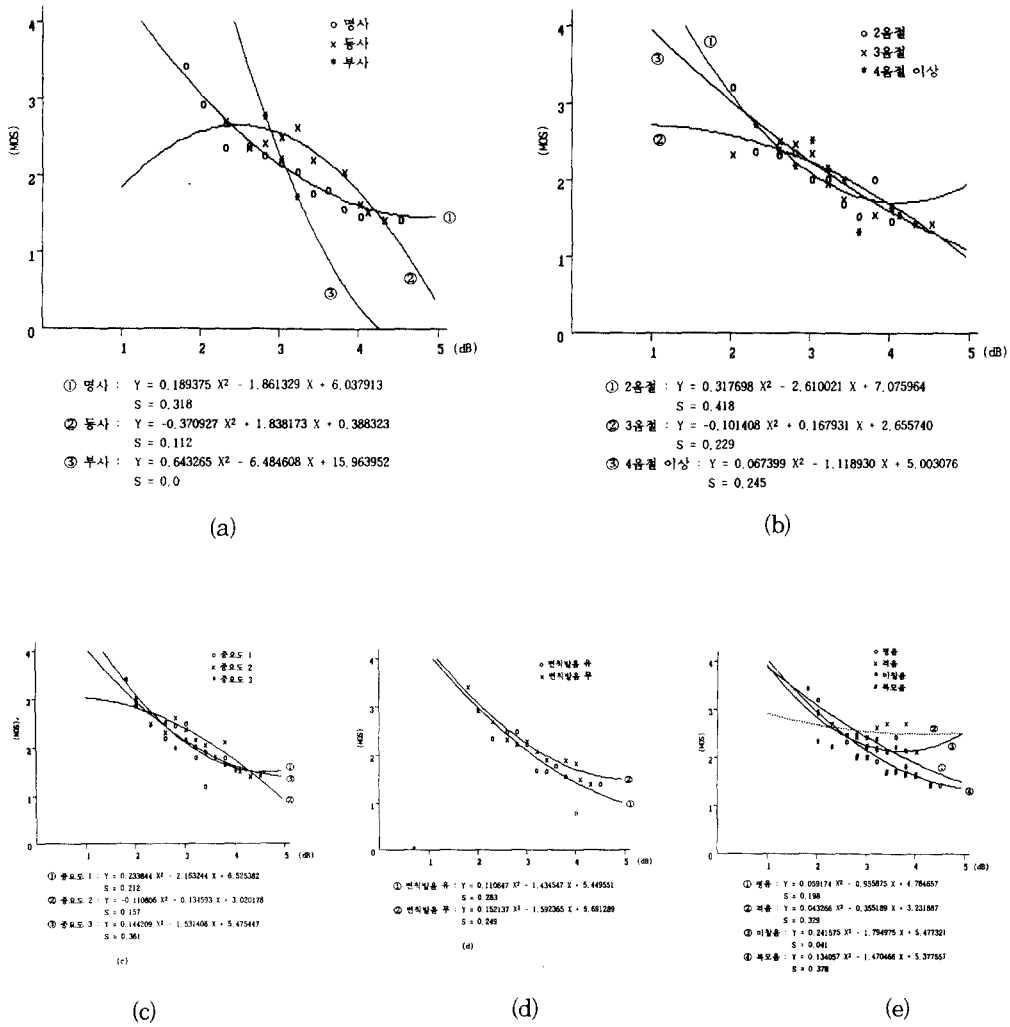


그림 6. 단어 속성별 MOS와 LPC CD와의 관계
Fig. 6. Relation between MOS and LPC CD in each word attributes.

를 평가하는 것에 따른 요인이라고 생각된다.

2) 단어의 속성별 LPC CD 와 MOS 와의 관계

규칙합성 음성의 객관적 평가를 위해 적용한 108개 단어에 대한 각 속성을 분석하여 각 단어 속성별 LPC CD 값과 MOS 와의 관계를 살펴보았으며 그 결과는 그림 6과 같다. 여기서 세로축은 MOS 값을, 가로축은 LPC CD 값을 의미하고, 2차곡선은 각각의 회귀 모형을 나타낸다.

a)는 품사류별 MOS 와 LPC CD 와의 관계를 나타낸 것이다. 품사류중 명사의 명사의 LPC CD 값이 가장 다양하게 분포되어 있는데 이것은 108개 단어중 명사가 차지하는 비율이 매우 높기 때문이다. 이 그림에서 MOS 값의 기준을 2.5 이상으로 잡을때 명사가 가장 짧은 LPC CD 값을 나타냈고, 그 다음 부사. 동사의 순이었다. 부사의 경우 LPC CD 값의 작은 변화에 대해 MOS 값이 급변하는 현상을 나타내고 있어 규칙합성 음성을 생성할때 특히 주위를 기울여야 할 것으로 판단된다. 형용사의 경우는 데이타가 적어서 분포를 알 수 없어 생략하였다. b)는 단어 길이별 (2음절, 3음절, 4음절이상 단어별) MOS와 LPC CD와의 관계를 나타낸다. 이 그림으로 부터 MOS값 2.5 이상에 대응하는 LPC CD 값은 단어 길이에 따른 영향이 거의 없음을 나타내고 있다. c)는 단어의 중요도별 (일상 생활에 많이 사용되는 순

서에 따라 중요도 1, 중요도 2, 중요도 3으로 분류) MOS와 LPC CD와의 관계를 나타낸다. 중요도별로 약간의 LPC CD값이 변화하고 있으나, 큰 영향을 미치지 않고 있음을 알 수 있다.

d)는 변칙 발음별 (단어와 발음이 다른것과 다르지 않은 것을 구분) MOS와 LPC CD와의 관계를 나타내고 있다. 전반적으로 변칙발음 현상이 있는 단어가 변칙 발음이 없는 단어에 비해 LPC CD 값이 작아진 것을 알 수 있다. 이것은 변칙발음 현상 (즉, 자음접변, 구개음화, 경음화 등의 법칙) 에 많은 고려사항을 부여하여 규칙합성 음성을 생성한 결과라고 생각된다. e)는 단어의 제 1 음운별중 ((평음; ㅂ, ㄷ, ㄱ), (격음; ㅍ, ㅌ, ㅋ), (마찰음; ㅅ, ㅎ, ㅆ), (복모음; ㅑ, ㅕ, ㅛ, ㅠ, ㅒ, ㅘ, ㅙ, ㅝ, ㅞ, ㅢ)) 에 대한 MOS와 LPC CD와의 관계를 나타낸다. 그림으로 부터 MOS값 2.5 이상에 대응하는 LPC CD의 값이 작은 순은 제 1 음절의 음운이 마찰음, 복모음, 평음, 격음의 순으로 나타나고 있음을 알 수 있다. 특히, 격음의 경우는 LPC CD 값이 작아져도 MOS 값이 증가하지 않는 현상을 보이고 있는데 이러한 현상으로 부터 격음에 해당되는 단어를 규칙합성 음성으로 생성할때 음의 억양이나 고조, 운율등에 많은 고려를 적용해야만 한다.

3) 문제점 도출 및 고려사항

본 논문에서 규칙 합성음성의 품질을 객관적으로 평가하기 위한 실험 결과로 부터 문제점 및 앞으로의 고려되야 할 사항들을 기술한다. 본 논문에서 규칙합성음성의 품질을 객관적으로 평가하기 위해 품질의 목표가 되는 기준 음성을 성인 남성이 발성한 자연 음성을 사용함으로써 발생되는 문제점이 있다. 앞으로 국내에 품질의 목표가 되는 기준 음성이 확보되야 정확한 객관적 평가를 수행할 수 있을 것으로 사료된다. 또한, 본 논문에서 시험하고자 하는 규칙 합성음성의 대상어가 108개로 한정되었기 때문에 각 단어 속성별 분석 자료가 충분히 확보되지 않아 불합리한 2차 회귀곡선 모델이 설정되는 경우가 있었다. 앞으로 실험 단어표를 다양하게 설정하여 많은 규칙 합성음성을 대상으로한 객관적 평가가 이루어져야 할 것이다.

V. 결론

본 논문에서는 규칙합성 음성의 품질을 정성적으로 또는 정량적으로 평가하기 위한 방법중 LPC 켑스트럼 거리를 적용한 객관적 품질 평가 방법 및 실험 결과에 대해 기술하였다. 또한, 객관적 품질 평가 결과와 주관적 품질 평가 결과와의 상관 관계를 살펴보기 위하여 주관적 품질 평가의 하나인 MOS 실험을 수행하였으며 다음과 같은 결과를 얻었다.

1. 규칙 합성음성의 품질을 정확히 평가하기 의해서는 주관적 평가와 객관적 평가가 종합적으로 수행되야 한다.
2. 본 논문에서 적용한 객관적 평가 시스템을 이용할 경우 양호한 품질의 규칙 합성음성에 대한 LPC CD 값은 2.6 ~ 2.9 dB 이하 이여야 한다.
3. 품사류중 부사의 경우 LPC CD 값의 작은 변화에 MOS 값이 급변하는 현상이 있어 규칙 합성음성을 합성할때 주위가 필요하다.
4. 규칙 합성음성의 생성에서 격음의 경우 LPC CD 값이 작아져도 MOS 값이 증가하지 않는 현상이 있으므로 음의 억양이나 고조, 운율등에 세심한 고려를 필요로 한다.

앞으로 규칙합성 음성의 품질을 더욱 정확히 평가하기 위한 기준 음성의 규정 및 체계화된 측정 방법 (주관적 평가와 객관적 평가에 대해) 이 확립되야 할 것이며 본 논문의 연구 결과가 국내의 규칙 합성음의 품질 향상을 위한 참고자료로 활용되기를 바란다.

감사의 말씀

본 연구의 수행에 많은 도움을 주신 ETRI 음향정보처리연구실의 강 성훈실장님 및 연구원, 그리고 주관 평가실험에 참여한 피험자들에게 감사드립니다.

參 考 文 獻

[1] 김영채, 한국어 어휘 빈도조사, 한국심리학회지, vol.5, no.3, pp216 -285, 1986.

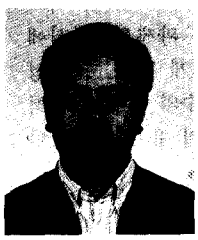
[2] 한국전자통신연구소, 통화품질 평가법 및 표준화에 관한 연구, 연구보고서, 1989. 12.

[3] A. H. Gray and D. Markel, "Distance Measure for Speech Processing," *IEEE Trans.* ASSP. 24, 1976.

[4] K. Itoh and N. Kitawaki, "Evaluation of Digital Coded Speech Quality ofLPC-Distance Measures," Nat. Conf. of

IECE Jpn., 1979.
[5] N. Kitawaki, K. Itoh, and K. Kakehi, "Speech Quality Measurement Methods for Synthesized Speech," Review of ECL, vol. 29, no. 9-10, Sep. 1981.
[6] Thomas P. Barnwell III , "Objective measures for speech quality testing", *J. Acoust. Soc.* Am. 66(6), Dec. 1979.
[7] N. Kitawaki, H. Nababuchi, "Quality Assessment of Speech Coding and Speech Synthesis Systems," *IEEE Comm.* Magazine, Oct. 1988.
[8] S. R. Quackenbush, T. P. Barnwell III, and M. A. Clements, Objective Measures of Speech Quality, Prentice Hall, 1988.
[9] 김 성한, 홍 진우, 김 순협, "규칙 합성음의 이해성 평가를 위한 단어표 구성 및 실험법," 대한전자공학회지, 제29권, 제1호, 1991.1.
[10] Kenzo ITOH, Nobuhiko KITAWAKI and Kazuhiko KAKEHIO, "Objective Quality Measures for Speech Waveform Coding Systems", REVIEW of the Electrical Communication Laboratories, vol. 32, no 2, 1984
[11] H. Fletcher, Speech and Hearing in Communication, Robert E. Krieger Publishing Co., 1972.

著者紹介



洪 鎭 祐(正會員)
1959年 4月 15日生. 1982年 2月 광운대학교 응용전자 공학과 졸업(공학사). 1984年 2月 광운대학교 대학원 전자공학과 졸업(공학석사). 1993年 8月 광운대학교 대학원 전자계산기공학과 졸업(공학박사). 1984年 3月 ~ 현재 한국전자통신연구소 음향정보처리연구실 근무(선임연구원). 주관심분야는 통화품질, 음성신호처리, 실감통신 등임.

金 淳 協(正會員) 第 29卷 B編 第 1號 參照
현재 광운대학교 전자계산기공학과 교수