

캡스트럼 계수를 이용한 전화 채널에서의 왜곡신호 보상기법

정회원 서호준*, 안정근**

Telephone channel compensation for using cepstrum coefficients

Ho-Joon Seo*, Jeong-Keun Ahn** *Regular Members*

요 약

최근 전화통화 이용한 개인 고유의 부가 서비스를 많이 이용하고 있다. 개인 고유의 서비스를 하기 위해서는 개인을 구별하는 기술이 필요하게 되는데, 전화통화 통해 개인을 구별할 수 있는 방법으로는 음성외 특징이 가장 많이 이용된다. 그러나 음성이 전화채널을 통과할 경우 여러 가지 왜곡요인으로 인해서 특징을 파악하는데 많은 어려움이 있다. 따라서 이러한 서비스를 제공하기 위해서는 전화 채널을 보상하여 음성의 특징 추출에 어려움이 없도록 해야 할 것이다.

본 논문에서는 음성 분석방법인 캡스트럼을 이용하여 보상하는 방법을 택하였는데 Klatt 방법이 가장 우수한 성능을 나타내었으며, 필터 계수의 보정을 통해 Klatt 방법보다 조금 더 성능이 우수한 결과를 얻었다.

ABSTRACT

Characteristic of speech is discriminate parameter to each individual. But speech signal distort to pass by telephone channel for channel bandwidth, jitter and nonlinear noise.

Since our application channel is the telephone network, we have investigating the impact of the network on speech and have developed compensation strategy. Approach to use cepstrum coefficients and Euclidean distance, Klatt method is superior performance.

This paper propose filter coefficients to difference Klatt method. Proposed method to appear best performance

I. 서 론

전화 한 통화로 본인임을 확인 받아 그 개인만의 고유한 서비스를 받을 수 있는 시대가 되었다. 이로 인해 개개인을 판별하는 여러 가지 기술들에 대한 관심이 고조되고 있는 시점이다. 개인을 구별하는 요소는 사람의 생체 정보이다. 지문, 손금, 홍채, 망막 음성 등 모두 개인의 특징을 나타낼 수 있는 정보들이다. 이들 중에서 음성은 다른 생체 정보들과는 달리 개인의 심리 상태나 환경에 따른 변화물

가지게 되는 문제를 갖는다. 그러나 다른 생체 정보들보다 어디서나 쉽게 개인 고유의 서비스를 받을 수 있는 장점을 갖고 있기도 하다.

다른 정보들보다 비교적 쉽게 개인 고유의 서비스를 받게 되는데 큰 역할을 하는 것으로 전화통화 생각할 수 있다. 그러나 전화는 열악한 채널환경과 음성이 전기적인 신호로 변화할 때의 음성의 손실이 생겨 전화통화 통한 개인 식별이 쉽지 않은 실정이다.

전화 채널은 사람이 낼 수 있는 음성 대역과 구리선(동선)의 여파 특성이 합쳐져서 채널 특성을 나

* 경희대학교 전기공학과(dsljgf@hitel.net)

** 김포대학 정보통신과(jkahn@kimpo.ac.kr)

논문번호 : 00011-0302, 접수일자 : 2000년 3월 2일

타낸다. 일반적으로 전화 채널의 대역폭은 300~3300Hz로 되어 있고, 이는 각 국가마다 조금씩 다르다.

음성 주파수에서 비교적 저주파 대역이 생체 정보를 많이 담고 있는 것으로 알려져 있다. 구리선은 비교적 저주파 대역에서 신호를 잘 통과시키기는 하지만, 전화 채널의 여러 가지 왜곡 현상에 의해 정보가 왜곡될 수 있다.

여기서는 음성신호 처리 기법을 이용하여 저주파 대역의 신호를 보상하는 방법에 대해 기술하기로 한다.

II. 전화 채널

2.1. 전화 채널 모델

전화 채널은 인간의 음성 대역폭을 기준으로 하고 있고, 음성 정보의 전달 매체인 구리선의 대역폭 특성을 함께 보인다. 일반적으로 전화 채널 대역폭은 300~3300Hz를 기준으로 하고 있고, 비교적 낮은 표본화율로 인해 화자를 식별하는데 어려움이 많다. 또한 전화 채널은 시변 시스템(time-varying) 특성을 나타내며, 여러 가지 왜곡 현상이 발생하므로 화자 식별에 앞서 전화 채널을 보상하는 방법에 대한 연구가 선행되어야 할 것이다.

2.1.1 왜곡 요소

손실을 보상하기 위해서는 손실의 원인을 알아야 할 것이다. 일반적으로 나타나는 전화 채널의 왜곡 요소와 보상방법은 다음과 같다.

이것들은 비선형 왜곡요소로서 실제로 보상이 상당히 까다롭거나, 거의 불가능한 것들이 대부분이다.

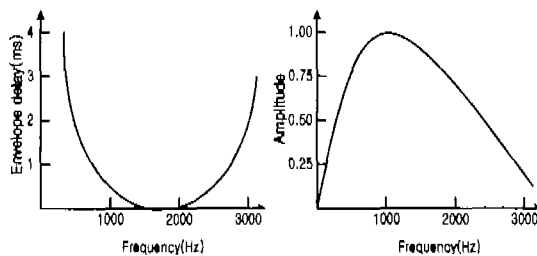


그림 1. 일반적이 전화 채널 특성

1) 주파수 offset(frequency offset) : 전화 채널에서 반송파에 의해 5Hz 이하의 주파수가 발생하는데 이는 복조단에서 carrier recovery loop의 방법으로

일부 보상이 가능하다.

2) 위상 지터(phase jitter) : 비교적 낮은 주파수가 변조될 때 고조파가 50~60Hz에 발생할 경우, 상용 전원 주파수와 간섭을 일으키게 된다. 복조단에서 위상을 확장하는 방법을 이용하여 보상할 수 있다.

3) 충격파 잡음(impulse noise) : 교환기에서 스위칭이 있을 때 발생한다. 이러한 경우의 잡음은 보상이 불가능하다. 여기에 열 잡음까지 존재한다면 신호의 레벨이 20~30dB 떨어지게 된다.

2.1.2 채널 모델링

전화 채널은 시변 시스템 특성을 갖고 있기 때문에 실제적인 분석은 매우 어렵고, 다만 수학적인 모델링을 이용한다. 입력 신호(음성)에 선형 여파기를 이용하여 여파하는 방법이 많이 이용되며 잡음원은 백색 잡음이 사용된다.

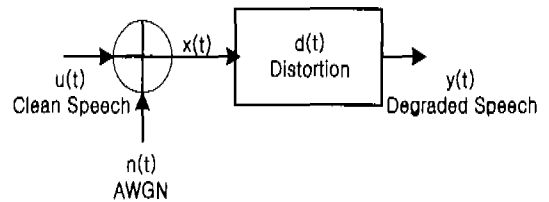


그림 2. 전화 채널 개요도

$$y(t) = \{s(t) + n(t)\} * d(t) \tag{1}$$

III. 채널 보상 시스템

3.1 시변 채널에서의 채널 보상

채널이 시간에 따라 변하는 시변 채널의 경우에는 bias를 adaptive 알고리즘을 이용하여 제거할 수 있으며, 등화기를 이용하는 경우는 음성의 평균 스펙트럼의 역함수를 취하여 등가적으로 여파를 할 수 있다.

음성의 평균 스펙트럼은 화자의 기본적인 특성음 가지고 있지만, 시간에 따른 화자의 상태가 다르기 때문에 여러 가지의 경우를 고려하여 혼란을 해야 보다 정확한 결과를 얻을 수 있다.

3.1.1 Cepstral 보상(cepstral compensation)

Cepstral 보상 기술은 채널에 의한 잡음 cepstral 벡터를 원래의 음성 cepstral 벡터에서 빼내는 원리를 이용하는 것이다. 즉, 기준신호와 이상적인 채널

로부터 실제 모델링 되어진 신호와 차이값 보상은 방법으로 수식적인 표현은 다음과 같다.

$$y_u(t) = s(t) * h_u(t) + n_u(t) \quad (2)$$

$y_u(t)$: 출력신호, $s(t)$: 입력신호,
 $n_u(t)$: 백색잡음 $h_u(t)$: 채널 u 의 임펄스 응답

이상적인 채널을 통과한 신호를 $y_u(t)$ 라 가정하면 (2)에서 cepstral 벡터를 구하는 식은 다음과 같이 나타낼 수 있다.

$$C_{yu} = IDFT(\log |Y_u(w)|) \\ = IDFT(\log |S(w)H(w) + N_u(w)|) \quad (3)$$

(3)에서 신호대 잡음비가 높아지면 백색잡음을 무시할 수 있게 되며 이는 다음과 같이 나타낼 수 있다.

$$C_{yu} \approx IDFT(\log |S(w)H(w)|) = C_{su} + C_{hu} \quad (4)$$

여기서 C_{su} 는 입력신호 cepstral 벡터를, C_{hu} 는 채널의 cepstral 벡터를 의미한다.

(4)에서 기준신호의 cepstral 벡터를 이용하여 그 차이값 구하면 다음과 같이 나타내어진다.

$$\Delta C_y = C_{yu} - C_{yu} \approx (C_{su} - C_{su}) + (C_{hu} - C_{hu}) \quad (5) \\ = \Delta C_s + \Delta C_h$$

C_{su} 는 이상적인 경우의 채널을 통과할 경우의 입력신호 cepstral 벡터, C_{hu} 는 이상적인 채널인 경우의 채널 cepstral 벡터를 나타낸다.

가장 이상적인 채널은 $\Delta C_h = 0$ 이 되는 환경이다. $\Delta C_s = 0$ 는 채널을 통과하기 이전의 신호 cepstral 벡터의 차이 값으로, 특별한 경우가 아니면 0에 수렴하게 된다. 그러므로 실제로 구하는 값은 ΔC_h 가 된다.

이러한 원리를 적용하는 cepstral 보상방법은 크게 3가지로 나누어지는데, SDCN, FCDCN, PDCN 방법이 있다.

(1) SDCN(Dependent Cepstral Normalization)

신호대 잡음비를 변화시켜 가면서 음성 데이터를 나누어 그룹으로 묶는 기술로 신호대 잡음비에 따라 훈련된 신호의 평균과 기준 신호의 차이값 계산하는 방법이다.

(2) FCDCN(Fixed Codeword-Dependent Cep-

stral Normalization)

SDCN 방법과 유사하나 훈련된 신호와 기준 신호의 차이값 벡터 양자화(VQ) 방법을 이용한다. 보상 벡터는 각각의 신호대 잡음비 클러스터를 벡터 양자화 하여 만들어 진다.

(3) PDCN(Phone-Dependent Cepstral Normalization)

실제 전화 채널 환경 변수가 많이 적용되는 경우로 위의 두 방법과 클러스터 구성하는 방법에 차이가 있다. 실제 환경과 비슷한 변수를 이용하므로 기주 stlsh호가 존재하지 않아도 보상을 할 수 있지만 각 환경이 독립적으로 발생하므로 변수가 많아져 계산량이 많아질 수 있는 단점이 있다. 위의 두 방법보다는 실제적인 방법이라고 할 수 있다.

3.1.2 채널 등화

전송 채널에서 신호의 왜곡을 갖게 하는 것은 전화 채널의 경우 대역 제한된 채널, 그리고 스피리어스 잡음의 발생 때문이다. 이러한 경우 비교적 적은 비용과 계산량으로 만족할 만한 보상효과를 내는 방법으로 cepstral normalization과 고역 통과 여파 방법이 있다.

(1) Cepstral normalization

각 신호 벡터로부터 평균 cepstral 계수를 추정하는 방법이다. 단지 음성 프레임만을 계산하므로 효율적이지만, 소리가 없는 경우 그대로 평균값을 쓰게 되어 오차가 커지는 단점을 갖는다.

앞서 설명한 cepstral compensation 방법은 백색잡음을 고려하지 않았고 이러한 경우 cepstral normalization 방법을 이용하면 보상 벡터는 null값을 갖게 된다.

(2) 고역 통과 여파(high-pass filtering)

환경조건을 미리 정의하여 만들어지는 보상 벡터 중에서 첫 번째 cepstral 계수를 감소시켜 나가는 방법이다.

(3) Spectral subtraction

음성 신호 처리에 많이 사용되는 방법으로 스펙트럼의 특성을 추정하는 것이다. 이는 잡음이 존재하지 않는 시간구간을 잡음이 많이 존재하는 구간에 입력으로 넣어 신호를 추정한다. 특히 소리가 없는 경우(silence sound) 뛰어난 성능 개선을 나타낸다. 짧은 구간의 스펙트럼 크기와 위상으로부터 각 표본에 잡음이 존재하는 것을 제거한다. 또한 짧은 구간의 스펙트럼을 강화(emphasis)시켜도 위상은 변화하지 않는 특성을 갖는다.

3.3.3 음성 분석 블록도

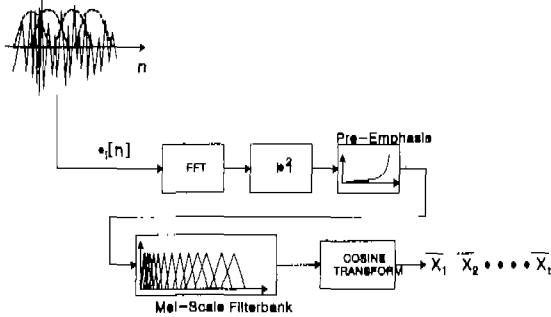


그림 3. Cepstral Analysis

IV. 모의 실험

4.1. 실험 환경

앞장에서 소개되었던 방법들 가운데 cepstral 계수를 구하여 채널이 보상됨을 모의 실험으로 보인다. $\Delta C_c = 0$ 으로 가정이 되었으며, 따라서 ΔC_c 만을 구하게 된다.

채널의 특성을 거리에 따라 다르게 하여, 먼 거리에서도 짧은 거리에서 통과하는 것과 마찬가지로 거리에 따른 채널 보상을 보인다.

(1) 전화 채널 모델링

그림 4와 그림 5는 거리에 따른 채널 모델을 나타낸 것이다. 거리가 멀어지면 구리선의 특성상 왜곡이 생기며 그 외에도 스위칭에 따른 왜곡, 에코 현상 등이 발생한다. 이를 모의실험에 적용하기 위해 원거리 채널의 경우 근거리 채널보다 15dB의 차이를 두었는데, 이는 입력신호의 강도를 15dB 낮췄다는 것과 같은 의미이다.

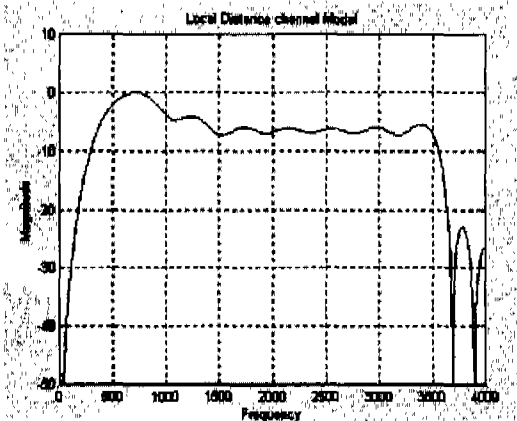


그림 4. Local distance channel model

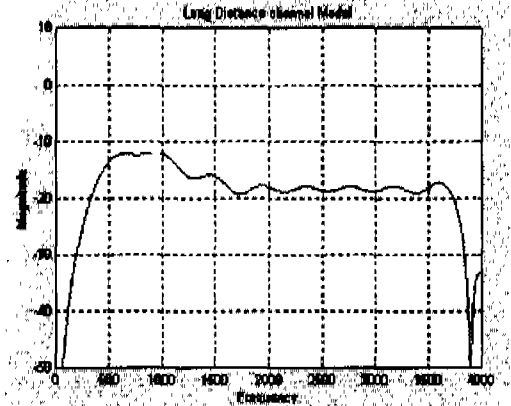


그림 5. Long distance channel model

4.2 입력 신호

입력 신호는 2500개로 표본화된 음성 신호를 사용하였으며, 이는 훈련된 신호이다. 기준 신호는 7500개의 표본화된 음성을 사용하기로 하였다. 입력 신호는 20번 훈련이 되었다.

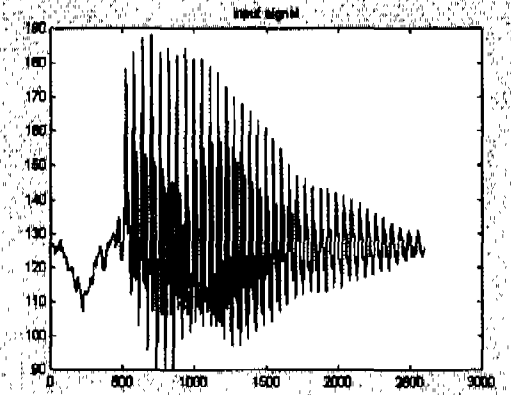


그림 6. 입력 신호

4.3 채널 보상의 결과

채널 보상의 결과를 얻기 위하여 선형 예측법, 전치 강세(pre-emphasis), Klatt의 3가지의 방법을 이용하였고, 이들의 cepstral 계수에 따른 기준 신호에 대한 상관값은 다음과 같다.

표 1. Cepstral 계수와 상관도

1.0000	1.0000	1.0000	1.0000
-0.8212	-0.3293	-0.0167	-0.5101
0.1826	-0.1404	-0.0475	-0.0623
0.0372	-0.1491	-0.0478	-0.0823
-0.2833	-0.1046	-0.0329	-0.0083
0.09397	-0.0264	-0.0208	0.0354
-0.0409	-0.0224	-0.0494	-0.0931

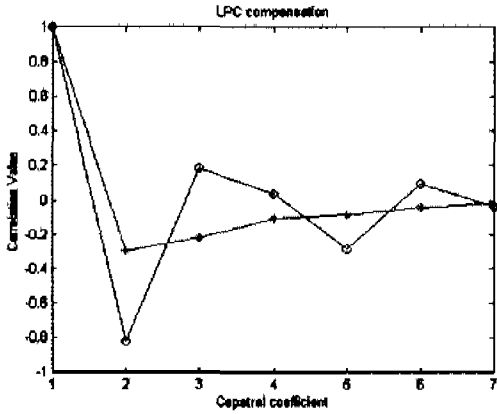


그림 7. LPC

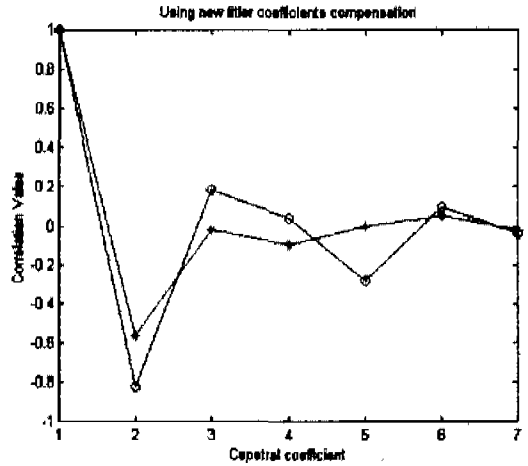


그림 10. 새로운 계수값 적용했을 경우

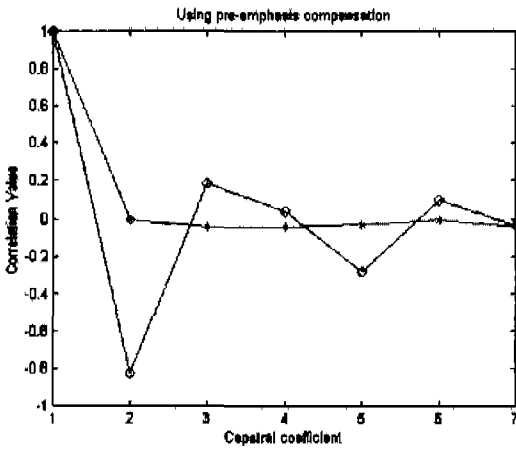


그림 8. Pre-emphasis

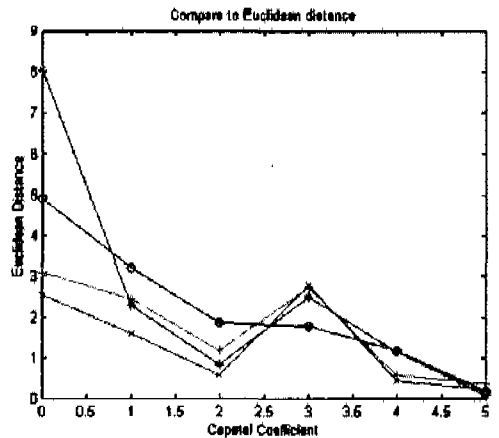


그림 11. 유클리드 거리 비교

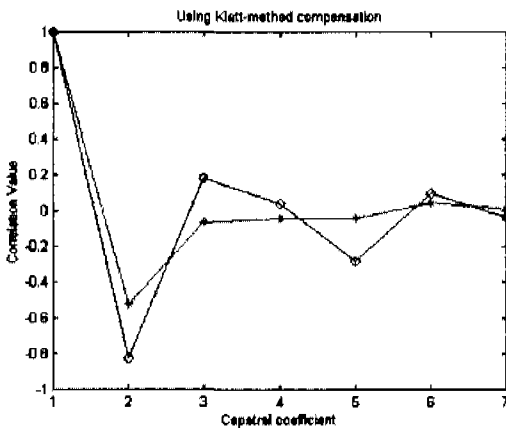


그림 9. Klatt-method

여기서 * 표시는 3가지 알고리즘의 결과값을 나타낸 것으로 기준신호를 추정한 값이다. 이들의 성능을 비교하기 위해 기준신호와의 유클리드 거리를 계산하였다. 유클리드 거리가 짧을수록 추정이 잘

되었음을 나타낸다.

그림 7.~그림 10은 이상적인 채널을 통과했을 경우와 각 알고리즘을 사용했을 경우의 그래프이다. Cepstral 계수와 상관값의 그래프이다. 이상적인 채널을 통과한 그래프와 유사할수록 좋은 성능을 나타내게 된다. 이에 대한 산술값은 표 1에 제시되어 있다. 표 1과 그래프를 통해서 본다면 Klatt-method가 이상적인 채널을 통과한 경우와 가장 유사한 값을 가짐을 알 수 있다. 이를 유클리드 거리로써 비교한 것이 그림 11이다. 유클리드 거리의 평균값으로 비교한 값도 역시 Klatt-method가 가장 좋은 성능을 나타내었다. 그러나 cepstral coefficient가 커지면 선형 예측법이나 전치강세보다 성능이 많이 떨어진다. 여기서, 새로 제안하는 방법은 mel scale을 Klatt-method에 적용하는 것이다.

Mel scale은 비선형적인 음성신호를 실험적 결과

