
개선된 선형예측 잔여를 이용한 음성의 잔향음 제거

박찬섭* · 김기만* · 강석엽*

Speech Dereverberation using Improved Linear Prediction Residual

Chan-Sub Park* · Ki-Man Kim* · Suk-Youb Kang*

요 약

배경 잡음과 실내 잔향음은 음성인식 시스템 성능 저하의 주요 이유이다. 많은 알고리즘이 음성의 잔향음 제거를 위해 개발되었다. 이 논문에서는 실내 환경에서 수정된 선형예측 잔여(Linear Prediction Residual)를 이용하여 음질 개선을 위한 잔향음 제거 방법을 제안한다. 제안된 잔향음 제거 방법은 음성에서 성문 경계의 순간에 발생한 성도(聲道) 시스템의 중요한 여기에 기반한다. 본 논문에서 제안한 방법은 3개의 센서로부터 수집한 반향신호로 각 센서에서의 시간지연 정보를 사용한다. 새로운 선형예측 잔여신호는 선형예측 잔여의 가중치와 힐버트 변환으로 얻은, 개선된 선형예측 잔여 조합을 사용한다. 코히런트하게 더해진 힐버트 포락선의 특징은 잡음과 반사로 인한 큰 진폭 피크를 가지는 것이다. 깨끗한 음성의 잔여는 개선된 음성을 얻는 시변전극 필터를 일으키는데 사용된다. 본 논문에서는 반향 환경에서 성능 분석을 위해 제안된 알고리즘의 시뮬레이션을 수행하였다. 제안된 알고리즘은 실내 잔향환경에서 기존의 알고리즘에 비해 반사된 음성의 품질 향상의 결과를 보였다.

ABSTRACT

Background noise and room reverberation are two causes of degradation in speech in listening situations. Many algorithms developed to enhance reverberant speech. In this paper we propose a dereverberation method for enhancement of speech using modified the linear prediction(LP) residual in reverberant room condition. The proposed dereverberation method based on the fact that the signification excitation of the vocal tract system takes place at the instant of glottal closure in voiced speech. Our method used delay information from each sensor, and we need reverberant signals from 3 sensors. We obtain a new LP residual signal using modified LP residual combination, which derived form weighting of the LP residual and the Hilbert transform of LP residual. The nature of the coherently added Hilbert envelop has several large amplitude spikes because of the effects of noise and reverberation. This residual of the clean speech is used to excite the time-varying all-pole filter to obtain the enhanced speech. We achieved simulation of proposed algorithm for performance analysis in reverberation environment. The proposed algorithm improves substantially the quality of reverberant speech.

키워드

LP, residual, dereverberation, SRR

I. 서 론

최근의 음성인식 기술은 연구 활동의 영역에서 벗어나 상용화되고 있다. 그러나 상용화된 음성인식 시스템은 정형화된 환경에서 비교적 좋은 성능을 보이나 설계 인식환경에서 성능이 저하될 수 있다. 잔향음과 부가 잡음은 음성인식 시스템의 가장 큰 성능 저하 요인이기 때문이다. 음성인식 기술은 연구단계와 비슷한 환경에서 좋은 성능을 보일지 모르나 전혀 다른 환경에서 성능의 저하가 일어날 수 있기 때문이다. 잔향음은 화자와 마이크로폰의 위치에 의해 좌우된다. 음성의 직접성분은 화자와 마이크로폰 사이의 거리가 멀어질수록 줄어들기 때문에, 음성의 직접 신호 대 잔향음성분비 (SRR : Signal to Reverberant component Ratio)는 줄어든다[1]. 부가적인 잡음에 의한 신호 대 잡음비 (SNR : Signal to Noise Ratio) 역시 화자와 마이크로폰의 거리가 증가할수록 줄어들지만, 이 저하는 음원의 음량이 커지면 보상된다. 그러나 SRR은 음량의 커짐으로 인해 보상되지 않는다.

음성강화의 어려운 점은 환경조건에 강하게 좌우된다는 것이다. 만약 화자와 마이크로폰이 가까우면 잔향음 효과는 최소가 될 것이고 기존의 방법으로 전형적인 중간템포의 잡음 레벨을 처리할 수 있을 것이다. 그러나 화자와 마이크로폰 사이가 멀어지면 눈에 띄는 잔향음과 많은 잡음을 포함하여 심하게 왜곡된다. 이런 조건에 서 음성의 잔향음과 잡음 제거는 매우 어려운 문제이다.

소스 필터 음성 제작 모델을 기반으로 하는 여러 잔향음 제거 알고리즘이 많은 저자에 의해 제안되었다. 이 방법들의 동기는 잔향음 환경에서 선형 예측 잔여가 다중 경로 반사에 기인한 다른 피크들에 의해 따르는 실제 임펄스를 포함한다는 정보이다[2].

일반적인 음성강화 방법은 두 카테고리로 나눌 수 있는데, 그것은 단일 마이크로폰을 이용한 방법과 다중 마이크로폰을 이용한 방법이다. 단일 마이크로폰 음성강화방법은 때때로 가우시안 잡음에 의해 저하된 싱글음성신호의 통계적인 모델의 개연적 구조를 사용한다[3, 4]. 이 모델들은 잔향음 제거나 다중 마이크로폰으로 연장되지 않는다.

다중 모델은 알고 있는 구조에서 마이크로폰 배열이 소리의 시공간 측정을 준비하여 배치한 마이크로폰 배열 처리에서 시작한다. 마이크로폰 배열은 단일 마이크로폰 방법과 비교하여 매우 큰 강점이 있다. 비적응형 알

고리즘은 제한된 길이의 방향에서 신호가 발생하는 동안 잡음을 쉽게 제거할 수 있다. 그러나 이 알고리즘은 잔향음 제거는 다루지 않는다. 적용 알고리즘이 어느 정도 잔향음 제거를 다루지만 기존의 방법들은 원칙적인 개연적 구조를 끌어내지 않으므로 최적화 되었다고 할 수 없다.

그러나 기존의 방법들에서 어떠한 시도도 개선을 위해서 소스 신호의 특징을 조사하지 않았다. 선형 예측 잔여 신호 샘플 같은 소스는 비상관적이므로 잔여 샘플은 신호보다 더 잡음에 가깝기 때문이다. 그러므로 잔여 신호는 음성 개선을 위해 유용한 특징을 가지고 있다고 기대하지 않았다. 이 논문에서 부가 잡음 및 잔향음 환경에서 강인한 음성인식을 위한 수정된 선형 예측 잔여 기법을 제안한다.

II절에서는 음성개선을 위한 제안된 접근법을 위한 기초로 선형예측 알고리즘을 서술하였고, III절에서는 잔향제거를 위한 새로운 선형예측 잔여 방법을 나타내었다. IV절에서는 시뮬레이션을 통한 실험결과를 비교하고, 마지막으로 V절에서는 결론을 내린다.

II. 선형예측 알고리즘

음성신호 처리에서 선형예측(LPC) 이론은 오랫동안 연구되었다. 선형예측 이론은 음성신호에서 특히 준불변 상태의 음성구간을 규정한다. 선형예측 이론은 합당한 성도(聲道) 시스템을 이끄는데, 실행하는데 있어서 수학적으로 정확하고 간단하며 분석적으로 다루기 쉬운 모델이다. 또한 인식과 코딩 적용 등에 용이하다. 음성 과정의 인접 샘플 사이의 높은 상관관계를 가지며 현재의 음성 샘플은 과거의 음성 샘플의 선형 조합으로서 예측할 수 있다는 장점을 가지고 있다[5].

인간은 스펙트럼과 시간 영역에서 높은 SNR 영역으로부터 약간의 특징을 포착하고, 낮은 SNR 영역의 특징을 추정하여 음성을 인식한다. 그러므로 음성 개선은 음성의 직접 신호 대 잔향음 성분비(SRR)을 높이는 한편, 부가점인 잡음에 의한 신호 대 잡음비(SNR)을 높여한다. SRR을 높이는 동시에 SNR을 높이는 한 가지 방법은 선형예측 잔여를 수정하는 것이다. 이는 다른 영역에 관하여 중요한 순간 영역 주위의 코히런트 부분을 개선하는 선형예측 잔여를 위한 가중치 함수를 만드는 것이다.

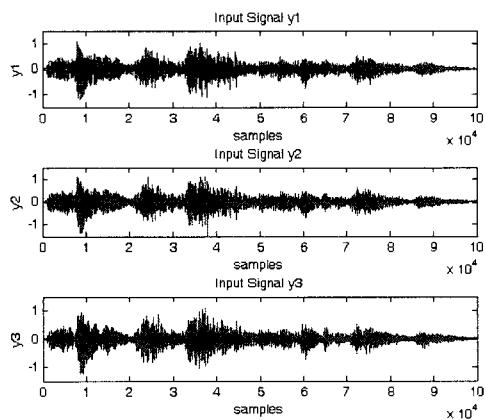


그림 1. 각 채널에 입력된 신호
Fig. 1 Input Signals of each channel

그림 1은 세 개의 마이크로폰으로 수신된 음성 데이터를 나타낸 것이다. 수신된 음성 데이터는 송신기에서 전달된 음성 데이터가 실내 잔향음과 부가잡음에 의해 저하된 데이터이다. 부가 잡음과 실내 잔향음에 의해 저하된 음성 신호 식 (1)과 같이 정의된다.

$$y = x + w \quad (1)$$

여기서 $y = [y(n-N+1), \dots, y(n-1), y(n)]^T$ 는 잡음이 섞인 음성의, $x = [x(n-N+1), \dots, x(n-1), x(n)]^T$ 는 깨끗한 음성의, $w = [w(n-N+1), \dots, w(n-1), w(n)]^T$ 는 잡음 샘플의 N 차 벡터를 각각 나타낸다. 입력신호 샘플의 p차 선형예측분석 실행에서, 선형예측 계수의 집합을 얻을 수 있다.

$$y(n) = -\sum_{i=1}^p a_i y(n-i) + v_y(n) \quad (2)$$

여기서 a_i 는 선형예측 계수이고, $v_y(n)$ 은 선형예측 오차이다. 선형예측 계수는 $N \times N$ 선형예측 행렬 A로 구성된다.

$$A = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ a_1 & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_p & a_{p-1} & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_p & \cdots & a_1 & 1 \end{bmatrix} \quad (3)$$

식 (2)는 행렬 A를 이용하여 다시 작성하면 다음식이 된다.

$$e_y = Ay \quad (4)$$

여기서 e_y 는 선형예측 잔여의 벡터이다. 식 (4)는 선형예측 잔여가 음성샘플의 역필터링에 의해 얻어진다는 것을 의미한다[6].

III. 제안한 잔향음 제거 방법

앞에서 설명한 선형예측 알고리즘을 이용하여 마이크로폰에 입력된 신호들의 선형예측 잔여신호를 얻을 수 있다. 그림 2는 각각의 마이크로폰에 입력된 신호의 선형예측 잔여를 구한 것이다. 선형예측 잔여는 위상에 좌우되는 양과 음의 샘플 모두를 가지고 있기 때문에, 각 순간에서의 선형예측 잔여의 강도는 선형예측 잔여 신호의 힐버트 포락선 계산에 의해 얻을 수 있다[7]. 선형예측 잔여 신호 $e(n)$ 의 힐버트 포락선 $\hat{e}(n)$ 은 다음 식으로부터 얻을 수 있다.

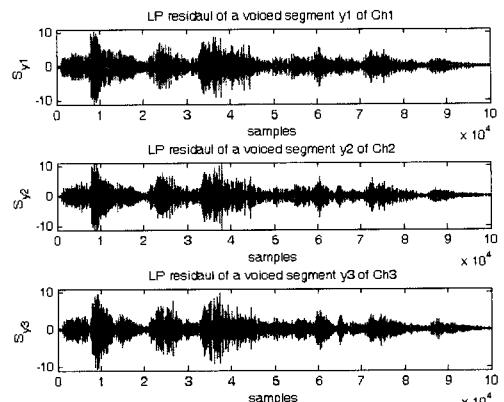


그림 2. 각 채널의 LP residual
Fig. 2 LP residuals of each channel

$$\hat{e}(n) = \sqrt{e^2(n) + e_H^2(n)} \quad (5)$$

여기서 $e_H(n)$ 은 선형예측 잔여 신호 $e(n)$ 의 헐버트 변환이다.

$$e_H(n) = \begin{cases} IDFT[jE(w)], & -\pi < w \leq 0 \\ IDFT[-jE(w)], & 0 < w \leq \pi \end{cases} \quad (6)$$

선형예측 잔여 신호 $e(n)$ 의 헐버트 변환은 $e(n)$ 의 DFT(Discrete Fourier Transform)의 실수와 허수부를 교환하고 난 후 IDFT (Inverse Discrete Fourier Transform) 계산을 하여 얻어진다. 여기서 $E(w)$ 는 $e(n)$ 의 DFT이다[8]. 이렇게 얻어진 헐버트 포락선이 그림 3에 나타나 있다.

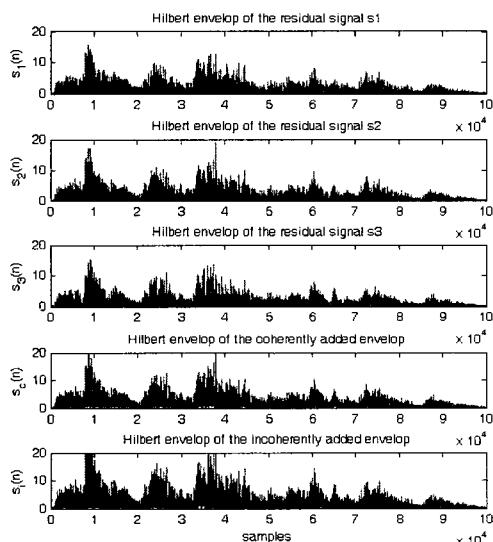


그림 3. 각 채널의 헐버트 포락선과 코히런트하게 그리고 인코히런트하게 더해진 헐버트 포락선
Fig. 3 Hilbert envelop of each channel and Hilbert envelop of the coherently added envelop and Hilbert envelop of the incoherently added envelop

잔향음과 잡음의 효과 때문에 선형예측 잔여의 헐버트 포락선에서 여러 큰 진폭 스파이크가 있다. 이런 스파이크의 효과를 줄이기 위해, 세 개의 마이크로폰으로부터 선형예측 잔여의 헐버트 포락선을 측정할 수 있고, 그 것들을 코히런트하게 더한다.

코히런트 덧셈을 위해 두 마이크로폰 사이의 음성 시

간지연이 계산되어야 한다. τ_{12} 는 마이크로폰 1과 2 사이의, τ_{13} 은 마이크로폰 1과 3 사이의 시간 지연이다. 마이크로폰 2와 3의 지연 보정된 헐버트 포락선은 첫 번째에 더해진다. 이 코히런트 덧셈을 위해 헐버트 포락선의 제곱이 고려된다. 그 결과 헐버트 포락선은 이 덧셈의 제곱 값이 된다.

$$\hat{e}_c = \sqrt{\hat{e}_1^2(n) + \hat{e}_2^2(n + \tau_{12}) + \hat{e}_3^2(n + \tau_{13})} \quad (7)$$

여기서 $\hat{e}_c(n)$ 은 코히런트하게 더해진 헐버트 포락선이고 $\hat{e}_1(n), \hat{e}_2(n), \hat{e}_3(n)$ 은 각각 마이크로폰 1, 2, 3의 선형예측 잔여의 헐버트 포락선이다. 반면에 인코히런트하게 더해진 헐버트 포락선은 식 (7)을 계산할 때 마이크로폰 사이의 시간 지연 τ_{12} 과 τ_{13} 이 더하지 않는다. 마이크로폰 사이의 시간지연 정보인 τ_{12} 과 τ_{13} 을 구하는 것 또한 매우 어려운 문제이므로 본 논문에서는 시간지연 정보를 미리 알고 있다는 가정하였다. 그림 3에 코히런트하게 더해진 헐버트 포락선과 인코히런트하게 더해진 헐버트 포락선을 비교해 놓은 것이다. 코히런트하게 더해진 헐버트 포락선의 특징은 잔여 가중치를 만들 수 있는 것이다. 선형예측 잔여 $e_1(n)$ 의 가중치는 다음에 사용된다[9].

$$e_{iM} = \frac{\sum_m e_i(n) \hat{e}_c(n)}{\sum_m \hat{e}_c(n)} \quad (8)$$

e_{1M}, e_{2M}, e_{3M} 은 각각 마이크로폰 1, 2, 3의 수정된 선형예측 잔여이다. 여기서 m 은 한 프레임의 길이를 나타낸다. 기존의 방법은 식 (8)에서 구한 한 개의 수정된 선형 예측 잔여 e_{iM} 을 이용하였지만, 본 논문에서는 각 센서의 수정된 선형예측 잔여의 조합을 이용하여 새로운 선형예측 잔여값을 구해내었다. 식 (8)을 통하여 얻은 각각의 센서의 수정된 선형예측 잔여에 각각 다른 가중치 w_1 과 w_2, w_3 을 주어 이들의 조합으로 개선된 선형예측 잔여를 구하였다.

$$e_M(n) = \sum e_{1M}(n) \times w_1 + e_{2M}(n) \times w_2 + e_{3M}(n) \times w_3 \quad (9)$$

IV. 실험 결과 및 고찰

반향 환경에서 제안된 알고리즘의 실험을 수행하였다. 1개의 송신기와 3개의 마이크로폰 배열을 사용하였고, 실험에 사용된 송신기는 DIATONE DS-7을, 마이크로폰은 Onkyo Sokki Corp.의 transducer Hoshiden KUC1333을 사용하였다. 마이크로폰 사이의 거리는 20cm이고, 송신기와 마이크로폰 사이의 거리는 2m이다. 실험에 사용된 음성 데이터의 RT(Reverberation Time)는 1.3초이다. 음성 데이터는 48kHz로 샘플링 되었으며, 16비트 양자화로 저장되었다. 실험 환경은 그림 4와 같은 에코룸에서 음성 데이터를 수집하였다.

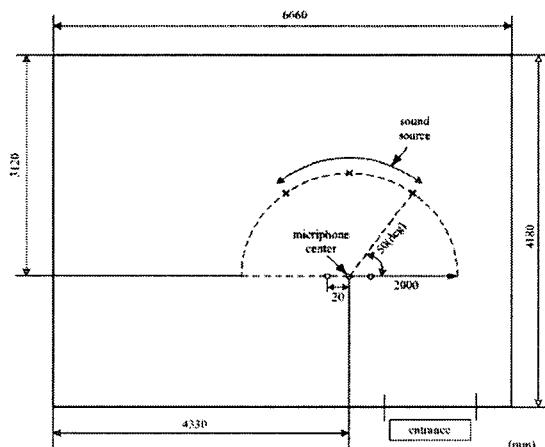


그림 4. 실험 환경
Fig. 4 Experimental environment

10차 선형 예측 분석은 세 개의 마이크로폰 출력에서 선형 예측 잔여의 힐버트 포락선을 얻는데 수행되었으며, 이렇게 구한 선형 예측 잔여를 이용하여 수정된 선형 예측 잔여를 얻을 수 있었고, 개선된 음성은 수정된 선형 예측 잔여로 시변 10차 전극(all-pole) 필터를 통해서 얻는다.

기존의 음성 개선 방법은 이렇게 수정된 하나의 선형 예측 잔여 $e_{iM}(n)$ 만을 이용하였지만, 본 논문에서 제안한 방법은 $e_{1M}(n)$, $e_{2M}(n)$, $e_{3M}(n)$ 의 조합을 이용하여 개선된 선형 예측 잔여를 이용하였다.

모의실험 결과를 그림 5에 나타내었다. (a)와 (b)는 각각 송신기와 수신 센서의 선형 예측 잔여를 나타내었다. (c)에서는 기존의 방법인 한 개의 $e_{iM}(n)$ 을 이용

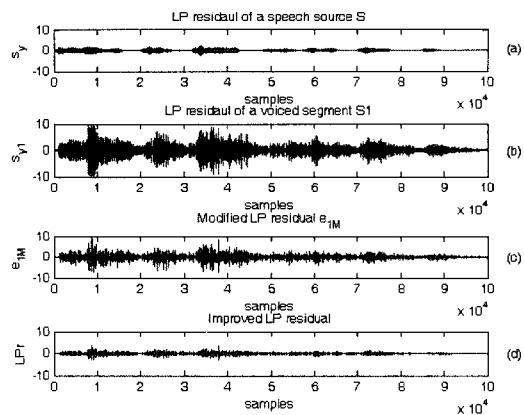


그림 5. (a) 원 음성 신호의 선형 예측 잔여, (b) LP residual of voiced segment S1, (c) 기존방법의 선형 예측 잔여, (d) 개선된 선형 예측 잔여
Fig. 5 (a) LP residual of clean source signal, (b) LP residual of voiced segment, (c) LP residual of existed method. (d) Improved LP residual

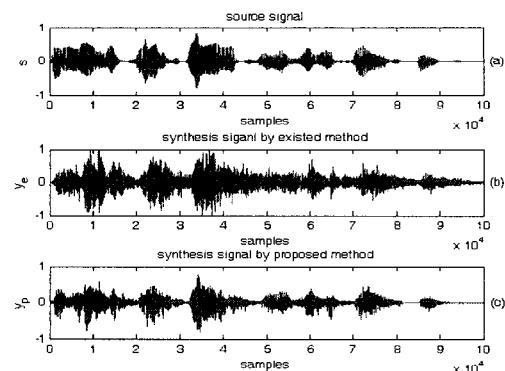


그림 6. (a) 원 음성 신호, (b) 기존의 선형 예측 잔여를 합성한 음성 신호, (c) 개선된 선형 예측 잔여를 합성한 음성 신호
Fig. 6 (a) clean speech signal, (b) synthesis by existed LP residual, (c) synthesis by improved LP residual

한 선형 예측 잔여를 볼 수 있으며, (d)는 제안된 방법인 선형 예측 잔여의 조합을 이용한 개선된 선형 예측 잔여를 볼 수 있다. (c)와 (d)를 비교해볼 때, 기존의 방법으로 얻은 선형 예측 잔여 (c)보다 제안된 방법으로 얻은 선형 예측 잔여인 (d)가 송신기에서 전파된 신호의 선형 예측 잔여 (a)에 매우 유사함을 알 수 있다. 기존의 방법과 제

안한 방법으로 합성한 신호를 깨끗한 음성 신호와 비교하여 그림 6에 나타내었다. 파형의 유사도를 육안으로 확인했을 때 및 주파적 척도에 의해 합성된 신호를 청취했을 때 본 논문에서 제안한 방법이 기존의 방법에 비해 원 신호에 더 가까움을 확인할 수 있었다. 또한 기존의 방법과 제안한 방법에 따른 결과를 객관적 척도로써 비교하기 위하여 각각의 방법으로 얻은 합성 신호와 원 신호 사이의 Spectral Distance (SD)를 구하였다. 음성의 품질을 평가할 수 있는 객관적 척도들 중 주파수 영역에서 음성 스펙트럼의 magnitude에서 야기된 편차를 측정하여 평가하는 방법들이 있는데 그 중 가장 대표적인 방법 중의 하나가 SD를 구하는 방법이다.

그림 7의 왼쪽은 기존의 선형예측 잔여를 합성한 음성 신호와 깨끗한 음성 신호의 SD를 구한 값이고 오른쪽은 개선된 선형예측 잔여 신호를 합성한 음성 신호와 깨끗한 음성 신호의 SD를 구한 값으로 제안된 방법으로 얻은 선형예측 잔여 신호로 합성한 음성 신호가 원 신호와 상관성이 높음을 알 수 있다.

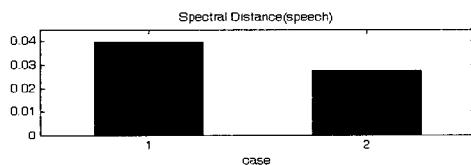


그림 7. 합성된 신호와 원 신호의 SD
Fig. 7 compare synthesis signals with source signal

V. 결 론

본 논문에서는 반사 환경에서의 음성 통신에서 센서 배열로부터 수집된 음성을 위한 새로운 방법을 제안하였다. 방법은 음성의 여기 정보의 특성에 기반하고 있다. 여기에서 가장 중요한 특성은 여기의 세기가 성문 경계의 주변에서 최대라는 것이다. 본 논문에서는 선형예측 잔여를 음원의 특징을 얻는데 사용하였다. 가중치 합수는 다른 센서로부터 코히런트하게 자연 보상된 선형예측 잔여의 헐버트 포락선 조합에 의해 얻었다. 새로운 선형예측 잔여 신호는 선형예측 잔여의 가중치와 헐버트 변환으로 얻은 수정된 조합들을 사용하였다. 모의실험 결과 제안된 방법을 통한 음질 개선 방법이 기존의 방법 보다 향상된 결과를 얻을 수 있음을 알 수 있었다.

감사의 글

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었으며, 이에 대해 깊이 감사를 드립니다.

참고문헌

- [1] P. Satyanarayana: "Short segment analysis of speech for enhancement", Ph.D. thesis, Dept of Electrical Engineering, IIT Madras, Chennai, India, Feb 1999.
- [2] J. Makhoul, "Linear Prediction: A tutorial review", *Proc. IEEE*, vol. 63, no. 4, pp. 651-580, Apr 1975.
- [3] A. Kounoudes, P. Naylor, and M. Brookes, "The DYPSA algorithm for estimation of glottal closure instants in voiced speech", in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. I-349-I-352, May 2002.
- [4] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Speech Communication*, vol. 9, no. 5-6, pp. 453-467, Dec. 1990.
- [5] Lawrence Rabiner, Biing-Hwang Juang, "Fundamentals of Speech Recognition", pp 97-122. Prentice-Hall International, Inc.
- [6] Wen Jin, Michael S. Scordilis, "Speech enhancement by residual domain constrained optimization", *Speech Communication*, vol. 48, pp. 1349-1364, 2006.
- [7] S. R. Mahadeva Prasanna, Suryakanth V. Gangashetty, B. Yegnanarayana, "Significance of vowel onset point for speech analysis", *Signal Processing and Communications (Biennial Conf., IISc Bangalore, India)*, pp. 81-88, July 2001.
- [8] B. Yegnanarayana, S. R. Mahadeva Prasanna, Ramani duraiswami, Dmitry Zoukin, "Processing of Reverberant Speech for Time-Delay Estimation", *IEEE Transaction on Speech and Audio Processing*, vol. 13, no. 6, pp. 1110-1118, November 2005.
- [9] B. Yegnanarayana, S. R. M. Prasanna, and K.S. Rao, "Speech enhancement using excitation source information", *International Conference on Acoustics, Speech and Signal Processing*, Orlando, FL, USA, May 2002.

저자소개



박 찬 섭(Chan-sub Park)

2006년 2월 한국해양대학교
전파공학과(공학사)
2006년~현재 한국해양대학교
전파공학과 석사과정

※ 관심분야 : 수중통신, DSP, ARM 및 마이크로프로세서 하드웨어 구현

김 기 만(Ki-man Kim)



1988년 2월 연세대학교
전자공학과(공학사)
1990년 8월 연세대학교
전자공학과(공학석사)
1995년 2월 연세대학교 전자공학과(공학박사)
1995년 3월~1996년 8월 : 연세대학교 의과대학
의용공학교실(Fellow)
1996년 9월~현재 : 한국해양대학교 전파공학과 교수

※ 관심분야 : 수중통신, 소나신호처리, 3차원 음향제어, 채널 추정, DSP 및 CPLD 구현

강 석 엽(Suk-Youb Kang)



1997년 2월 인천대학교
전자공학과(공학사)
1999년 2월 인하대학교
전자공학과(공학석사)
2005년 8월 인하대학교 전자공학과(공학박사)
2001년 7월~2006년 8월 (주)아이엔텍 대표이사
2006년 8월~현재 한국해양대학교 BK21 연구교수