工學碩士 學位論文

개선된 선형예측 잔여를 이용한 음성의 잔향음 제거

Speech Dereverberation using Improved LP Residual Combination

指導敎授 金 基 萬

*2008*年 *2*月

韓國海洋大學校 大學院

電 波 工 學 科

朴 贊 燮

本 論文을 朴贊燮의 工學碩士 學位論文으로 認准함.

委員長 : 工學博士 鄭 智 元 (印)

委 員 : 工學博士 姜 錫 燁 (印)

委 員 : 工學博士 金 基 萬 (印)

2008年 2月

韓國海洋大學校 大學院

電 波 工 學 科

朴 贊 燮

차 례

그 림 차 례

표 차 례

Nomenclature

$A(z)$	오차 전달 함수
a_i	선형예측 계수
$E(w)$	선형예측 잔여신호의 DFT
$e_H(n)$	힐버트 변환
$e_{im}(n)$	수정된 선형예측 잔여신호
$e_M(n)$	제안된 선형예측 잔여신호
$e_y(n)$	선형예측 잔여신호
$\hat{e}(n)$	힐버트 포락선
$\hat{e}_c(n)$	코히런트하게 더해진 힐버트 포락선
p	선형예측 차수
V_n	평균제곱 오차
$v_y(n)$	선형예측 오차
$w(n)$	잡음 신호
$x(n)$	깨끗한 음성 신호
$y(n)$	잡음 섞인 음성 신호

Abbreviations

DFT	Discrete Fourier Transform
IDFT	Inverse Discrete Fourier Transform
LPC	Linear Prediction Coding
RT	Reverberation Time
SNR	Signal to Noise Ratio
SRR	Signal to Reverberant component Ratio

Abstract

Speech Dereverberation using Improved LP Residual Combination

Recently, information of document form is flooded with advancement of the internet, and the demand to converts the document in voice is increasing. Advanced countries are securing a voice-activated technology service for various text information in numerous industries such as multimedia and communications.

The speech signal received from speaker in an acoustical environment is corrupted both by additive noise as well as room reverberation. When people listen to a lecture in a large auditorium, it is difficult to comprehend speech. This is because reverberant sound, witch is reflected from the walls or ceiling, masks in direct sound. So we need a robust algorithm for reverberant speech enhancement.

The difficulty of speech enhancement depends strongly on environmental conditions. The reverberation effect is critically dependent on the position of the microphone and the speaker. The direct component of speech is reduced with increasing

distance of the microphone from the speaker, hence the direct signal to reverberant component ratio (SRR) of speech decreases. The signal to noise ratio (SNR) due to additive noise also decreases with increasing distance of the microphone from speaker, but this reduction can be compensated by increasing the volume of the source of speech. But the SRR is unaffected by the increase in volume.

If a speaker is close to a microphone, reverberation effects are minimal and traditional methods can handle typical moderate noise level. However, if the speaker is far away from a microphone, there are more severe distortions, including large amounts of noise and noticeable reverberation. Denoising and dereverberation of speech in this condition has proven to be a very difficult problem. Therefore, we need a robust speech processing technique in reverberation effect by distance between speaker and microphone.

In this thesis, an improved LP residual method is proposed for a speech recognition in reverberation and noise condition from multiple microphones. The enhanced speech is significantly better compared to the coherently added speech signal, in the sense that the reverberation effects are reduced significantly. The proposed method showed the improved ability than the existed one for more than 10% in numerical value.

제 1 장 서 론

최근, 인터넷의 발달로 문서형태의 정보는 홍수를 이루고 있으며, 이를 음성으로 변환하고자 하는 수요는 급격히 늘어나고 있다. 특히 선진국에서는 멀티미디어·통신 등 여러 분야에서 다양한 문자정보를 음성으로 서비스하기 위해 경쟁적으로 음성 인식 기술을 확보하고 있다. 또한 음성처리 분야의 근간이 되는 음성인식 기술은 연구 활동의 영역에서 벗어나 상용화되고 있다. 그러나 상용화된 음성인식 기술은 연구단계와 같은 환경에서는 비교적 좋은 성능을 보이나 실제 인식환경에서는 성능이 저하될 수 있다. 전형적인 환경 조건에서 수집된 음성 신호는 보통 잔향음과 부가잡음에 의하여 저하되는데, 실제로 강당, 화상회의 회의실, 강의실 등과 같은 닫혀진 공간에서 마이크로폰 시스템에 수신된 신호는 직접 전달파와 벽면에 의한 반사파들이 더해진다. 결국 실제 발생음을 명확히 얻기가 어렵기 때문에 음질 향상 기법의 적용이 필요하다.

음질 향상의 어려운 점은 환경조건에 크게 영향 받는다는 것이다. 잔향음은 화자와 마이크로폰의 위치에 의해 크게 좌우된다. 음성의 직접성분은 화자와 마이크로폰 사이의 거리가 멀어질수록 줄어들기 때문

에, 신호 대 잔향음 성분비(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 커지더라도 보상되지 않는다. 만약 화자와 마이크로폰이 가까우면 잔향음 효과는 최소가 될 것이고 기존의 방법으로 전형적인 중간 정도의 잡음 레벨을 처리할 수 있을 것이다. 그러나 화자와 마이크로폰 사이가 멀어지면 눈에 띄는 잔향음과 부가 잡음을 포함하여 신호가 심하게 왜곡된다. 이러한 조건에서의 잔향음 제거와 음성의 잡음제거는 매우 어려운 문제이다. 그러므로 화자와 마이크로폰 사이의 거리에 의한 잔향음 효과에 강인한 음성 처리 기술을 필요하다.

음질 향상을 위한 방법들은 잡음 요인에서 구하고 이 요인들로부터 음성을 합성하는 것에 기초하여 발전되어왔다[2]. 저하된 음성의 all-ploe 모델링은 이런 방법들의 하나이다[3]. All-pole 모델링에서, 만약 잘못된 피크가 구해진다면 이 피크들이 향상될 수 있다. 또한 자연스러운 음성의 부드러운 윤곽선과 비교할 때, 이러한 피크들의 일시적인 시퀀스는 스펙트럼 피크의 윤곽선에서 불연속을 일으킨다.

일반적으로 음질 향상의 방법은 short-time spectral 포락선의 수정에 달려있는 것처럼 보인다. 만약 스펙트럼 포락선의 특징 추출에 오

차가 있거나, 오차가 스펙트럼 특징의 temporal contours 수정으로 인한 스펙트럼 포락선에 나타난다면, 처리된 음성은 부자연스럽게 들리는 왜곡을 나타낼 수도 있다.

음질 향상을 위한 피치의 주기에 기반한 방법도 제안되었다[4]. 이런 음질의 향상을 위한 방법들은 잡음이 포함된 음성 신호를 얼마나 정확하게 찾아내느냐에 따라 성능이 좌우된다. 또한, 이러한 경우에 합성 여기 신호는 음성을 만드는데 사용된다. 따라서 소음의 효과가 줄어든다 해도, 음성의 질은 낮아지게 된다.

피치 음조곡선과 음정의 길이 같이 여러 구분을 넘는 변수는 강한 특징이다. 그러나 비록 향상된 음성 신호를 만들기 위하여 스펙트럼 포락선과 각 분석 프레임의 여기를 모두 필요하다 할지라도, 이러한 특징은 음질 향상을 위해 유용하지 않다.

그러나 위에서 언급한 몇 가지 방법들에서 음질 향상을 위한 음원 신호의 특징을 조사하지 않았다. 이와 같은 가장 큰 이유는 음원 신호에서 선형예측 잔여 신호 같은 샘플은 비상관적이므로 잔여 샘플은 음원 신호보다 잡음에 더 가깝기 때문이다[5]. 그러므로 잔여 신호는 음질 향상을 위한 어떠한 유용한 특징을 가지고 있다고 예상되지 않았다. 이 논문에서는 부가 잡음 및 잔향음 환경에서 강인한 음성인식을 위한 수정된 선형예측 잔여 기법을 제안한다.

본 논문에서 제안된 음질 향상을 위한 방법은 멀티 마이크로폰 환경에서 선형예측 잔여신호의 시간에 따른 변화의 신호해석을 위해 힐버트 포락선을 사용하여 수정된 선형예측 잔여 신호를 구하고, 수정된 선형예측 잔여 신호의 조합을 통하여 음성의 질을 향상시키고자 하였다.

제 2 장에서 음질 향상을 위한 기존의 방법에 대해 간략히 살펴본 후, 제 3 장에서 잔향효과 제거 및 음질 향상을 위한 제안된 접근법을 서술하였다. 제 4 장에서는 시뮬레이션을 통한 실험결과를 비교하고, 마지막으로 제 5 장에서 결론을 내린다.

제 2 장 잔향음 제거 연구 동향

잔향음에 대해서 설명하면 잔향음이란 공간 때문에 생기는 일종의 반사음들로 청취 환경의 크기나 모양, 재질에 따라 소리의 반사되는 각도나 질이 변하면서 미묘한 시간차를 두고 우리의 귀에 도착하면서 생긴다.

현재까지 많은 저자들에 의하여 음질 향상을 위한 알고리즘들이 개발되어왔다. 일반적인 음질 향상방법은 두 종류로 나눌 수 있는데, 그것은 단일 마이크로폰을 이용한 방법과 다중 마이크로폰을 이용한 방법이다. 첫 번째로 단일 마이크로폰을 이용한 음질 향상방법으로 가우시안 잡음에 의해 저하된 단일 마이크로폰 음성신호의 통계적인 모델을 사용하는 방법이 있다[6, 7]. 그러나 이 모델들을 이용한 음질향상 방법은 잔향음 제거나 다중 마이크로폰을 이용한 방법으로 연장되지 않는다.

두 번재로 다중 마이크로폰을 이용한 음질 향상방법은 알고 있는 마이크로폰 배열 구조에서 소리의 시공간 측정을 준비하는 배치한 마이크로폰 배열 처리에서 시작한다. 마이크로폰 배열을 이용한 방법은 단일 마이크로폰을 이용한 방법과 비교하여 신뢰도를 높일 수 있다는 장점이 있으며, 그 중에서 비적응형 알고리즘은 가장 단순한 방법에 속하며 제

한된 방향에서 발생한 신호에서 잡음을 쉽게 제거할 수 있다[8]. 본 논문에서는 신뢰도를 높이기 위해 다중 마이크로폰 배열을 구조를 사용하며 수신된 신호의 선형예측 알고리즘으로 얻은 잔여신호를 이용하여 잔향음을 제거 하고자 하였다.

제 2-1 절 단일 마이크로폰 잔향음질 향상을 위한 알고리즘

소음이 없는 조건하에서, 잔향 음성의 질은 두 개의 뚜렷한 성분인 coloration 효과와 긴 반사 잔향음에 좌우된다. 이것들은 두 개의 물리적인 변수, 신호 대 잔향음 성분비(SRR)과 잔향 시간에 각각 대응되는데, 이 방법에서는 이러한 정보에 의하여 단일 마이크로폰을 이용한 잔향음질 향상을 위한 알고리즘을 제안하였다. 여기서 coloration 효과란, 반사음의 지연시간이 수㎳~수십㎳로 짧은 경우, 직접음과 간접음 간에 위상 간섭이 생겨 음색이 변화되는 현상을 말한다. 첫 번째 단계에서는 coloration 효과를 줄이거나 신호 대 잔향음 성분비를 늘리기 위하여 역필터를 계산하였다. 두 번째 단계에서는 긴 반사 잔향의 효과를 최소화하기 위하여 스펙트럼 차감(subtraction)을 사용하였다. 그림 2-1은 이 알고리즘의 블록다이어그램을 나타낸 것으로 (a)와 (b)의

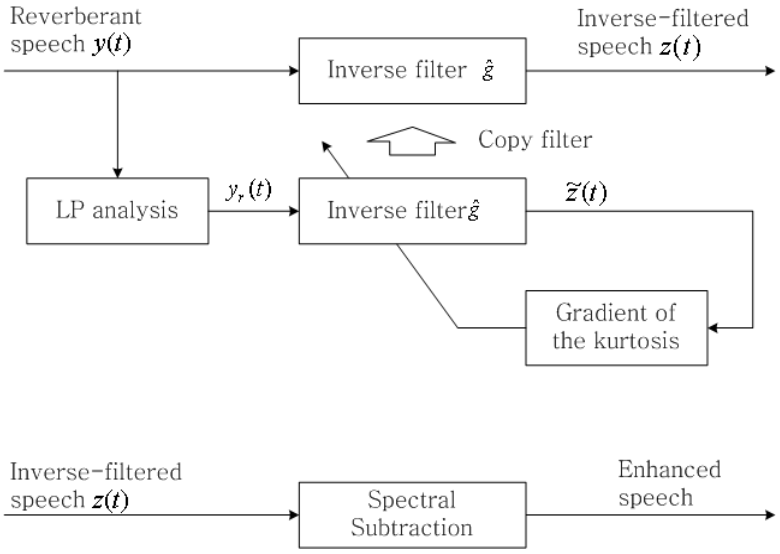

그림 2-1. 단일 마이크로폰 잔향음질 향상을 위한 2단 알고리즘의 블록다이어그램

Fig 2-1. Block diagram of a two-stage algorithm for one-microphone reverberant speech enhancement

두 단계로 나타나있다. Berkley와 Allen은 두 개의 물리적인 변수인 잔향시간 T_{60}과 화자-청자 사이의 거리가 잔향 음질에 중요하다는 것을 확인하였다[9]. 늦은 반사는 음성 스펙트럼을 왜곡하고 명료도와 음질을 떨어뜨린다. 이른 반사는 coloration이라 불리는 왜곡을 야기하는데 이른 반사의 nonflat 주파수 응답은 음성 스펙트럼을 왜곡시킨다. coloration효과는 주파수 응답의 표준편차로 정의된 스펙트럼 편차로 간주된다. Jetzt는 스펙트럼 편차가 신호 대 잔향음 성분비에 의하여

결정됨을 확인하였다[10]. 게다가 상대적인 잔향 에너지는 화자청자의 위치에 관계없이 대략 일정하므로, 스펙트럼 편차는 직접신호의 힘을 결정하는 화자 마이크의 거리에 의해 결정된다.

여기서 coloration 효과와 긴 반사 두 가지 요소를 다룰 두 단계의 알고리즘을 제안하는데, 첫 번째로 coloration 효과를 제거하거나 신호 대 잔향음 성분비를 높이기 위한 역필터 계산을 하고 두 번째 단계에서 긴 반사의 영향을 줄이기 위한 스펙트럼 차감 방법을 제안하였다.

우선 coloration 효과를 제거하거나 신호 대 잔향음 성분비를 높이기 위한 역필터를 구하는데, Gillespie에 의하여 제안된 다중 마이크로폰 역필터 알고리즘을 사용하였다. 역필터는 역필터된 신호의 선형예측 신호의 Kurtosis의 Maximizing에 의해 찾을 수 있다[11]. 임펄스 응답은 이른 반사와 느린 반사 두 부분으로 분해할 수 있다. 이때 늦은 반사가 역필터된 음성의 질을 저해할 수 있으므로 늦은 반사의 효과를 계산하고 제거함으로서 음성의 질을 향상시킬 수 있다. 결과적으로 첫 번째 단계에서 이른 반사에 의한 coloration 효과를 줄이고 역필터로 잔향 음성의 파워 스펙트럼을 개선하고 두 번째 단계에서 늦은 반사의 효과를 줄이는 스펙트럼 차감은 비상관적인 잡음에 대해 향상된 음성을 산출한다[12].

제 2-2 절　시공간 평균을 이용한 다중 마이크로폰 음성 잔향음 제거

잔향 음성의 강화를 위한 방법에서 음원 필터 음성 제작 모델의 사용은 지난 몇 년 동안 상당히 주목받았을 뿐만 아니라, 최근에는 선형예측 계수의 공간영역 평균이 알고리즘의 이들 타입의 실행에 있어서 정확도를 높일 필요가 있는 것으로 나타났다. 이 방법에서는 공간영역 평균과 상호 주기 공간영역 평균에 기반한 새로운 접근법의 조합으로 구성된 선형예측 잔여신호로 동작하는 새로운 다중채널 음성 잔향효과 제거 접근법을 제시하였는데, [13]에서 볼 수 있듯이, 선형예측 계수의 정확도 향상은 잔향환경에서 만들어진 음성 신호의 시간 맞춤 정보인 공간영역 평균으로부터 얻을 수 있다고 실험적으로 보여준다. 따라서 인접한 larynx-cycles 사이의 파형이 천천히 바뀐다는 사실을 사용하는 것은 선형예측 잔여신호 끝의 향상을 위해 자체적으로 또 그것의 가장 가까운 주기의 평균에 의해 대체된다는 것이다.

이 방법에서는 공간과 시간영역 평균 둘로 구성된 새로운 다중 마이크로폰 음성 잔향 알고리즘을 제안 하였는데[14], 이 알고리즘의 흐름도는 그림 2-2에 나타나있다. 이 방법은 공간영역 평균 마이크로폰 어레이 입력의 선형예측과 선형예측 잔여신호의 larynx synchronous

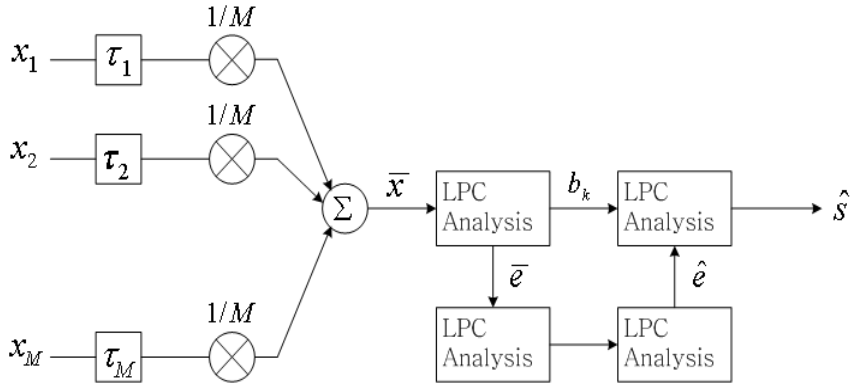

그림 2-2. 시공간 평균을 이용한 다중 마이크로폰 음성 잔향음 제거의 블록다이어그램

Fig 2-2. Block diagram of multi-microphone speech dereverberation using spatio-temporal averaging

temporal averaging에 기반하고 있다. 하지만 이 알고리즘은 유성음 부분에만 초점을 맞추고 있고 무성음 부분에 대해서는 고려하고 있지 않는다. 그림 2-2에서 알 수 있듯이 지연시간의 평균의 합을 이용함으로서 선형예측 계수의 정확도 향상을 위해 잔향환경에서 만들어진 음성신호의 시간 맞춤 정보인 공간영역 평균으로부터 얻어야 하는 복잡한 과정을 거쳐야 하므로 실제 시스템 구현에 있어서 어려운 점을 가지고 있다.

제 2-3 절　잔향 음질 향상을 위한 피치 기반 단일 마이크로폰 알고리즘

이 방법에서 제안된 모델은 두 가지 단계로 구성되어 있는데 그 첫 번째 단계는 잔향시간의 계산을 위한 피치 기반 지연 계산이다. 그리고 모델의 두 번째 단계는 잔향음 향상의 방법으로 첫 번째 단계에서 계산된 잔향 시간을 이용하여 늦은 잔향 성분을 차감하는 과정이다.

피치 기반 측정법은 최근의 다중 피치 추정 알고리즘으로부터 얻는다[16]. 피치 추적 알고리즘은 네 단계로 구성되어있다. 첫 번째 단계에서 입력신호는 16㎑로 샘플되고, 4차 gammatone 필터에 의하여 128 주파수 채널로 필터링 된다[17]. 800㎐보다 낮은 중심 주파수의 채널은 저주파로 분류되고 나머지는 고주파로 분류된다. 포락선은 고주파 채널에서 얻는다. 첫 번째 단계의 끝에서 정규화된 correlogram은 모든 채널에서 16㎳의 윈도우 사이즈를 사용하여 계산된다. 채널과 피크의 선택은 두 번째 단계에서 구성된다. 잡음에 의해 저하된 채널만 선택된 후 다음 단계로 넘어간다. 세 번째 단계는 모든 채널의 주기성 정보를 통합하고 마지막 단계는 hidden Markov 모델을 사용하여 연속된 피치 추정을 형성한다. 피치 기반 측정법은 잔향 시간을 계산하기 위하여 사용되었다.

잔향음에는 이른 잔향음과 늦은 잔향음이 존재하는데 이 연구에서는 이전 단계에서 계산된 잔향 시간을 이용하여 늦은 잔향 성분을 계산하고 빼는 방법으로 음질을 향상하였다[18].

본 논문에서는 다중 마이크로폰 배열 구조를 사용하여 수신된 신호의 선형예측 알고리즘으로 얻은 잔여신호를 이용하여 잔향음을 제거 하고자 하였다.

제 3 장 제안한 잔향음 제거 방법

제 3-1 절 선형예측 분석 모델

음성신호 처리에서 선형예측(LPC : Linear Prediction Coding) 이론은 오랫동안 연구되었다. 음성의 특징을 추출하는 방법 중 하나인 선형예측분석은 음성의 기본적인 파라미터를 음성발생의 선형적인 모델에 기초하여 추출해내는 방법으로 1970년 Itakura 등에 의해서 발표되었고 오늘날 가장 널리 사용되고 있는 분석법으로서 기본 아이디어는 그림 3-1에서 보는 것과 같이 '주어진 시점 n에서의 음성 신호 샘플 값은 지나간 p개의 음성출력 샘플 값과의 선형조합으로 근사화할 수 있다는 가정에서 시작된 분석법이다[19]. 선형예측을 통한 특징 추출방법은 근본적으로 음성발생 모델과 밀접한 연관이 있으며, 음성에 관한 특징을 비교적 정확히 묘사할 수 있을 뿐 아니라, 그 정확도와 계산 속도 면에서도 비교적 좋은 성능을 보이므로 음성처리를 위하여 많이 사용되어 왔다. 이외에도 선형예측분석의 장점은 이 모델은 크게 음원에 해당되는 부분과 합성필터에 해당되는 부분으로 분리해서 모델화 할 수 있

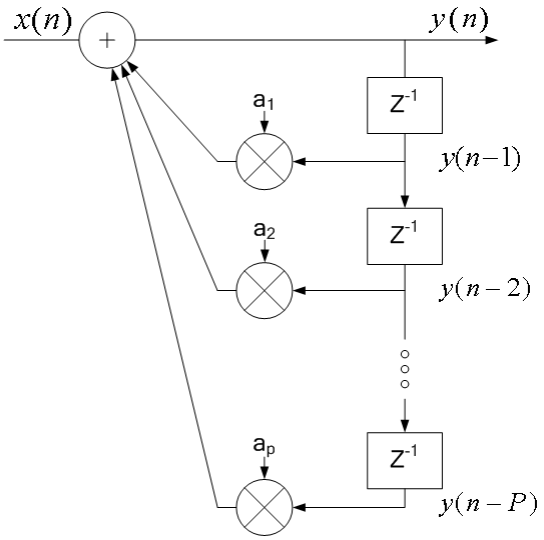

그림 3-1. 이산시간 시퀀스의 발생을 위한 전극 모델

Fig 3-1. All-pole model for the generation of a discrete-time sequence.

으므로 음성분석 및 합성을 훨씬 용이하게 해주는 데에 있다. 또한 음성을 코딩하는 작업은 시간영역에서 이루어지지만 음성의 분석 및 합성은 주파수영역에서 행하여지는 것인데, 이 선형예측분석은 이것을 만족하고 있다는 것이다.

즉, 선형예측분석은 시간영역에서 코딩해 주면서 formant 주파수, 대역폭, 진폭 등의 주파수 영역에서의 파라미터 값들을 추출해 낼 수 있다는 것이다. 성도의 특징을 나타내는 전달함수는 일반적으로 전극 필터를 가정하는데, 실제적으로 성도의 특징을 보다 정확하게 묘사하기 위해서는 pole뿐만 아니라 zero도 필요하다. 그러나 zero를 전달함수

에 포함시키면 필터의 계수를 구하는 일이 대단히 복잡해질 뿐만 아니라, 또 pole의 수가 충분히 많아지면 pole로 zero를 어느 정도 나타낼 수 있으므로 보통의 경우 그냥 전극 필터를 가정하게 된다.

위와 같은 장점에 비해, 분석차수가 낮으면 peak 즉, formant가 제대로 찾아지지 않으며, 차수를 높이면 계산 속도가 느려지기 때문에 분석차수를 효과적으로 설정해야하는 문제가 있고, 코딩을 위해서 양자화하는 과정에서 에러발생시 불안정한 필터계수가 얻어질 수 있다. 주어진 시점 n에서의 음성 신호 샘플 값은 지나간 p개의 음성출력 샘플 값과의 선형결합을 식으로 나타내면 다음 식 3-1로

$$y(n) \approx a_1 y(n-1) + a_2 y(n-2) + \cdots + a_p y(n-p) \qquad (3\text{-}1)$$

여기서 $y(n)$은 n 시점에서의 음성 샘플을 나타내고, $y(n-k)$는 n 시점에서 k 만큼 이전의 음성 샘플을 나타낸다. 그리고 a_k는 해당 프레임에서 정해지는 계수이다. 식 3-1을 보면 비선형 성분이 전혀 포함되지 않은 선형 조합으로 나타나 있으며, 과거 값을 가지고 미래의 값을 예측하는 모델이므로, 이를 선형예측 모델이라고 한다.

시간축의 구간과 주파수축의 대역폭이 반비례한다는 것을 상기한다

면 이 모델이 주파수축의 전체 모양을 모델링함을 예측할 수 있다. 선형예측 알고리즘에서 zero를 사용하지 않고 pole만을 사용해서 성도를 모델링 하는 이유는 zero의 추가에 따른 계산량의 증가에 비해서 모델링의 정교도가 좋아지지 않기 때문이며 pole만으로도 무리 없이 모델링이 되기 때문이다.

전극 합성기는 신형예측 분석으로부터 얻는다. 그림 3-2는 선형예측 분석으로부터 유래한 구성으로 p차 차분방정식에 상응하는 디지털 필터이다[19]. 식 3-2는 음성을 합성하기 위한 현재의 출력 $y(n)$의 과거 출력 함수식이다.

$$y[n] = \sum_{k=1}^{p} a_k y[n-k] \tag{3-2}$$

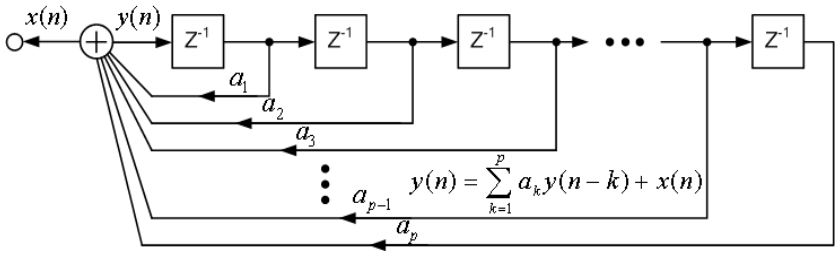

그림 3-2. 선형예측 분석 기반 All-pole 합성기

Fig 3-2. All-pole synthesizer based on LPC analysis

제 3-2 절 선형예측 분석 알고리즘

수신된 신호를 $s(n)$이라고 나타내면 식 3-3으로 나타낼 수 있다.

$$\bar{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{3-3}$$

이때의 원 신호와 추정 신호 사이의 차이인 예측 오차는 식 3-4와 같다.

$$e(n) = s(n) - \bar{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{3-4}$$

그리고 오차 전달함수(transfer function)는 식 3-5와 같이 나타난다.

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{3-5}$$

입력으로 받은 한 프레임의 음성신호에 대해서 그 신호를 가장 잘 모

델링하는 a_k를 구하는 것이 목적이기 때문에 식 3-6와 같은 표기법을 정의한다. 이는 곧 n 번째 프레임의 m 번째 샘플을 의미한다.

$$y_n(m) = y(n+m),\ v_n(m) = v(n+m) \tag{3-6}$$

수신된 신호에 관한 식 3-3에 근거한 합당한 a_k를 구하는 방법에 대해서 살펴보자. 한 프레임 내에서 원 신호와 추정 신호의 차에 해당하는 평균제곱 오차를 최소화 하는 a_k를 찾아내는 것을 기본적인 접근방법으로 사용한다.

n 번째 프레임 내에서의 평균제곱 오차를 식 3-6의 표기법을 근거로 정의하면 식 3-7과 같이 나타난다.

$$V_n = \sum_m v_n^2(m) = \sum_m [y_n(m) - \sum_{k=1}^{p} a_k y_n(m-k)]^2 \tag{3-7}$$

평균제곱 오차를 최소화 하는 기준을 만족하는 a_k를 구하는 것은 곧 식 3-8을 만족하는 a_k를 구하는 것과 같다.

$$\frac{\partial V_n}{\partial a_k}=0,\ \ k=1,2,\cdots,p \tag{3-8}$$

이 식을 풀어보면 식 3-9의 해를 구하는 것과 같다.

$$\sum_m y_n(m-i)y_n(m)=\sum_{k=1}^{p} a_k\sum_m y_n(m-i)y_n(m-k),\ \ i=1,2,\cdots,p \tag{3-9}$$

식 3-9를 간단히 하면 식 3-10이 되는데

$$\Phi_n(i,k)=\sum_m y_n(m-i)y_n(m-k) \tag{3-10}$$

이를 이용하면 오차를 최소화하는 a_k를 구하는 식은 다음 식 3-11과 같이 간단하게 표시된다.

$$\Phi_n(i,0)=\sum_{k=1}^{p} a_k\Phi_n(i,k),\ \ i=1,2,\cdots,p \tag{3-11}$$

결국 해당 프레임에서 우리가 원하는 a_k를 구하는 방법은 식 3-11의 p개의 미지수를 가진 p개의 방정식을 푸는 것과 같다. 한편 평균제곱 오차는 식 3-12와 같이 나타난다.

$$E_n = \sum_m y_n^2(m) - \sum_{k=1}^{p} a_k \sum_m y_n(m) s_n(m-k) = \Phi_n(0,0) - \sum_{k=1}^{p} a_k \Phi_n(0,k) \quad (3\text{-}12)$$

제 3-3 절 선형예측 계수 추출

선형예측 계수를 추출하는 방법으로 자기상관 방법과 공분산 두 방법이 있는데, 두 가지 측면에서 서로 장·단점을 가진다. 우선 계산량 면에서는 자기상관 방법이 p^2(여기서 p는 선형예측의 길이)에 비례하고, 공분산 방법이 p^3에 비례함으로써 자기상관 방법이 좋음을 알 수 있다. 공분산 방법이 음성신호에 윈도우를 사용하지 않음으로 해서 계산량이 조금 적어지기는 하지만 자기상관 방법에 비하면 훨씬 많은 계산량을 가진다. 그리고 다른 하나의 측면인 선형예측 계수를 양자화 할 때 생기는 안정성 문제에서는 자기상관 방법에 의해서 구해진 계수가 공분산 방법에 의해 구해진 계수보다 양자화 시에 발산할 가능성이 높게 나타

난다.

두 방법의 성능 면에서는 공분산 방법이 우수하지만 큰 차이를 보이지 않는 것으로 알려져 있다. 따라서 자기상관 방법이 성능 면에서 공분산 방법에 비해 그리 뒤떨어 지지 않는 반면 계산량이라는 측면에서 많은 이점을 가지고 있으므로 본 논문에서는 선형예측 계수를 추출하는 방법으로 자기상관 방법을 사용한다.

제 3-4 절 선형예측 잔여신호

잠깐 잔여신호에 대해서 설명하자면, 잔여신호란 선형예측 알고리즘으로 신호를 선형예측 하였을 때 출력으로 나오는 신호의 차이로 남는 신호이다. 잡음 혹은 잔향음과 다르게 잔여신호에는 여전히 음성을 판단하기 위해 필요한 정보들이 남아있어 이러한 정보를 살리기 위한 적절한 잔여신호 모델링을 동반하여야 한다. 잔여신호는 바꾸어 말하면 음성 신호를 출력하는 선형시스템으로의 입력신호이다. 실제로 잔여신호를 식의 디지털 필터로 입력하면 음성신호가 출력된다.

아래에 보이는 그림 3-3은 세 개의 마이크로폰으로 수신된 음성 데이터를 나타낸 것이다. 마이크로폰에 수신된 음성 데이터는 송신기에

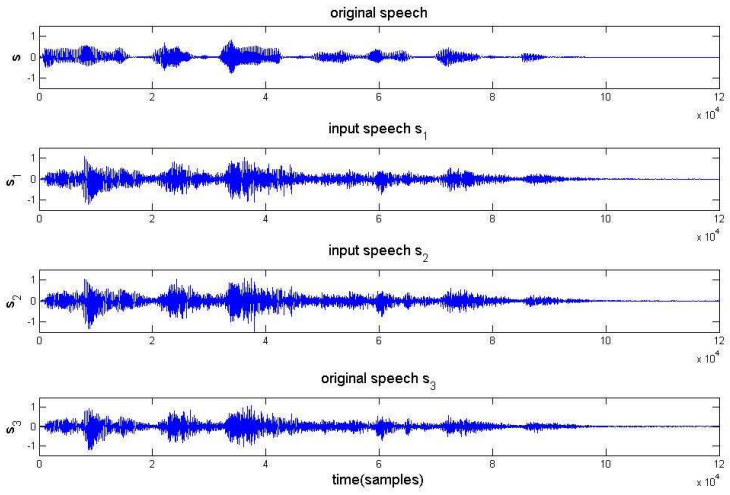

그림 3-3. 음원 신호와 각각의 마이크로폰에 입력된 신호
Fig 3-3. Source signal and input signals of each microphone

서 전달된 음성 데이터가 실내 잔향음과 부가잡음에 의해 저하된 데이터로 그림에서 보는 바와 같이 잡음과 잔향음의 효과로 저하된 모습을 확인할 수 있다. 부가잡음에 의해 저하된 음선신호는 식 3-13과 같이 정의된다[20].

$$y = x + w \tag{3-13}$$

여기서 $y = [y(n-N+1), \cdots, y(n-2), y(n-1), y(n)]^T$는 잡음이 섞인

음성의, $x=[x(n-N+1),\cdots, x(n-2),x(n-1),x(n)]^T$는 깨끗한 음성의, $w=[w(n-N+1),\cdots, w(n-2),w(n-1),w(n)]^T$는 잡음 샘플의 N차 벡터를 각각 나타낸다. 입력신호 샘플의 p차 선형예측분석 실행에서, 선형예측 계수의 집합을 얻을 수 있다.

실내 잔향음에 의해 저하된 음성 신호는 식 3-14와 같이 정의된다.

$$y(n)=-\sum_{i=1}^{p} a_i y(n-i)+v_y(n) \tag{3-14}$$

여기서 a_i는 선형예측 계수이고, $v_y(n)$은 선형예측 오차이다. 선형예측 계수는 N×N 선형예측 행렬 A로 구성된다.

$$A=\begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_1 & 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & & & \vdots \\ a_p & a_{p-1} & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & & & \ddots & \vdots \\ 0 & \cdots & a_p & \cdots & a_1 & 1 \end{bmatrix} \tag{3-15}$$

저하된 음성신호에 대한 식 3-14는 행렬 A를 이용하여 다시 작성하면 식 3-16이 되는데

$$e_y = Ay \tag{3-16}$$

여기서 e_y는 선형예측 잔여신호의 벡터이다. 식 3-16은 선형예측 잔여신호가 음성샘플의 역필터링에 의해 얻어진다는 것을 의미한다[20].

제 3-5 절 제안한 잔향음 제거 방법

음성의 직접성분은 화자와 마이크로폰 사이의 거리가 멀어질수록 줄어들기 때문에, 신호 대 잔향음 성분비(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 직접 신호 대 잔향음 성분비는 음량의 커짐으로 인해 보상되지 않는다. 그러므로 음질 향상은 음성의 직접 신호 대 잔향음 성분비를 높이는 한편, 부가접인 잡음에 의한 신호 대 잡음비를 높여야한다. 직접 신호 대 잔향음 성분비를 높이는 동시에 신호 대 잡음비를 높이는 한 가

지 방법은 선형예측 잔여신호를 수정하는 것이다. 이는 다른 영역에 관하여 중요한 순간 영역 주위의 코히런트 부분을 개선하는 선형예측 잔여신호를 위한 가중치 함수를 만드는 것이다. 선형예측 잔여신호는 위상에 좌우되는 양과 음의 샘플 모두를 가지고 있기 때문에, 각 순간에서의 선형예측 잔여신호의 강도는 선형예측 잔여 신호의 힐버트 포락선 계산에 의해 얻을 수 있다[21].

그림 3-4는 본 논문에서 제안하는 음질 향상을 위한 알고리즘에 대한블록다이어그램을 나타낸 것이다. 세 개의 마이크로폰으로부터 입력된 수신신호는 선형예측 분석을 거쳐 잔여신호를 구한다. 이렇게 구한

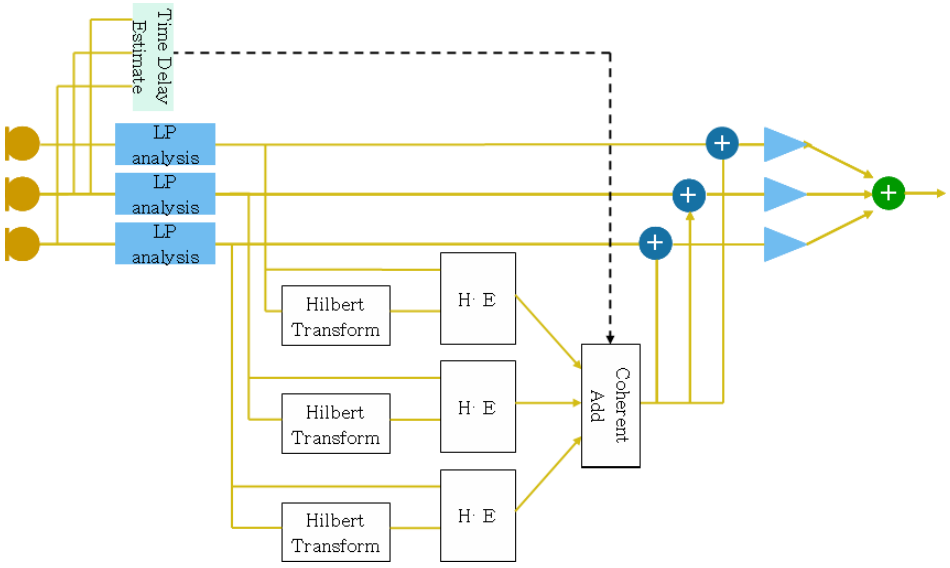

그림 3-4. 제안된 알고리즘의 블록 다이어그램

Fig 3-4. Block Diagram of proposed Algorithm

선형예측 잔여신호로부터 코히런트하게 더해진 힐버트 포락선을 얻고 수신신호의 선형예측 잔여신호와 조합하여 개선된 잔여신호를 구한다.

그림 3-5는 음원 신호의 선형예측 잔여신호와 각각의 마이크로폰에 입력된 신호의 선형예측 잔여신호를 구한 것이다. 선형예측 잔여 신호 $e(n)$의 힐버트 포락선 $\hat{e}(n)$은 식 3-17로부터 얻을 수 있고, 여기서 식 3-18과 같은 $e_H(n)$은 선형예측 잔여 신호 $e(n)$의 힐버트 변환이다.

$$\hat{e}(n) = \sqrt{e^2(n) + e_H^2(n)} \tag{3-17}$$

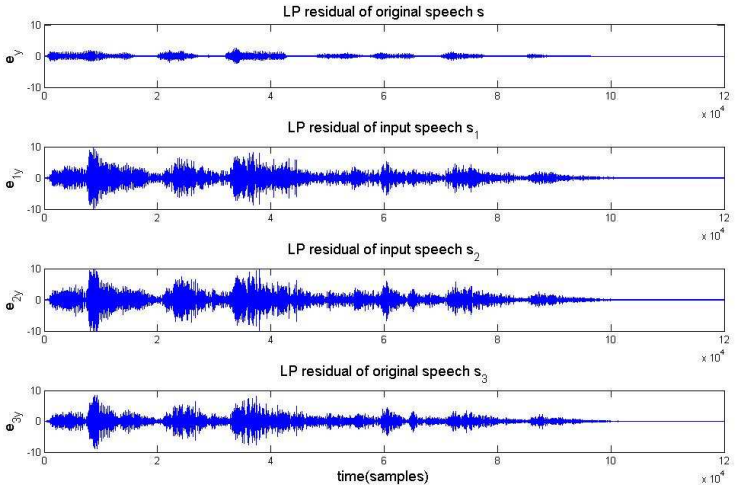

그림 3-5. 음원 신호와 각각의 마이크로폰에 입력된 신호의 선형예측 잔여
Fig 3-5. LP residual of source signal and input signal of each microphone

$$e_H(n) = \begin{cases} IDFT[jE(w)], & -\pi < w \le 0 \\ IDFT[-jE(w)], & 0 < w \le \pi \end{cases} \quad (3\text{-}18)$$

선형예측 잔여 신호 $e(n)$의 힐버트 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[22]. 이렇게 구해진 힐버트 포락선은 그림 3-6에 나타나있다.

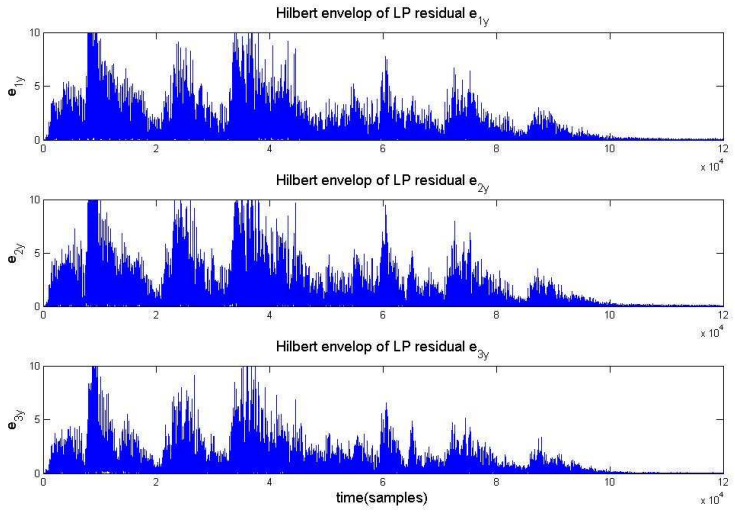

그림 3-6. 각각의 선형예측 잔여의 힐버트 포락선

Fig 3-6. Hilbert envelop of each LP residuals

잡음과 잔향음의 효과 때문에 선형예측 잔여신호의 힐버트 포락선에서 여러 큰 진폭 스파이크가 있다. 이런 스파이크의 효과를 줄이기 위해, 세 개의 마이크로폰으로부터 선형예측 잔여신호의 힐버트 포락선을 측정할 수 있고, 그것들을 코히런트하게 더한다.

코히런트 덧셈을 위해 두 마이크로폰 사이의 음성 시간지연이 계산되어야 한다. τ_{12}는 마이크로폰 1과 2 사이의, τ_{13}은 마이크로폰 1과 3 사이의 시간 지연이다. 마이크로폰 2와 3의 지연 보정된 힐버트 포락선은 첫 번째에 더해진다. 이 코히런트 덧셈을 위해 힐버트 포락선의 제곱이 고려된다. 그 결과 힐버트 포락선은 이 덧셈의 제곱 값이 된다. 이를 수식으로 표현하면 식 3-19와 같고

$$\widehat{e_c}(n) = \sqrt{\widehat{e_1}^2(n) + \widehat{e_2}^2(n+\tau_{12}) + \widehat{e_3}^2(n+\tau_{13})} \qquad (3\text{-}19)$$

여기서 $\widehat{e_c}(n)$은 코히런트하게 더해진 힐버트 포락선이고 $\widehat{e_1}(n)$ $\widehat{e_2}(n)$ $\widehat{e_3}(n)$은 각각 마이크로폰 1, 2, 3의 선형예측 잔여신호의 힐버트 포락선을 나타낸다. 반면에 인코히런트하게 더해진 힐버트 포락선은 식 3-19을 계산할 때 마이크로폰 사이의 시간 지연 τ_{12}과 τ_{13}이 더하지 않는다. 그림 3-7은 코히런트하게 더해진 힐버트 포락선과 인코히런트하

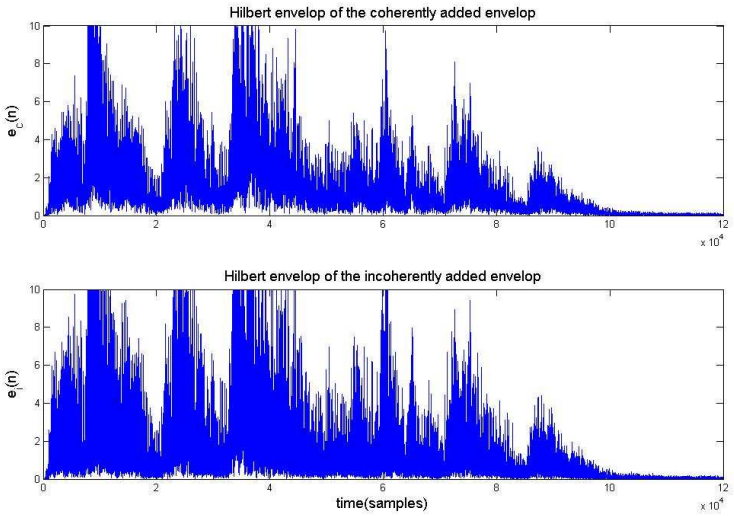

그림 3-7. 코히런트하게 더해진 힐버트 포락선과 인코히런트하게 더해진 힐버트 포락선

Fig 3-7. Hilbert envelop of the coherently added envelop and Hilbert envelop of the incoherently added envelop

게 더해진 힐버트 포락선을 비교해 놓은 것이다.

코히런트하게 더해진 힐버트 포락선의 특징은 잔여신호 가중치를 만들 수 있는 것이다. 선형예측 잔여신호 $e_1(n)$의 가중치는 식 3-20에 사용된다.

$$e_{iM} = \frac{\sum_m e_1(n)\,\hat{e}_c(n)}{\sum_m \hat{e}_c(n)} \qquad (3\text{-}20)$$

e_{1M} e_{2M} e_{3M}은 각각 마이크로폰 1, 2, 3의 수정된 선형예측 잔여신호이다. 여기서 m은 한 프레임의 길이를 나타낸다. 이렇게 하여 얻은 수정된 선형예측 잔여신호를 그림 3-8에 나타내었다.

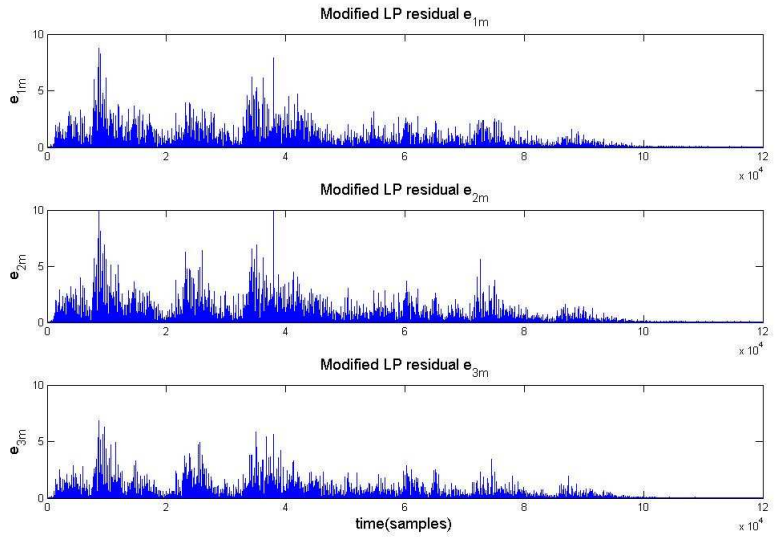

그림 3-8. 기존의 방법으로 얻은 선형예측 잔여
Fig 3-8. LP residuals by existed method

식 3-20에서 구한 한 개의 수정된 선형 예측 잔여신호 e_{1M}을 합성하여 음성을 만들어 개선된 음성을 얻을 수 있지만, 본 논문에서는 이미 계산된 잔여신호를 이용하여 계산량의 증가 없이 각 마이크로폰의 수정된 선형예측 잔여신호의 조합을 이용하여 새로운 선형예측 잔여값을 구해냄으로서 음질의 향상과 더불어 신뢰성을 높이고자 하였다. 식 3-20

를 통하여 얻은 각각의 마이크로폰의 수정된 선형예측 잔여신호에 각각 다른 가중치 w_1과 w_2, w_3을 주어 이들의 조합으로 새로운 선형예측 잔여신호를 구하였다.

$$e_M(n) = \sum e_{1M}(n) \times w_1 + e_{2M}(n) \times w_2 + e_{3M}(n) \times w_3 \quad (3\text{-}21)$$

식 3-21의 가중치 값은 실험을 통하여 합성된 음성의 질에 있어서 가장 좋은 결과를 나타는 새로운 선형예측 잔여신호를 얻을 때 사용된 가중치 값으로 실험에 의하여 얻어진 값이다.

제 4 장 실험 결과

제 4-1 절 실험 환경

잔향 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. 각각의 마이크로폰 사이의 거리는 20cm이고, 송신기와 마이크로폰 사이의 거리는 2m이다. 실험에 사용된 음성 데이터

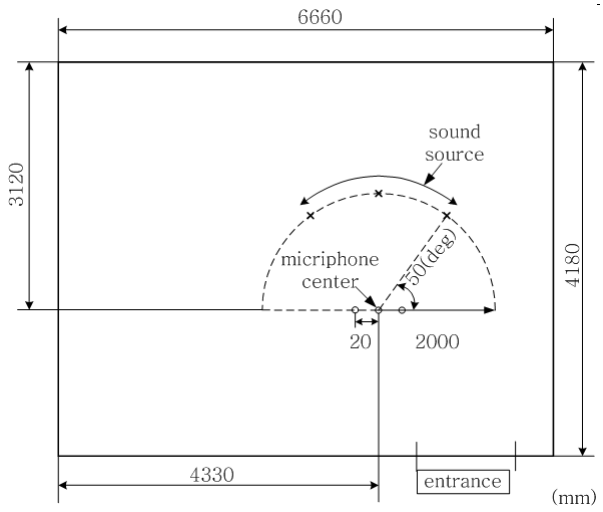

그림 4-1. 실험 환경

Fig 4-1. Experimental environment

의 RT(Reverberation Time)는 1.3초이다. 음성 데이터는 48㎑로 샘플링 되었으며, 16비트 양자화로 저장되었다. 실험 환경은 그림 4-1와 같은 에코룸에서 음성 데이터를 수집하였다.

10차 선형예측 분석은 세 개의 마이크로폰 출력에서 선형예측 잔여신호의 힐버트 포락선를 얻는데 수행되었으며, 이렇게 구한 선형예측 잔여신호를 이용하여 수정된 선형예측 잔여신호를 얻을 수 있었고, 향상된 음성은 수정된 선형예측 잔여신호로 시변 10차 전극(all-pole) 필터를 통해서 얻는다.

수정된 하나의 선형예측 잔여신호 $e_{iM}(n)$만을 이용하여 음질을 개선한 방법과 본 논문에서 제안한 수정된 선형예측 잔여신호 $e_{1M}(n)$, $e_{2M}(n)$, $e_{3M}(n)$의 조합을 이용하여 얻은 새로운 선형예측 잔여신호를 이용하여 음질을 개선한 방법을 비교해 보았다.

제 4-2 절 실험 결과

4-2-1. 입력신호와 잔여신호의 비교

그림 4-2에서는 원음 신호와 수신된 신호와 함께 각각의 잔여신호를 비

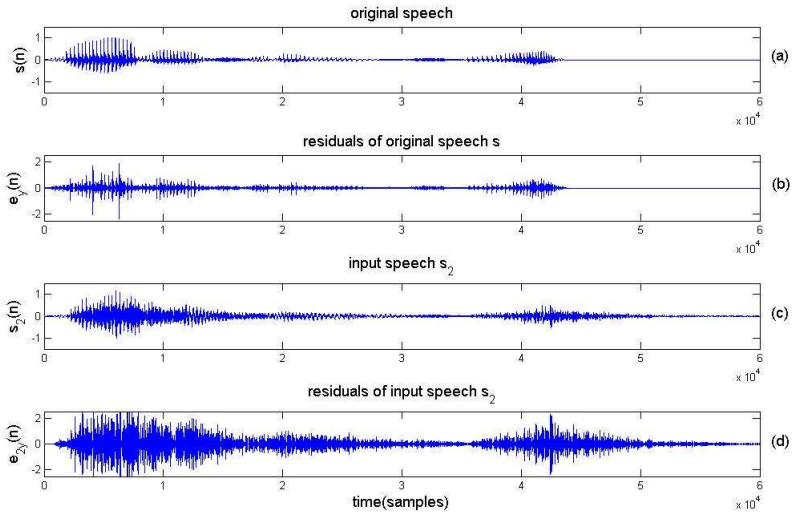

그림 4-2. 음성과 잔여신호: (a)와 (b)는 원음과 원음의 잔여신호, (c)와 (d)는 수신된 신호와 수신된 신호의 잔여신호

Fig 4-2. speech and residual signal: (a) and (b) is clean speech and residual, (c) and (d) is input signal and residual

교하였다. 원음 신호와 수신된 신호를 비교했을 때 수신된 신호의 경우 음절과 음절 사이의 끊김 없이 잔향효과가 음절과 음절 사이에 포함되어 수신되었음을 알 수 있다. 또한 원음과 비교해 잔향효과가 있는 수신된 신호의 잔여신호에는 음성을 판단하기 위한 정보인 잔여신호에 잔향효과에 의하여 잔여정보가 많이 남아있음을 그림 4-2 (d)의 피크치를 통하여 알 수가 있다.

4-2-2. 각 단계의 잔여신호의 비교

그림 4-3에서는 각각의 단계의 잔여신호를 비교하였다. (a)는 원음을 선형예측 하였을 때 얻은 잔여신호이고, (b)수신된 신호의 잔여신호로 원음의 잔여신호와 상당한 차이가 있음을 알 수 있다. (c)에 나타낸 것이 코히런트하게 더해진 잔여신호와 한 개의 수신된 신호의 잔여신호 (b)를 이용하여 얻은 수정된 잔여신호로 (b)에 비하여 비교적 원음의 잔여신호에 가까움을 볼 수 있다. 마지막에 있는 (d)는 이 논문에서 제안한 방법인 코

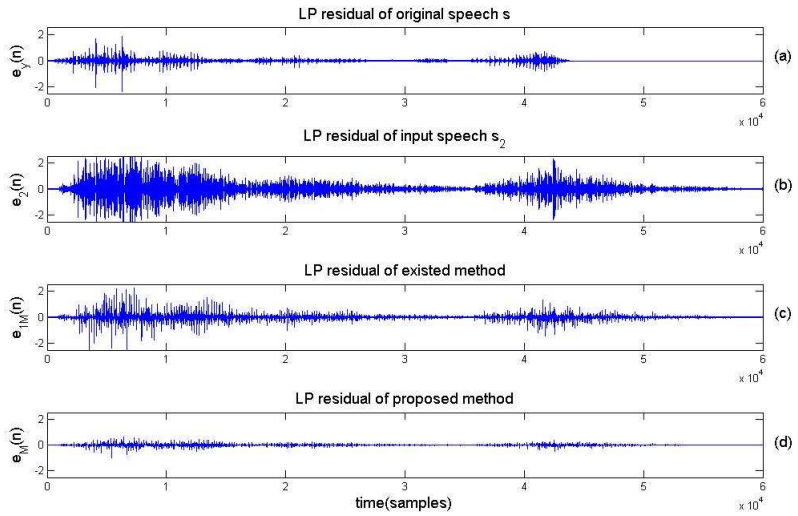

그림 4-3 각 단계의 잔여신호

Fig 4-3. residual of each procedure

히런트하게 더해진 잔여신호와 코히런트하게 더해진 잔여신호를 얻기 위하여 이미 구해놓은 세 개의 수신된 신호의 잔여신호를 이용하여 얻은 세 개의 수정된 잔여신호의 가중치 합으로 얻은 잔여 신호로 (c)에서 얻은 잔여신호보다 비음성 구간에서의 zero 부분이나 음절과 음절 사이의 잔향음 효과가 줄어있음을 알 수 있다.

4-2-3. 각 시간지연의 잔여신호의 비교

코히런트하게 더해진 힐버트 포락선을 구하려면 마이크로폰에 수신된 신호들 사이의 시간지연 정보가 필요하다. 하지만 각 신호들 사이의 시간지연 정보를 정확하게 계산 하는 것은 매우 어려운 일이다. 그림 4-4에서는 시간지연 정보의 오차가 미치는 잔여신호의 차이를 나타내었는데, 정확한 시간지연 정보를 가지고 코히런트하게 더해진 잔여신호와 정확하지 않은 시간지연 정보를 가지고 코히런트하게 더해진 잔여신호를 비교 하였다. 그림 4-4에서 볼 수 있듯이 정확한 시간지연 정보를 가진 잔여신호와 정확하지 않은 시간지연 정보를 가진 잔여신호 사이에 차이가 거의 없음을 알 수 있다. 그 이유는 마이크로폰 어레이의 간격이 매우 좁아 수신된 신호들 사이의 시간지연이 매우 작기 때문이다. 만약 본 논문에서 제안한 알고리즘을 마이크로폰의 간격이 좀 더 큰 시스템에 적용하여 사용하였을

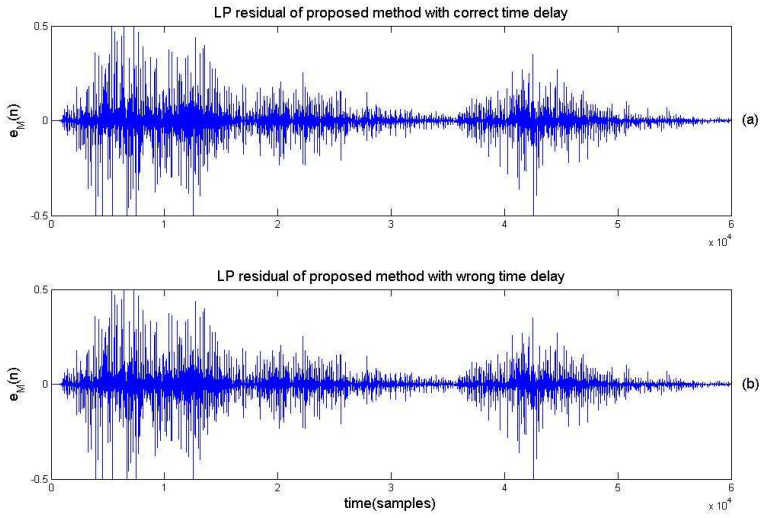

그림 4-4. 각 시간지연의 잔여신호

Fig 4-4. residual of each time delay

경우 각각의 마이크로폰에 수신된 신호의 시간지연 차에 의한 영향이 잔여신호의 차이를 만들 것이라 판단되며, 이를 실험을 통하여 확인해 보았을 때 좀 더 완벽한 논문이 될 것이다.

4-2-4. 방식이 다른 잔여신호의 비교

그림 4-5에서는 코히런트하게 더해진 힐버트 포락선을 구하는 식 3-19에서 정확한 시간지연 정보를 더한 코히런트하게 더한 힐버트 포락선 (a)

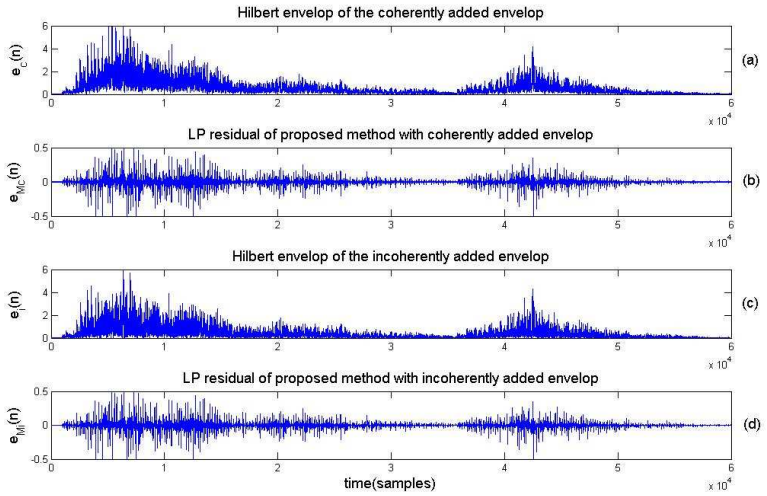

그림 4-5. 코히런트하게 더한 잔여와 인코히런트하게 더한 잔여
Fig 4-5. residual of coherently added and incoherently added

와 이를 이용한 잔여신호 (b)와 시간지연 정보를 더하지 않은 인코히런트하게 더한 힐버트 포락선 (c)과 이를 이용한 잔여신호 (d)를 각각 나타내었다. 인코히런트하게 더한 힐버트 포락선 (c)가 코히런트하게 더한 힐버트 포락선 (a)에 비해 피크치가 조금씩 크다는 것을 알 수 있다. 하지만 (a)와 (c)의 두 포락선은 사용하여 얻은 잔여신호의 경우 코히런트하게 더한 힐버트 포락선을 사용하여 얻은 잔여신호와 인코히런트하게 더한 힐버트 포락선을 사용하여 얻은 잔여신호 사이에 뚜렷한 차이를 보이지 않았는데 이는 코히런트하게 더한 힐버트 포락선을 구할 때 사용되는 시간지연정보가 매우 작기 때문에 두 잔여신호의 차이가 뚜렷하지 않은 것으로

보인다. 하지만 인코히런트하게 더한 힐버트 포락선을 사용하여 얻은 잔여신호를 사용하는 경우 단일 마이크로폰을 사용하는 것과 차이가 없고, envelop을 구할 때 지연보상을 해주기 때문에 정확한 샘플에서 높은 값을 가질 수 있지만, 인코히런트하게 더한 힐버트 포락선을 사용할 경우 정확하지 않은 샘플에서 높은 값을 가져 오차를 낼 수 있다.

4-2-5. 가중치에 따른 잔여신호의 비교

그림 4-6은 3개의 수정된 선형예측 잔여신호를 이용한 제안된 방법을 사용하였을 경우 식 3-21의 가중치의 변화에 따른 잔여신호를 차이를 나

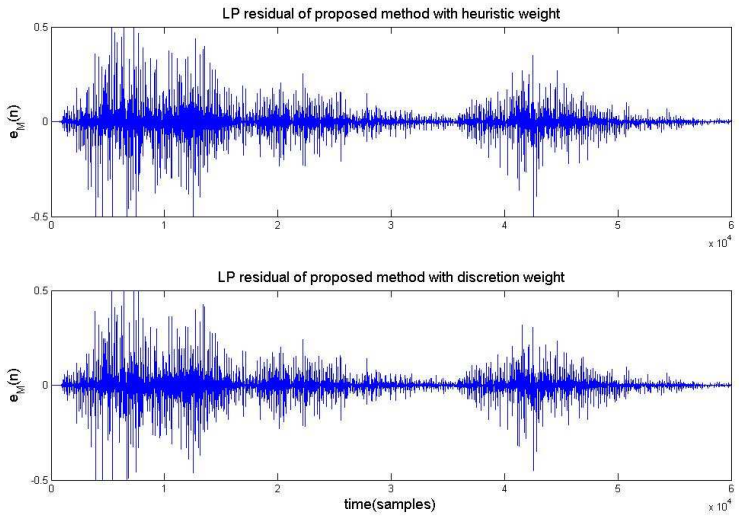

그림 4-6. 가중치에 따른 잔여(음원의 위치 50°)
Fig 4-6. residual of each weight(source position 50°)

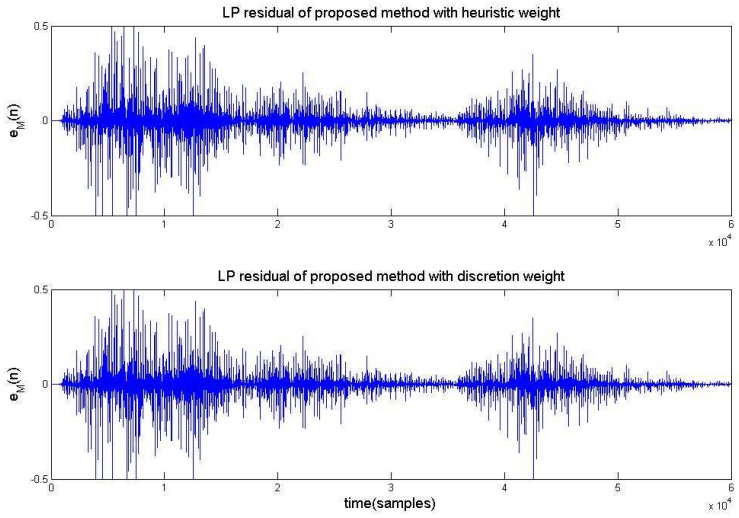

그림 4-7. 가중치에 따른 잔여(음원의 위치 130°)
Fig 4-7. residual of each weight(source position 130°)

타내고 있다. 잔향성분이 임의의 가중치를 준 잔여신호와 실험을 통해 얻은 가중치를 준 잔여신호와 비교하여 임의의 가중치를 준 잔여신호가 높은 피크치가 줄어든 것처럼 보인다. 하지만 음원 신호의 위치가 반대일 경우에는 그림 4-7에서 보는 것과 같이 실험을 통해 얻은 가중치를 준 잔여신호는 그림 4-6과는 동일한 결과를 보이나 임의의 가중치를 준 잔여신호의 경우 결과가 다르게 나옴을 통하여 임의의 가중치를 주었을 때보다 실험을 통해 얻은 가중치를 주었을 경우 다양한 환경에서 고른 잔여신호를 가질 수 있다는 것을 알 수 있다.

제 4-3 절 결과 비교

한 개의 수정된 잔여신호를 이용한 방법과 3 개의 수정된 잔여신호를 모두 이용한 제안한 방법으로 얻은 잔여신호를 합성하여 음성신호를 얻을 수 있었다. 이렇게 합성한 신호를 깨끗한 음성 신호와 비교하여 그림 4-8에 나타내었다. 파형의 유사도만을 육안으로 확인 했을 때 음절과 음절 사이의 잔향음 성분이 줄어 음절과 음절의 구분을 쉽게 할 수 있다는 면에서 볼 때, 본 논문에서 제안한 방법이 기존의 방법에 비해 원 신호에 더

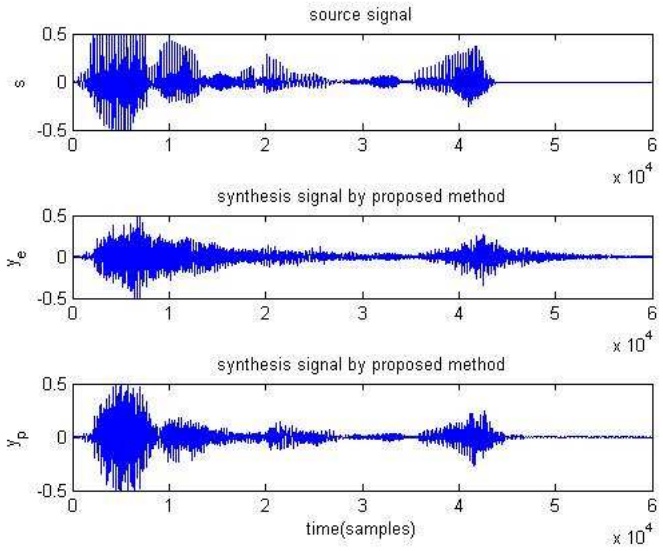

그림 4-8. 합성신호 : (a) 원 음성 신호, (b) 기존의 선형예측 잔여를 합성한 음성 신호, (c) 개선된 선형예측 잔여를 합성한 음성

Fig. 4-8 synthesis signal : (a) clean speech signal, (b) synthesis by existed LP residual, (c) synthesis by improved LP residual

가깝도록 잔향음을 제거함을 확인할 수 있다. 하지만 단순한 신호의 그래프를 보는 것을 떠나 좀 더 정확한 비교를 위하여 음성의 품질을 평가하는 객관적인 방법과 주관적인 방법 두 가지를 통하여 두 음성의 차이를 수치적으로 나타내었다.

4-3-1. Spectral Distance(SD) 평가법

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 상대적으로 높은 상관성을 나타내는 것으로 알려져 있는 Spectral Distance(SD)를 구하는 방법이다. 시간 영역 척도와는 달리 스펙트럼 영역 촉도로 Fourier 변환에 근거하여 주파수 영역으로 변환하는 객관적 품질 척도로서 SD는 식 4-1과 같이 정의된다. SD에서 전체의 특성을 포함하는 음석 스펙트럼은 FFT에 의해 계산되고 입·출력에서의 스펙트럼 차이에 대한 특성을 보여주는 척도이며 식 4-1과 같이 계산된다.

$$SD=[\frac{1}{w}\int_0^w S_x(w)-S_y(w)^2 dw]^{1/2} \qquad (4-1)$$

여기에서, $S_x(w)$와 $S_y(w)$는 입력과 출력 음성 스펙트럼이고, w는 신호 주파수 대역을 의미한다. 기존의 방법과 제안한 방법에 따른 결과를 객관적 척도로써 비교하기 위하여 각각의 방법으로 얻은 합성 신호와 원 신호 사이의 Spectral Distance (SD)를 구하였다.

그림 4-9의 왼쪽은 기존의 선형예측 잔여신호를 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.

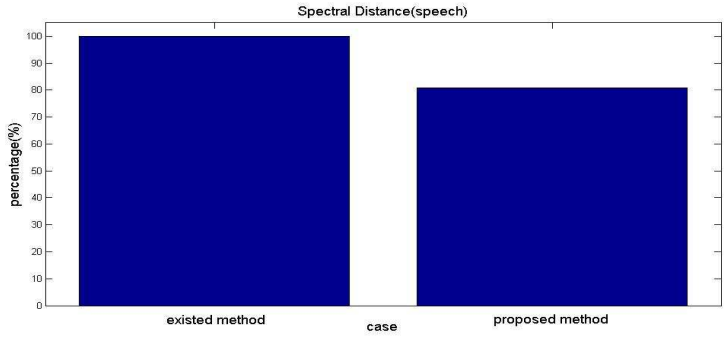

그림 4-9. 합성된 신호와 원 신호의 SD

Fig. 4-9 compare synthesis signals with source signal

4-3-2. Mean Opinion Score(MOS) 평가법

MOS는 청취자의 반응 평가시 5단계 scale에 따라 5~1점의 점수를 주어 청취 시험에 참가한 다수 청취자에 의한 반응 의견에 대해서 가중 평균치를 구하는 척도이다. MOS는 분류상 청취자들의 의견을 종합적으로 평가하고자하는 opinion test에 속하며, 절대 평가에 속하며 상대 평가로는 Differential MOS가 있으며 평가 대상에 따라 명료성이나 자연성 평가법도 사용된다. MOS 점수가 4.0 이상이면 toll quality 품질이라고 하는데 이는 200~3200㎐의 아날로그 전화 음성과 거의 구분할 수 없을 정도의 음질을 의미한다. MOS 점수가 3.5~4.0이면 communication quality라 하는데, 이는 자연스러운 통화를 하기에는 충분한 정도의 음질을 의미한다. MOS 점수가 2.5~3.5이면 synthetic quality라 하는데 이는 통화는 가능하지만, 자연성이 부족하고 화자를 식별할 수 없을 정도의 음질을 의미한다. 주관적 음질 평가를 위해서 화자가 발음하여 얻은 약 2초 길이의 음성 신호를 사용하였다. MOS 시험에 있어서, 청취자들은 원래의 음성 표본을 듣지 못하고, 왜곡된 음성 표본의 전체적인 음질에 등급을 매긴다. 20~30대의 남녀가 MOS 테스트를 위해 피험자로 참여하였다. 전체 테스트 인원은 남자 10명, 여자 10명으로 총 20명이며, 전문적인 훈련을 받지는 않은 사람들을 대상으로 하였지만 가능한 많은 인원에게 테스트하여

음질 평가의 객관성을 높이고자 하였다. 정확한 테스트를 위하여, 피험자에게 각 실험 샘플별로 두 가지의 자료를 제시하였다. 먼저 기존의 방법으로 얻은 음성을 들려주었고, 곧바로 본 논문에서 제안한 방법으로 얻은 음성을 들려주었다. 이때 피험자가 원할 경우 같은 음성을 반복적으로 듣도록 하였다. MOS 테스트는 평가 환경의 변화에 영향을 받지 않도록, 피험자들은 HIFI 오디오용 헤드폰을 사용하여 동시에 시행하였다.

표 4-1. MOS 테스트 결과
table. 4-1 MOS test results

구 분	기존의 방법으로 얻은 음성	제안된 방법으로 얻은 음성
20대 남 (5명)	2.2	4.4
20대 여 (5명)	2.1	4.2
30대 남 (5명)	1.8	4
30대 여 (5명)	2.4	4
평 균	2.125	4.15

표 4.1에 각 샘플에 대한 MOS 테스트의 평균값을 보였다. MOS 테스트의 평균값은 기존의 방법으로 얻은 음성의 경우 2.125로 나타났으며 본 논문에서 제안한 방법으로 얻은 음성의 경우 4.15로 나타남으로써 본 논문에서 제안한 방법이 더 낳은 성능을 가짐을 알 수 있다.

제 5 장 결 론

최근의 음성 인식 분야는 이전의 멀티미디어와 통신의 분야를 넘어 인터넷을 통한 사용자 제작 콘텐츠 (UCC: User Creadted Contents)로의 활용과 디지털 음원 필터링 기술, 더 나아가 음성을 인식하여 발언자의 의도까지 파악하는 인공지능 기술 등 매우 광범위한 분야에 사용되어지고 있다. 더불어 음성인식 기술 시장의 규모는 작년 $6억에서 2009년까지 두 배로 성장할 전망이다[23]. 음성인식 분야는 사실상 모든 전자제품에 적용될 수 있으므로 제대로 된 음성인식 기술을 필요로 하고 있다. 하지만 음성인식 기술은 연구 계발단계와 같은 환경에서는 좋은 성능을 나타내지만 실제 인식환경에서는 성능이 저하될 수 있다. 특히 수신된 음성의 질은 잔향음과 부가잡음에 의해 쉽게 왜곡된다. 이에 본 논문에서는 반사환경에서의 음성 인식에서 마이크로폰 배열로부터 수집된 음성의 음질 향상을 위한 새로운 방법을 제안하였다. 제안된 방법은 음성의 여기 정보의 특성에 기반하고 있다. 여기에서 가장 중요한 특성은 여기의 세기가 성문 경계의 주변에서 최대라는 것이다. 음성신호 처리에서 가장 널리 사용되는 선형예측 분석을 사용하여 선형예측 잔여신호를 통해 음원의 특징을 얻었다. 선형예측 잔여신호의 가중치 함수는 다른 마이크로폰으로부터 코

히런트하게 지연 보상된 선형예측 잔여신호의 힐버트 포락선 조합에 의해 얻었다. 새로운 선형예측 잔여 신호는 선형예측 잔여신호의 가중치와 힐버트 변환으로 얻은 수정된 조합들을 사용하였다. 실험 결과 제안된 방법을 통한 음질 향상 방법이 기존의 방법보다 향상된 결과를 얻을 수 있음을 알 수 있었다. 기존의 방법과 제안한 방법으로 얻은 잔여신호를 비교해 놓은 그림에서 보는 바와 같이 제안된 방법을 통해 얻은 잔여신호의 크기가 기존의 방법을 통해 얻은 잔여신호의 크기보다 눈에 띄게 음원 신호의 잔여신호에 가까움을 알 수 있었다. 잔여신호를 합성하여 복원된 음성 신호를 나타낸 그림에서 복원된 신호 또한 제안된 방법을 통해 합성한 음성 신호가 기존의 방법으로 얻은 음성 신호보다 음원 신호에 더 가까움은 물론이고 잔향효과가 줄었음을 눈으로 확인할 수 있다. 주관적인 평가방법인 MOS 테스트 결과와 객관적인 방법인 SD 그래프에서 보았던 바와 같이 제안된 방법으로 얻은 음성 신호가 기존의 방법으로 얻은 음성신호보다 음원신호에 가까움을 알 수 있다. 수치적으로 기존의 방법에 비해 약 20%정도의 음질 향상 효과를 가짐을 알 수 있었다.

앞으로 다양한 잔향시간과 다양한 음성 및 환경, 그리고 움직이는 음원에 대한 실험을 통하여 제안된 방법의 성능 검증에 관한 연구를 계속 하고자 한다. 또한 제안된 알고리즘의 가중치 벡터에 대하여 수치적인 정의를 정리하면 더욱 완성도 높은 연구가 될 것이라고 판단된다.

참 고 문 헌

[1] P. Satyanarayana, "Short segment analysis of speech for enhancement," *Ph.D. thesis, Dept of Electrical Engineering, IIT Madras, Chenai, India,* Feb 1999.

[2] V. R. Ramachandran and B. Yegnanarayana, “Coarticulation rule for a text-to-speech system for Hindi,” *in Proc. Workshop on Speech Technology, Madras,* pp. 211-219, Dec. 1992.

[3] Jae Lim and A. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. on Speech and Audio Processing,* vol. 26, pp. 197-210, Jun. 1978.

[4] A. Erell and M. Weintraub, "Estimation of noise-corrupted speech DFT -spectrum using the pitch period," *IEEE Trans. on Speech and Audio Processing,* vol. 2, pp. 1-8, Jan. 1994.

[5] B. Yegnanarayana, C. Avendano, H. Hermansky and P. Satyanarayana Murthy, "Speech enhancement using linear prediction residual," *Speech Communication,* vol. 28, pp. 25-42, May 1999.

[6] A. Dembo and O. Zeitouni, "Maximum a posteriori estimation of time-varying ARMA processes from noisy observations," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. 36, pp. 471−476, Apr. 1988.

[7] Y. Ephraim, "Statistical-model-based speech enhancement systems," *in Proc. IEEE,* vol. 80, pp. 1526-1555. Oct. 1992.

[8] Hagai Attias, John C. Platt, Alex Acero and Li Deng, "Speech Denoising and Dereverberation Using Probabilistic Models,” *Advance in NIPS, Denver, USA,* vol. 13, pp. 758-764. 2000.

[9] D. A. Berkley and J. B. Allen, “Normal listening in typical rooms: the physical and psychophysical correlates of reverberation,” *in Acoustical*

Factors Affecting Hearing Aid Performance, 2nd ed, G. A. Studebaker and I. Hochberg, Eds. Needham Heights, MA: Allyn and Bacon, pp. 3–14, 1993.

[10] J. J. Jetzt, “Critical distance measurement of rooms from the sound energy spectral response,” *J. Acoust. Soc. Amer.*, vol. 65, pp. 1204–1211, May. 1979.

[11] B. W. Gillespie, H. S. Malvar and D. A. F. Florencio, “Speech dereverberation via maximum-kurtosis subband adaptive filtering,” *in Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, vol. 6, pp. 3701–3704, 2001.

[12] Mingyang Wu and DeLiang Wang, "A Two-Stage Algorithm for One-Microphone Reverberant Speech Enhancement", *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, pp. 774-784, MAY. 2006.

[13] N. Gaubitch, P. A. Naylor and D. B. Ward, "On the use of linear prediction for dereverberation of speech", *in Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC-03), Kyoto, Japan*, pp. 99-102, Sept. 2003.

[14] N. D. Gaubitch, P. A. Naylor and D. B. Ward, “Multi-microphone speech dereverberation using spatio-temporal averaging,” *in Proc. 12th European Signal Processing Conf. (EUSIPCO-04), Vienna, Austria*, pp. 809–812, Sept. 2004.

[15] X. Huang, A. Acero and H.W. Hon, *Spoken Language Processing: a Guide to Theory, Algorithm, and system development*, Upper Saddle River, NJ: Prentice Hall, 2001.

[16] M. Wu, D.L. Wang and G. J. Brown, “A multi-pitch tracking algorithm for noisy speech,” *in Proc. IEEE ICASSP*, pp. 369-372, 2002.

[17] R.D. Patterson, I. Nimmo-Smith, J. Holdsworth and P. Rrice, *APU Report 2341: An Efficient Auditory Filterbank Based on the Gammatone Function*, Cambridge: Applied Psychology Unit, 1988.

[18] M. Wu and D. L. Wang, “A one-microphone algorithm for reverberant speech enhancement,” *in Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, pp. 844–847, 2003.

[19] Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing." pp. 280-294, JOHN WILEY & SONS, INC.

[20] Wen Jin and Michael S. Scordilis, “Speech enhancement by residual domain constrained optimization,” *Speech Communication*, vol. 48, pp. 1349-1364, 2006.

[21] T. V. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction from linear prediction residual for identification of closed glottis interval,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 4, pp. 309–319, Aug. 1979.

[22] B. Yegnanarayana, S. R. M. Prasanna and K. S. Rao, “Speech enhancement using excitation source information,” *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, vol. I, FL, pp. 541–544, May 2002.

[23] CNN News, http://money.cnn.com/magazines/business2/business2_archive/2007/02/01/8398978/index.htm

[24] 박찬섭, 김기만, 강석엽, “개선된 선형예측 잔여를 이용한 음성의 잔향음 제거,” 한국해양정보통신학회논문지, vol. 10, Nov. 2007.

[25] Chan Sub Park, Hyung Jun Ju, Ji Won Jung and Ki Man Kim, "Multi-Sensor Speech Dereverberation using LP Residual Combination," *International Symposium on Electrical-Electronics Engineering, Hochiminh, Vietnam*, Track 2, pp. 59-63, Oct. 2007.

[26] 박찬섭, 주형준, 김기만, 서익수, 오원천, “수중 아날로그 통신에서 선형예측 잔여 조합을 이용한 음질 개선,” 수중음향학회 학술발표회 논문집, 진해, Aug. 2007.

감사의 글

끝없는 사랑과 관심으로 지난 2년간 제게 가르침을 주셨던 김기만 교수님께 진심으로 감사드립니다.
바쁘신 와중에도 미비한 논문을 정성으로 심사해주신 정지원 교수님, 강석엽 교수님께도 진심으로 감사드립니다.

오늘의 이 논문이 있기까지 실로 많은 분들의 아낌없는 배려와 도움이 있었습니다. 인자한 성품으로 고민을 들어준 외형이형, 사랑으로 후배를 대해준 세영이형, 언제나 다정다감한 정우형, 입학동기로서 동고동락하며 버팀목이 되어준 형준이와 도움을 주신 그 외 많은 분들에게 이 글을 빌어 감사의 마음을 전합니다. 항상 조언을 아끼지 않으셨던 아버지, 자식 걱정에 잠 못 드시는 어머니, 끝없는 손주 사랑 할머니, 하나뿐인 소중한 동생 찬희 그리고 다른 모든 우리 가족 분들에게도 감사드립니다.

나의 가장 큰 후원자였던 독수리 5형제 형준, 진우, 덕우, 성훈에게 고마움을 표시합니다. 몸은 떨어져 있지만 마음만은 항상 같이 있었던 상철, 재범, 대학에서 만난 소중한 99동기 병협, 동우, 동진, 민군, 성민, 도훈, 승빈, 대식, 영진, 선혜, 영아, 은진, 희경, 현주, 소영, 힘들 때 큰 힘이 되어주었던 수정, 먼 이국땅에서 인연을 맺은 민석, 가혜, 경은, 현진, 일권, 기선, 혜선, 미경, 장희, 혜성, 진훈, 정아, 중현, 경자, 만준, shingos, Ai, 나의 영원한 사랑 넵튠 선후배님들, 사랑하는 동기 은실, 어려운 부탁 들어주시던 희연누나, 그 외 대응연 식구들까지 모두 감사드립니다. 위성통신, 안테나, MMIC, 마이크로파, 이동통신 실험실의 모든 선후배님들에게 감사의 말을 전합니다. 나의 영원한 형제들 경민, 재진, 보람, 소이, 영주, 혜란, 수년에 한번 만나지만 마음만은 내편인 은정, 보혜 화영, 경주, 상호, 용주, 리나, 누군가를 가르친다는 경험을 함께 나눈 재은, 지숙, 희진, 정대, 주연, 진우, 윤선, 가끔 나의 귀찮은 부탁도 항상 들어주던 미국에 있는 우일, 나를 먹여살려준 은영누나, 민정누나, 순영누나, 지온누나, 간지의 보경, 소영, 우연, 은영형, 현성, 명진누나, 명희, 동석, 은경, 지선, 성은, 동현형, 경훈형, 아름, 현아, 일우, 건, 미화, 은수, 다은누나, 우주누나, 경민, 선영누나, 소영, 인택, 보람, 경용형, 은정, 은영, 시훈형, 그리고 전파공학과의 명승형, 준오형, 익수형, 동식형, 철승형, 호승형, 재국형, 건희누나, 민지누나, 희정누나, 길수누나, 은정누나, 주희누나, 지영누나, 경화누나, 혜진누나, 유리누나, 경미, 제현, 희선, 수경, 은혜, 지현, 성화, 은희, 다운, 경지, 예준, 진경, 혜란, 태우, 철희, 승목, 수훈, 상길, 남수, 윤성, 경학, 경수, 경관, 태두, 은상, 해인, 선관, 승배, 광호, 승민, 병철, 광훈, 마니, 진화 외 모든 선후배님들께 감사의 말을 전합니다.

어느새 시간이 흘러 헤어짐의 순간이 다가왔습니다. 길다면 길고 짧다면 짧았던 대학생활이 정말 소중한 시간이었습니다. 저는 부자가 되었습니다. 배움은 물론 좋은 분들과의 추억과 인연은 머리와 가슴에 가득 차 세상 무엇과도 바꿀 수 없는 귀한 재산으로 남았습니다. 그동안 저에게 보내주신 질책과 격려 그리고 사랑 절대 잊지 않고 실망시켜드리지 않는 박찬섭이 되기 위해 앞으로도 열심히 살아가겠습니다.