음성통신을 위한 잡음처리 기술

신종원 | 장준혁 | 김남수
서울대학교 · 인하대학교 · 서울대학교

요 약

음성 통신을 할 때 배경 잡음이 존재하게 되면 일반적으로 음질이 저하된다. 이것은 잡음 자체가 듣기 싫다거나 음성을 더 작게 들리게 만들기 때문이기도 하고 음성 코덱이 잡음이 섞이지 않은 깨끗한 음성에 최적화되어 있어서 잡음이 섞인 음성에 대한 코딩 효율이 떨어지기 때문이기도 하다.

이 논문에서는 잡음에 의한 음성 통신의 품질 저하를 막기 위한 방법으로서 음성 향상(speech enhancement) 기술과 음성 강화(speech reinforcement) 기술에 대해 소개한다.

음성 향상 기술이란 전송부의 마이크에서 녹음된 잡음과 음성이 섞인 입력 음성으로부터 깨끗한 음성을 추정하는 기술을 말한다. 음성 향상 기술은 상당히 오랜 기간 동안 연구되어 온 기술이며, 최근에는 각 파라미터의 분포에 의존하는 방법보다 확률 모델에 기반한 방법이 각광을 받고 있으며 인간의 청각 특성을 고려한 음성 향상 방법도 제안되고 있다. 음성 강화 기술이란 수신단에서 주변 잡음에 따라 전송되어 온 음성을 주파수별로 증폭하여 더 잘 들리도록 만드는 기술이다.

음성 향상이 내 주위의 잡음이 상대방에게 들리는 음성에 미치는 영향 혹은 상대방 주변의 잡음이 나에게 들리는 소리에 미치는 영향을 줄여주는 기술이라면 음성 강화는 내 주위의 잡음이 나에게 들리는 음성에 미치는 영향을 상쇄해 주는 기술이다.

이 경우 주변 잡음은 어떤 전자 시스템도 거치지 않고 귀로 직접 들어오기 때문에 잡음 자체를 줄여 주는 것은 힘들고 전송되어 온 음성을 적절히 증폭 혹은 변형함으로써 귀에 들리는 음질 또는 명료성을 개선하게 된다. 이 논문에서는 통계 모델을 기반으로 한 음성 향상 기법과 인간의 청각 특성을 고려한 음성 향상 기법, 그리고 음성 강화 기법에 대해 설명한다.

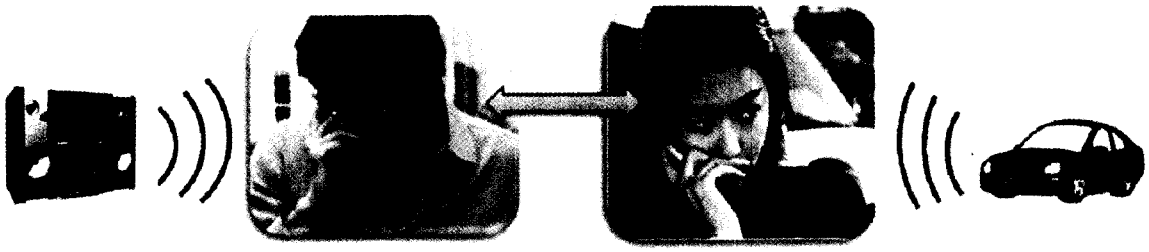
1. 서 론

멀티미디어가 발달한 지금도 음성은 가장 편리한 통신 수단으로 자리매김하고 있다. 특히 이동 통신의 발달과 함께 현재는 휴대폰 가입자수가 4천만을 돌파하여 1인 1휴대폰 시대가 열린 상황이다. 이와 함께 더 나은 음질의 음성 통화를 원하는 목소리가 더욱 높아지고 있다.

잡음이란 우리가 듣고자하는 소리 외의 모든 소리를 말한다. 곧, 음성 통화에 있어서는 상대방의 목소리를 제외한 상대방이나 내 주변에서 발생하는 모든 소음들이 잡음이 된다. 이러한 잡음이 존재할 경우 음성통신 시스템을 거친 소리의 음질은 떨어지게 되는데, 이것은 단순히 잡음 자체가 듣기 싫은 것 때문만이 아니라 잡음이 존재하면 음성이 더 작게 들리고(마스킹 효과), 또 휴대폰의 경우 음성 코덱이 잡음이 없는 음성에 최적화되어 있어서 잡음 섞인 음성은 효율적으로 코딩하지 못하기 때문이기도 하다.

이러한 잡음에 의한 음질 저하 문제에 대한 해결책으로 제시된 것이 바로 음성 향상(speech enhancement) 기술과 음성 강화(speech reinforcement) 기술이다.

음성 향상 기술은 전송부의 입력 마이크에 들어온 잡음 섞



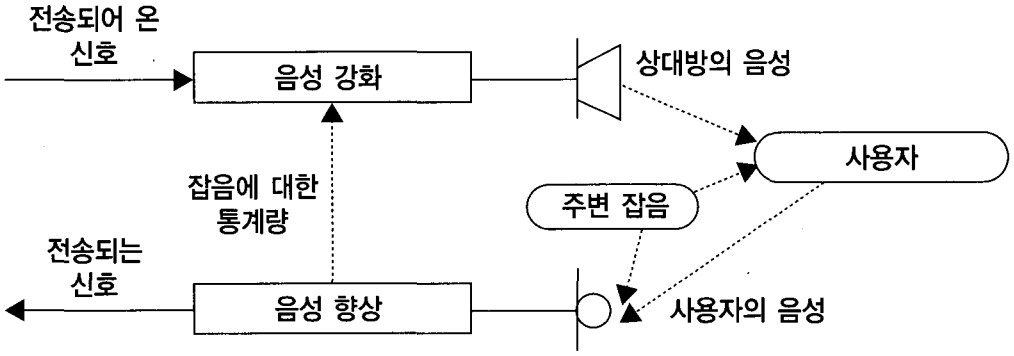
(그림 1) 잡음 환경에서의 음성 통신

인 신호로부터 잡음 없는 깨끗한 신호를 추정하는 기술이다. 음성 향상 기술은 크게 마이크가 한 개인 경우인 단채널 음성 향상(single channel speech enhancement)와 마이크가 둘 이상인 경우인 다채널 음성 향상(multi channel speech enhancement)로 나누어진다. 다채널 음성 향상은 신호 분리나 신호원 위치 추정 같은 기법들을 활용할 수 있어 성능이 더 좋지만 실제로 복수의 마이크가 존재하는 응용례가 많지 않아 적용이 제한적이다. 그래서 현재 음성 통신 시스템에 적용되어 있는 것은 이 논문에서 설명할 단채널 음성 향상 방법이다.

단채널 음성 향상 기법은 6, 70년대부터 연구되어 왔다. 예전에는 음성의 주기성에 기초한 방법이나 음성 발성 모델에 기초한 방법도 사용되었으나 [1] 최근에는 통계 모델에 기반을 둔 스펙트럼의 크기를 추정하는 방법이 주로 사용되고 있다. Ephraim과 Malah의 논문에서 제시한 음성 향상 알고리즘 [2]의 놀라운 성공 이후 통계 모델을 기반으로 한 음성 향상 기법은 발전을 거듭해 오고 있으며, 최근에는 인간의 청각 특성을 고려한 음성 향상 기법들이 속속 개발되고 있는 상황이다.

음성 향상 기술이 음성 통신의 입장에서 보면 내 주변의 잡음(near-end noise)이 상대방(far-end speaker/listener)에게 미치는 영향 혹은 상대방 주변에 있는 잡음(far-end noise)이 나(near-end speaker/listener)에게 미치는 영향을 줄여주는 기술이라면 음성 강화 기술은 내 주위의 잡음이 나에게 미치는 영향을 보상해 주는 기술이다.

음성 향상에서와는 달리 이 경우 주변 잡음이 어떠한 전자 시스템도 거치지 않고 직접 귀로 들어오기 때문에 잡음 자체를 제어하는 것은 거의 불가능하다. 따라서 주변 잡음에 따라 상대 쪽에서 전송되어 온 음성을 적절히 증폭하거나 변형함으로써 인간이 느끼는 음질이나 음성의 명료성을 개선하는 것이 음성 강화의 목표이다. 음성 향상과 음성 강화 알고리즘이 적용된 음성 통신 시스템이 (그림 2)에 나타나



(그림 2) 음성 향상과 음성 강화 기법을 적용한 음성 통신 시스템

있다.

이 논문에서는 최근 각광을 받고 있는 음성 향상 기법들 및 음성 강화 기법들을 소개한다. II장에서는 통계 모델을 기반으로 한 음성 향상 기법들에 대해 설명하고 III장에서는 인간의 청각 특성을 고려한 음성 향상 기법들에 관해 소개한다. 음성 강화 기법들에 관해서는 IV장에서 다루기로 한다.

II. 통계 모델을 기반으로 한 음성 향상 기법

음성 신호를 표현하는 방법에는 여러 가지가 있지만 가장 대표적인 것 중 하나가 바로 스펙트럼이다. 스펙트럼이란 어떤 신호를 주파수에서 표현한 것을 말한다. 음성의 경우에는 시간에 따른 스펙트럼의 변화가 중요하기 때문에 흔히 스펙트럼이라고 하면 10ms에서 30ms 정도의 길이의 음성에 대한 이산 푸리에 변환(discrete Fourier transform, DFT) 계수를 말한다. 음성 향상도 여러 영역에서 수행할 수 있지만 스펙트럼 영역에서 수행하는 것이 구현이 간단하고 적용 범위가 넓어서 가장 널리 쓰이고 있다.

가장 전통적인 형태의 주파수 영역에서의 음성 향상 방법은 Wiener 필터이다 [1]. Wiener 필터는 각 주파수 성분이 통계적으로 독립이라는 가정 하에 잡음 없는 음성의 DFT 계수 자체를 최소 평균 제곱 에러(minimum mean square error, MMSE) 방법으로 추정했을 때의 해로서 k째 DFT 계수에 곱해져야 할 이득(gain) $H(k)$가 다음과 같은 간단한 꼴로 구해진다.

$$H(k) = \frac{\lambda_s(k)}{\lambda_s(k) + \lambda_n(k)}$$

여기서 $\lambda_s(k), \lambda_n(k)$는 각각 깨끗한 음성과 잡음의 분산(variance)이다. Wiener 필터는 계산이 간단하고 입력 신호의 확률 분포에 영향을 받지 않으므로 많은 다른 기술들이 발달한 지금도 아직 그 효용성이 다하지 않고 있다.

음성 향상 기법을 적용하면 잡음이 줄어드는 대신에 잃는 것이 있는데 그것이 음성이 왜곡되는 것(speech distortion)과 잔여 잡음(residual noise)이 귀에 거슬리게 되는 것이다. 음성 향상 기법의 적용 때문에 발생하는 잡음(혹은 추정 에러)을 뮤지컬 노이즈(musical noise)라고 한다. 자연계에 존재하는 모든 소리들은 인간의 귀에 그리 거슬리지 않는 데에 비해 인공적인 처리에 의해 생긴 이런 뮤지컬 노이즈는 상당히 듣기 싫기 때문에 잡음 레벨이 높지 않은 곳에서는 차라리 음성 향상 기법을 적용하지 않은 음성을 선호하는 사람들도 있다. 이러한 뮤지컬 노이즈를 획기적으로 줄인 음성 향상 기법이 바로 1984년 발표된 Ephraim과 Malah의 스펙트럼 크기 추정 기법이다 [2].

Ephraim과 Malah는 Wiener 필터와 마찬가지로 MMSE 추정을 수행하되 DFT 계수 자체가 아닌 DFT 계수의 크기(amplitude or magnitude)를 추정했다. 이것은 인간의 귀가 스펙트럼의 크기에는 민감하고 위상(phase)에는 민감하지 않다는 점에 착안한 것이다. 음성과 잡음의 DFT 계수의 분포는 실수부와 허수부가 각각 정규 분포(Gaussian distribution)을 따르는 것으로 가정하고 그 분산 값을 추정했다. 또, 신호 대 잡음비(signal-to-noise ratio, SNR)를 추정하는 방법으로 결정-인도(decision-directed, DD) 방법을 써서 현재 입력으로부터 구한 SNR 값과 이전의 추정치로부터 구한 SNR 값을 적절히 배합함으로써 SNR의 추정을 강인하게 했다.

이들은 이듬해 발표한 논문에서 사람의 귀가 크기를 log 축척으로 느낀다는 점에 착안해 DFT 계수의 크기의 log 값을 추정하는 방법을 써서 성능을 약간 더 추가적으로 향상시키기도 했다 [3]. 이 Ephraim과 Malah의 두 논문은 이후 음성 향상 논문들에 많은 영향을 미쳤으나 최종 결과식이 Wiener 필터의 경우처럼 단순하지 않고 modified Bessel function을 구해야 하므로 modified Bessel function의 값을 표 형태로 저장해 놓아야 하는 부담이 있다. 이 정규 분포를 바탕으로 한 음성 향상 방법은 soft decision 방식을 가미하여 음성의 존재 확률을 강인하게 모델링함으로써 그 성능이 향상되기도 했다 [4].

Wiener 필터의 경우와 같이 DFT 계수 자체를 MMSE 추정하는 경우에는 그 해가 음성이나 잡음의 확률 분포와 관계없이 일정하지만 DFT 계수의 크기를 추정하는 경우에는 그 해가 확률 분포에 따라 달라진다. 또한 MMSE 외의 최대 유

사도 추정(maximum likelihood estimation, MLE)이나 최대 사후 확률(maximum a posteriori) 추정 방법을 사용해도 결과는 달라진다.

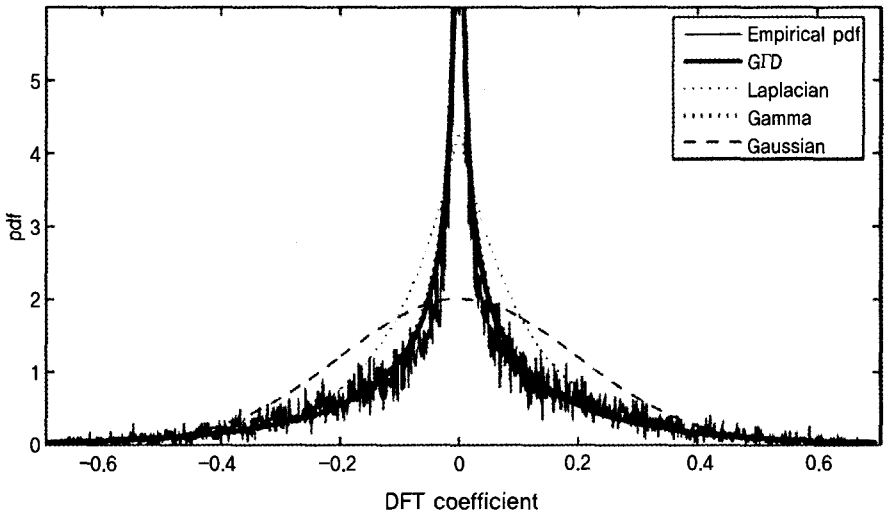
잡음의 DFT 계수의 확률 분포는 어떻게도 변할 수 있으므로 정규 분포로 추정하는 것이 타당하지만 음성의 DFT 계수의 확률 분포에 관해서는 정규 분포보다 잘 맞는 분포가 있다는 것이 이미 알려져 있다. 음성의 스펙트럼의 확률 분포는 정규 분포보다 '꼬리 부분' 이 큰(heavy-tailed), 보다 sparse한 분포라고 알려져 있다. 얼마나 작은 신호도 음성으로 인정하느냐에 따라 둘 사이에는 어느 것이 더 잘 맞는 분포냐가 달라지지만 Laplacian 분포와 Gamma 분포 모두 정규 분포보다 음성의 스펙트럼의 분포를 잘 모델링하는 것으로 알려져 있다 [5]-[8].

또한 Laplacian 분포와 정규 분포를 포함하는 일반적인 분포인 generalized Gaussian 분포(GGD)나 정규 분포, Laplacian, Gamma, GGD까지 포함하는 더욱 일반적인 분포인 generalized gamma 분포(GΓD)를 쓴다면 음성의 분포를 더욱 정확하게 모델링할 수 있다. 부가적으로 이 일반적인 분포들은 음성 및 잡음의 분포 모양이 시간에 따라 변할 경우에도 잘 따라갈 수 있는 유연성을 갖추고 있다. (그림 3)에는 실제 음성의 DFT 계수의 분포와 각 확률 모델들이 나타나 있다. 여기서 각 확률 모델들의 파라미터들은 MLE를 통해 전체 음성 파일에 대해 최적화되도록 추정되었다. 그림에서 보는 바와 같이 정규 분포(Gaussian)는 음성의 DFT 계수의 분포를 표현하는 데에 그리 좋은 모델이 아니며, Laplacian이 더 좋은 모델이고 그보다는 Gamma 분포가 더 좋은 모델이며 GΓD가 음성의 분포를 가장 잘 모델링한다. 물론 이 그림에는 시간에 따른 변화를 따라 갈 수 있는 유연성에 대한 부분은 나타나 있지 않다.

이러한 더 나은 통계 모델들을 이용한 음성 향상 방법에 대한 연구들이 최근 들어 속속 발표되고 있다 [5], [8]-[11]. 보다 정확한 통계 모델을 이용하는 것은 음성 향상에 있어 어느 정도의 성능 향상을 가져다주었다.

III. 인간의 청각 특성을 고려한 음성 향상 기법

앞서 말한 것과 같이 음성 향상 기법을 적용하면 잡음이 줄어드는 대신에 잃는 것이 있는데 그것이 음성이 왜곡되는 것(speech distortion)과 잔여 잡음(residual noise)이 귀에 거



(그림 3) 실제 음성 DFT 계수의 분포(Empirical pdf)와 파라미터들이 MLE로 구해진 각 확률 모델들의 분포

슬리게 되는 것이다. 이 두 가지 문제점의 해결은 상충하는 것이어서 음성의 왜곡을 없애려 하면 남아 있는 잡음이 많게 되고 잔여 잡음을 없애려 하면 음성이 많이 왜곡된다. Signal subspace 음성 향상 기법은 잔여 잡음 혹은 잔여 잡음의 주파수 성분을 특정 값 아래로 유지하는 조건 하에서 음성의 왜곡을 최소화하는 것을 목표로 한다[12], [13].

Signal subspace 음성 향상이라는 이름은 이런 목표에 따른 음성 향상의 결과가 입력 신호를 음성 신호가 형성하는 subspace와 잡음이 형성하는 subspace로 분리하여 음성 향상을 수행하는 것으로 주어지기 때문에 붙여졌다. 이 방법은 음성 향상의 성능 자체는 꽤 좋지만 기본적으로 correlation matrix의 고유 벡터(eigen vector)들을 구해야 하므로 계산량이 상당히 많기 때문에 그리 널리 쓰이지는 못했다. 이 접근 방법의 계산량을 줄이기 위한 방법들도 연구되고 있다 [14].

이와 관련하여 인간의 청각 특성을 고려한 음성 향상 기법들도 제안되고 있다. 가장 먼저 제안된 방법들은 인간의 청각의 마스킹 효과(masking effect)를 이용한 방법들이다. 마스킹 효과란 어떤 큰 신호(masker)가 존재할 때 같이 존재하는 다른 작은 신호(maskee)가 들리지 않는 효과를 말한다. 이 마스킹 효과는 현재 특히 mp3 등의 오디오 코딩에서 많이 이용되고 있는데, 거기에서는 적은 데이터 양으로 음악을 표현함으로써 생기는 에러(양자화 잡음, quantization noise)이 원래의 음악에 의해 마스킹되게 함으로써 그 에러가 들리지 않게 만든다. 어떤 신호가 만드는 마스킹 임계치(masking threshold)란 함께 존재하는 다른 신호가 그 레벨 이하이면 들리지 않게 되는 레벨을 말한다.

이런 인간의 청각의 마스킹 효과를 이용하여 각 주파수별로 잔여 잡음이 음성이 만드는 마스킹 임계치보다 작아지게 만듬으로써 잔여 잡음이 들리지 않도록 하는 기법들이 제안되었다[15], [16].

이 방법들은 인간의 청각 특성을 고려한 최초의 시도들이고, 성능도 꽤 좋지만 기본적으로 계산량이 많고, 잡음이 클 경우에는 잔여 잡음을 마스킹 임계치 이하로 만들 경우 음성의 왜곡이 커질 수 있다는 단점이 있다.

인간의 지각을 고려한 또 다른 접근 방식으로는 잔여 잡음을 안 들리게 하는 대신 듣기에 좋은 잡음으로 바꿔 주는 것이 있다 [17]. 이 방법은 인간이 귀에 듣기에 좋은 잡음이 있고 듣기 싫은 잡음이 있다는 점에 착안한다. 일반적으로 자연계에 존재하는 소리들은 합성음에 비해 듣기에 편안하고 저주파가 많은 소리가 고주파가 많은 소리에 비해 듣기에 편하다. 여기에는 음성이 저주파 성분이 많으므로 마스킹 효과의 영향도 있지만 그것만이 전부는 아니다. 이 방법에서는 이미 편안한 잡음이라고 알고 있는 스펙트럼이 주어져 있을 때 음성 왜곡의 크기, 잔여 잡음의 크기, 잔여 잡음의 스펙트럼의 모양이 이미 알고 있는 편안한 잡음의 스펙트럼의 모양과 얼마나 다른가 이 세 가지를 줄여주도록 음성 향상을 수행한다.

이 방법은 잔여 잡음의 스펙트럼을 주파수별로 특정 임계치로 정확히 만드는 것이 아니라 전체적인 모양만을 융통성 있게 통제한다는 특징이 있다. 또, 여기에 몇 가지 가정을 할 경우 결과 식이 Wiener 필터와 비슷한 꼴로 나와서 계산량이 극히 적다는 장점도 있다.

IV. 음성 강화 기법

서론에서 설명한 바와 같이 음성 강화 기법은 내 주위의 잡음이 나에게 미치는 영향을 보상해 주는 기술이다. 음성 향상에서와는 달리 이 경우 주변 잡음이 어떠한 전자 시스템도 거치지 않고 직접 귀로 들어오기 때문에 잡음 자체를 제어하는 것은 거의 불가능하다. 능동 소음 제어(active noise control) 기법은 잡음 자체를 제어하는 쪽이지만 적용에 제한이 많다.

잡음이 심할 때 귀로 듣는 소리의 크기가 작게 느껴지는 것은 누구나 한번쯤 일상생활에서 경험해 봤을 것이다. 지하철 같은 곳에서 음악을 들을 때, 혹은 시끄러운 곳에서 휴대폰으로 통화를 할 때 조용한 곳에서보다 같은 크기의 소리가 훨씬 작게 느껴지는 것을 경험적으로는 이미 다들 알고 있을 것이다. 이런 일이 일어나는 이유가 심리 음향학(psychoacoustics)에서 말하는 부분 마스킹 효과 (partial masking effect)이다 [18], [19].

앞 장에서 작은 신호가 큰 신호 때문에 아예 안 들리게 되는 마스킹 효과에 대해 설명했었다. A라는 신호가 존재할 때 B라는 신호가 A보다 아주 크다면 A는 아예 안 들리고 B

는 A가 없을 때와 같은 크기로 들릴 것이다. 반대로 B가 A보다 아주 작다면 A는 B가 없을 때와 같은 크기로 들리고 B는 아예 안 들릴 것이다. 그렇다면 B의 크기가 그 중간일 때는 어떨까? 실험 결과는 직관적으로, 혹은 경험적으로 예상하듯이 두 신호 모두 원래보다 작게 들리게 된다는 것이다.

즉, A는 B를, B는 A를 부분적으로 마스킹하게 된다. 이것이 부분 마스킹 효과이고, 이에 따라 잡음이 있으면 음성의 크기가 작게 들리게 된다.

이 내 주위의 잡음이 내가 듣는 데에 미치는 영향에 대해서는 60~70년대에 약간의 기초적인 연구가 이루어진 후 상당 기간 주목을 받지 못해 왔다. 70년대 Niederjohn의 연구에서는 음성의 명료성에 고주파 성분이 더 중요하다고 하여 high pass filtering을 하고 자음 등 크기가 작은 부분이 없어지는 것을 막기 위해 크기가 작은 부분은 크게 하고 큰 부분은 작게 하는 등의 임기응변적인 방법을 통해 음성의 명료도를 향상시켰다 [20]. 그 후 오랫동안 연구가 이루어지지 않다가 최근에 와서야 다시 체계적인 연구가 이루어지기 시작한 것이 현실이다.

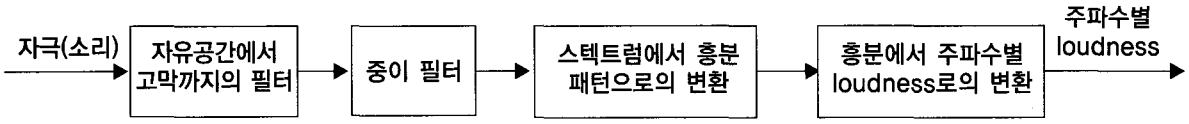
따라서 주변 잡음에 따라 상대 쪽에서 전송되어 온 음성을 적절히 증폭하거나 변형함으로써 인간이 느끼는 음질이나 음성의 명료성을 개선하는 것이 음성 강화의 목표이다. 전송되어 온 신호를 증폭시키면 더 명료하게 들리고 음질도 좋아지는 것은 당연하다. 그러나 얼마나 증폭시킬 것인가, 그리고 주파수별로 어떻게 다르게 증폭시킬 것인가의 문제가 남는다. 가장 쉬운 방법은 잡음 레벨에 따라 전송되어 온 신호에 일정한 수를 곱해주는 것이다. 다시 말해 모든 주파수 성분에 같은 수를 곱해주는 것이다. 그러나 음성과 잡음의 스펙트럼 모양을 생각할 때 이 방법이 최적이 아님은 명백하다.

또한 이 경우에도 각 주파수 별로 잡음이 있는 정도가 다르므로 음색은 원래의 전송되어 온 신호와는 다르게 된다. 어느 정도로 크게 증폭시켜야 할 것인지도 불분명한 문제이다. 또 다른 가능성은 앞 장에서 설명한 마스킹 효과를 이용하여 잡음이 완전히 들리지 않을 때까지 주파수별로 전송되어 온 신호를 증폭시켜 주는 것이다 [21]. 그러나 이 방법은 잡음 레벨이 어느 정도 이상일 경우에는 증폭된 신호가 엄청나게 커지게 된다. 잡음이 아예 안 들릴 때까지 음성을 키운 것이니 얼마나 큰 것일지는 짐작이 갈 것이다. 또 다른 가능성은 주파수별로 신호 대 잡음 비(Signal-to-Noise Ratio, SNR)가 일정하게 유지되도록 전송되어 온 음성을 키워주는 것이다.

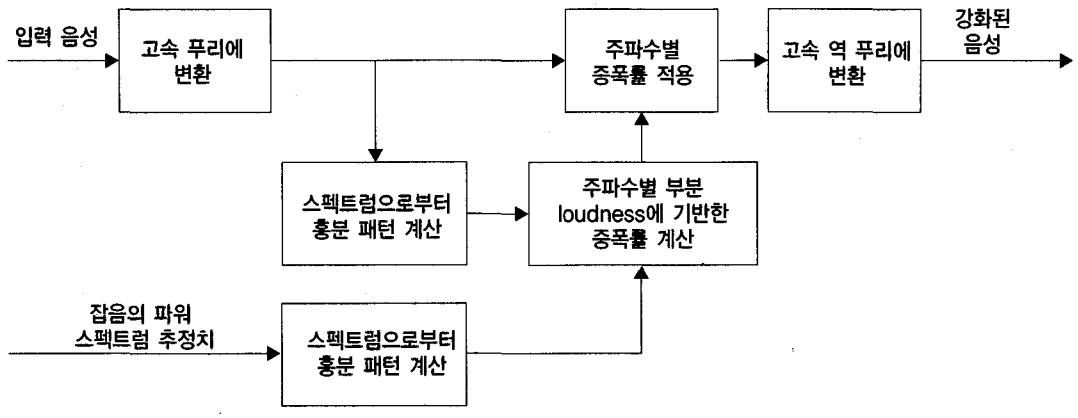
[22], [23]. 이 방법은 모든 주파수를 똑같이 키워 준 경우에 비해서는 좋은 음질과 명료도를 얻을 수 있지만 SNR이라는 척도는 인간의 청각 특성과 들어맞지는 않는다. SNR을 얼마로 맞추어 주어야 하는지가 불분명한 것도 마찬가지이다. 또한 위의 두 방법 모두 원래 전송되어 온 신호의 크기가 얼마였나 하는 것은 고려하지 않는다. 즉, 원래 작게 들려야 할 신호와 크게 들려야 할 신호가 있다면 모두 같은 정도로 들리게 되는 것이다.

최근에는 (그림 4)에 나타난 인간 청각의 loudness 지각 모델을 이용한 음성 강화 방법이 제안되었다 [24]. loudness란 인간이 느끼는 소리의 크기를 말한다. 입력 신호의 음압 레벨과 loudness의 관계는 물론 정의 관계이지만 아주 nonlinear하다. 이 알고리즘에서는 심리 음향학 분야에서 개발된 loudness 지각 모델을 이용하여 주파수 별로 원래 잡음이 없는 상태에서 전송되어 온 신호가 들렸을 크기와 같은 크기로 느껴지도록 잡음이 있을 때 그에 맞춰 전송되어 온 신호를 키워 준다. 이용한 loudness 지각 모델은 물론 부분 마스킹 효과를 고려한 것이다.

이 방법은 인간의 소리 크기 지각에 대한 최근의 연구 결과를 반영하여 원래의 잡음이 없는 소리와 음색이 거의 비슷하게 느껴질 수 있게 한다. 이 알고리즘의 블록도는 (그림



(그림 4) loudness 지각 모델



(그림 5) loudness 지각 모델에 기반을 둔 음성 강화 알고리즘의 블록도 [24]

5)에 나타나 있다.

기본적으로는 지금까지 말한 모든 음성 강화의 방법들이 강화하기 전의 신호에 비해 많은 파워를 필요로 한다. 하지만 이것을 꼭 파워의 과다한 소모로만 볼 수 없는 것이 사용자가 볼륨을 조정할 수 있기 때문이다. 현재의 휴대폰의 경우 음성 강화 기능이 없기 때문에 잡음이 많은 곳에서도 원활히 통화할 수 있고 잡음이 없는 곳에서도 너무 시끄럽지 않으려면 사용자가 조절할 수 있는 볼륨의 범위가 상당히 넓어야 했다. 하지만 음성 강화 기능이 있을 경우 사용자는 볼륨을 일일이 조정하지 않아도 잡음의 크기의 변화에 상관없이 거의 같은 크기의 음성을 들을 수 있으며 평균적으로 소모하는 파워는 오히려 조용할 때에 적게 소모되므로 더 적을 수도 있을 것이다.

V. 결 론

지금까지 음성 통신을 위한 잡음 처리 기술로서 최근에 제안되고 있는 음성 향상과 음성 강화의 기술들에 대해 살펴보았다. 음성 향상 기술은 그 역사도 꽤 오래되었고 현재 2세대, 3세대 혹은 그 이후의 이동통신에 쓰이고 있거나 쓰이게 될 음성 코덱들에 이미 포함되어 있다. 음성 강화 알고리즘을 음성 향상 알고리즘과 함께 사용할 경우 내 쪽에 있는 잡음이 상대방에 미치는 영향과 나에게 미치는 영향, 상대방 쪽에 있는 잡음이 나에게 미치는 영향과 상대방에게 미치는 영향을 모두 감소시켜 줄 수 있어서 잡음 환경에 강인한 더 나은 음성 통신 환경을 조성하는 데에 일조하게 될 것이다.

또한 현재 많은 연구 개발이 이루어지고 있는 더 정확한 확률 모델과 인간의 청각 특성에 바탕을 둔 접근 방식은 음성 향상과 음성 강화 알고리즘의 성능을 보다 향상시킬 수 있을 것으로 전망된다.

[1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, Dec. 1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 6, Dec. 1984.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 2, pp. 443-445, Apr. 1985.

[4] N. S. Kim and J. -H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, no. 5, pp. 108-110, May 2000.

[5] R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed priors," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, FL, USA, vol. 1, pp. I-253 - I-256, May 2002.

[6] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204-207, Jul. 2003.

[7] J. W. Shin, J. -H. Chang and N. S. Kim, "Statistical modeling of speech signals based on generalized gamma distribution," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 258-261, Mar. 2005.

[8] J. -H. Chang, S. Gazor, N. S. Kim and S. K. Mitra, "Multiple statistical models for soft decision in noisy speech enhancement," *Pattern Recognition*, vol. 40, no. 3, pp. 1123-1134, Mar. 2007.

[9] I. Cohen, "Speech enhancement using super-Gaussian speech models and noncausal a priori SNR estimation," *Speech Communication*, vol. 47, issue 3, pp. 336-350, Nov. 2005.

[10] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 845-856, Sep. 2005.

[11] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 7, pp. 1110-1126, 2005.

[12] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," IEEE Transactions on Speech and Audio Processing, vol. 3, no. 4, pp. 251-266, Jul. 1995.

[13] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," *IEEE Signal Processing Letters*, vol. 10, no. 4, pp. 104-106, Apr. 2003.

[14] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95, Feb. 2001.

[15] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 700-708, Nov. 2003.

[16] Y. Hu and P. C. Loizou, " Incorporating a psychoacoustical model in frequency domain speech enhancement," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 270-273, Feb. 2004.

[17] J. W. Shin, S. Y. Lee, H. S. Yun and N. S. Kim, "Speech enhancement based on residual noise shaping, " *Interspeech 2006*, pp. 1415-1418, September 2006.

[18] E. Zwicker and H. Fastl, *Psychoacoustics-Facts and Models*, Berlin: Springer, 1990.

[19] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of Audio Engineering Society*, vol. 45, no. 4, pp. 224-240, Apr. 1997.

[20] R. J. Niederjohn and J. H. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, no. 4, Aug. 1976.

[21] M. Tzur (Zibulski) and A. A. Goldin, "Sound equalization in a noisy environment," *Audio Engineering Society 110th Convention*, preprint no. 5364, May 2001.

[22] A. A. Goldin, A. Budkin and S. Kib, "Automatic volume and equalization control in mobile devices," *Audio*

Engineering Society 121th Convention, Preprint No. 6960, Oct. 2006.

[23] B. Sauert and P. Vary, "Near end listening enhancement: Speech intelligibility improvement in noisy environments," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. I-493-I-496, 2006.

[24] J. W. Shin and N. S. Kim, "Perceptual reinforcement of speech signal based on partial specific loudness," *IEEE Signal Processing Letters*, to appear.

약 력

신 종 원

2002년 서울대학교 전기공학부 학사
2002년 ~ 현재 서울대학교 전기컴퓨터공학부 박사과정
관심 분야 : 통계적 신호 처리, 음성 향상, 음성 검출, 음성 코딩, 음성 강화

장 준 혁

1998년 경북대학교 전자공학과 학사
2000년 서울대학교 전기공학부 석사
2004년 서울대학교 전기공학부 박사
2000년 ~ 2005년 (주)넷더스 수석 엔지니어
2004년 ~ 2005년 U. C. Santa Barbara 박사 후 과정
2005년 한국과학기술연구원(KIST) 연구원
2005년 ~ 현재 인하대학교 전자공학과 교수
관심 분야 : 음성 코딩, 음성 향상, 음성 인식, 오디오 코딩, 적응 신호 처리

김 남 수

1988년 서울대학교 전자공학과 학사
1990년 한국과학기술원(KAIST) 전기 및 전자공학과 석사
1994년 한국과학기술원(KAIST) 전기 및 전자공학과 박사
1994년 ~ 1998년 삼성종합기술원 선임연구원
1998년 ~ 현재 서울대학교 전기컴퓨터공학부 교수
관심 분야 : 음성 인식, 잡음 처리, 음성 합성, 음성 코딩, 통계 신호 처리, 오디오 코딩