

근단 배경 잡음 환경에서 G.729A 음성부호화기 파라미터에 기반한 새로운 음성 강화 기법

Speech Reinforcement Based on G.729A Speech Codec Parameter Under Near-End Background Noise Environments

최 재 훈*, 장 준 혁*
(Jae-Hun Choi*, Joon-Hyuk Chang*)

*인하대학교 전자공학부

(접수일자: 2009년 2월 16일; 수정일자: 2009년 4월 1일; 채택일자: 2009년 4월 7일)

본 논문에서는 근단 (Near-End) 잡음 환경에서 ITU-T의 표준 음성부호화기인 G.729A CS-ACELP 기반의 효과적인 음성 강화 기법을 제시한다. 일반적으로 다양한 배경 잡음이 존재하는 근단 환경에서 수신하는 원단 화자 음성의 명료도가 매우 감소하므로, 이를 극복하기 위한 원단 화자 음성 강화 기법이 필요하다. 기존의 음성강화 시스템과는 대조적으로, 다양한 배경 잡음이 존재하는 근단 환경에서 음성부호화기에 기반하여, 원단으로부터 수신된 비트스트림 파라미터 중 여기신호 (excitation signal)를 강화하는 알고리즘을 제시한다. 구체적으로, 다양한 배경 잡음이 존재하는 근단 환경에서 G.729A CS-ACELP의 부호화기를 통해 배경 잡음의 여기신호를 추정하고, 추정된 배경 잡음의 여기신호를 기반으로 원단 화자로부터 전송된 음성 신호의 여기신호를 강화시키는데, 특별히 G.729A 복호화기내에서 원단의 음성 신호를 직접 강화하는 알고리즘을 제안한다. 제안된 음성 강화 기법의 성능은 다양한 잡음 환경 하에서 ITU-T P.800의 주관적 음질 측정 방법인 CCR (Comparison Category Rating) 테스트에 의해 평가되었으며, 기존의 SNR 복구 기법과 비교해서 우수한 성능을 보여주었다.

핵심용어: G.729A, 마스킹 효과, SNR 복구, 근단 배경 잡음, 음성 강화

투고분야: 음성처리 분야 (2)

In this paper, we propose an effective speech reinforcement technique base on ITU-T G.729A CS-ACELP codec under the near-end background noise environments. In general, since the intelligibility of the far-end speech for the near-end listener is significantly reduced under near-end noise environments, we require a far-end speech reinforcement approach to avoid this phenomena. In contrast to the conventional speech reinforcement algorithm, we reinforce the excitation signal of the codec's parameters received from the far-end speech signal based on the G.729A speech codec under various background noise environments. Specifically, we first estimate the excitation signal of ambient noise at the near-end through the encoder of the G.729A speech codec, reinforcing the excitation signal of the far-end speech transmitted from the far-end. we specially propose a novel approach to directly reinforce the excitation signal of far-end speech signal based on the decoder of the G.729A. The performance of the proposed algorithm is evaluated by the CCR (Comparison Category Rating) test of the method for subjective determination of transmission quality in ITU-T P.800 under various noise environments and shows better performances compared with conventional SNR Recovery methods.

Keywords: G.729A, Masking effect, SNR recovery, Near-end background noise, Speech reinforcement

ASK subject classification: Speech Signal Processing (2)

I. 서론

일반적으로 배경 잡음은 음성 품질에 상당한 영향을

미치게 되며, 배경 잡음이 존재하는 상황에서 휴대폰으로 상대방과 통화를 하게 되면 배경 잡음이 음성의 명료도를 감소시키고, 음질을 저하시키게 때문에 상대방의 음성을 듣는데 있어서 상당한 지장을 초래한다. 이렇게 배경 잡음이 음성의 명료도 및 음질을 저하시키는 가장

큰 이유는 마스킹 효과에 의해 설명될 수 있다.

마스킹 효과란 주변을 에워싼 배경 잡음으로 소리의 역치 (threshold)가 변하는 현상으로, 시간 또는 주파수 영역에서 특정 강한 신호 (Masker)가 존재할 때 함께 존재하는 약한 신호 (Maskee)는 마스킹 효과에 의해 거의 들을 수 없게 되며, 강한 신호인 Masker에 의해 약한 신호인 Maskee가 완전히 마스킹 되지 않더라도, 부분 마스킹 효과 (Partial masking effect)에 의해 두 신호가 동시에 존재하면 서로 부분적으로 마스킹하게 되므로 원래 소리의 크기는 감소되어 작게 들리게 된다 [1]. 따라서 배경 잡음이 증가함에 따라 화자가 듣게 되는 음성의 명료도 또한 급격하게 감소하게 된다.

배경 잡음을 고려한 음성 처리 기법은 다양하며, 특히 원단과 근단 환경에 따라 분류될 수 있다. 먼저 원단 화자 환경에서 배경 잡음을 고려한 음성 처리 기법은 잡음 제거 및 음성 향상 기법으로 나눌 수 있다 [2][3]. 잡음 제거 및 음성 향상 기법은 원단의 입력 마이크로부터 들어온 잡음이 섞인 오염된 음성 신호로부터 통계적으로 추정된 잡음 자체를 제거하거나 혹은 잡음이 제거된 깨끗한 음성 신호만을 추정하여 근단으로 전송하는 기법이다. 그러나 원단의 배경 잡음 환경을 고려한 음성 처리 기법은 근단의 다양한 배경 잡음 환경을 고려하지 못하는 한계를 가지게 된다. 만약 근단 화자가 조용한 환경에서 원단으로부터 전송되어온 잡음이 제거된 음성 신호를 듣게 된다면 음성의 명료성이 크게 저하되지 않지만, 근단 화자를 에워싼 주변 환경이 배경 잡음 환경이라면 근단의 배경 잡음은 근단 화자의 귀에 직접적으로 영향을 미치게 되고, 이러한 근단 배경 잡음으로 인해 원단의 잡음이 제거된 깨끗한 음성의 명료도는 저하되게 된다.

이렇듯 배경 잡음이 근단 화자에 직접적으로 영향을 미치는 주변의 다양한 배경 잡음 환경을 고려하여, 원단으로부터 전송되어 온 음성 신호를 근단 화자를 에워싼 주변의 배경 잡음에 따라 증폭시키는 음성 강화기법이 제시되었다. 가장 대표적으로 부분 마스킹 효과에 의해 원단의 음성 신호를 강화하는 기법 (Speech reinforcement approaches)과 근단의 배경 잡음의 추정을 통하여, 주파수별 신호 대 잡음 비 (Signal-to-Noise Ratio, SNR)가 일정하게 유지되도록 전송 되어온 음성의 증폭을 통하여 음성을 강화하는 기법 (Near-end listening enhancement)으로 나눌 수 있다 [4-6]. 그러나 이러한 근단의 다양한 배경 잡음을 고려하여 원단으로부터 전송된 음성 신호를 증폭시키는 음성 강화 기법의 경우, 먼저 원단의 부호화에 의해 전송된 비트스트림 (bitstream)을 근단의 복호

화에 의해 복호화 되어야 하고, 아날로그 신호로 복원된 음성 신호는 추가적으로 기존의 음성 강화 기법에 의해 처리되기 위해서 고속 푸리에 변환 (Fast Fourier Transformation) 및 내부의 다양한 음성 강화 처리 블록을 거쳐야만 한다. 그림 1에는 기존의 제시된 음성 강화 기법들의 전체적인 블록도를 보여준다. 이러한 주파수 변환에 의한 고속 푸리에 변환 (FFT) 방법은 계산량의 증대와 함께, 내부적으로 많은 음성 처리 블록을 거쳐 처리되기 때문에 복잡성이 증가하는 단점을 가지게 되고, 추가적으로 정교한 고속 푸리에 변환 회로를 필요로 하게 되어 실시간 디지털 음성 처리 (Real-time digital speech signal processing)에 있어서 많은 문제점을 가지게 된다.

따라서 본 논문에서는, 이러한 복잡한 음성 신호 처리의 내부 처리 블록 및 고속 푸리에 변환 (FFT)을 거치지 않고, 다양한 배경 잡음이 존재하는 근단 환경에서 근단 화자에 전송된 원단 음성 신호를 추정된 근단 배경 잡음에 기반하여, 근단의 음성부호화기의 복호화 단계에서 직접적으로 강화하는 효과적인 음성 강화 기법을 제시한다. 그림 2은 제안된 G.729A 음성부호화기 기반의 음성 강화 시스템을 보여준다. 구체적으로, 제안된 기법은 광대역 VoIP 서비스에 가장 널리 사용되고 있으며, 최근의 와이브로 (WiBro)와 같은 휴대 무선 인터넷 환경의 Mobile

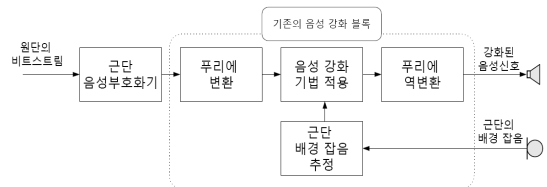


그림 1. 기존의 음성 강화 기법들의 전체 블록도 [4-6]
Fig. 1. Block diagram of the conventional speechreinforcement techniques [4-6].

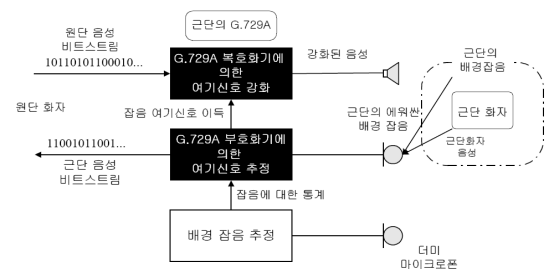


그림 2. 근단 화자를 에워싼 다양한 배경 잡음 환경에서 제안된 근단의 G.729A 음성부호화기 기반의 음성 강화 시스템 개략도
Fig. 2. Block diagram of the proposed speech reinforcement system based on near-end G.729A speech codec for near-end listener with ambient noise under various background noise environments.

VoIP망에서 사용될 ITU-T 표준 음성부호화기인 CELP 계열의 G.729A 음성부호화기를 사용하였다. G.729A 음성부호화기 기반으로 원단 음성 신호를 다양한 배경 잡음이 존재하는 근단 환경에서 근단 화자가 듣는 원단 음성 신호의 명료성을 강화를 통해, 기존에 제시된 음성 강화 기법에서의 고속 푸리에 변환 (FFT) 없이 계산 량 감소의 이점과 보다 효율적으로 음성 신호를 강화시키는 기법을 제시한다 [7-9].

제안된 알고리즘은 성능은 다양한 배경 잡음이 존재하는 환경에서 ITU-T P.800의 표준 음질 측정 방법인 CCR (Comparison Category Rating) 테스트로 실험을 진행하였으며, 기존의 G.729A 음성부호화기에 의해 복호화된 음성 신호 및 기존에 시간 축에서 제시된 SNR 복구 기법과 비교, 테스트 결과 본 논문에서 제시된 기법에 의해 강화된 음성 신호의 명료성 및 음질이 매우 향상된 결과를 나타내었다. 특히 기존의 Sauert와 Vary의 고속 푸리에 변환을 통한 주파수축 기반의 SNR 복구 기법과 동등한 음성 품질을 보여주었다 [4].

II. G.729A 음성부호화기 개요

G.729A는 분석/합성 방법 (Analysis-by-synthesis: AbS)의 CELP 구조를 이용하며, 8 kHz로 샘플링된 16비트 PCM 입력 음성 신호를 10 ms의 프레임과 각각 두 개의 5 ms 부프레임으로 나누어 부호화한다 [7-9]. 10 ms 프레임마다 입력 음성 신호로부터 CELP 모델 파라미터가 추출되어 분석되며, 주요 파라미터로는 선형 예측 (linear Prediction) 필터 계수, 적응 코드북 인덱스 (adaptive-codebook index), 고정 코드북 인덱스 (fixed-codebook index), 그리고 각각의 이득 (gains)으로 구성된다. 선형 예측 계수 (LPC)의 추출을 위해, 30 ms 비대칭 윈도우를 사용하게 되는데, 이로 인해 5 ms의 예견 구간이 필요하기 때문에 전체적으로 15 ms의 알고리즘 지연시간을 갖는다.

G.729A는 전체적으로 크게 LPC (linear predictive coefficient) 분석과 양자화 (vector quantization) 부분, 피치 (Pitch) 분석 부분, 적응 코드북 (adaptive-codebook) 과 고정코드북 (fixed-codebook) 분석 부분으로 나뉘어진다. 여기신호 (excitation signal)는 적응 코드북 벡터 (adaptive codebook vector)와 고정 코드북 벡터 (fixed codebook vector)의 합으로 구성되며, 사람의 성도특성

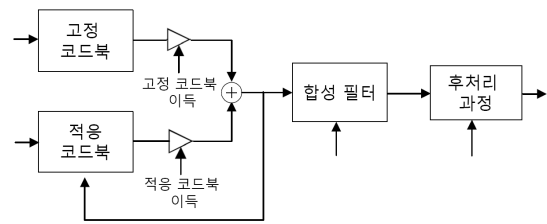


그림 3. G.729A 음성부호화기의 복호화 과정
Fig. 3. Decoding principle of the G.729A decoder.

을 모델링하는 10차의 All-Pole 필터를 통과한 출력으로 합성된다. 또한 여기신호는 매 5 ms (40샘플)의 부프레임마다 추출되어진다.

복호화기는 원단으로부터 전송된 비트스트림 (bit-stream)을 입력으로 받아서 음성 신호를 재생하게 된다. 비트스트림으로부터 추출되는 파라미터는 선형 예측 계수, 피치 정보, 여기신호 및 각각에 대한 이득 값들이다. 그림 2은 복호화기로 전송된 비트 스트림으로부터 음성을 재생하는 원리를 개략적으로 보여준다. 그림 3에서 보는 바와 같이, 보간된 선형 예측 계수가 매 부프레임마다 선형 예측 필터의 계수로 사용되며, 매 5 ms의 부프레임마다 여기신호가 각각의 고정 코드북과 적응 코드북에 이득이 곱해진 후에 합으로 구해진다. 구해진 여기신호는 선형 예측 필터에 의해 음성 신호로 재생되어지고, 복원된 음성 신호는 음질의 개선을 위해 후처리 (post-processing)과정을 거치게 되며, 마지막으로 고주파필터와 스케일링 과정을 거친 후에 최종 음성 신호로써 근단 화자에게로 전달된다.

III. G.729A 음성 부호화 기반의 제안된 음성 강화 기법

G.729A 음성부호화기 기반의 효과적인 음성 강화 기법은 원단으로부터 전송된 비트스트림에서 근단의 G.729A 복호화기에 의해 추출된 원단 음성 신호에 대한 여기신호 (excitation signal)를 근단으로부터 추정된 배경 잡음을 기반으로 증폭함으로써, 다양한 배경 잡음이 존재하는 근단 환경에서 명료도가 떨어진 원단 음성 신호의 명료성을 강화시키는 것이다. 구체적으로, 음성부호화기의 여기신호에서 적응 코드북 이득은 주기적인 성분을 나타내는 피치 (pitch) 성분이고, 고정 코드북 이득은 실제 음성의 에너지 (Energy)이므로, 근단의 배경 잡음을 통해 마스크 되는 부분은 잡음 신호의 에너지에 해당되기 때문

에, 본 논문에서는 에너지에 해당하는 고정 코드북 이득 값의 비를 구하여 원단 음성의 여기신호중 에너지 부분을 증폭시키는 기법을 제시한다. 그림 4에는 제안된 G.729A 음성부호화기 기반의 음성 강화 시스템의 전체 블록 도를 보여준다. 근단의 배경 잡음은 통신시스템의 전송부, 더미 마이크, 또는 추가적인 다른 장비로부터 추정할 수 있다고 가정한다. 먼저 근단의 G.729A 음성부호화기로부터 원단의 비트스트림에 대한 원단 음성 신호의 여기신호와 배경 잡음에 대한 여기신호를 구한다 [5]. 배경 잡음으로부터 구해진 여기신호와 원단의 비트스트림으로부터 구해진 원단 음성에 대한 여기신호의 연산을 통해 새롭게 여기신호 이득 비 (excitation gain ratio)를 구하게 된다. 구해진 여기신호 이득 비를 기반으로 원단 음성 신호의 여기신호에 여기신호 이득 비를 곱하여 증폭시킴으로써, 다양한 배경 잡음의 근단 화자 환경에서 근단 화자에 위산 배경 잡음으로 인해 음질이 저하된 원단의 음성 신호의 명료도를 근단의 G.729A 음성부호화기의 복호화 단계에서 강화하게 된다.

구체적으로, 원단으로부터 전송된 비트스트림으로부터 선형 예측 계수, 적응 코드북 벡터, 고정 코드북 벡터 및 적응 코드북 이득, 고정 코드북 이득을 추출한 후, 각 부프레임에 대한 여기신호 $u(n)$ 을 다음과 같이 정의한다.

$$u(n) = \hat{g}_p(n)v(n) + \hat{g}_c(n)c(n) \quad (1)$$

여기신호 $u(n)$ 은 적응 코드북과 고정 코드북의 합으로 구성되기 때문에, 식 (1)에서 $v(n)$ 은 적응 코드북 벡터 값을 나타내고, $c(n)$ 은 고정 코드북 값을 나타낸다. 또한 \hat{g}_p 와 \hat{g}_c 는 각각 적응 코드북 이득 값과 고정 코드북 이득 값을 나타낸다. 제안된 G.729A 음성부호화기 기반의 음성 강화 기법은 원단으로부터 추출된 음성의 여기신호를 근단의 G.729A 음성부호화기로부터 추정된 배경 잡음의 여기신호와 연산을 통해 구해진 새로운 여기신호 이득 비를 곱하여 증폭시키는 것이기 때문에, 원단의 비트스

트림으로부터 추출된 여기신호는 새롭게 정의된 여기신호 이득 값 $G_R(n)$ 에 의해 다음과 같이 강화된다.

$$\hat{u}_R(n) = \hat{g}_p(n)v(n) + G_R(n)c(n) \quad (2)$$

여기서, $\hat{u}_R(n)$ 은 여기신호 이득값 $G_R(n)$ 에 의해 강화된 근단의 여기신호를 나타낸다. 여기신호는 각 부프레임마다 구해지기 때문에, 식 (2)의 n 은 전체 80샘플에 대한 각각의 부프레임에 대한 성분을 나타낸다. 따라서 각각의 부프레임에 대해, n 값은 0부터 39를 값을 가지게 된다.

또한 근단의 배경 잡음에 대한 여기신호를 구하기 위하여, 그림 4에서의 음성 강화 블록도로부터 이미 가정한 바와 같이 통신시스템의 전송부 혹은 더미 마이크로로부터 구해진 배경 잡음 신호로부터 다음과 같이 근단의 배경 잡음에 대한 여기신호를 나타낼 수 있다.

$$\hat{u}_N(n) = \hat{g}_N(n)v_N(n) + G_N(n)c_N(n) \quad (3)$$

식 (3)에서 $\hat{u}_N(n)$ 은 근단에 존재하는 배경잡음에 대한 여기신호를 나타내고, $G_N(n)$ 은 추정된 배경잡음 신호에 대한 여기신호의 고정 코드북 이득 값을 의미한다.

강화된 근단 음성 신호의 여기신호 이득 값과 근단 배경 잡음 여기신호에 대한 이득 값의 비가 $\xi = 15dB$ 보다 같거나 크게 유지하도록 다음과 같이 나타낼 수 있다 [4].

$$\frac{G_R(n)}{G_N(n)} \geq \xi \quad (4)$$

식 (4)에서 강화된 근단 음성 신호의 여기신호 이득 값 $G_R(n)$ 은 원단에서 전송된 비트스트림으로부터 추출된 여기신호 이득 값 \hat{g}_c 에 근단으로부터 추정된 배경 잡음의 여기신호 이득 값 $G_N(n)$ 과 연산을 통해 구해진 여기신호 이득 비 (excitation gain ratio) $g_r(n)$ 을 곱해줌으로써 증폭되기 때문에, 식 (4)은 다음과 같이 표현될 수 있다.

$$\frac{G_R(n)}{G_N(n)} = \frac{g_r(n) \cdot \hat{g}_c(n)}{G_N(n)} \geq \xi \quad (5)$$

또한, 식 (5)을 여기신호 이득 비 $g_r(n)$ 에 의한 식으로 표현하면 다음과 같다.

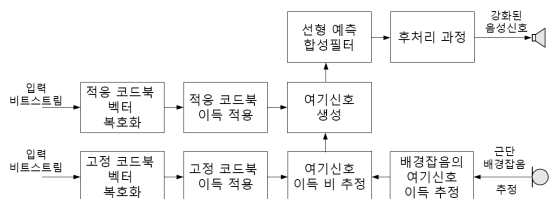


그림 4. 제안된 G.729A 음성부호화기 기반 음성 강화 전체 블록도
Fig. 4. Overall block diagram of the proposed speech reinforcement approach based on the G.729A codec.

$$g_r(n) \geq \xi \cdot \frac{G_N(n)}{\hat{g}_c(n)} \quad (6)$$

원단으로부터 전송된 음성 신호는 근단 배경 잡음 환경에서 약화되지 않아야 하기 때문에, 다음과 같은 최소 여기 신호 이득 비 $g_r(n)$ 에 대한 경계 조건을 도입할 수 있다.

$$g_r(n) \geq 1 \quad (7)$$

그러므로, 원단의 전송된 음성 신호의 여기신호를 강화하기 위한 여기신호 이득 비 $g_r(n)$ 은 (6)식과 (7)식의 경계 조건을 결합하면 다음과 같이 표현될 수 있다.

$$g_r(n) = \max\left\{\xi \cdot \frac{G_N(n)}{\hat{g}_c(n)}, 1\right\} \quad (8)$$

원단의 음성 신호에 대한 여기신호가 여기신호 이득 비 $g_r(n)$ 에 의해 과도하게 증폭되지 않도록 여기신호 이득 비 $g_r(n)$ 에 대한 최대값은 $g_{r,max} \cong 30 \text{ dB}$ 로 제한한다 [4]. 따라서 (8)식과 $g_r(n)$ 에 대한 최대값을 결합하면 최종적으로 원단 음성 신호의 여기신호를 강화할 수 있는 여기신호 이득 비 $g_r(n)$ 을 다음과 같이 구할 수 있다.

$$g_r(n) = \min\left\{\max\left\{\xi \cdot \frac{G_N(n)}{\hat{g}_c(n)}, 1\right\}, g_{r,max}\right\} \quad (9)$$

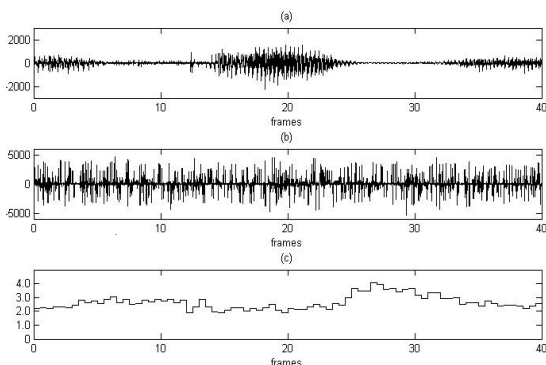


그림 5. 여기 신호의 비교. (a) 원단 음성 신호에 대한 여기 신호 (b) 추정된 배경 잡음에 대한 여기신호 (c) 계산된 여기 신호 이득 비.

Fig. 5. The comparison of the excitation signal. (a) The estimated excitation signal of the far-end speech (b) The estimated excitation signal of the near-end ambient white noise (c) The computed excitation gain ratio between the near-end ambient noise's excitation and the far-end speech's excitation signal

그림 5에는 40 프레임에 대해서 원단 음성 신호에 대한 여기신호 및 근단의 배경 잡음에 대한 여기신호, 여기신호 이득 비를 나타내었다.

IV. 실험 방법

본 논문에서는 제안된 G.729A 음성부호화기 기반의 음성 강화 알고리즘의 성능을 평가하기 위해서 ITU-T P.800의 주관적 음질 선호도 측정 방법 CCR (Comparison Category Rating) 테스트로 실험을 실시하였다 [10]. 음성 신호에 대한 여기신호를 추출하기 위하여 남성 화자 6명, 여성 6명이 발음한 음성을 8초 길이로 구성한 후, 8 kHz로 샘플링을 하였다. 각 음성 파일은 두 개의 문장으로 구성되는데, G.729A 음성부호화기를 이용해 부호화하여 음성 샘플에 대한 비트스트림을 수집하였다. 제안된 음성 강화 기법의 성능을 테스트하기 위해, 수집된 비트 스트림을 음성 강화 알고리즘이 적용되지 않은 G.729A 음성부호화의 복호화기를 통하여 원단 음성 신호를 복원하였다. 또한, 제안된 알고리즘이 적용된 G.729A 음성부호화기의 복호화기를 기반으로 수집된 비트스트림에 적용하여 강화된 원단 음성 신호로 데이터베이스를 만들어 비교 테스트를 진행하였다. 근단의 다양한 배경 잡음 환경을 구축하기 위해, 근단의 배경 잡음으로써 NOISEX-92 데이터베이스 white gaussian noise (WGN), babble와 vehicle을 잡음 레벨이 SNR 5 dB, 10 dB, 15 dB, 20 dB로 원단의 음성 신호에 섞이도록 하였으며, 헤드폰을 통해 들을 수 있게 실험을 진행하였다 [5]. 또한, 배경 잡음에 대한 여기신호를 구하기 위해, 음성검출기를 통해 식별된 비음성 구간에서의 배경 잡음 white, babble, vehicle 신호를 음성부호화기의 입력 신호로 통과시켜 각각의 배경 잡음에 대한 여기신호 중 고정 코드북 이득 값을 추출하였다. 이렇게 추출된 배경 잡음의 여기신호에 대한 고정 코드북 이득 값을 원단 음성 신호의 여기신호에서 구해진 고정 코드북 값의 연산을 통해 여기신호 이득 비를 구하는데 사용하였다.

알고리즘의 성능을 테스트하기 위해, 주관적 음질 선호도 측정 방법인 CCR 테스트에서 청자들은 참고용 파일로써 잡음이 섞이지 않은 깨끗한 음성 신호를 듣게 된다. 그리고 테스트에 참가한 청자들은 2개의 음성 파일을 듣게 되는데, 하나는 제안된 알고리즘이 적용된 음성 파일이고, 다른 하나는 비교 대상이 되는 음성 파일이다. 참고용 음성 파일을 제외한, 비교 대상이 되는 2개의 음성 파

일은 순서 없이 재생되며, 청자들은 연속적으로 들은 음성 파일 중에서 음질이 좋은 파일을 선택하여 점수를 매기게 된다. 참고용 음성 파일을 기준으로 2개의 파일 중 어떤 음성 신호가 얼마나 음질이 좋은지, 배경 잡음에 대해 음성 신호가 얼마나 명료하게 들리는지, 얼마나 크게 잘 들리는지, 왜곡은 없는지, 어떤 음성 신호가 더 명료하게 들리는지를 평가하게 된다. 평가 기준으로 -3점부터 +3점까지 점수를 부여하게 되는데, 평가의 세부적인 기준은 Much Better (+3), Better (+2), Slightly Better (+1), About the Same (0), Slightly Worse (-1), Worse (-2), Much Worse (-3)으로 나뉜다. 실험 결과를 통하여 나타난 표의 각각의 결과 점수는 평균값으로써, 표에 나타난 결과의 평균 점수가 0보다 큰 양수 값을 가질수록 제안된 알고리즘이 성능이 더 좋다는 것을 의미하고, 반대로 0보다 작은 음수의 평균값을 가지게 된다면 강화된 음성 신호의 성능이 나쁘다는 것을 의미하게 된다.

첫 번째 실험의 목적은 제안된 G.729A 음성부호화기 기반의 음성 강화 기법이 기존의 방법과의 알고리즘 성능을 비교하기 위한 것으로서, 제안된 G.729A 음성부호화기 기반의 음성 강화 기법에 의해 처리된 음성 신호와 어떠한 음성 처리 기법도 적용되지 않은 음성 신호를 비교하였다. 어떠한 음성 처리 기법도 적용되지 않은 음성 파일은 기존의 G.729A 음성부호화기에 의해 부호화 및 복호화된 음성 파일이다. 표 3에는 white, babble, vehicle noise에 따라 CCR 테스트의 결과를 나타내었다. 표 1의 결과에 따르면 white, babble, vehicle의 잡음에 대해서 제안된 G.729A 음성부호화기 기반의 음성 강화 기법에 의한 음성 신호가 기존의 G.729A 음성부호화기에 의해 부호화 및 복호화된 음성 신호와 비교해서 평균 점수가 0보다 큰 점수를 가짐을 볼 수 있다. 표 1에 나타난 평균 점수를 비교하면 평균 95% 신뢰 구간 (± 0.12)에서

white 잡음의 평균 점수는 1.910, babble 잡음의 경우 평균 95% 신뢰구간 (± 0.11)에서 평균 1.988, vehicle 잡음의 경우 평균 95% 신뢰 구간 (± 0.11)에서 평균 점수 1.960을 나타냄으로써, CCR 테스트 평가 기준 Better (2)에 매우 근접한 점수를 보여주었으며, 이를 통해 제안된 G.729A 음성 부호화기 기반의 음성 강화 알고리즘이 우수하다는 것을 보여주었다.

두 번째 실험의 목적은 제안된 G.729A 음성부호화기 기반의 음성 강화 알고리즘과 기존의 Sauert와 Vary에 의해 제안된 시간 축에서의 SNR 복구 기법 (Frequency Independent SNR Recovery)을 비교하기 위한 것이다. 표 2에는 제안된 G.729A 음성부호화기 기반의 음성 강화 기법에 의해 처리된 음성 신호와 시간 축에서의 SNR 복구 기법에 의해 처리된 음성 신호를 비교한 결과이다. 표 2의 결과에 의하면 white, babble, vehicle 잡음에 대해서 제안된 G.729A 음성부호화기 기반의 음성 강화 기법에 의한 음성 신호가 white 잡음의 경우 평균 95% 신뢰 구간 (± 0.14)에서 평균 점수 1.469, babble 잡음에서 평균 95% 신뢰 구간 (± 0.16)에 대해 평균 점수 1.690이었으며, vehicle 잡음을 살펴보면 평균 95% 신뢰 구간 (± 0.16)에 평균 점수 1.803을 가짐을 볼 수 있다. white, babble, vehicle 잡음에서 1보다 큰 양수 값을 가짐을 볼 수 있는데, 이를 통해 제안된 G.729A 음성부호화기 기반의 음성 강화 알고리즘이 성능이 더 우수함을 알 수 있다. 두 번째 실험에서 주목할 점은, 기존의 시간 축에서 제안된 SNR 복구 기법의 경우 저주파 대역에 내재된 잡음이 음성 신호와 함께 증폭되어 오히려 음질의 저하를 초래하였다 [4]. 표 2의 결과에 의하면 낮은 SNR에서 오히려 점수가 감소됨을 볼 수 있는데, 이것은 시간 축에서의 SNR 복구 기법에 의해 증폭된 저주파 대역의 잡음이 white, babble, vehicle 잡음에 의해 심하게 마스킹 되기 때문에 실험에

표 1. 다양한 배경 잡음 환경에서 제안된 음성 강화 알고리즘 대비 음성 처리 알고리즘이 적용되지 않은 경우에 대한 CCR 테스트 결과 (95% 신뢰 구간)

Table 1. The CCR results for the proposed reinforced algorithm with respect to the unprocessed under various background noise (With 95% confidence interval).

noise	white	babble	vehicle
5 dB	1.750±0.13	2.038±0.11	2.025±0.10
10 dB	2.000±0.11	2.050±0.09	1.938±0.12
15 dB	1.963±0.12	1.913±0.11	1.975±0.11
20 dB	1.925±0.12	1.950±0.11	1.900±0.10
average	1.910±0.12	1.988±0.11	1.960±0.11

표 2. 다양한 배경 잡음 환경에서 제안된 음성 강화 기법 대비 시간 축에서 제안된 SNR 복구 기법 (Frequency Independent SNR Recovery)에 대한 CCR 테스트 결과 (95% 신뢰 구간) [4]

Table 2. The CCR results for the proposed reinforced algorithm with respect to the Frequency Independent SNR Recovery algorithm under various background noise (With 95% confidence interval).

noise	white	babble	vehicle
5 dB	1.225±0.12	1.575±0.16	1.763±0.17
10 dB	1.325±0.13	1.700±0.17	1.788±0.16
15 dB	1.600±0.15	1.775±0.16	1.825±0.16
20 dB	1.725±0.16	1.713±0.16	1.838±0.16
average	1.469±0.14	1.690±0.16	1.803±0.16

표 3. 다양한 배경 잡음 환경에서 제안된 음성 강화 기법과 주파수 축에서 제안된 SNR 복구 기법에 대한 CCR 테스트 결과 음성 처리 알고리즘 적용되지 않은 경우 대비 제안된 음성 강화 알고리즘에 대한 CCR 테스트 결과 (95% 신뢰 구간)

Table 3. The CCR results for the proposed reinforced algorithm with respect to the Frequency dependent SNR Recovery algorithm under various background noise (With 95% confidence interval).

noise	white	babble	vehicle
5 dB	0.238±0.13	-0.075±0.15	-0.113±0.11
10 dB	0.375±0.12	0.150±0.14	0.025±0.12
15 dB	0.175±0.13	-0.063±0.14	0.038±0.14
20 dB	0.200±0.13	0.013±0.13	-0.013±0.13
average	0.247±0.13	0.006±0.14	-0.016±0.13

참가한 청자가 높은 SNR 잡음에 대해서는 인식을 하지만, 낮은 SNR 잡음에서는 쉽게 인식하지 못한 결과로 볼 수 있다. 이 결과로부터, 제안된 G.729A 음성부호화기 기반의 음성 강화 기법이 다양한 배경 잡음 환경에서 기존의 시간 축에서 제시된 SNR 복구 기법보다 저주파 대역의 잡음의 증폭 없이 음질을 향상시킨다고 결론지을 수 있다.

마지막으로 세 번째 실험의 목적은 제안된 G.729A 음성부호화기 기반의 음성 강화 기법과 고속 푸리에 변환 (FFT)을 한 후, 근단의 배경 잡음의 추정을 통하여 주파수 밴드별로 신호 대 잡음비가 일정하게 유지되도록 전송되어온 음성을 증폭시키는 주파수축 기반의 SNR 복구 기법 (Frequency Dependent SNR Recovery)을 비교하는 것이다. 표 3에는 근단의 다양한 배경 잡음 환경에서 제안된 G.729A 음성부호화기 기반의 음성 강화 기법과 Sauer와 Vary에 의해 제안된 주파수축 기반의 SNR 복구 기법을 비교한 결과이다. 표 3의 결과를 분석해 보면 white, babble, vehicle 잡음에 대해서 제안된 음성 강화 기법에 의한 음성 신호와 주파수축 기반의 SNR 복구 기법에 의한 음성 신호의 경우 CCR 테스트 평가 기준 항목 About the Same인 0점에 근접한 점수를 나타냄을 볼 수 있다. 구체적으로 살펴보면, white 잡음에 대해 평균 95% 신뢰 구간 (± 0.247)에서 평균 점수 0.247, babble 잡음의 경우 평균 95% 신뢰 구간 (± 0.14)에서 평균 점수 0.006, vehicle 잡음 환경에 대해 평균 95% 신뢰 구간 (± 0.13)에서 평균 점수 -0.016을 나타내었다. 평균 점수가 0점에 근접한 점수의 의미는 제안된 G.729A 음성부호화기 기반의 음성 강화 기법에 의한 음성 신호의 음질이 음성 신호를 고속 푸리에 변환 (FFT)을 통하여 주파수 영역으로 변환시켜서 처리하는 주파수축 기반의 SNR 복구 기법과 동등하거나 그 이상의 음질 향상이 이루어진다

표 4. 제안된 알고리즘과 시간축 및 주파수축 기반의 SNR 복구 기법과의 프레임 마다 복잡도 비교.

Table 4. The comparison of computation cost for the proposed algorithm with respect to the frequency independent and dependent SNR Recovery methods (per frame).

	SNR Recovery		Proposed (G.729A Dependent)
	Frequency domain	Time domain	
Computation Cost	28,945	2,880	1,634

는 것을 결론지을 수 있다.

제안된 G.729A 음성부호화기 기반의 음성 강화 기법은 기존의 방법과 비교하여 시간축 및 주파수축 변환 없이 근단의 음성부호화기에서 직접적으로 강화되기 때문에 계산 량 및 복잡도를 상당히 줄일 수 있는 장점을 가진다. 알고리즘의 계산 량 및 복잡도에 대한 객관적인 비교로써, 표 4에는 기존의 방법과 제안된 방법과의 계산 량을 비교한 결과를 보여준다. 계산 량의 비교를 위해서, 알고리즘 수행시 각각의 연산 과정에 가중치를 적용하여, 가중치에 따른 연산의 횟수를 조사하였다. 연산에 대한 가중치로써, 덧셈 및 뺄셈 연산자는 1, 곱셈 연산자는 1, 나눗셈 연산자는 5, 로그 연산자는 10을 각각 적용하였다. 표 4에 나타난 결과에 따르면, 제안된 G.729A 음성부호화기 기반의 음성 강화 기법은 기존의 시간축 및 주파수축 기반의 SNR 복구 기법과 비교하여 계산량이 현저하게 줄일 수 있으면서, 동시에 시간축 SNR 기법보다는 더 좋은 명료도를 나타내는 음성 신호의 품질을 보여주고,

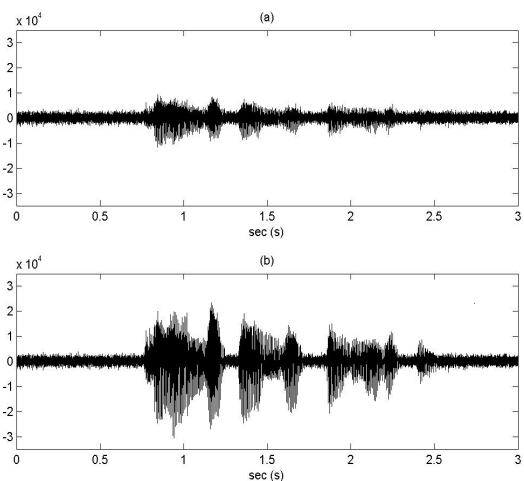


그림 6. 음성신호의 비교 (a) 근단의 배경 잡음에 의해 오염된 원단 음성 신호 (b) 제안된 음성 강화 기법에 의해 명료성이 강화된 음성 신호

Fig. 6. The comparison of the speech signal (a) Noisy far-end speech signal by the near-end ambient noise (b) the reinforced speech signal based on the proposed reinforcement algorithm.

FFT에 의한 주파수축 기반의 SNR 기법과는 동등한 음성 품질을 보여줌을 볼 수 있다.

제안된 음성 강화 알고리즘은 그림 6을 통해서 확인할 수 있다. 그림 6에는 원단의 음성 신호가 근단의 배경 잡음에 의해 마스킹 되었을 때의 경우로써, (a)는 비교 대상이 되는 음성 파일을 나타내고, (b)에는 제안된 알고리즘이 적용된 음성 파일이다. 그림에서 보는 바와 같이, (a)의 경우 배경 잡음이 부가된 음성 신호는 잡음에 의해 심하게 마스킹되지만, (b)의 음성 파일은 제안된 음성 강화 알고리즘이 적용되어 근단 배경 잡음 환경에서 명료성이 향상된 것을 알 수 있다.

V. 결론

본 논문에서는 근단 배경 잡음이 존재하는 환경 하에서 고속 푸리에 변환 (FFT) 변환을 거치지 않고, 근단의 G.729A 음성부호화기를 이용해 추정된 근단 화자의 배경 잡음에 대한 여기신호를 기반으로 근단 배경 잡음으로 인해 명료도가 저하된 원단 음성 신호의 여기신호를 강화하는 효과적인 알고리즘을 제안하였다. 근단의 G.729A 음성부호화기로 근단에 존재하는 배경 잡음의 여기신호를 추정하였고, 추정된 여기신호를 기반으로 근단의 G.729A 복호화기에 의해 추출된 원단 음성 신호의 비트 스트림에 대한 여기신호와의 연산을 통해 여기신호 이득 비를 새롭게 정의하였다. 이를 바탕으로 배경 잡음으로 인해 명료성이 저하된 원단 음성 신호의 여기신호를 새롭게 구해진 여기신호 이득 비를 곱하여 증폭시킴으로써, 근단의 배경 잡음 환경이 근단 화자의 귀에 직접적으로 영향을 미치는 환경에서 근단 화자가 듣게 되는 원단 음성 신호의 명료도를 향상시킬 수 있었다. 실험 결과는 제안된 G.729A 기반의 음성 강화 알고리즘이 기존의 음성 부호화기에 의해 복원된 음성 신호 및 시간 축에서 제안된 Sauert와 Vary의 SNR 복구 기법보다 음성의 명료도가 향상된 결과를 보여주었다. 또한 기존의 고속 푸리에 변환 (FFT)에 의한 주파수 영역에서 제안된 주파수축 기반의 SNR 복구 기법과 비교해 동등한 음성 품질을 보여주었으며, 이러한 결과를 통해 고속 푸리에 변환 (FFT) 과정을 거치지 않으므로써 제안된 기법이 복잡성 및 계산량 증대 없이 보다 효율적으로 음성 신호를 강화시킴을 보여준다.

감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업 [2008-F-045-01]과 또한 본 연구는 지식경제부와 한국산업기술재단의 전략기술인력양성사업으로 수행된 연구결과임.

참고 문헌

1. E. M. Danaher and J. M. Pickett, "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss," *Journal of Speech and Hearing Research*, vol. 18, pp. 261-271, 1975.
2. 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.
3. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.
4. B. Sauert and P. Vary, "Near end listening enhancement: Speech intelligibility improvement in noisy environments," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. 1-493-1-496, 2006.
5. J. Sohn, N. S. Kim, "Perceptual reinforcement of speech signal based on partial specific loudness," *IEEE Signal Processing Letters*, to appear
6. 최재훈, 장준혁, "원단 잡음 환경에서 Soft Decision에 기반한 새로운 음성 강화 기법," *한국음향학회지*, 27권, 7호, 379-385쪽, 2008.
7. International Telecommunication Union, *ITU-T G.729A: Coding of speech at 8 kbit/s using conjugate structure algebraic code excited linear prediction (CS-ACELP)*, Mar. 1996.
8. R. Salami, C. Laflamme, B. Bessette, and J. P. Adoul, "ITU-T G.729 Annex A: reduced complexity 8kb/s CS-ACELP codec for digital simultaneous voice and data," *IEEE Communication Magazine*, vol. 35, pp. 56-63, Sep. 1997.
9. R. Salami, C. Laflamme, J-P. Adoul and D. Massaloux, "A toll quality 8kb/s speech codec for the personal communication system (PCS)," *Proc. IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 808-816, 1994.
10. ITU-T P. 800, *Methods for subjective Determination of Transmission Quality*, Aug. 1996.

저자 약력

• 최 재 훈 (Jae-Hun Choi)



2007년 2월: 인하대학교 전자공학과 학사
2007년 1월~2008년 2월: 삼성전자 정보통신 총괄
연구원
2008년 3월~현재: 인하대학교 전자공학부 석사과정

• 장 준 혁 (Joon-Hyuk Chang)



1998년 2월: 경북대학교 전자공학과 학사
2000년 2월: 서울대학교 전기공학부 석사
2004년 2월: 서울대학교 전기컴퓨터공학부 박사
2000년 3월~2005년 4월: 쥘넷더스 연구소장
2004년 5월~2005년 4월: 캘리포니아 주립대학,
산타바버라 (UCSB) 박사후연구원
2005년 5월~2005년 8월: 한국과학기술연구원 (KIST)
연구원
2005년 9월~현재: 인하대학교 전자공학부 조교수