

Push-to-talk 통신을 위한 진폭 및 위상 복원 기반의 단일 채널 음성 향상 방식

A single-channel speech enhancement method based on restoration of both spectral amplitudes and phases for push-to-talk communication

조혜승,¹ 김형국^{1†}

(Hye-Seung Cho¹ and Hyoung-Gook Kim^{1†})

¹광운대학교 전자공학과

(Received August 24, 2016; accepted January 25, 2017)

초 록: 본 논문에서는 PTT(Push-To-Talk) 기반의 무선 통신을 위한 진폭 및 위상 복원 기반의 단일 채널 음성 향상 방식을 제안한다. 제안한 방식은 신호의 진폭만을 대상으로 음성 향상을 진행했던 기존의 방식들과 달리, 음성 신호의 진폭과 위상을 분리하여 각각 향상시켜 다시 결합함으로써 더욱 양질의 음성을 제공한다. 본 논문에서 제안하는 방식의 성능을 평가하기 위해 동적 잡음 환경에서의 단계별 비교 실험을 실시하였으며, 실험 결과를 통해 제안한 방식이 다양한 잡음 환경에서 양질의 음성을 제공하는 것을 확인할 수 있다.

핵심용어: PTT(Push-To-Talk), Log-spectral 진폭 추정, 위상 분해

ABSTRACT: In this paper, we propose a single-channel speech enhancement method based on restoration of both spectral amplitudes and phases for PTT (Push-To-Talk) communication. The proposed method combines the spectral amplitude and phase enhancement to provide high-quality speech unlike other single-channel speech enhancement methods which only use spectral amplitudes. We carried out side-by-side comparison experiment in various non-stationary noise environments in order to evaluate the performance of the proposed method. The experimental results show that the proposed method provides high quality speech better than other methods under different noise conditions.

Keywords: PTT(Push-To-Talk), Log-spectral amplitude estimation, Phase decomposition

PACS numbers: 43.72.Dv, 43.72.Kb

1. 서 론

PTT(Push-To-Talk)는 인스턴트 음성 메시지를 전달하는 반이중통신 방식으로 개인 또는 그룹의 사용자가 유용하게 사용할 수 있는 통신 서비스이다. PTT는 별도의 네트워크 구성 없이 현존하는 통신 네트워크를 기반으로 이용 가능하며 일반 통화와는 다

르게 대기 시간이 없기 때문에 최근에는 위급 상황에서의 실시간 통신을 대상으로 PTT에 대한 가치가 재정의 되고 있다. 실제의 위급상황에서는 다양한 배경 잡음이 동시 다발적으로 발생하므로 이러한 상황에서 PTT를 사용할 경우, 강인한 음성 통신을 보장하기 위해 음성 향상 방식이 필요하다.

기존의 단일 채널 음성 향상 방식들은 입력되는 오염된 신호에 STFT(Short-Time Fourier Transform)을 적용하여 시간 축의 신호를 시간-주파수축의 신호로 변환하여 처리하였다. Wiener 이득 값 적용방식,

†Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)
Department of Radio Science and Engineering, Kwangwoon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea

(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

Bayesian STSA(Short-Time Spectral Amplitude) estimators^[1] 등으로 대변되는 지난 수십 년간의 대표적인 단일 채널 음성 향상 방식들은 위상을 배제한 채 진폭만을 대상으로 수행되었다. 이러한 방식들은 정적 잡음에는 성공적으로 동작하는 반면 동적 잡음에는 약한 특성을 나타내는데, 실생활의 잡음 환경에는 정적 잡음과 동적 잡음이 혼재하므로 이러한 방식들로 실제 환경의 잡음을 제거하는데 충분하지 않다는 한계가 있다.

이와 함께 최근에는 위상 추정에 대한 연구^[2]를 통해 음성 신호에 대한 위상 추정이 음성 향상에서 중요한 역할을 한다는 사실이 밝혀졌다. 하지만 아직까지 단일 채널 음성 향상 방식에서는 진폭 추정과 위상 추정이 각각 독립적으로 적용되어 오고 있으며 이러한 방식들의 독립적인 적용으로는 다양한 잡음 환경에서 오염된 음성 신호에 대한 향상이 충분히 이루어지지 않는다는 한계가 존재한다.

이에 본 논문에서는 입력신호에 대한 음성 구간 검출을 기반으로 음성 존재 불확실성 계수를 반영한 적응적 Log-spectral 진폭 추정 방식(Adaptive Log-Spectral Amplitude estimation: ALSA) 과 위상 분해를 이용한 위상 추정방식을 함께 사용하여 낮은 연산량을 갖는 효과적인 단일 채널 음성 향상 방식을 제안한다.

II. 적응적 단일 채널 음성 향상 방식

Fig. 1은 본 논문에서 제안하는 단일 채널 음성 향상 방식의 전체 알고리즘을 나타낸다. 제안하는 알고리즘은 STFT 적용부, ALSA 기반의 진폭 추정부, 위상 분해를 이용한 위상 추정부, ISTFT(Inverse Short-Time Fourier Transform) 적용부로 구성된다.

제안한 방식에서, PTT 동작 여부에 관계없이 마이크는 언제나 켜져 있는 상태이다. PTT가 동작하지 않는 상태에서는 사용자의 주변 환경잡음이 선형적 잡음 신호로서 입력되며 PTT가 동작할 때는 잡음에 오염된 음성신호에 대해 음성 향상이 적용된다.

PTT가 동작할 때 입력되는 오염된 음성신호 $y(t)$ 는 음성신호 $s(t)$ 에 부가적 잡음 $n_a(t)$ 가 더해진 것으로 가정된다. 또한 이때의 $n_a(t)$ 는 앞서 입력된 선형적 잡음신호 $n_p(t)$ 와 유사하다고 가정할 수 있다.

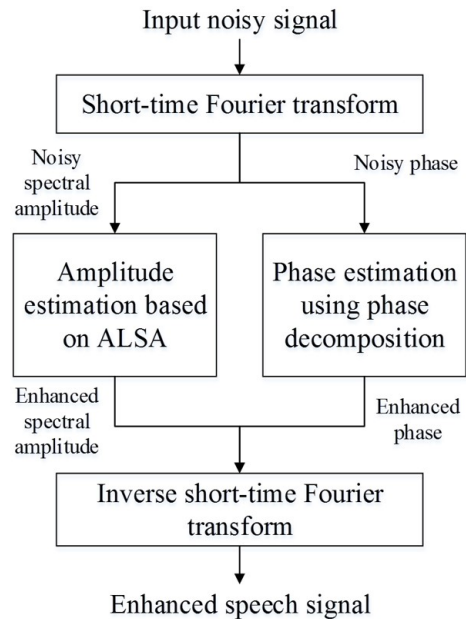


Fig. 1. Overall flow chart of proposed single-channel speech enhancement method.

$n_p(t)$, $n_a(t)$ 그리고 $y(t)$ 에 대해 STFT를 적용하여 다음과 같이 나타낼 수 있다.

$$\begin{aligned} Y(k,l) &= |Y(k,l)|\exp[j\phi(k,l)], \\ &= A(k,l)\exp[j\psi(k,l)] + N_a(k,l), \\ &= R(k,l)\exp[j\phi(k,l)]. \end{aligned} \quad (1)$$

$$N_p(k,l) = D(k,l)\exp[j\varphi(k,l)],$$

여기서 k 와 l 은 주파수와 프레임 인덱스를 나타내며, $R(k,l)$, $A(k,l)$, $D(k,l)$ 은 각각 오염된 음성신호, 클린 음성신호, 잡음 신호의 진폭성분을 나타낸다. $\phi(k,l)$, $\psi(k,l)$, $\varphi(k,l)$ 은 오염된 음성신호, 클린 음성신호, 잡음 신호의 위상성분을 나타낸다.

PTT가 동작하지 않는 동안 입력되는 선형적 잡음 신호는 계속해서 갱신되어 저장되며, 이는 PTT의 동작 시에 ALSA 기반의 진폭 추정부에 적용되어 진폭 추정 이득 값 $G(k,l)$ 을 구하는데 반영된다.

오염된 신호의 진폭 성분 $R(k,l)$ 에 진폭 추정 이득 값을 적용하여 다음과 같이 잡음이 제거된 음성 진폭 성분 $A_e(k,l)$ 을 얻을 수 있다.

$$A_e(k,l) = G(k,l) \cdot R(k,l). \quad (2)$$

이와 함께 잡음에 오염된 위상성분 $\phi(k,l)$ 에 대해 위상 분해를 적용한 뒤 위상의 시간적 평활화를 적용하여 향상된 위상성분 $\psi_e(k,l)$ 을 얻을 수 있다. 향상된 진폭성분과 위상성분에 대해 ISTFT를 적용하여 시간 축의 향상된 음성 신호 $s_e(t)$ 를 얻을 수 있다.

III. ALSA 방식 기반의 진폭 추정

Log-spectral 진폭 추정 방식(LSA)^[3]은 인간의 청각적 지각 특성을 고려하여 기존의 MMSE(Minimum-Mean Square Error) STSA 추정 방식의 비유함수에 로그를 적용한 방식으로 기존의 방식보다 잡음 제거에 적합한 것으로 알려져 있다. 본 논문에서 제안하는 진폭 향상 과정에서는 약한 음성 성분을 보존하고 잔여 잡음을 더욱 효과적으로 제거하기 위해 음성 구간 검출을 기반으로 음성 존재 불확실성 계수 $T(k,l)$ 을 반영한 적응적 Log-spectral 진폭 추정 방식(ALSA)을 사용한다. ALSA방식의 이득함수 $G(k,l)$ 은 다음과 같이 정의된다.

$$\begin{aligned} \gamma(k,l) &= \frac{R^2(k,l)}{\lambda(k,l)}, \\ \xi(k,l) &= \max \left[(1-a_d) \max(\gamma(k,l) - 1, 0), \right. \\ &\quad \left. + a_d \frac{A_e^2(k,l-1)}{\lambda(k,l-1)}, \xi_{\min} \right], \\ v(k,l) &= \frac{\xi(k,l)}{1+\xi(k,l)} \gamma(k,l), \end{aligned} \quad (3)$$

$$G(k,l) = \frac{\xi(k,l) \cdot T(k,l)}{1+\xi(k,l)} \exp \left[\frac{1}{2} \int_{v(k,l)}^{\infty} \frac{e^{-t}}{t} dt \right], \quad (4)$$

여기서 $\lambda(k,l)$ 은 잡음 진폭 스펙트럼, $\xi(k,l)$ 은 a priori SNR(Signal-to-Noise Ratio), $\gamma(k,l)$ 는 a posteriori SNR를 나타낸다. $\xi_{\min} = -20$ dB를 사용하며 스무딩 파라미터 a_d 는 0부터 1 사이의 값을 가진다.

사용자가 PTT를 작동시키지 않을 때에는 마이크를 통해 음성이 포함되지 않은 환경 잡음이 입력된다. 이때 입력되는 선형적 잡음신호는 계속해서 갱신되며 갱신된 선형적 잡음신호를 $D_u(k,l)$ 이라고 한다. 사용자가 PTT를 작동시켜 오염된 음성신호

가 입력되면, $D_u(k,l)$ 에 대해 전체 프레임에 대한 각 주파수 bin의 최소 진폭 값 $D_{\min}(k)$ 와 평균 진폭 값 $D_{avr}(k)$ 가 계산된다. $D_{avr}(k)$ 와 입력신호의 진폭 스펙트럼 $R(k,l)$ 을 이용하여 다음과 같이 코사인 유사도 $C(l)$ 을 계산한다.

$$C(l) = \frac{\sum_{k=1}^K R(k,l) D_{avr}(k)}{\sqrt{\sum_{k=1}^K R^2(k,l)} \sqrt{\sum_{k=1}^K D_{avr}^2(k)}}. \quad (5)$$

$C(l)$ 은 입력신호와 선형적 잡음 신호의 프레임 별 유사도를 나타내며 입력신호에서 음성구간과 비음성구간을 판별하는데 사용된다. $C(l)$ 을 기반으로 유사도 문턱값 δ_n 을 이용해 다음과 같이 음성의 유무를 나타내는 표지함수 $I(k,l)$ 을 구할 수 있다.

$$M(k,l) = \begin{cases} D_{avr}(k) & \text{if } C(l) > \delta_n \\ D_{\min}(k) & \text{otherwise} \end{cases}, \quad (6)$$

$$I(k,l) = \begin{cases} 1 & \text{if } R(k,l) > \delta_s M(k,l) \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

여기서 δ_s 는 k 번째 주파수 bin에서의 $I(k,l)$ 의 문턱값을 나타내며 시간-주파수 블록에서의 약한 음성 성분을 보존하기 위해 $D_{\min}(k)$ 를 적용하였다.

$I(k,l)$ 을 기반으로 조건부 음성 존재 확률 $p(k,l)$ 은 다음과 같이 추정된다.

$$p(k,l) = a_p p(k,l-1) + (1-a_p) I(k,l), \quad (8)$$

여기서 a_p 는 스무딩 파라미터를 나타낸다.

추정된 $p(k,l)$ 을 기반으로 다음과 같이 시변 평활화 변수 $\sigma(k,l)$ 를 이용해 현재의 잡음 진폭 스펙트럼 $\lambda(k,l)$ 을 추정할 수 있다.

$$\sigma(k,l) = a_n + (1-a_n) p(k,l), \quad (9)$$

$$\lambda(k,l) = \sigma(k,l) \lambda(k,l-1) + [1-\sigma(k,l)] R^2(k,l), \quad (10)$$

여기서 a_n 은 노이즈 플로어를 나타낸다.

또한, Cohen^[4]이 보인 것과 같이 주파수 도메인에서 지역적, 전역적 평균 윈도우를 적용하여 $p_{local}(k, l)$ 과 $p_{global}(k, l)$ 를 구할 수 있다. 이와 함께 앞서 구한 $C(l)$ 의 최소값 C_{min} 과 최대값 C_{max} 을 바탕으로 $p_{frame}(l)$ 을 계산하여 음성 부재 확률 $q(k, l)$ 를 다음과 같이 도출할 수 있다.

$$p_{frame}(l) = \begin{cases} 1 & \text{if } C(l) \leq C_{min} \\ 0 & \text{else if } C(l) \geq C_{max} \\ \frac{\log(C(l)/C_{min})}{\log(C_{max}/C_{min})} & \text{otherwise} \end{cases}, \quad (11)$$

$$q(k, l) = 1 - p_{frame}(l)p_{local}(k, l)p_{global}(k, l). \quad (12)$$

유사도 $C(l)$ 을 이용해 도출한 $p_{frame}(l)$ 을 적용함으로써 현재 잡음 환경에 적응적인 음성 부재 확률 $q(k, l)$ 의 추정이 가능하다.

음성 존재 불확실성 계수 $T(k, l)$ 는 다음과 같이 표현된다.

$$T(k, l) = \frac{1 - q(k, l)}{1 - q(k, l) + q(k, l)\{1 + \xi(k, l)\exp[-v(k, l)]\}}. \quad (13)$$

이때 $q(k, l)$ 의 시변 값에 대한 정확한 추정이 이뤄진다면 음성 신호의 추정을 더욱 원본에 가깝게 수행할 수 있다.

IV. 위상 분해 기반의 위상 향상

일반적으로 음성 신호는 진폭과 위상으로 구성된 하모닉스의 합으로 모델링 된다. 음성 신호가 잡음에 의해 오염될 때 위상 또한 오염되는데, 오염된 위상을 향상시키기 위해 본 논문에서는 위상 분해 방식을 사용한다. 제안하는 방식은 하모닉스에서의 위상 정보를 이용하며 입력 신호의 하모닉 h 에서의 위상을 $\phi(h, l) = \Omega[\phi(k', l)]$ 으로 표시한다. k' 는 하모닉 인덱스 h 와 $h : k' = hf_oK$ 의 관계를 가진다. 하모닉스의 주파수가 두 주파수 빈 사이에 위치할 때 $\Omega(\cdot)$ 는 $\phi(k', l)$ 의 가중 합을 반환한다.

잡음에 오염된 유성음 프레임의 각 하모닉 위상은 음성 위상 $\phi_s(h, l)$ 과 포락선 위상 $\phi_e(h, l)$ 로 구성되며, 음성 위상은 다시 선형 위상 $\phi_l(h, l)$ 과 여기 위상 $\theta(h, l)$ 성분으로 구성된다.

$$\begin{aligned} \phi(h, l) &= \phi_s(h, l) + \phi_e(h, l), \\ &= \theta(h, l) + \phi_l(h, l) + \phi_e(h, l). \end{aligned} \quad (14)$$

최근, 음성 신호의 ZZT(Zeros of Z-Transform)에 대한 연구를 통해 유성음의 ZZT에 대해 별도의 패턴이 존재하는 것이 밝혀졌다. 따라서 다른 추가적인 모델링 없이 ZZT 분해 기반의 분리 알고리즘을 통한 위상 분해가 수행될 수 있으므로, 본 논문에서는 먼저 ZZT^[5]를 사용하여 입력 위상을 포락선 위상과 음성 위상으로 분해한다.

ZZT의 전극 모델 파라미터 $a_j(j = 1, 2, \dots, J)$ 를 사용하여 포락선 위상을 다음과 같이 추정한다.

$$\begin{aligned} Re(h, l) &= 1 + \sum_{j=1}^J a_j \cos[j \times \rho(h, l)], \\ Im(h, l) &= - \sum_{j=1}^J a_j \sin[j \times \rho(h, l)], \\ \rho(h, l) &= [(h-1) \times I_f + 1] \times \frac{\pi}{S_f + 1}, \end{aligned} \quad (15)$$

$$\begin{aligned} \tilde{\phi}_e(h, l) &= (-2) \tan^{-1} \left[\frac{Im(h, l)}{Re(h, l)} \right] \\ &h = 1, 2, \dots, N_h, \end{aligned} \quad (16)$$

여기서 N_h 는 유성음 하모닉스의 개수, I_f 는 보간 계수를 나타내며 $S_f = (N_h - 1)I_f + 1$ 는 보간 벡터의 크기를 나타낸다.

다음으로, 오염된 입력 위상에서 추정한 포락선 위상을 차감하여 음성 위상을 구할 수 있다.

$$\phi_s(h, l) = \phi(h, l) - \tilde{\phi}_e(h, l) \cong \theta(h, l) + \phi_l(h, l). \quad (17)$$

음성 위상을 구성하는 선형 위상은 인접 프레임의 선형 위상과 프레임의 기본 주파수에 의존적이며 다음과 같이 표현된다.

$$\phi_l(h,l) = \begin{cases} (\phi_l(h,l) \times R_F + f_v \times M) \times h & \text{if } (l-1) \text{ is voiced,} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

여기서 R_F 는 R_1/R_2 를 나타내고($R_1, R_2 \in 1, 2, 3, 4$) 이는 이전 피치 주기 p_{prp} 와 현재 피치 주기 p_{cup} 사이의 $|p_{cup} \times R_1 - p_{prp} \times R_2|/p_{cup} \times R_1$ 로 구해지는 변화의 폭을 최소화하는 값으로 구해진다. f_v 는 이전 프레임과 현재 프레임의 기본 주파수 $f_o(l)$ 의 가중합으로 주어지며 M 은 샘플 단위의 프레임 시프트 크기를 나타낸다.

선형 위상은 프레임 간격과 하모닉 인덱스에 따른 wrapped 위상이며 음성 신호의 고유한 위상을 포함하는 반면 여기 위상은 unwrapped 위상으로 환경에 민감하게 반응하며 변화가 많은 성분이다. 음성 신호가 잡음에 의해 오염될 때 여기 위상은 오염되며 이로 인해 여기 위상의 순간 위상 차 또한 왜곡된다. 따라서 이러한 위상 왜곡을 추정하여 평활화를 적용함으로써 변질된 여기 위상을 복원시킬 수 있다.

잡음에 오염된 음성 위상에서 선형 위상을 분리함으로써 잡음에 의해 변질된 여기 위상을 얻을 수 있다. 여기 위상 차, 즉 위상 왜곡 $P_D(h,l)$ 은 다음과 같이 계산된다.

$$P_{RS}(h,l) = \theta(h,l) - h\theta(1,l), \quad (19)$$

$$P_D(h,l) = \Delta_h P_{RS}(h,l) = \theta(h+1,l) - \theta(h,l) - \theta(1,l), \quad (20)$$

여기서 Δ 는 유한 차분 연산자를 나타낸다.

계산한 여기 위상의 왜곡 성분에 대해 다음과 같이 위상 평활화 윈도우를 적용함으로써 왜곡이 향상된 여기 위상차 $\tilde{P}_D(h,l)$ 를 얻을 수 있다.

$$\tilde{P}_D(h,l) = \left\{ \sum_{l'=l-R}^{l+R} w_p(l') \exp[P_D(h,l')] \right\}, \quad (21)$$

여기서 R 은 20 ms의 시간 간격에 들어가는 프레임의 개수를 의미하며 w_p 는 위상 평활화 윈도우를 나타낸다.

향상된 여기 위상 차, 추정된 선형 위상과 포락선

위상을 모두 합하여 음성 신호의 하모닉한 구조가 복원된 향상된 위상 $\phi_E(h,l)$ 을 얻을 수 있다. 향상된 위상은 STFT 도메인의 위상으로 변환된다.

$$\phi_E(h,l) = \tilde{P}_D(h,l) + \tilde{\phi}_l(h,l) + \tilde{\phi}_e(h,l), \quad (22)$$

$$\psi_e[(hf_oK) + i, l] = \phi_E(h,l). \quad (23)$$

이렇게 변환된 향상된 위상 $\psi_e(k,l)$ 은 ISTFT 적용 부에서 향상된 진폭과 함께 적용되며 이를 통해 최종적으로 향상된 음성 신호가 출력된다.

V. 실험 결과

제안한 방식의 성능을 측정하기 위해서 단계별 음성 향상 실험을 진행하였다: 1) LSA: Ephraim과 Malah^[3]가 제안한 LSA 방식; 2) ALSA: 음성 존재 불확실성 계수를 반영한 적응적 LSA 방식; 3) ALSA-PE: 위상 향상 방식을 포함한 ALSA 방식.

실험을 위해 TIMIT 데이터베이스에서 200개의 음성 데이터를 대상으로 NOISEX-92 데이터베이스에서 버블 잡음(BN), 레스토랑 잡음(RN), 공장 잡음(FN) 등 3개의 동적 잡음을 segmental SNR 0 dB부터 15 dB까지 5 dB의 간격으로 혼합하여 사용하였다. Segmental SNR은 다음과 같이 정의된다.

$$SegSNR = \frac{10}{|L|} \sum_{l \in L} \log \frac{\sum_{k=0}^{N/2} |X(k,l)|^2}{\sum_{k=0}^{N/2} |D(k,l)|^2}, \quad (24)$$

여기서 L 은 음성을 포함하는 프레임을 나타내고 $|L|$ 은 L 의 개수를 의미한다.

실험에 사용된 음성과 잡음신호는 8 kHz의 샘플링 레이트, 16 bit의 해상도로 구성되었으며 고속 푸리에 변환 시 프레임 크기는 256개 샘플을 사용하였고 75%의 오버래핑을 적용하였다. 본 논문에서 사용한 선형적 잡음은 PTT가 실행되지 않는 상태에서의 5 s의 잡음신호에 대해 수집하였다. 성능을 객관적으로 평가하기 위해 PESQ(Perceptual Evaluation of Speech

Table 1. Averaged PESQ Scores and 95 % confidence interval.

Method	Noise type			
	BN	RN	PN	Average
Unprocessed	1.95±0.14	1.98±0.12	2.01±0.13	1.98±0.13
LSA	2.18±0.13	2.23±0.12	2.28±0.12	2.23±0.12
ALSA	2.37±0.11	2.49±0.14	2.54±0.13	2.47±0.13
ALSA-PE	2.61±0.15	2.74±0.13	2.83±0.11	2.73±0.13

Quality)를 사용하였다.

Table 1은 평균 PESQ 결과를 나타낸다. 실험 결과를 통해 ALSA 방식의 성능이 기존의 방식보다 뛰어나며, 본 논문에서 제안한 방식인 ALSA-PE가 가장 높은 성능을 보이는 것을 알 수 있다.

VI. 결 론

본 논문에서는 사용자 디바이스에 포함된 PTT의 단일 채널 음성 향상 방식에 대해 제안하였다. 제안한 방식은 기존의 단일 채널 음성 향상 방식과는 다르게 입력 신호에 대해 진폭 향상과 위상 향상을 각각 수행한 뒤 이를 결합함으로써 향상의 성능을 더욱 높이고자 하였다. 진폭 향상 과정에서는 Log-spectral 진폭 추정 방식을 기반으로 음성 존재 불확실성 계수를 반영한 ALSA 방식을 적용하였으며 이를 통해 현재의 잡음 환경을 반영하여 음성왜곡을 감소시키면서 효과적으로 잡음을 제거할 수 있었다. 또한 위상 분해 방식 기반의 위상 추정을 적용하여 위상 왜곡을 향상시키고 음성 신호의 고유한 하모닉 구조를 복원함으로써 더욱 명료한 음성 신호를 얻을 수 있었다. 향후 본 논문을 기반으로 실제 디바이스에서 사용가능한 PTT 어플리케이션에 대한 연구를 수행할 예정이다.

감사의 글

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2015R1D1A1A01059804).

References

1. E. Plourde and B. Champagne, "Auditory-based spectral amplitude estimators for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.* **16**, 1614-1623 (2008).
2. J. Kulmer and P. Mowlaee, "Phase estimation in single channel speech enhancement using phase decomposition," *IEEE Signal Process. Lett.* **22**, 598-602 (2015).
3. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.* 443-445 (1985).
4. I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Process. Lett.* **9**, 113-116 (2002).
5. B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit, "Zeros of Z-transform representation with application to source-filter separation in speech," *IEEE Signal Process. Lett.* **12**, 344-347 (2005).

저자 약력

▶ 조 혜 승 (Hye-Seung Cho)



2015년: 광운대학교 전자융합공학과
학사
2015년 ~ 현재: 광운대학교 전자공학과
석박통합 과정

▶ 김 형 국 (Hyoung-Gook Kim)



1999년 ~ 2002년: 독일 SIEMENS/ Cortologic
AG 책임연구원
2002년 ~ 2005년: 독일 베를린 공과대학교
Assistant Professor
2005년 ~ 2007년: 삼성종합기술원 수석
연구원
2007년 ~ 현재: 광운대학교 전자융합공학과
교수