혼합여기모델을 이용한 대역 확장된 음성신호의 음질 개선

최무열(부산대), 김형순(부산대)

<차 례>

<Abstract>

Quality Improvement of Bandwidth Extended Speech Using Mixed Excitation Model

Mu Yeol Choi, Hyung Soon Kim

The quality of narrowband speech can be enhanced by the bandwidth extension technology. This paper proposes a mixed excitation and an energy compensation method based on Gaussian Mixture Model (GMM). First, we employ the mixed excitation model having both periodic and aperiodic characteristics in frequency domain. We use a filter bank to extract the periodicity features from the filtered signals and model them based on GMM to estimate the mixed excitation. Second, we separate the acoustic space into the voiced and unvoiced parts of speech to compensate for the energy difference between narrowband speech and reconstructed highband, or lowband speech, more accurately. Objective and subjective evaluations show that the quality of wideband speech reconstructed by the proposed method is superior to that by the conventional bandwidth extension method.

* Keywords : Wideband speech reconstruction, Bandwidth extension, Mixed excitation model, Gaussian mixture model.

1. 서 론

아날로그 전화망과 이동 통신망을 포함해 현존하는 대부분의 음성통신 시스템은 0.3-3.4 kHz 대역의 협대역(narrowband) 음성신호를 전송한다. 이러한 협대역 음성은 0-8kHz의 광대역 신호와 비교할 때 0-300Hz의 저대역과 3.4-8kHz의 고대역 성분이 제거된 특성으로 인해 명료도가 감소되고 억눌린(muffled) 음질을 갖는다. 이로 인해 협대역 음성에 대한 청취자들의 선호도가 광대역에 비해 매우 떨어진다[1]. 대역폭 확장(bandwidth extension)은 협대역 음성신호의 음질을 향상시키는 기술로서 제거된 저대역과 고대역의 음성 신호를 추정하여 복원함으로써 대역폭을 확장한다.

협대역 음성으로부터 광대역 음성을 복원하는 방법에 대한 다양한 연구들이 진행되어 왔다[2-5]. 광대역 복원을 위해 추정할 파라미터로는 스펙트럼 포락선, 에너지, 그리고 여기신호 등이 있다. 스펙트럼 포락선의 추정은 VQ[2], GMM[3], HMM[4], 그리고 정현파[5]를 이용하는 방법이 있으며, 음성부호화기로 선형예측(LPC)방법을 사용할 경우 여기신호의 추정방법으로 전통적인 주기적 임펄스 입력에서부터 spectral folding[6], BP-MGN[7]방법 등이 사용되고 있다.

본 논문에서는 GMM을 이용하여 스펙트럼 포락선을 추정하는 대역폭 확장 방식[3]의 음질 개선을 위해 mixed excitation 방법을 도입하였다[8]. 기존 LPC 합성에 사용되는 임펄스 입력은 합성시 음질 저하를 나타내는데, 이는 자연음의 여기신호가 갖는 특성을 임펄스로는 충분히 표현하지 못하는데 원인이 있다[8]. 특히 고대역의 여기신호는 유성음인 경우 주기적인 성분뿐만 아니라 비주기적인 성분도 함께 나타나기 때문에 음질 개선을 위해 이를 모델링 할 필요가 생겨났다[8]. 본 논문에서는 이점에 착안하여 고대역 음성의 복원을 위해 sub-band별로 주기성을 추정하여 이를 여기신호에 반영하였으며, 추정하는 대역의 에너지보상을 위해 유무성음을 구분한 GMM 방법을 이용하여 에너지를 추정함으로써 합성음의 음질을 개선시켰다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 GMM을 이용한 대역폭 확장 방법에 대해 설명하고, 3장에서 음질개선을 위해 제안된 여기신호와 에너지 모델링에 대해 설명한다. 4장에서는 실험 및 결과를 보여 주고, 5장에서 결론을 맺는다.

2. GMM을 이용한 대역폭 확장

협대역 신호로부터 광대역 신호를 추정하는 GMM방법은 협대역 신호와 광대역 신호가 어느 정도 서로 상관관계에 있다는 가정 하에 출발한다.

협대역 신호를 $x \in R^n$이라 하고 추정할 광대역 신호를 $y \in R^n$ 하면 $z=(x,y)^T$는 Q 개의 Gaussian 확률밀도함수를 이용한 GMM으로 모델링 된다.

$$p(z|\lambda)=\sum_{i=1}^{Q}\frac{\alpha_i}{(2\pi)^{\frac{1}{2}}|C_i|^{\frac{1}{2}}}\exp[-\frac{1}{2}(z-\mu_i)^T C_i^{-1}(z-\mu_i)] \quad (1)$$

여기서, α_i, μ_i 그리고 C_i는 i번째 밀도함수의 가중치, 평균벡터, 그리고 공분산행렬을 나타내며, 다음 식과 같이 자승오차를 최소화하는 mapping 함수를 통해 광대역 스펙트럼 포락선을 추정한다.

$$\varepsilon_{mse}=E[\,\|y-F(x)\|^2] \quad (2)$$

여기서 $E[.]$는 기대값을 나타내며, $F(x)$는 추정될 광대역 스펙트럼 포락선이 된다.

최소자승오차를 만족하는 mapping 함수는 다음과 같이 표현된다.

$$F(x)=E[y|x]=\sum_{i=1}^{Q}h_i(x)\;\mu_i^y \quad (3)$$

여기서

$$h_i(x)=\frac{\frac{\alpha_i}{(2\pi)^{\frac{1}{2}}|C_i^{xx}|^{\frac{1}{2}}}\exp[-\frac{1}{2}(z-\mu_i^x)^T C_i^{xx\,-1}(z-\mu_i^x)]}{\sum_{i=1}^{Q}\frac{\alpha_i}{(2\pi)^{\frac{1}{2}}|C_i^{xx}|^{\frac{1}{2}}}\exp[-\frac{1}{2}(z-\mu_i^x)^T C_i^{xx\,-1}(z-\mu_i^x)]}$$

및

$$C_i=\begin{bmatrix} C_i^{xx} & C_i^{xy} \\ C_i^{yx} & C_i^{yy} \end{bmatrix},\quad \mu_i=\begin{bmatrix} \mu_i^x \\ \mu_i^y \end{bmatrix} \quad (4)$$

이며, $h_i(x)$는 i번째 Gaussian 밀도 함수의 사후 확률을 나타낸다.

3. 음질개선을 위해 제안한 방법

본 논문에서는 GMM 기반의 대역폭 확장된 합성음의 음질을 개선하기 위하여 여기신호의 추정과 유무성음을 구분한 에너지를 사용하였다.

3.1. GMM 기반의 여기신호 추정

LPC 모델에 의한 음성합성은 합성 필터와 여기신호를 통하여 음성을 합성해 낸다. 여기신호는 일반적으로 주기적인 임펄스열과 비주기적인 잡음으로 모델링하는데, 임펄스열은 고대역 유성음 구간에서의 합성음의 자연스러움을 감쇄시키는 원인이 된다. 자연음의 여기신호는 부분적으로 주기성과 비주기성이 혼합된 신호의 형태를 나타낸다. 따라서 혼합된 여기신호를 임펄스 대신 사용하는 것은 LPC 합성의 음질 개선에 이미 사용되어 온 방법이다[8]. 본 논문에서는 이를 응용하여 고대역의 주기성과 비주기성을 통계적인 방법을 통해 추정하고 임펄스와 잡음 성분에 가변적인 비율을 적용하는 방식을 제안한다. 식 (5)는 임펄스열과 잡음신호에 추정된 주기성의 가중합으로 표현된 여기신호 $e(k)$를 나타낸다.

$$e(k) = \sum_{i=1}^{M} [\, m_i(k) * W_i + n_i(k) * (1 - W_i)] \tag{5}$$

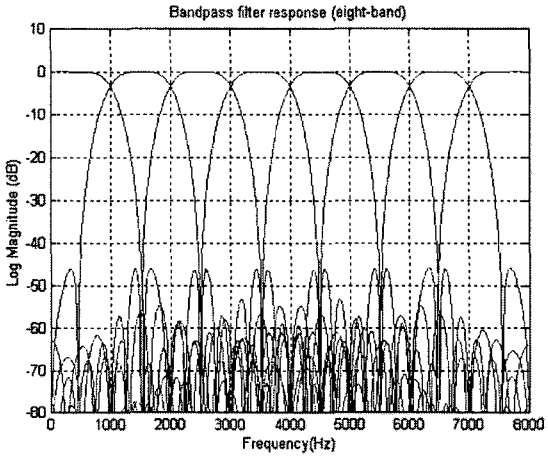
여기서 M은 필터뱅크의 개수이고, $m_i(k)$와 $n_i(k)$는 각각 i번째 필터를 통과한 임펄스열과 잡음신호이다. W_i는 i번째 대역에서의 주기성을 나타내며, 식 (6)와 같이 필터를 통과한 신호의 정규화된 자기상관함수를 이용하여 구한다.

$$W_i = \max_{\tau_{min} \le \tau \le \tau_{max}} \frac{\sum_{n=0}^{N-1} s_n s_{n+\tau}}{\sqrt{\sum_{n=0}^{N-1} s_n^2 \sum_{n=0}^{N-1} s_{n+\tau}^2}} \tag{6}$$

여기서 s_n은 i번째 대역 필터를 통과한 신호이며, N은 신호의 길이, τ는 지연시간이다. W_i 값을 구하기 위해 2ms~10ms 범위에서 상관계수를 구한다. 훈련 데이터로부터 필터뱅크를 통과한 신호의 상관계수를 구하여 GMM 파라미터로 모델링하게 되면, 협대역 입력 신호로부터 광대역 스펙트럼 포락선 파라미터와 더불어 여기신호의 주파수 대역별 상관계수를 식 (7)과 같이 추정할 수 있게 된다.

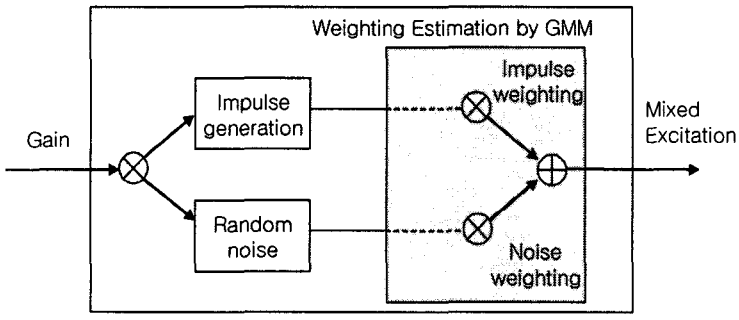
$$F(\boldsymbol{x}) = \hat{\boldsymbol{c}} = E[\boldsymbol{c}|\boldsymbol{x}] = \sum_{i=1}^{Q} h_i(\boldsymbol{x}) \quad \mu_i^{\boldsymbol{c}} \tag{7}$$

여기서 $\mu_i^{\boldsymbol{c}}$는 i번째 Gaussian 밀도 함수의 평균 상관계수 벡터이고, $\hat{\boldsymbol{c}}$는 추정된 필터뱅크의 상관계수 벡터이다.



<그림 1> 필터 뱅크의 주파수 응답

필터뱅크는 균일한 대역폭을 갖는 8개의 필터로 구성했으며, 그 주파수응답은 <그림 1>과 같다. 추정된 상관계수는 <그림 2>에서와 같이 주기적인 임펄스와 비주기적인 잡음성분에 가중치로 사용되어 혼합 여기신호를 발생시킨다.



<그림 2> Mixed excitation 구성도

3.2. GMM기반의 에너지 추정

대역폭 제한된 신호의 에너지는 특히 무성음처럼 고대역에 큰 에너지를 갖는 경우 대부분의 에너지를 잃게 되어 명료성이 저하된다. 본 논문에서는 고대역 및 저대역 복원신호의 에너지를 보다 정교하게 추정하기 위해, 음향공간을 유성음 구간과 무성음 구간으로 나누고, 각각 GMM으로 모델링 하였다. 에너지를 모델링 하기 위한 파라미터로는 협대역에 대한 고대역, 저대역의 에너지 비율을 2차원 벡터로 만들어 훈련하였다. 훈련된 GMM모델에 의한 에너지 추정식은 다음과 같다.

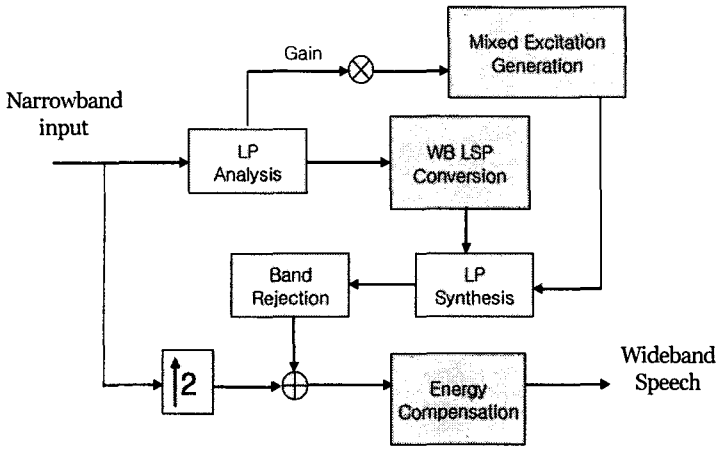
$$F(\boldsymbol{x}) = \widehat{G} = E[G|\boldsymbol{x}] = \sum_{i=1}^{Q} h_i(\boldsymbol{x}) \quad \mu_i{}^G \tag{8}$$

$$G = [\log \frac{G_H}{G_N}, \log \frac{G_L}{G_N}]^T$$

여기서 $\mu_i{}^G$는 i번째 Gaussian 밀도 함수의 평균 에너지 벡터이고, 추정된 에너지 $\widehat{G}$는 협대역에너지 G_N과 고대역 에너지 G_H, 그리고 저대역 에너지 G_L의 로그 에너지 비율을 나타내는 2차원 벡터이다.

3.3 대역폭 확장 시스템 구성

<그림 3>은 대역폭 확장의 구성도 이다. 우선 전화망을 통해 대역 제한된 음성이 들어오면 선형 예측 필터를 통해 협대역 스펙트럼 포락선을 구한 뒤 GMM을 이용한 대역폭 확장 방법을 통해 광대역 스펙트럼 포락선을 추정한다. 또한 제안된 방식으로 혼합여기신호를 발생시켜 추정된 광대역 스펙트럼 포락선과 함께 선형 필터를 통해 광대역 음성을 합성한다. 합성된 음성신호는 대역차단(band rejection) 필터를 통해 고대역 신호와 저대역 신호로 분리된 뒤 전송된 음성과 합하여 광대역 신호를 합성하며 이후 제안된 에너지 보상 방법을 통해 합성음의 음질을 개선한다.



<그림 3> 대역폭 확장의 시스템 구성도

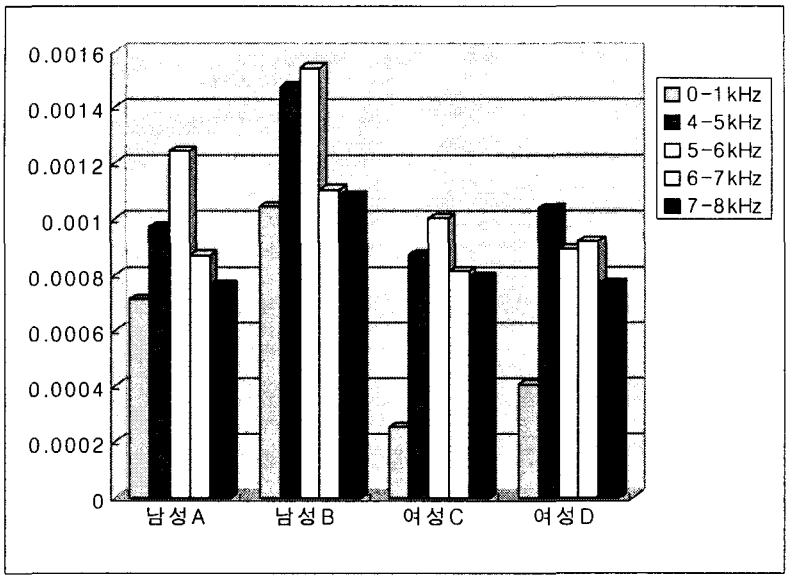
4. 실험 및 결과

4.1. 실험환경

실험을 위해 국어공학센터에서 구축한 PBS 589문장에 대한 남녀 50명분의 발성 데이터 약 13시간 분량을 사용하였다. 스펙트럼 변환과 에너지 추정 그리고 여기신호 가중치 추정을 위해서 256 mixture를 갖는 GMM 모델을 훈련하였다. 또한 복원된 광대역 음성의 음질을 평가하기 위한 객관적 척도로서 스펙트럼 왜곡(spectral distortion (SD))을 계산하였고, 주관적 척도로서 청취 평가를 수행하였다.

4.2 대역별 주기성 측정

<그림 4>는 필터뱅크에 의한 주파수 대역별 주기성을 추정하는 GMM모델의 성능평가이다. 추정된 고대역 신호의 주기성을 평가하기 위해 원음의 주기성과의 최소자승오차법으로 계산한 결과를 나타냈다. 남성 2명과 여성 2명의 음성으로 추정된 주기성을 평가해 본 결과 주기성이 강한 저대역이 비주기성이 큰 고대역 보다 추정 오차가 적은 것을 알 수 있다.



<그림 4> 추정된 주파수 대역별 주기성의 MSE

4.3. 스펙트럼 왜곡

합성음의 객관적인 음질평가를 위한 SD는 식 (9)와 같이 구하였고, 그 결과를 <표 1>에 나타냈다.

$$SD=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left(20\log\frac{g}{|A_n(\omega)|}-20\log\frac{\hat{g}}{|\hat{A}_n(\omega)|}\right)^2 d\omega} \qquad (9)$$

여기서 $A_n(\omega)$, $\hat{A}_n(\omega)$는 각각 n번째 프레임에 대한 원음과 합성음으로 구한 선형 필터의 주파수 응답이고, g와 $\hat{g}$는 각 신호의 선형예측 이득이다.

스펙트럼 왜곡 평가 결과를 나타낸 <표 1>에서 기존 방식이란 GMM모델을 통하여 저대역과 고대역의 스펙트럼 포락선을 예측하는 방법에 주기적 임펄스열과 비주기적인 잡음을 여기신호로 사용하고, 에너지 보상은 VQ코드북 방법을 사용한 합성방식이다. 그리고, 제안 방식은 스펙트럼 포락선을 예측하는 동일한 GMM모델에 통계적인 방법으로 주파수 대역별 주기성을 추정하여 만든 mixed excitation을 사용하고, 유무성음 구간의 에너지를 각각 GMM으로 모델링하여 에너지를 보상한 방법을 적용한 것이다.

<표 1> 합성음과 원음 사이의 스펙트럼 왜곡

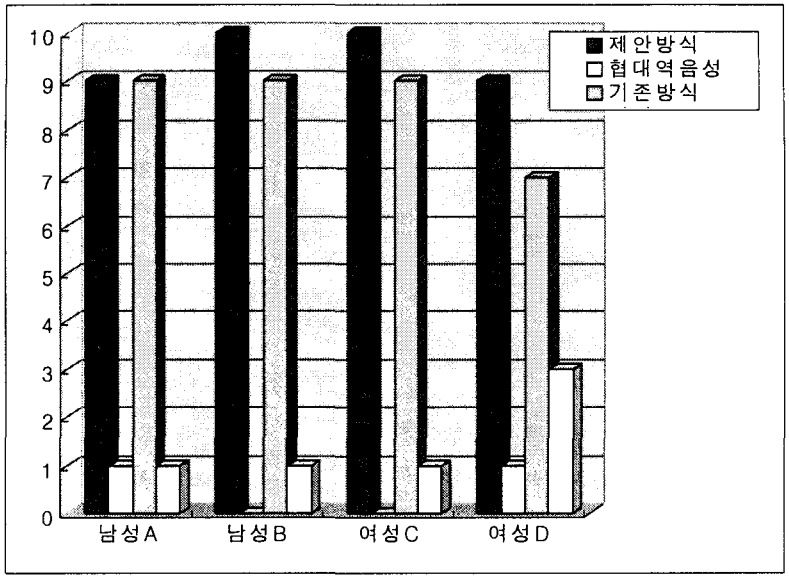
	성별	구현방법	유성음(dB)	무성음(dB)
협대역음성	남성	-	12.3	-
	여성	-	14.7	-
복원된 광대역음성	남성	기존 방식	6.91	8.18
		제안 방식	6.57	7.96
	여성	기존 방식	7.21	15.55
		제안 방식	6.38	12.19

<표 1>에서 보는 바와 같이 합성된 광대역 신호는 협대역 신호와 비교하여 왜곡이 현저히 줄어들었으며, VQ 방법과의 비교에서도 성능이 향상되었다.

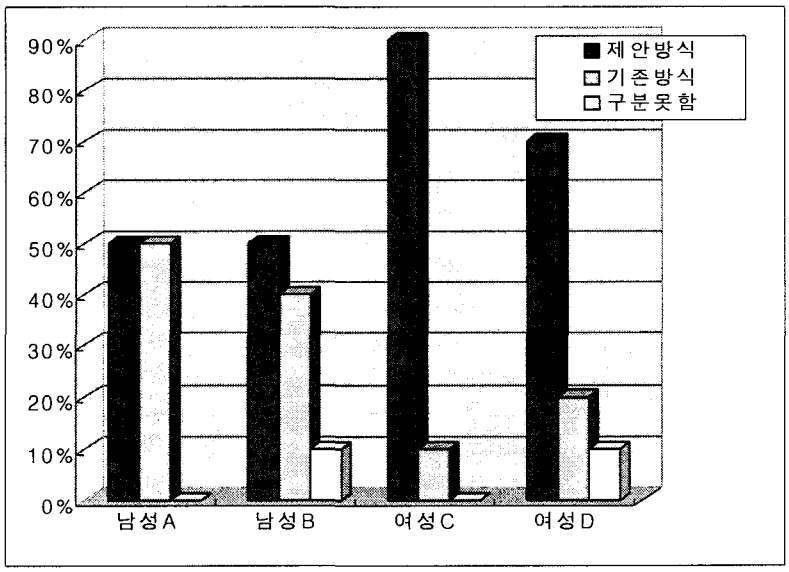
4.4. 청취 평가

주관적 평가의 첫 번째 단계로 원음을 기준으로 기존 방식에 의한 합성음과 협대역 신호, 그리고 제안 방식에 의한 합성음과 협대역 신호와의 비교실험을 통해 어느 음성의 대역폭이 광대역에 더 가까운지를 평가하였다. <그림 5>의 결과에서 보는 바와 같이 다수의 청취자가 합성음의 대역폭이 확장되었다고 판단했다. 여성D 음성의 경우, 기존방식에 의한 합성음의 대역폭이 협대역과 비교하여 확장되지 않았다고 응답한 평가자가 30%에 달했는데 이는 기존 방식의 합성음이 제안 방식의 합성음에 비해 음질 저하가 나타남을 간접적으로 보여주는 결과이다.

<그림 6>은 두 가지 합성음에 대한 주관적 평가로 합성음의 선호도를 ABX 테스트로 측정한 결과이다. 실험에 사용된 음성은 남성 2명과 여성 2명의 발성 문장을 사용했으며, 두 가지 합성 방식에 따라 합성한 뒤 10명의 청취자에게 헤드셋을 통해 청취하도록 하였다. 그림에서 보는 바와 같이 제안한 방식의 결과가 기존 방식보다 여성음성에서 선호도가 높았고, 남성음성은 동일하거나 약간 더 선호되는 양상을 보였다.



<그림 5> 두 합성음과 협대역의 대역폭 비교 결과



<그림 6> 선호도 평가 결과

5. 결 론

본 논문에서는 GMM을 이용하여 대역 제한된 전화음성에서 광대역 신호를 복원하는 방법을 통한 복원 음성의 음질 개선을 위하여, 고대역 합성음의 자연성에 영향을 미치는 혼합 여기신호의 도입과 유/무성음을 구분한 음향공간에서의 GMM 기반의 에너지 추정방식을 제안하였다. 혼합여기 모델 방식을 구현하기 위해 균등한 대역을 가진 필터뱅크를 사용하였고, 필터를 통과한 신호에 통계적 모델을 적용하여 대역별 주기성을 추정하였다. 음질 평가를 위하여 기존 방식과 제안된 방식에 대한 객관적 및 주관적 평가를 수행하였다. 기존 방식과 제안된 방식의 합성음 모두 대역 제한된 음성보다 선호되는 결과를 보였고, 제안된 방식에 의한 합성음이 기존 방식에 의한 합성음보다 더 선호되는 결과를 나타내었다. 그러나 남성음성의 경우 여성음성에 비해 선호도 개선 정도가 상대적으로 저조하였으며, 이는 복원된 음성의 고주파 영역에서의 잡음 특성이 남성음성의 경우 비교적 크게 나타났기 때문으로 풀이된다. 제안된 방식에서의 추정 오류는 합성음의 음질에 영향을 미치는 잡음 특성의 원인이 되므로, 추정 오류를 줄이기 위한 모델 개선 방안에 대해 추가적 연구가 필요하다.

참고문헌

[1] S. Voran, "Listener ratings of speech passbands", *in Proc. of IEEE Workshop on Speech Coding*, pp. 81-82, Pocono Manor, 1997.

[2] Y. Yoshida and M. Abe, "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping", *in Proc. of ICSLP*, pp. 1591-1594, 1994.

[3] K. Y. Park and H. S. Kim, "Narrowband to wideband conversion of speech using GMM based transformation", *in Proc. of ICASSP*, vol. 3, pp. 1843-1864, 2000.

[4] P. Jax and P. Vary, "Wideband extension of telephone speech using a hidden markov model", *in Proc. of IEEE Workshop on Speech Coding*, 2000.

[5] J. Epps, Wideband extension of narrowband speech for enhancement and coding, Ph. D. thesis, University of New South Wales, Sep, 2000.

[6] J., Makhoul, M., Berouti, "High-frequency regeneration in speech coding systems", *in Proc. of ICASSP* '79, Volume: 4 , Apr 1979 pp. 428 - 431

[7] Y. Qian and P. Kabal, "Dual-mode wideband speech recovery from narrowband speech", *in Proc. of EUROSPEEH*, Geneva, Swiss pp. 1433-1437, Sep. 2003.

[8] A. V. McCree and T. P. Barnwell III, "A mixed excitation LPC Vocoder Model for low bit rate speech coding", *in Proc. of IEEE Trans. on Speech and Audio Processing*, Volume: 3 , Issue: 4 , July 1995. pp.242-250

[9] 박진수, 김형순, "4800 bps CELP 음성 부호화기에 적용한 대역폭 확장에 관한 연구", 대한음성학회 학술대회 발표 논문집, pp. 175-178, 2002년.

접수일자 : 2004년 11월 15일
게재결정 : 2004년 12월 10일

▶ 최무열 (Mu Yeol Choi)
주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과
소속: 부산대학교 전자공학과 음성통신연구실
전화: 051) 516-4279
E-mail: mychois@pusan.ac.kr

▶ 김형순 (Hyung Soon Kim)
주소: 609-735 부산시 금정구 장전동 산30번지 부산대학교 공과대학 전자공학과
소속: 부산대학교 전자공학과 음성통신연구실
전화: 051) 510-2452
E-mail: kimhs@pusan.ac.kr