

LSP 파라미터의 분포특성을 이용한 EVRC의 음질개선에 관한 연구

민소연^{1*}, 나덕수²

¹서일대학교 정보통신과, ²(주)보이스웨어 연구소

A Study on the Improvements of the Speech Quality by using Distribution Characteristics of LSP parameters in the EVRC(Enhanced Variable Rate Codec)

So-Yeon Min^{1*} and Deok-Su Na²

¹Dept. of Information and Communication, Seoil College

²Voiceware co. Ltd, R&D center

요 약 EVRC에서는 채널 스펙트럼의 효율을 높이고 시스템의 소비 전력을 줄이기 위하여, 통화시간 중에서 사용자가 말을 할 때만 음성신호를 압축하여 전달하고, 말을 하지 않을 때는 음성신호를 전달하지 않는다. 또한, EVRC에서는 음성 프레임을 1, 1/2, 1/8의 세 가지 전송률로 구분하여 다르게 처리 하고 있으며, 예를 들어, 1/8 전송률은 입력 신호가 묵음구간인 것을 의미한다. 본 연구에서는 LSP 파라미터의 분포특성을 이용한 유성음 구간, 무성음 구간, 묵음 구간을 구분하는 방법을 사용하여, 유성음인 경우에 대해 1 rate으로 부호화하고, 무성음 구간의 경우는 1/2 rate, 묵음의 경우에는 1/8 rate으로 전송하는 방법에 대하여 제안하였다. 즉, EVRC에서 full rate으로 보내는 부분에 대해서는 기존의 방식을 그대로 적용하며, half rate은 유성음, 무성음을 구분하여 유성음일 경우 full rate으로 바꾸어 전송하였고, 묵음에 대해서는 EVRC 기본 rate을 적용하였다. 실험과정에서는, SNR, ASDM, 전송률을 측정하였으며, 제안한 알고리즘을 사용하는 경우 EVRC에 비해 음성품질이 향상됨을 증명하였다.

Abstract To improve the efficiency of the channel spectrum and to reduce the power consumption of the system in EVRC, the voice signal is compressed and transmitted only when the user speaks to. In addition to this, voice frames are divided into three rates 1, 1/2 and 1/8 and each frame is handled differently. For example, we assumed that the input is silence region if the 1/8 rate is used. In this paper, the sections are firstly separated into the voiced speech signal region, unvoiced speech signal region, and silence region by using distribution characteristics of LSP parameters. Then the paper suggested to encode 1 rate for the voiced speech signal, 1/2 rate for the unvoiced speech signal region, 1/8 rate for the silence region. In other words, traditional way of transmission is used when sending full rate in the EVRC. However, when sending half rate, the voice is firstly distinguished between voiced and unvoiced. If the voice is distinguished as voiced, voice is converted into full rate before the transmission. If it is distinguished as silence, EVRC's basic rate is applied. In the experimental results with SNR, ASDM, transmission bit rate measurement, we have demonstrated that voice quality was improved by using the proposed algorithm.

Key Words : EVRC, LSP, Speech Quality, Transmission Bit Rate

1. 서론

에서 사용되어지고 있는 대표적인 음성압축 방식으로, 채널 스펙트럼의 효율을 높이고 시스템의 소비 전력을 줄이기 위해 사람이 말 할 때만 음성을 압축하여 전달하고 말

EVRC(Enhanced variable rate codec)는 이동통신 시스템

본 논문은 2010년도 서일대학 학술연구비에 의해 연구되었습니다.

*교신저자 : 민소연(symin@seoil.ac.kr)

접수일 11년 11월 25일

수정일 11년 12월 12일

게재확정일 11년 12월 13일

을 하지 않을 때는 음성을 전달하지 않으므로 저전송률이 가능하다. 또한 EVRC에서는 패킷 타입별로 1, 1/2, 1/8의 세 가지 전송률로 구분하여 다르게 처리를 하고 있다. 예를 들어, 1/8 전송률은 입력신호가 음성신호가 아닌 잡음만이 존재하는 묵음구간이라는 것을 의미하고 있다[1].

EVRC 부호화방식에서 사용되는 LSP 파라미터는 일정한 스펙트럼 민감도와 저전송률 부호화 특성을 지니며 낮은 스펙트럼 왜곡, 좋은 선형보간 특성을 지닌다[2]. 본 연구에서는 LSP 파라미터를 이용하여 음성신호의 성분분리를 수행하고, 이를 적용한 부호화기에서의 음질향상을 목적으로 한다.

본 논문의 구성은 다음과 같다. 2장에서는 EVRC와 LSP 파라미터에 대해 분석하고, 3장에서는 LSP 파라미터의 분포특성을 이용하여 유/무성음 구간, 묵음구간을 판별하고, 효율적 성분분리를 수행하여 rate 결정을 수행함으로써 음질을 향상하는 제안한 방법에 대해 설명한다. 4장에서는 제안한 알고리즘의 실험결과를 살펴보고 5장에서는 결론을 맺는다.

2. EVRC와 LSP 파라미터

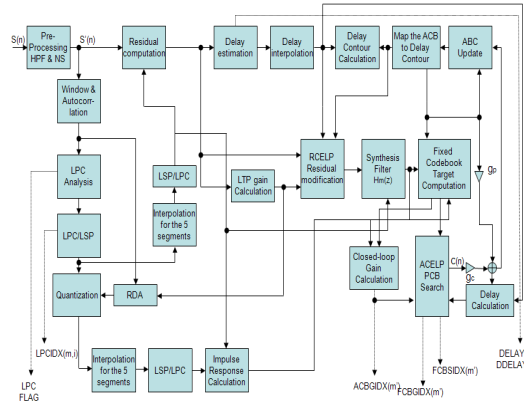
2.1 EVRC 부호화 방식

EVRC 부호화 방식은 채널용량을 효율적으로 사용하기 위해 세 가지의 전송률을 가변적으로 사용한다. 즉, 최대 전송률인 Rate 1, 중간 전송률인 Rate 1/2, 그리고 최소 전송률인 Rate 1/8이다[1].

EVRC에 입력된 음성 신호는 전처리 단을 통과하여 코덱의 음질을 저하시킬 수 있는 매우 낮은 저주파 성분과 주변 잡음에 대해 입력 신호를 조절하는 과정을 거치게 된다. 이 과정은 DC 성분을 제거하기 위해 120Hz의 차단 주파수를 갖는 고역 통과 필터링과 스펙트럼 차감법을 이용한 잡음 억제 과정으로 이루어진다. 다음으로 20ms의 프레임 단위로 선형 예측 계수를 구하여 LSP(Line Spectrum Pairs) 파라미터로 변환된다. 또한 전처리된 음성 신호를 현재와 이전 프레임에서 얻은 LSP 파라미터를 보간(interpolation)하여 얻은 역필터(inverse filter)에 통과시켜 잔차(residual) 신호를 얻은 후, 개회로 검색으로 지연을 추정하고 장구간 예측이득을 구한다. 그리고 RDA(Rate Determination Algorithm)는 앞의 과정에서 구한 17개의 자기 상관계수와 장구간 예측 이득을 이용하여 세 가지 전송률 중에서 하나를 선택한다[1].

프레임 전송률은 추정 잡음 레벨에 비례하는 전송률 결정 임계치(Rate Determination Threshold)와 프레임 에너

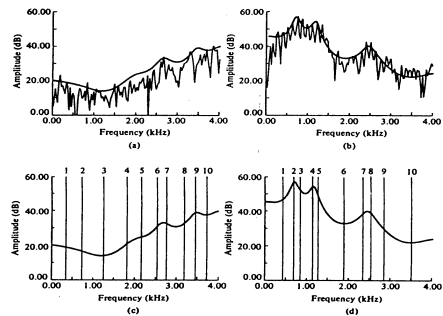
지를 비교하여 결정되는데 추정 잡음 레벨이 높아지면 같은 프레임 에너지를 갖는 프레임이 낮은 전송률로 부호화될 확률이 높아진다. 그림 1에서 위에서 설명되어진 EVRC에 대한 전체 처리과정을 도시화하고 있다.



[그림 1] EVRC의 전체 블록도
[Fig. 1] Block diagram of EVRC

2.2 LSP 파라미터

다양한 종류의 저전송률 음성부호화기에 사용되는 LSP 파라미터는 일정한 스펙트럼 민감도와 저전송률 부호화 특성을 지니며 낮은 스펙트럼 왜곡, 좋은 선형보간 특성을 지닌다.



[그림 2] LSP 파라미터의 예
(a) 자음 /s/
(b) 모음 /a/
(c),(d) /s/와 /a/에 대한 LPC 분석과 LSP
[Fig. 2] Example of the LSP parameter
(a) Consonant /s/
(b) Vowel /a/
(c),(d) The analysis of LPC and LSP for /s/, /a/

그림 2에서 음성신호의 주파수 스펙트럼과 LPC 분석 그리고 LSP 파라미터를 보여주고 있다[2]. 즉 영문 /s/ 발

음에 해당하는 스펙트럼은 (a)와 같이 저주파수 영역의 에너지가 낮고 고주파수 영역의 에너지가 높다. 반면에 유성음의 스펙트럼은 (b)와 같이 1kHz 이내에서 제 1 포만트가 나타나고 3개 이상의 포만트가 존재한다. 이와 같은 스펙트럼 모양으로 LSP의 모양도 유성음과 무성음에서 뚜렷한 차이를 가진다. 먼저 유성음 구간에서는 포만트로 인해 저주파수 영역에서 보다 많은 LSP가 분포하고 그 간격도 고주파수의 LSP 간격 보다 좁게 나타난다. 반면에 무성음 구간에서는 고주파수 영역에서 LSP가 많이 분포하고 그 간격도 저주파수영역의 LSP 간격보다 좁게 나타난다[2~4].

3. 제안한 알고리즘

EVRC의 추정잡음 레벨은 주 피치의 검출 여부에 따라 갱신되는데, 주 피치가 검출되지 않으면 추정 잡음레벨이 점차 높아지게 된다. 또한, 이전 프레임과 현재 프레임의 관계에 의해서 1/2 rate을 결정하고 있다. 이전 프레임이 1 rate 이고 현재 프레임이 배경잡음(background noise) 으로 결정되면 현재 프레임을 1/2 rate으로 선택된다.

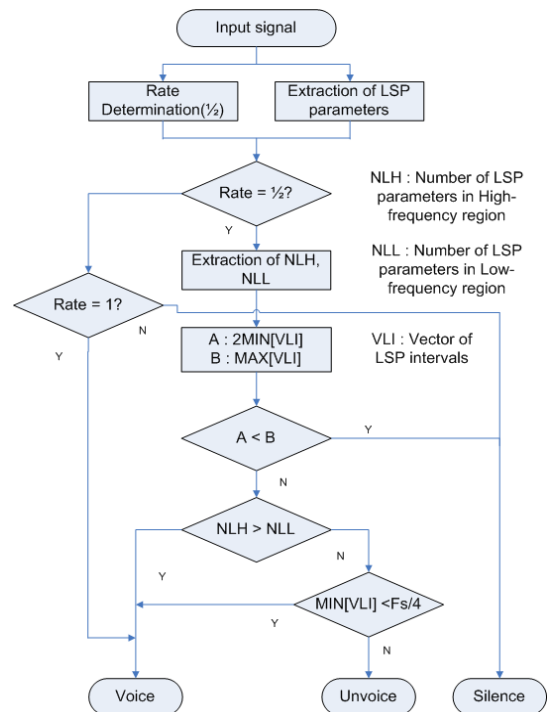
무성음(unvoiced speech signal) 경우는 잡음구간으로 판단되어 frame rate 결정에 오류를 범할 수 있다. 그렇기에 음성의 불연속성으로 인하여 1/2 rate으로 결정이 되는 경우 rate 결정 오류를 범할 수 있어 유/무성음 구간을 구분하지 않으면 음성이 존재하는 구간에서도 낮은 전송률로 보내 음질 열하를 초래하게 된다.

또한, 기존 성분분리 과정에서는 파열음이나 천이구간이 존재하는 음성인 경우 안정된 주기 측정이 어렵고, 에너지가 낮은 유성 자음구간의 경우 배경에 있어서 잡음 에너지와 구분이 어렵다. 그러므로 주변 잡음과 임펄스 성 잡음에도 강인하며, 음절 간에 생기는 묵음구간도 음성신호에 포함시켜서 검출하여야한다. 그림 3은 LSP 파라미터의 분포특성을 이용한 EVRC에서의 음질개선 알고리즘의 블록도이다.

먼저 LSP parameter를 이용하여 유, 무성음, 묵음구간을 구분하며, 유성음인경우에 대해 1 rate으로 encoding하며, 무성음인 경우는 1/2 rate, 묵음의 경우에는 1/8 rate으로 전송한다.

이 과정에서 샘플링 주파수를 F_s 라고 가정할 때 $F_s/4$ 이하의 주파수영역에 존재하는 LSP 개수를 NLL이라 하고 $F_s/4$ 이상의 주파수영역에 존재하는 LSP 개수를 NLH라 하면 NLL이 NLH보다 큰 경우는 음성신호의 스펙트럼이 저주파 쪽에서 봉우리가 많이 나타나는 모양이어서 유성음 프레임 구간이라고 판단한다. 왜냐하면,

유성음의 제 1 포만트와 제 2 포만트가 주로 저주파수 영역에 존재하기 때문이다. 이와 반대로 NLH가 NLL보다 큰 경우는 무성음 프레임이라고 판단한다. 무성음 프레임에 해당하는 스펙트럼은 주요 포만트가 고주파영역에 나타나기 때문이다. 또 한 가지 정보로는 제 1 포만트의 존재 여부로써 무성음인지 유성음인지를 결정한다. 즉, LSP 파라미터들의 간격을 조사하여 $F_s/4$ 이하의 영역에서 좁은 간격을 나타내는 LSP들이 존재하면 유성음으로 간주한다[5,6].



[그림 3] 제안한 알고리즘의 블록도
[Fig. 3] Block diagram of proposed algorithm

묵음 구간에서는 LSP들이 비교적 일정한 간격으로 나타나고 무성음에서는 고주파 영역에서 좁은 간격을 나타내는 LSP들이 존재한다. 이러한 특징을 이용하여, 묵음 구간과 무성음 구간의 결정을 위해 $F_s/4$ 이상의 LSP들의 간격 중 최소 값과 최대 값이 2배 이상 차이가 있는지를 계산하였다. 만일 차이가 2배 이상 존재한다면 무성음이고 그렇지 않다면 묵음으로 간주한다.

EVRC full rate으로 보내는 부분은 그대로 적용하며, half rate 은 유/무성음을 구분하여 유성음일 경우 full rate으로 바꾸어 전송한다.

묵음에 대해서는 EVRC에서 사용하는 원래의전송률을 적용하여 동작하므로 문제 삼지 않는다.

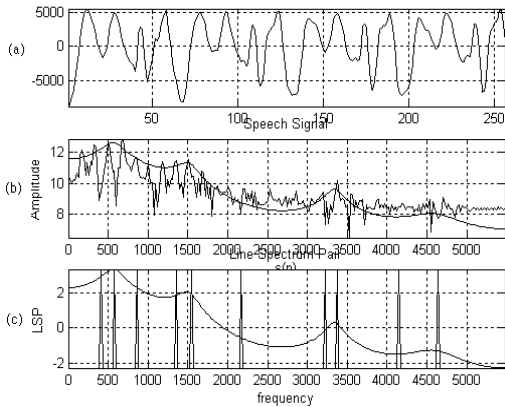
4. 실험결과

LSP 파라미터의 분포특성을 이용한 성분분리 알고리즘을 C++로 구현하여 부호화기에서의 전송률을 측정하였고, 음성신호 각 프레임 구간에서 LSP 파라미터의 분포특성을 분석하기 위해 Matlab으로 구현하였다.

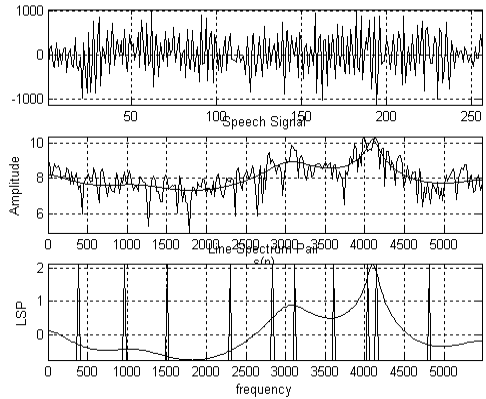
그림 4에서는 임의의 유성음 구간 음성 프레임에 대한 시간영역 음성파형, 주파수 영역 분석결과, LSP 파라미터 분포 특성을 나타내었다.

그림 5에서는 무성음 구간에 해당하는 임의의 음성프레임에 대해 시간영역 음성파형, 주파수 영역 분석결과, LSP 파라미터 분포 특성을 나타내었다. 즉, 그림 4와 그림 5의 실험결과를 통해 유성음 구간은 저주파수 영역에서 LSP 파라미터가 많이 분포하고 있으며, 이 결과 저주파수 쪽에서 봉우리(pole)가 많이 나타남을 알 수 있다. 이에 반하여 무성음 구간은 고주파 영역에서 LSP 파라미터와 봉우리가 많이 존재함을 알 수 있다. 그러므로 3장에서 기술한 LSP 파라미터의 분포특성을 이용한 성분분리에 의한 EVRC rate 결정이 효율적임을 알 수가 있다.

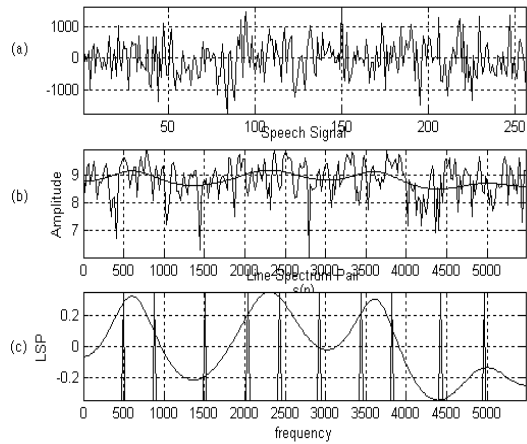
그림 6은 묵음 구간에서의 LSP 파라미터의 분포를 나타내고 있고, 비교적 일정한 분포특성을 분석할 수 있다. 그러므로 고주파 영역에서 주로 봉우리가 나타나는 무성음 구간의 LSP 파라미터의 분포특성과는 차이가 남을 알 수가 있다.



[그림 4] 유성음에 대한 LSP 파라미터의 분포
 (a) 유성음 파형 (b) 스펙트럼 분석
 (c) LPC 분석과 LSP 파라미터
[Fig. 4] LSPs distribution of voiced speech signal
 (a) Waveform of voiced speech signal
 (b) Spectrum analysis
 (c) LPC analysis and LSPs



[그림 5] 무성음에 대한 LSP 파라미터의 분포
 (a) 무성음 파형 (b) 스펙트럼 분석
 (c) LPC 분석과 LSP 파라미터
[Fig. 5] LSPs distribution of unvoiced speech signal
 (a) Waveform of unvoiced speech signal
 (b) Spectrum analysis
 (c) LPC analysis and LSPs



[그림 6] 묵음구간에 대한 LSP 파라미터의 분포
 (a) 시간영역 파형 (b) 스펙트럼 분석
 (c) LPC 분석과 LSP 파라미터
[Fig. 6] LSPs distribution of silence region
 (a) Waveform of silence region
 (b) Spectrum analysis
 (c) LPC analysis and LSPs

제안한 알고리즘의 성능평가를 위해 부호화기를 통한 합성음의 음질비교와 전송률 비교를 수행하였다. 즉, 음질 비교과정에서는 객관적인 음질평가방법인 SNR (Signal to Noise Ratio)과 선형 예측 계수에 기반을 둔 음성 오차를 측정하는 평균 스펙트럼 왜곡 측정(Average Spectral Distortion Measure) 방법을 사용하였으며, EVRC

기본 rate을 사용하는 경우와 제안하는 방법을 사용하는 경우의 전송률을 비교하여 실험결과를 나타내었다.

식(1)을 사용하여 기존 EVRC와 제안하는 알고리즘이 적용된 음성 부호화기에서의 프레임별 SNR을 구하였다. 또한 식(2)를 사용하여 전체 음성의 평균 SNR을 구하였으며, 여기서 N은 총 프레임 수를 의미한다.

$$SNR(i) = 10\log(s^2(n)/n^2(n)) \quad (1)$$

$$n^2(n) = \sum_{n=0}^{Frame-1} (s(n) - \hat{s}(n))^2$$

$$s^2(n) = \sum_{n=0}^{Frame-1} (s(n))^2$$

$$Ave_{SNR} = \frac{1}{N} * \sum_{i=1}^N SNR(i) \quad (2)$$

평균 스펙트럼 왜곡 측정을 위해 식(3)을 사용하였으며, 여기서 $S(\omega)$ 는 full rate으로 부호화/복호화된 음성신호의 LPC 스펙트럼이고, $\hat{S}(\omega)$ 은 기본 rate을 사용하는 EVRC와 제안한 알고리즘을 적용한 부호화/복호화된 음성 스펙트럼에 해당하는 LPC 스펙트럼을 의미한다. 또한, N은 프레임의 총 개수를 의미한다.

$$ASDM = \frac{1}{N} \sum_{m=1}^N \sqrt{\frac{1}{\pi} \int_0^{\pi} [10\log_{10} S(\omega) - 10\log_{10} \hat{S}(\omega)]^2 d\omega} \quad (3)$$

표 1과 표 2에서는 객관적 음질평가를 수행한 결과를 나타내었으며, 신호 대 잡음비의 측정결과에서는 제안한 알고리즘이 평균 53%가 증가하였다. 그리고, 평균 스펙트럼 왜곡의 측정결과는 약 57%가 감소하여 음질향상을 가져옴을 입증하였다. 제안한 알고리즘을 사용한 음성부호화기에서 음질개선 효과가 큰 이유는 LSP 파라미터의 분포특성을 이용한 정확한 성분분리를 수행하기 때문이다. 예를 들어, 무성음 구간이 잡음구간으로 판단되는 등의 오류를 줄여 정확한 rate 결정을 수행하게 된다.

[표 1] SNR 실험 결과

[Table 1] Experimental results of SNR

	EVRC	제안한 방법	증가율[%]
발성 1	36.93	50.24	36.04
발성 2	38.95	53.97	38.56
발성 3	33.98	56.31	65.71
발성 4	38.15	66.88	75.31

[표 2] 평균스펙트럼 왜곡 측정 실험 결과

[Table 2] Experimental results of ASDM

	EVRC	제안한 방법	감소율[%]
발성 1	1.2106	0.8377	30.8
발성 2	1.1152	0.5290	52.6
발성 3	0.8458	0.0925	89.1
발성 4	0.8529	0.3624	57.2

표 3에서는 전송률 측정결과를 나타내었다. 즉, LSP 파라미터의 분포특성을 통해 성분분리 과정을 거친 제안한 알고리즘을 적용한 음성부호화기에서 평균 3.82%가 증가하였으며, 이는 성분분리 과정의 오버헤드가 추가되었음을 의미한다.

[표 3] 전송률에 대한 실험결과[단위:kbps]

[Table 3] Experimental results of transmission bit rate

	EVRC	제안한 방법	증감율[%]
발성 1	7.51	7.99	+6.39
발성 2	8.86	9.04	+2.03
발성 3	9.25	9.57	+3.46
발성 4	8.87	9.17	+3.38

5. 결론

EVRC 부호화 방식은 이동통신 시스템에서 채널용량을 효율적으로 사용하기 위해 세 가지의 전송률을 가변적으로 사용한다. 본 연구에서는 LSP 파라미터의 분포특성을 이용하여 유성음 구간, 무성음 구간, 묵음구간을 구분하며, 유성음인 경우에 대해 1 rate, 무성음 구간의 경우는 1/2 rate, 묵음의 경우에는 1/8 rate으로 전송하는 방법에 대하여 제안하였다. 즉, EVRC full rate으로 보내는 부분은 그대로 적용하며, half rate은 유/무성음 구간을 구분하여 유성음 구간에 대해서는 full rate으로 바꾸어 전송하였다. 실험결과에서 신호 대 잡음비의 경우, 제안한 방법이 약 53%가 증가하고 평균 스펙트럼 왜곡의 측정결과는 약 57%가 감소하여 음질개선의 효과가 높았지만, 전송률에서는 오버헤드가 추가되어 약 3.8% 증가함을 알 수 있었다. 본 논문에서 제안한 알고리즘의 음질개선 효과가 큰 이유는 성분분리를 효과적으로 수행하여 무성음 구간을 잡음구간으로 판단되는 오류를 줄여 정확한 rate 결정을 하였기 때문이라고 판단되어진다. 또한, 에너지가 낮은

유성 자음구간의 경우에도 효율적으로 성분분리를 수행하였음을 알 수가 있었다. 향후, 성분분리 알고리즘에 대한 다양한 연구를 통해, 지속적인 EVRC의 음질개선과 전송률 감소를 위한 연구를 수행할 계획이다.

References

- [1] TIA/EIA-IS-127 Enhanced Variable Rate Codec, 1997.
- [2] S. Saito and K. Nakata, Fundamentals of Speech Signal Processing, pp. 126~132, Academic Press, 1985.
- [3] A.M. Kondoz, Digital Speech, pp. 84~96, John & Sons Ltd, 1994.
- [4] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, pp. 367, Prentice Hall, 1978.
- [5] S.Y. Min and H.C. Kim, "The V/UV Decision Algorithm for a Reduction of the Transmission Bit Rate in the CELP Vocoder," The Journal of Korea Navigation Institute, Vol.11, No.3, pp. 87~92, 2007.
- [6] S.Y. Min, "Complexity Reduction Algorithm of Speech Coder(EVRC) for CDMA Digital Cellular System," Journal of Korea Multimedia Society, Vol.10, No.12, pp. 1551~1558, 2007.
- [7] H. Y. Gong, "A Novel AOCG-OFDM Modulation Technique for Variable-high-bit-rate", Journal of The Institute of Webcasting, Internet and Telecommunication, Vol 10, No 2, pp. 152~158, 2010.
- [8] J. C. Kim, "Channel Capacity of Additional Data Transmission System for Spectrum Efficiency Improvement", Journal of The Institute of Webcasting, Internet and Telecommunication, Vol 10, No 4, pp. 145~150, 2010.

나 덕 수(Deok-Su Na)

[정회원]



- 1998년 2월 : 송실대학교 정보통신공학과(공학사)
- 2000년 2월 : 송실대학교 정보통신공학과(공학석사)
- 2009년 2월 : 송실대학교 정보통신공학과(공학박사)
- 2001년 11월 ~ 현재 : (주)보이스웨어 연구소 책임보

<관심분야>

음성합성 시스템, 음성 신호처리

민 소 연(So-Yeon Min)

[중신회원]



- 1994년 2월 : 송실대학교 전자공학(공학사)
- 1996년 2월 : 송실대학교 일반대학원 전자공학과(공학석사)
- 2003년 2월 : 송실대학교 일반대학원 전자공학과(공학박사)
- 2005년 3월 ~ 현재 : 서일대학교 정보통신과 부교수

<관심분야>

통신 및 신호처리, 임베디드 시스템