

비균등 표본화 기법과 음성 부호화로의 응용

A Nonuniform Sampling Technique and Its Application to Speech Coding

임병관

Byeong-Gwan Iem

강릉원주대학교 전자공학과

Department of Electronic Engineering, Gangneung-Wonju National University

요 약

국소적으로 선형적인 특성을 보이는 음성신호와 같은 신호의 데이터율을 감소시키기 위하여 변곡점 검출에 기반한 비균등 표본화 방법을 제안한다. 국소적인 최대값과 최소값 검출에 기반하여 비균등 표본화를 수행하는 기존의 방법에 비하여 변곡점에 기반한 비균등 표본화는 신호의 기하학적인 특징을 충실히 활용한다. 결과로서, 변곡점 검출에 기반하여 비균등 표본화된 데이터를 보간법으로 처리하면 기존의 방법보다 원시신호를 정밀하게 복원할 수 있다. 컴퓨터 모의실험을 통하여 기존의 최대값/최소값 검출 방법에 비교해서 제안된 변곡점 검출 기반의 비균등 표본화가 약 9~23dB의 신호대 잡음비 개선효과가 있음을 확인하였다. 제안된 변곡점 검출 기반의 비균등 부호화의 유용성을 보이기 위하여 음성신호의 부호화에 적용하였으며, Continuously variable slope delta modulation (CVSD) 방법과 성능을 비교하였다. 변곡점 표본은 1로 설정된 플래그와 함께 이진수로 전송되며, 비 변곡점은 플래그만 0으로 설정된다. 음성신호에 따라 약 0.3 ~ 9dB의 신호대 잡음비 개선효과가 있으며, 주관적인 성능지표인 Mean Opinion Score (MOS)는 약 0.5 ~ 1.3 개선되었다.

키워드 : 비균등 표본화, 변곡점 검출, 음성의 가변 부호화, 최대값/최소값 검출

Abstract

For a signal such as speech showing piece-wise linear shape in a very short time period, a nonuniform sampling method based on the inflection point detection (IPD) is proposed to reduce data rate. The method exploits the geometrical characteristics of signal further than the existing local maxima/minima detection (MMD) based sampling method. As results, the reconstructed signal by the interpolation of the IPD based sampled data resembles the original speech more. Computer simulation shows that the proposed IPD based method produces about 9~23 dB improvement over the existing MMD method. To show the usefulness of the IPD technique, it is applied to speech coding, and compared to the continuously variable slope delta modulation (CVSD). The nonuniformly sampled data is binary coded with one bit flag set "1". Noninflection samples are not sent, but only flag bits set 0 are sent. The method shows 0.3 ~ 9 dB SNR and 0.5 ~ 1.3 mean opinion score (MOS) improvements over the CVSD.

Key Words : Nonuniform Sampling, Inflection Point Detection, Variable Bitrate Speech Coding, Maxima/Minima detection.

1. 서 론

신호의 부호화는 신호 전송, 저장 등에서 폭 넓게 활용되는 기술이며, 전송 및 저장되는 신호의 데이터율을 감소시키기 위하여 다양한 방법이 연구되어 왔다 [1]. 이러한 부호

화는 신호의 균등 표본화 이후에 적용된다. 데이터율의 추가적인 감축을 위하여 표본화 단계에서도 정보량을 감소시킬 필요가 있으며, 기존의 균등 표본화의 대안으로 비균등 표본화 방법이 지속적으로 연구되어 왔다 [2,3]. 비균등 표본화의 주요 응용 분야로는 음성 부호화, 음성인식 등을 들 수 있다 [4-11]. 음성 부호화로의 간단한 응용 예로는 비균등 표본화를 한 후 이진수로 변환하여 전송하는 기법을 들 수 있으며, 수신측에서는 표본화 되지 않은 샘플을 보간법을 이용하여 복구한다 [3,4]. 또 다른 응용 예로서, 비균등 표본화를 한 후 이를 보간법으로 복구한 신호와 원시신호를 비교하여 그 차이 신호를 미리 준비된 코드들과 비교하여 패턴이 가장 가까운 신호의 주소를 전송하는 보코딩 방식의 부호화 응용이 있다 [9]. 이러한 비균등 표본화에서 활용되는 표본화 방식은 신호의 국소적인 최대값과 최소값을 검출하여 표본화하는 것이다. 따라서, 국소적인 최대값과 최소값

접수일자: 2013년 10월 18일

심사(수정)일자: 2013년 12월 10일

게재확정일자 : 2013년 12월 11일

† Corresponding author

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

사이의 값은 선형적인 특성을 갖는다고 가정한다. 그러나, 음성의 경우 국소적인 최대값과 최소값 사이에서 단순한 선형적인 특성을 보이지 않는 경우가 대부분이다. 따라서 본 논문에서는, 비균등 표본화 방법으로 국소적인 최대값과 최소값에 더하여 국부적으로 신호의 기울기가 변화하는 위치, 즉 변곡점을 검출하여 비균등 표본화를 구현한다. 기존의 비균등 표본화와 비교하여 음성의 기하학적인 특성을 충실히 반영하기 때문에 신호 복구시 신호대 잡음비의 성능이 향상될 것으로 기대된다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존의 비균등 표본화와 제안된 비균등 표본화를 설명하며, 3장에서는 두 비균등 표본화 기법을 음성표본에 적용시 성능을 비교한다. 4장에서는 다양한 음성표본에 적용한 예와 성능을 비교하며 5장에서 결론을 맺는다.

2. 비균등 표본화 방법

시간에 따라 특성이 변하는 신호의 경우, 매우 짧은 시간 구간에서 관찰하면 그 특성이 선형적인 모습을 보일 수 있다. 그림 1은 시간에 따라 특성이 변화하는 대표적인 신호인 음성 신호의 유성음과 무성음의 매우 짧은 구간을 각각 확대한 그림이다. 그림에서 볼 수 있듯이 짧은 구간에서는 기하학적으로 선형적인 특성을 보인다. 따라서, 비균등하게 표본화하면 균등하게 표본화할 때와 비교하여 데이터의 양을 줄일 수 있다 [4].

대표적인 비균등 표본화 방법으로는 구간별로 음성의 국소적인 최대값과 최소값만을 표본화하는 방법이 있다 [3]. 그림 2는 이와 같이 표본화하고 이를 다시 보간법을 이용하여 복구한 예를 보인다. 그림 2 (a)는 원 음성 신호를 보이며, (b)는 국지적인 최대값과 최소값을 검출하여 비균등 표본화를 수행한 결과를 원 음성신호에 중첩해서 보여준다. 그리고 그림 2 (c)는 비균등 표본화 결과를 보간법으로 처리하여 복구한 신호를 보인다. 균등 표본화에 비교해서, 비균등 표본화로 약 62%의 표본 데이터 수를 절감하는 효과가 있다. 복구된 신호는 대체로 원음성의 특성을 보여주고 있다. 그러나, 위 그림 1에서 관찰할 수 있듯이 음성의 국소적인 선형적인 특성은 단순히 국소적인 최대값과 최소값의 연결로만 볼 수 없으며, 최소값과 최대값 사이에 변곡점이 있는 경우가 대부분이다. 그림 3은 국소적인 최대값과 최소값을 포함하여 가능한 모든 변곡점의 형태를 보여준다. 따라서, 본 논문에서는 변곡점을 모두 검출하여 비균등 표본화하는 방법을 제안한다.

변곡점을 모두 검출하기 위하여 신호의 순간 기울기의 변화를 측정하는 방법을 사용한다. 신호의 기울기는 아래 식과 같이 연속하는 신호 값의 차로 근사된다 [12].

$$y[n] = \frac{1}{2}(x[n] - x[n-1]) \quad (1)$$

국소적인 최대값과 최소값은 기울기 $y[n]$ 의 부호가 변화할 때 발생하며, 단순 변곡점은 기울기 $y[n]$ 값이 부호는 변하지 않는 가운데 값만 미리 설정된 임계값보다 크게 변화할 때 발생한다고 생각할 수 있다. 이를 정리하면 아래와 같다.

$$\text{조건 i) } y[n] \cdot y[n-1] < 0 \quad (2)$$

$$\text{조건 ii) } |y[n] - y[n-1]| > \text{임계값} \quad (3)$$

즉 조건 i)은 국소적인 최대값 혹은 최소값을 검출하는 조

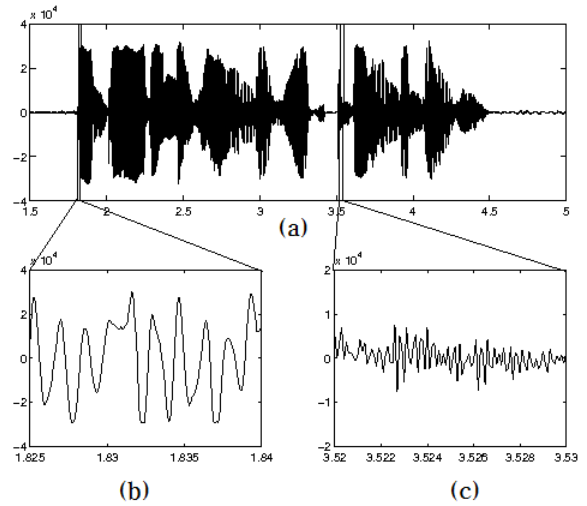


그림 1. 음성의 유성음 구간과 무성음 구간을 확대한 그림 (a) 음성 신호 (b) 유성음 구간 확대그림 (c) 무성음 구간 확대 그림

Fig. 1. Enlarged plot of voiced and unvoiced parts of a speech signal (a) Speech signal (b) Enlarged plot of the voiced (c) Enlarged plot of the unvoiced

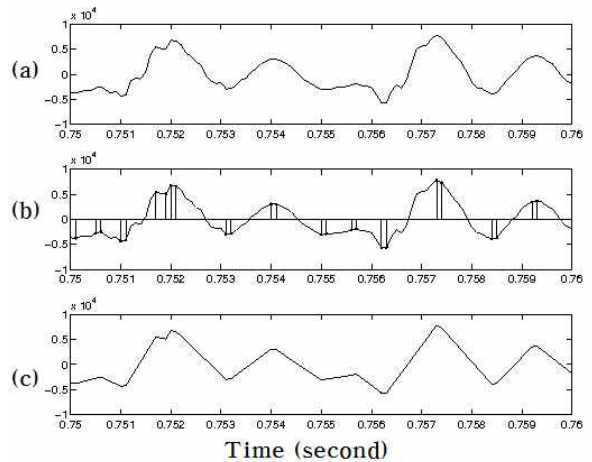


그림 2. 최소값/최대값 검출에 기반한 비균등 표본화 a) 원시음성 b) 비균등 표본화 결과 c) 보간법으로 복구된 신호

Fig. 2. Local maxima/minima detection based nonuniform sampling a) Original signal b) Nonuniformly sampled data and c) Reconstructed signal by interpolation

건이며, 조건 ii)는 순수한 변곡점을 검출하는 조건이다. 위 조건 가운데 하나라도 충족이 되면 변곡점으로 판정하여 표본화한다. 조건 ii)에서 임계값을 조정하여 검출되는 변곡점의 양을 조절할 수 있다. 조건 i)과 ii)가 모두 만족되지 않는 경우는 비변곡점으로 판정되어 표본화 하지 않는다. 그림 4는 이러한 변곡점 검출 알고리즘을 흐름도로 표현한 것이다. 그림 5는 그림 2에서 사용한 동일한 음성신호를 제안된 변곡점 검출 알고리즘으로 처리하여 비균등 표본화한 결과이다. 그림 5 (a)는 원 음성 신호를 보이며, (b)는 모든 변

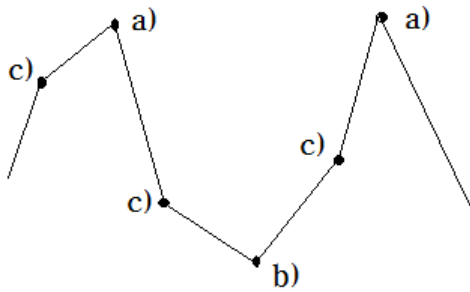


그림 3. 변곡점의 다양한 형태 a) 국소적인 최대값 b) 국소적인 최소값 c) 단순 변곡점
 Fig. 3. Various types of inflection points a) Local maxima b) Local minima and c) Pure inflection points

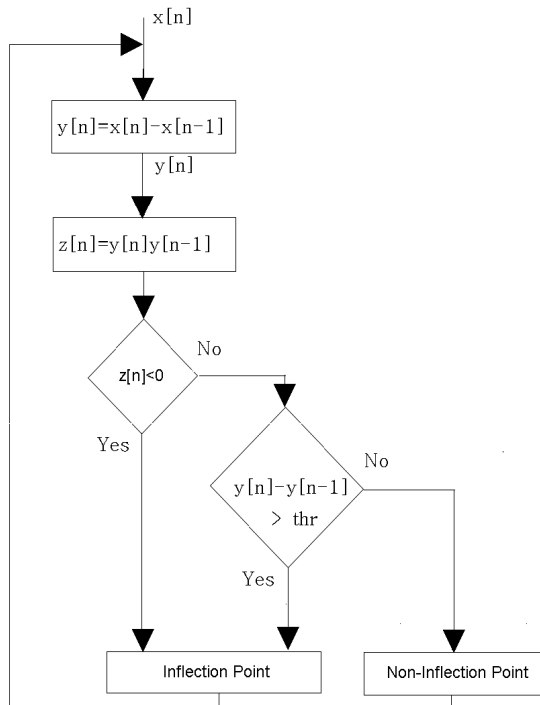


그림 4. 변곡점 검출 알고리즘
 Fig. 4. Inflection point detection algorithm

곡점을 검출하여 비균등 표본화를 수행한 결과를 원 음성신호에 중첩해서 보여준다. 그리고 그림 5 (c)는 비균등 표본화 결과를 보간법으로 처리하여 복구한 신호를 보인다. 그림 2의 국소적인 최대값 혹은 최소값에서만 표본화하는 비균등 표본화와 비교해서, 모든 변곡점에서의 데이터를 표본화하므로 더 많은 표본샘플을 검출할 수 있다. 그 결과로 보간법으로 복구된 신호가 그림 2의 경우보다 충실하게 원 신호를 근사함을 볼 수 있다. 구체적으로 그림 2와 그림 5의 약 7.51~7.52초 구간, 0.756~0.757 구간 등을 비교하면 그림 5의 변곡점을 추출한 경우가 원 신호를 충실히 복구함을 볼 수 있다. 다음 장에서는 모의 실험을 통해 정량적으로 기존의 비균등 표본화와 제안된 비균등 표본화의 성능을 비교한다.

3. 비균등 표본화 성능 분석

국소적인 최대값/최소값 검출에 기반한 비균등 표본화와 변곡점 검출에 기반한 비균등 표본화를 컴퓨터 모의실험을 통해 비교한다. 사용된 신호는 음성 신호로 유성음 구간과 무성음 구간이 적절히 조합된 신호이다. 비교 척도는 신호 대 잡음비이며, 잡음은 원신호와 복구된 신호의 차로 정의된다. 신호는 10kHz로 표본화된 데이터를 사용하였다. 제안된 변곡점 추출에 기반한 비균등 표본화 기법은 식(3)의 조건 ii)에서 임계값을 조정하여 변곡점의 양을 조절할 수 있다. 즉 임계값이 커지면 검출되는 변곡점의 양이 줄어들며, 임계값이 작아지면 변곡점의 양이 증가한다. 변곡점 검출에 기반한 비균등 표본화는 기존의 최대값 및 최소값 검출에 기반한 비균등 표본화보다 표본 수가 증가할 것이며, 표본수의 증가량을 비교키 위하여 임계값에 따른 데이터 율도 제시한다. 임계값에 따른 변곡점 수의 변화는 신호 형태에 의존하며, 음성이 시간에 따라 특성이 변화하기 때문에 임계값과 변곡점 수 사이에는 선형적인 관계가 있지는 않다. 표본화된 신호는 8비트로 양자화 된다고 가정한다. 두 비균등 표본화 방법으로 처리한 결과가 그림 6에서 보인다. 제안된 변곡점 검출에 기반한 비균등 표본화 기법이 국소적인 최대값/최소값 검출에 기반한 비균등 표본화 보다 약 9 ~ 23 dB의 신호대 잡음비 개선을 보인다. 그에 따른 비용으로 기존의 비균등 표본화 대비 약 7 ~ 12 kbps의 데이터 율 증가를 보인다. 그러나 통상의 Pulse Code Modulation (PCM) 부호화에 비교하여 낮은 데이터 율을 보인다. 그림 6에서 임계값이 클수록 기존의 최대값/최소값 검출 방법에 비교해서 적은 데이터 율의 증가 속에 신호대 잡음비를 개선할 수 있음을 알 수 있다.

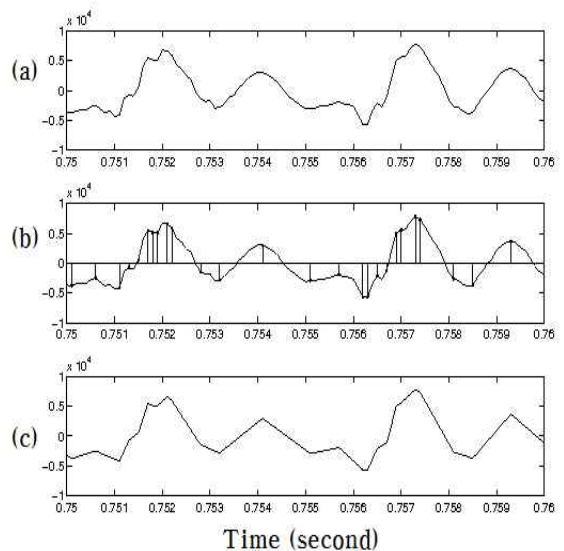


그림 5. 변곡점 검출에 기반한 비균등 표본화 처리 결과 a) 원시음성 b) 비균등 표본화 결과 c) 보간법으로 복구된 신호

Fig. 5. Inflection point detection based nonuniform sampling a) Original signal b) Nonuniformly sampled data and c) Reconstructed signal by interpolation

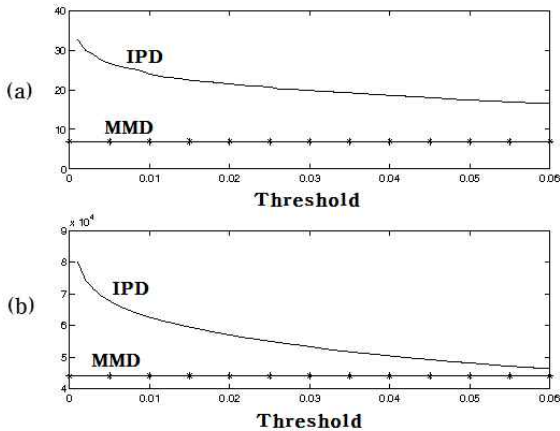


그림 6. 최대값/최소값 검출 비균등 표본화기법 (*표된 실선)과 변곡점 검출 비균등 표본화기법 (실선)의 비교 (a) 임계값에 대한 신호대 잡음비 (dB) (b) 임계값에 따른 데이터율 (kbps)

Fig. 6. Comparison of the maxima/minima detection(MMD) based nonuniform sampling (solid line with *) and the inflection point detection(IPD) based nonuniform sampling techniques (solid line) (a) SNR vs. threshold (b) number of inflection points vs. threshold

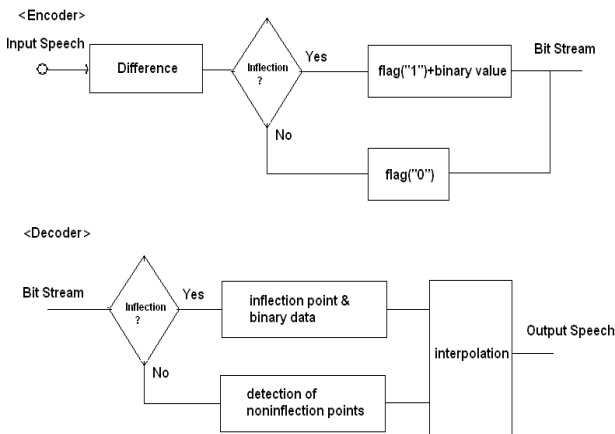


그림 7. 변곡점 검출에 기반한 비균등 표본화를 활용한 음성부호화기와 복호화기 a) 음성부호화기 b) 복호화기
Fig. 7. Speech encoder and decoder based on the inflection point detection a) Encoder b) Decoder

4. 음성부호화 응용 예

본 장에서는 제안된 변곡점 검출에 기반한 비균등 표본화의 유용성을 보이기 위하여 음성 부호화에 적용하는 예를 보인다. 기존의 최대값/최소값 검출에 기반한 비균등 표본화의 경우, 다양한 음성부호화로 구현되었다. 대표적인 예로는 단순히 검출된 표본과 위치를 PCM부호화 하고 수신측에서 보간법으로 복원하는 방법 [4], 보코딩 형태의 부호화기에 적용하는 방법 [9]을 들 수 있다. 본 논문에서는 변곡

점으로 검출된 표본의 위치와 값을 전송하고 수신측에서 선형 보간법으로 신호를 복원하는 방법을 적용한다 [13]. 따라서 부호화기와 복호화기는 그림 7과 같은 구조를 갖는다.

그림 7의 부호화기에서는 음성신호의 변곡점을 검출하여 비균등 표본화를 수행한다. 검출된 변곡점은 그 위치를 나타내기 위하여 1 비트 플래그를 “1”로 설정하고 그 뒤에 변곡점에서의 신호값을 이진수로 양자화한다. 비변곡점의 경우 “0”으로 1 비트 플래그를 설정하여 전송한다. 수신측에서는 “1”로 설정된 플래그 비트에서의 데이터 값에 보간법을 적용하여 “0”으로 설정된 플래그 비트에서의 데이터 값을 복원한다.

표 1은 이렇게 처리된 음성 신호의 개관적인 성능지표인 신호대 잡음비와 주관적인 성능지표인 Mean Opinion Score (MOS)를 보인다. 신호대 잡음비 계산시 잡음은 원 신호와 복구된 신호의 차로 정의한다. 주관적인 평가인 MOS는 성인 134명을 대상으로 얻었다. MOS는 가장 나쁜 음질을 1로, 가장 좋은 음질을 5로 채점했을 때의 평균치이다 [14]. 비교를 위하여 32kbps CVSD (Continuously Variable Slope Delta modulation) 방식의 델타 변조의 성능을 산출하였다 [1,14]. 변곡점 검출에 기반한 부호화기는

표 1 변곡점 검출 비균등 표본화 기법의 성능분석

Table 1. Performance analysis of the inflection point detection(IPD) based nonuniform sampling techniques

sample speech	IPD based coder			32kbps CVSD	
	data rate (kbps)	SNR (dB)	MOS	SNR (dB)	MOS
#1	44	17	4.03	12.22	2.71
#2	41.4	21.1	4.06	12.38	3.03
#3	44.2	16.8	3.99	11.88	2.63
#4	38.6	16.68	3.85	16.3	3.38

음성에 따라 변곡점의 검출량이 다르기 때문에 가변적인 데이터율을 보인다. 대체로 무성음이 많은 음성에서 변곡점이 많으므로 유성음이 많은 음성보다 데이터율이 높은 결과를 보인다. CVSD방식의 부호화보다 6 ~ 12 kbps 만큼 높은 데이터율을 보이며, 신호대 잡음비는 0.3 ~ 9 dB까지 개선된다. 주관적인 평가의 경우 0.5 ~ 1.3점까지의 개선을 보인다. 데이터율의 경우 부호화 방식을 개선하여 줄일 수 있는 여지가 있다고 사료된다.

5. 결론 및 향후 연구

본 논문에서는 변곡점 검출에 기반한 비균등 표본화 방법을 제안하였다. 제안된 변곡점 검출에 따른 표본화는 기존의 국소적인 최대값과 최소값 검출에 기반한 비균등 표본화에 비교해서 약 9 ~ 23dB의 신호대 잡음비를 개선함을 컴퓨터 모의실험을 통하여 확인했다. 제안된 방법이 음성부호화에 활용될 때, PCM부호화에 비교해서 데이터율이 압축됨을 보였고 기존의 CVSD보다 신호에 따라 약 0.3 ~ 9dB까지 신호대 잡음비가 개선됨을 보였다. 아울러 주관적인 성능지표인 MOS에 있어서는 약 0.5~1.3 개선됨을 보였다. 향후 연구과제로는 음성 부호화에 적용시, 잡음환경에서도 내성을 갖도록 각 표본을 부호화하는 방법을 연구할 필요가 있다.

References

- [1] A.M. Kondoz, *Digital Speech*, John Wiley & Sons, England, 1994.
- [2] L. D. Davisson, "Data compression using straight line interpolation," *IEEE Trans. on Information Theory*, vol. IT-14, No.3, pp. 390-394, 1968.
- [3] J. W. Mark, and T. D. Todd, "A nonuniform sampling approach to data compression," *IEEE Trans. on Communications*, vol. COM-29, No.1, pp. 24-32, 1981.
- [4] M. Budaes, and L. Goras, "On speech signal reconstruction from local extreme values," *Proc. of ISSCS*, vol. I, pp. 315-318, 2005.
- [5] S. Elramly, S. G. Foda, and M. El-shafie, "Continuous variable sampling rate, application on speech," *Proc. of IEEE ISCC*, pp. 189-193, 1997.
- [6] M. R. Nakhai, and F. A. Marvasti, "Application of extremum sampling in speech coding," *Proc. of IEEE ICASSP*, vol. 6, pp. 3842-3845, 2000.
- [7] T. Fjällbrant, "Method of data reduction of sampled speech signals by using nonuniform sampling and a time-variable digital filter," *Electronics Letters*, vol. 13, No.11, pp. 334-335, 1977.
- [8] P. K. Ghosh, and T. V. Sreenivas, "Dynamic programming based optimum non-uniform samples for speech reconstruction and coding," *Proc. ICASSP*, vol. I, pp. 1221-1224, 2006.
- [9] M. Bae, W. Lee, and S. Im, "On a new vocoder technique by the nonuniform sampling," *Proc. of IEEE MILCOM*, vol.2, pp. 649-652, 1996.
- [10] G. Lee and W. Kim, "Robust speech parameters for the emotional speech recognition," *Journal of the Korea Institute of Intelligent Systems*, vol. 22, pp. 681-686, 2012.
- [11] W. Kim, "Emotion robust speech recognition using speech transformation," *Journal of the Korea Institute of Intelligent Systems*, vol. 20, pp. 683-687, 2010.
- [12] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-Part 2: Algorithms and applications," *Proc. IEEE*, vol. 80, pp. 540-568, 1992.
- [13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The art of scientific computing*, Cambridge University Press, London, U.K., 1986.
- [14] L. Rabiner and R. Schafer, *Digital processing of speech signals*, Prentice-Hall, NJ, 1978.

저 자 소 개



임병관(Byeong-Gwan Iem)

1988년 : 연세대학교 전자공학과 공학사

1990년 : 연세대학교 대학원 전자공학과 공학석사

1998년 : Univ. of Rhode Island, 전기공학과 공학박사

2002년 ~ 현재 : 강릉원주대학교 전자공학과 부교수

관심분야 : 디지털신호처리, 음성신호처리

Phone : +82-33-640-2426

E-mail : ibg@gwnu.ac.kr