

비균일 표본화된 음성 신호에서의 기본적인 신호처리

*임재열, 나경민, 안수길
서울대학교 전자공학과 통신이론실

Fundamental Signal Processing in NonUniformly Sampled Speech Signal

JaeYeol RHEEM, KyungMin NA, and SouGiil ANN
Dept. of Electronics Engr., Seoul National University, Seoul 151-742, KOREA

요약 - 극점에서 비균일 표본화된 음성 신호는 크기열과 간격열의 이중구조로 표현되어, 균일 표본화된 신호에 근거한 기존의 신호처리 방법을 그대로 적용할 수 없다. 본 논문에서는 비균일 표본화된 음성 신호에서 에너지, 크기, 영교차율 함수의 관계를 직접 유도하고, 특징을 살펴보면, 비균일 표본화된 음성 신호에서도 균일 표본화된 신호에 해당하는 에너지, 크기, 영교차율과 같은 전처리과정 파라미터의 추정이 가능함을 확인 한다.

1. 서 론

최근, 음성신호의 효율적인 저장이나 전송을 위하여, 표본간의 상관관계(correlation)가 적은 방향으로 표본화(sampling)를 수행하는, 극점에서의 비균일 표본화 방법(nonuniform sampling method)이 제안되었다 [1-3]. 제안된 비균일 표본화 방법은, 음성의 인지가 기본적으로 음성 파형의 극점 간의 간격에 의존하고 있다는 과거의 실험적 사실에 근거하여, 음성신호를 극점에서 비균일 표본화하는 방법이다.

음성을 극점에서 비균일 표본화하여 나타내면, 균일 표본화에 근거한 기존의 디지털 표현기법과는 다르게, 크기열과 간격열이라는 이중구조로 나타난다 [1]. 따라서, 균일 표본화에 근거한 기존의 디지털 신호처리기법을 비균일 표본화된 신호에 직접 적용할 수 없다.

비균일 표본화된 음성신호에서의 디지털 신호처리 방법에는 다음의 두 가지를 고려할 수 있다. 첫째는 비균일 표본화된 음성신호를 재구성(reconstruction)하여 균일 표본화시에 해당하는 신호를 추정하여 얻고, 이것에 기존

의 디지털 신호처리 방법을 적용하는 간접적인 방법이며, 두 번째 방법은 비균일 표본화된 신호에서 균일 표본화된 신호에서의 파라미터를 직접 추정하는 직접적인 방법이다. 비균일 표본화된 신호로부터 원래의 균일 표본화된 신호를 재구성하는 방법이 간단하므로 첫 번째 방법도 매우 유용한 방법이나, 비균일 표본화된 신호와 균일 표본화된 신호의 상호관계 규명 및 특성을 파악하기 위해서는 두 번째 방법이 고찰되어야 한다.

따라서 본 논문에서는, 비균일 표본화된 신호와 균일 표본화된 신호의 상호관계 규명 및 특성을 파악하기 위해서, 음성 신호처리의 기본적인 파라미터라 할 수 있는 에너지, 크기, 영교차율 함수를 비균일 표본화된 신호에서 유도하여 균일 표본화된 신호에서의 상호비교하기로 한다.

II. 비균일 표본화된 음성 신호

표본화주파수 F_s 로 표본화되고 L 레벨로 양자화된 디지털 신호를 $x(n)$, $-\infty < n < \infty$ 이라 하면, $x(n)$ 을 극대점 및 극소점 시각 n_k 에서 비균일 표본화하여 얻은 신호 $x_{nks}(n)$ 은

$$x_{nks}(n) = x(n) \sum_k \delta(n - n_k) \quad (1)$$

으로 나타낼 수 있다. 여기서 $\sum_k \delta(n - n_k)$ 는 비균일 표본화 임펄스열이며, n_k 는 비균일 표본화 시각을 나타내고, $\delta(\cdot)$ 는 Kronecker delta 함수이다 [4].

식 (1)을 살펴보면, 그림 1에서와 같이, 비균일 표본화

비균일 표본화된 음성신호에서의 기본적인 신호처리

원 신호 $x_{analog}(n)$ 은 균일 표본화된 신호 $x(n)$ 에서 비균일 표본화된 극대점 및 극소점 시각 n_k 에서만 표본값을 유지하고 그 외의 점에서는 표본 값이 0인 신호가 됨을 알 수 있다. 이렇게 얻어진 비균일 표본화된 신호 $x_{non}(n)$ 는, 효율적인 저장 및 전송을 위하여, 그림 1의 (c)에서와 같이, 시각 n_k 에서의 크기 $a(k)$ 와 이전 비균일 표본점과의 간격 $i(k)$ 로 나타내어져 저장 또는 전송되며, 크기와 간격을 $\{a(k), i(k)\}$ 로부터 다음의 관계식으로 $x_{non}(n)$ 을 얻을 수 있다.

$$a(k) = x(n_k), \quad i(k) = n_k - n_{k-1}, \quad n_k = \sum_{p=1}^k i(p). \quad (2)$$

디지털화된 음성 신호에서의 극대 및 극소점의 결정은 연속한 두 표본의 기울기의 곱을 이용하여 결정할 수 있으며, 비균일 표본화된 신호에서의 재구성(reconstruction)은 각각의 비균일 표본화 점들의 값을 간격을 고려하여 선형보간(linear interpolation) 또는 삼각함수 보간(sinusoidal interpolation)을 이용하면 된다 [1][2].

비균일 표본화된 k 번째 표본의 크기와 간격을 각각 $a(k)$ 와 $i(k)$ 라 하면, 이전 표본 크기 $a(k-1)$ 이후에서 $a(k)$ 까지 선형보간에 의해서 재구성된 신호 $y'_k(n+m)$ 와 삼각함수 보간법을 이용하여 재구성된 신호 $y''_k(n+m)$ 은

$$y'_k(n+m) = \left[\frac{a(k) - a(k-1)}{i(k)} \cdot m + a(k-1) \right]_L \quad (3)$$

$$y''_k(n+m) = \left[\frac{a(k-1) - a(k)}{2} \cdot \cos\left(\frac{\pi}{i(k)} \cdot m\right) + \frac{a(k-1) + a(k)}{2} \right]_L \quad (4)$$

으로 표현된다. 여기서 n 은 $k-1$ 번째 비균일 표본의 균일 표본화시의 시각이며, $a(k-1) = y_{k-1}(n)$ 이고, $1 \leq m \leq i(k)$ 이며, $[\]_L$ 은 L 레벨의 양자화함수이다.

III. 비균일 표본화된 신호에서의 에너지, 크기, 영교차율 함수

균일 표본화로 얻은 디지털 신호 $x(n)$, $-\infty < n < \infty$ 에서, 구간 $[m-N+1, m]$ 의 N 개 표본에 대한 시각 m 에서의 표

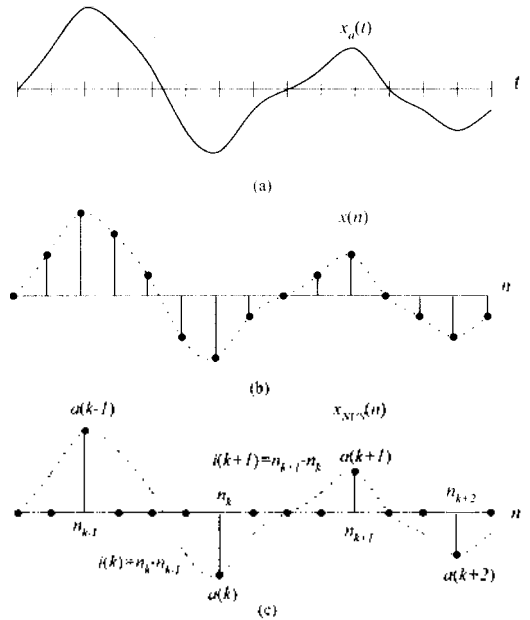


그림 1. 균일 표본화된 신호와 비균일 표본화된 신호의 비교; (a) 아날로그 신호, (b) 균일 표본화된 신호, (c) 비균일 표본화된 신호

본수에 의한 구간 평균 에너지(short-time average energy)는

$$E_{avg}(m) = \frac{1}{N} \sum_{n=m-N+1}^m x^2(n) \cdot w(m-n) \quad (5)$$

로 정의된다. 여기서 $w(n)$ 은 윈도우 함수로 직각 윈도우(rectangular window)인 경우에는

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{elsewhere} \end{cases}$$

로 정의된다 [4].

크기 함수(magnitude function)는 에너지 함수가 신호를 제곱함으로써 큰 신호 값에 민감하다는 단점을 줄이기 위하여 사용된다. 구간 평균 크기 함수(short-time average magnitude function)는

$$M_{avg}(m) = \frac{1}{N} \sum_{n=m-N+1}^m |x(n)| w(m-n) \quad (6)$$

로 정의된다.

구간 영교차율(short-time zero-crossing rate)은 신호에 대한 거친 주파수특성을 추정하게 하는 파라미터로 다음과 같이 정의된다.

$$Z(m) = \sum_{n=-\infty}^{\infty} \text{sgn}[x(n)] - \text{sgn}[x(n-1)] u(m-n). \quad (7)$$

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

$$u(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{elsewhere} \end{cases}$$

균일 표본화된 신호 $x(n)$ 을 극대점 및 극소점 시간 n_k 에서 비균일 표본화하여 얻은 신호 $x_{NIS}(n)$ 에 대한 식 (1)을 식 (5)와 (6)에 대입하면 다음과 같이 비균일 표본화된 신호에서의 구간 평균 에너지와 크기를 구할 수 있다.

$$E_{avg}^{NIS}(m) = \frac{1}{N} \sum_{n=-\infty}^{\infty} x_{NIS}^2(n) u(m-n)$$

$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} \left[x(n) \sum_k \delta(n-n_k) \right]^2 u(m-n) \quad (8)$$

$$= \frac{1}{N} \sum_k x^2(n_k) u(m-n_k)$$

$$M_{avg}^{NIS}(m) = \frac{1}{N} \sum_{n=-\infty}^{\infty} |x_{NIS}(n)| u(m-n)$$

$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} \left| x(n) \sum_k \delta(n-n_k) \right| u(m-n) \quad (9)$$

$$= \frac{1}{N} \sum_k |x(n_k)| u(m-n_k)$$

그러나 비균일 표본화된 신호에서의 영교차율은 식 (1)을 식 (7)에 직접 대입하여 유도할 수 없다. 왜냐하면, 균일 표본화된 신호 $x(n)$ 을 비균일 표본화하여 얻은 신호 $x_{NIS}(n)$ 이 비균일 표본점 n_k 를 제외한 점에서는 모두 그 값이 0이기 때문에, 식 (1)을 식 (7)에 대입하여 얻은 식은 결과적으로 영교차율 대신에 한 윈도우 내의 음의 값을 갖는 극점 수의 비율을 얻게 되기 때문이다.

그런데 연속한 두 극점의 부호가 다를 경우, 그 사이 반드시 하나의 영교차가 존재해야 하므로, 비균일 표본화된 신호에서의 영교차율은 비균일 표본화된 극점과 극점, 즉 시간 n_{k-1} 와 n_k 에서의 신호 값의 부호를 이용하여 영교차율을 구할 수 있다. 따라서 비균일 표본화된 신호 $x_{NIS}(n)$ 에서의 영교차율은

$$Z^{NIS}(m) = \sum_k \left[\text{sgn}[x(n_k)] - \text{sgn}[x(n_{k-1})] \right] u(m-n_k) \quad (10)$$

$$u(n) = \begin{cases} \frac{1}{2N}, & 0 \leq n \leq N-1 \\ 0, & \text{elsewhere} \end{cases}$$

로 정의할 수 있다.

식 (8) - (10)을 살펴보면, 비균일 표본화된 신호에서의 구간 평균 에너지, 크기, 영교차율이 구간 내의 극점에만 관계됨을 관찰할 수 있다. 그런데 한 구간 내에서 극점의 수는 일반적으로 윈도우의 크기 N 보다 작으므로 식 (8)과 식 (9)로 구한 구간평균 에너지와 크기 값은 식 (5)와 식 (6)으로 구한 값보다 상대적으로 작아진다. 따라서 비균일 표본화된 신호에서 균일 표본화된 신호의 에너지 및 크기에 해당하는 값을 추정하기 위하여는 한 윈도우 내의 균일 표본점의 표본수 N 대신에 비균일 표본점의 개수로 평균을 취하여 주어야 한다. 따라서 추정된 구간 평균 에너지와 크기는

$$\hat{E}_{avg}^{NIS}(m) = \frac{1}{N_{NIS}(m)} \sum_k x^2(n_k) u(m-n_k) \quad (11)$$

$$= \frac{1}{N_{NIS}(m)} \sum_k a^2(k) u(m-n_k)$$

$$\hat{M}_{avg}^{NIS}(m) = \frac{1}{N_{NIS}(m)} \sum_k |x(n_k)| u(m-n_k) \quad (12)$$

$$= \frac{1}{N_{NIS}(m)} \sum_k |a(k)| u(m-n_k)$$

이 된다. 여기서 $N_{NIS}(m)$ 는 구간 $[m-N+1, m]$ 에 포함되는 비균일 표본화된 극점의 수이며, $a(k) = x(n_k)$ 는 그 극점의 값이고, $n_k = \sum_{p=1}^k j(p)$ 는 비균일 표본화 시간이다.

영교차율의 경우에는 한 구간 내의 균일 표본점의 개수 분에 영교차 표본점의 개수로 정의되므로 윈도우 함수에서는 $2N$ 대신 $2N_{NIS}(m)$ 으로 나누어줄 필요가 없다. 따라서 식 (10)을 이용하면 비균일 표본화된 신호에서의 영교차율을 추정할 수 있다.

$$\hat{Z}^{NIS}(m) = \sum_k \left[\text{sgn}[x(n_k)] - \text{sgn}[x(n_{k-1})] \right] u(m-n_k) \quad (10')$$

$$= \sum_k \left[\text{sgn}[a(k)] - \text{sgn}[a(k-1)] \right] u(m-n_k)$$

IV. 실험 및 결과

그림 2에는 8 kHz로 균일 표본한 발성/안녕하세요/에서, 구간 길이 256 표본으로 192 표본씩 이동시켜서 균일 표본화된 신호와 비균일 표본화된 신호 그리고 선형보간으로 재구성한 신호에서 구간, 구간 평균 에너지, 크기, 및 영교차율을 보여 준다.

그림 2를 살펴보면, 비균일 표본화된 신호에서 직접 구한 구간 평균 에너지, 크기 및 영교차율의 값들이 균일 표본화된 신호에서와 같이 일관되게 얼어짐을 관찰할 수 있다. 따라서 에너지 및 영교차율을 이용한 기존의 음성음/무성음/무음 구분 알고리즘 또는 끝점 검출(end-points detection) 알고리즘을 별다른 수정 없이 그대로 비균일 표본화된 신호에도 적용할 수 있음을 알 수 있다.

V. 결론

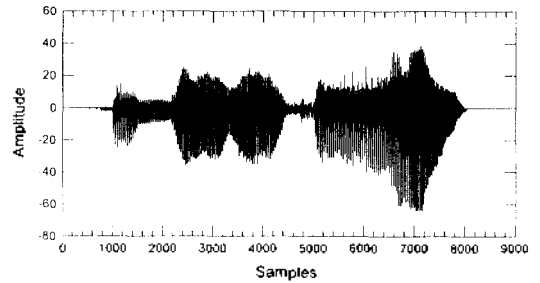
본 논문에서는 비균일 표본화된 음성 신호에서의 기본적인 신호처리로써 구간평균 에너지, 크기, 영교차율을 살펴보았다. 음성신호를 비균일 표본화하여 나타내면, 크기열과 간격열의 이중구조를 가지고 있기 때문에, 균일 표본화에 근거한 기존의 디지털 신호처리 방법을 직접 적용할 수 없어 구간 평균 에너지, 크기, 영교차율 등은 식 (10), (11), (12)에서와 같이 새로이 정의되어야 한다. 이렇게 비균일 표본화된 신호에서 정의된 값들은 균일 표본화된 신호에서의 값에 대하여 근사값이 되며, 균일 표본화시와 같이 음성 신호에 대하여 일관성을 모두 보이기 때문에, 기본적인 음성 파라미터로써 사용가능함을 확인하였다.

참 고 문 헌

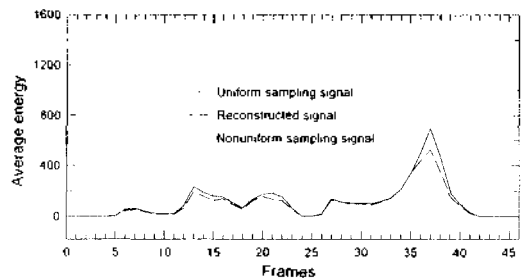
[1] JaeYeol RHEEM, BumHoon KIM, and SouGuil ANN, "A Nonuniform Sampling Method of Speech Signal and Its Application to Speech Coding," *Signal Processing* vol. 41, no. 1, pp. 43-48, Jan. 1995.
 [2] 임재열, 백성준, 안수길, 김범훈, "극대 및 극소점에서의 비균일 표본화에 의한 음성 압축," 한국음향학회지, 11권 4호, pp. 36-44, 1992.
 [3] 임재열, 김정일, 안수길, "국소 극대-극소점간의 간격 정보를 이용한 시간 영역에서의 음성인식을 위한 파라미터 추출 방법," 대한전자공학회논문지, vol. 31, no.

2, pp. 81-87, 1994.

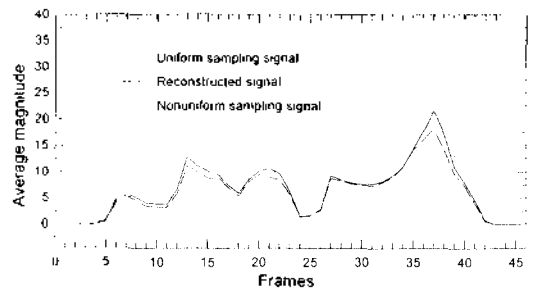
[4] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.



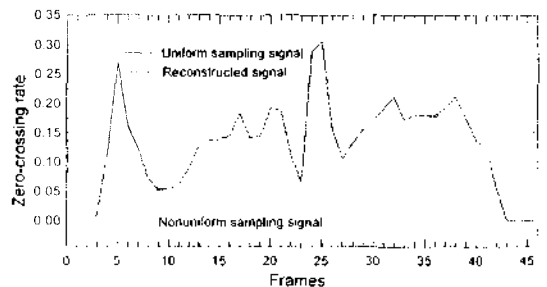
(a)



(b)



(c)



(d)

그림 2. 균일표본화된 신호와 비균일표본화된 신호에서의 에너지, 크기, 영교차율 비교: (a) 균일 표본화된 신호와 (b) 평균 에너지 (c) 평균 크기 (d) 영교차율