음성신호의 실시간 처리 기법에 관한 연구

○ 이 택수 *, 안 창 *, 김 성락 **, 이 상범 *

* 단국대학교 전자공학과, ** 관동대학 정보처리학과

A Study on the Real Time Processing Technique of Speech Signal

Taek-Soo Lee *, Chang Ahn *, Sung-Nak Kim **, Sang-Burm Rhee *

* Dankook University, ** Kwandong College

ABSTRACT

Zero-crossng analysis techniques have been applied to speech recognition. Zero-crossing rate, level-crossing rate and differentiated zero-crossing rate in time domain are used in analyzing speech signals. Speech samples could be stored in memory buffer in real time.

I. 서론

최근 음성신호를 인식하기 위하여 많은 연구가 진행되고 있다[1]. 이러한 음성을 처리하기 위해서는 신뢰도, 정확도, 처리 속도등이 중요한 요인이 된다. 특히 표본화된 데이타를 메모리에 저장하는 알고리즘이 효율적으로 구성되어야 한다.

본 논문에서는 alarm 방법이나 tape 방법 과는 달리 메모리 버퍼를 시작 버퍼와 나머지 버퍼로 나누어 peak를 이용한 영교차 간격(ZCI) 의 면적을 측정하여 Pitch를 구한다음, 이 Pitch를 연속 세번 발생하면 유성음이 검출된 것으로 인식하고 나머지 버퍼에 음성 데이터를 저장하도록 하였다. 특히 추출시간에 중점을 두어 영교차율과 이를 확장한 레벨 교차율(LCR), 미분 영교차율(DZCR)을 이용하였다. 영교차율은 신호의 크기에 무관하며 주파수 영역 데이타보다 화자에 덜 의존적이고 쉽게 추출할 수 있다는 이점이 있다[2]. 이러한 방법을 이용하여 메모리를 효율적으로 사용할 수 있고 음성의 실시간 처리에 적합한 알고리즘을 구성하고자 한다.

II. 신호 추출 방법

1. 영교차율 추출 방법

이산 신호에 표본값이 서로 다른 대수적 부호를 갖게 되면 영교차가 발생하며 이 영교차율을 이용하여 협대역 음성신호의 주기를 추출할 수 있다. 이러한 spectral 특성을 단구간 평균 영교차율(zero crossing rate)로 표현하면 다음과 같다[3].

$$Zn = \sum_{m=-\infty}^{\infty} \left| Sgn[S(m)] - Sgn[S(m-1)] \right| W(n-m) \quad \cdots(1)$$

이때,

$$Sgn[S(n)] = \begin{cases} 1 & S(n) \geq 0 \\ -1 & S(n) < 0 \end{cases}$$

$$W(n) = \begin{cases} \frac{1}{2N} & 0 \leq n \leq N-1 \\ 0 & \text{the others} \end{cases}$$

식 (1)에 해당하는 블럭도는 아래와 같다.

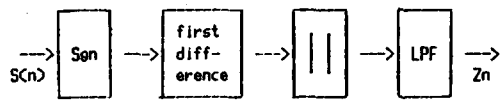


그림 1. 단구간 영교차 추출 블럭도

Fig. 1. Block diagram of short-time average zero crossings

유성음의 에너지는 glottal 파형에 의해 스펙트럼 감소 (fall-off) 때문에 3KHz 이하에 집중되어 있다. 반면 무성음은 대부분의 에너지가 3KHz 이상에서 분포되어 있어 영교차율과 에너지 분포와는 밀접한 관계가 있다.

2. 레벨 교차율

음성신호는 main peak의 에너지가 여기된 후, 성도의 전달 함수에 따라 변하는 응답곡선으로 감쇠진동을 하게된다. 따라서 음성파형이 임의의 레벨을 통과하는 수는 성도함수에 대한 정보를

나타낼 수 있다. 즉 영교차율에서 기본으로 하는 영축(zero-level)에서 데이타의 평균값을 기준레벨로 확장하여 정보를 표현하게 된다[4].

기준레벨 Li 는

$$Li = \left(\frac{1}{N} \sum_{j=1}^{N} \left\| Sj \right\| \right) * i \qquad \text{---(2)}$$

여기서

$i = 0.1, 0.2, \ldots, 5$

$j = 1, \ldots,$ 측정구간 데이타 수

즉 레벨 교차율(level-crossing rate)은

$$LCR = \sum_{j=1}^{N} [1 - Sgn(Sj - Li) * Sgn(Sj-1 - Li)] / 2 \qquad \text{---(3)}$$

이 된다. 단,

$Sgn(Sj - Li) = 1, \quad Sj - Li >= 0$

$= -1, \quad Sj - Li < 0$

3. 미분 영교차율

단구간 내에서 영교차율과 입력된 신호의 peak 수의 관계에 따라 모음 간의 구별을 하게 된다. 즉 peak 수의 측정은 신호를 1차 미분하여서 이 미분된 함수의 영교차율로 표시할 수 있게 된다.

$$DZCR = \sum_{j=1}^{N} [1 - Sgn(Sj+1 - Sj + 1) * Sgn(Sj+1 - Sj)] / 2 \qquad \text{---(4)}$$

이상에서 얻은 영교차율, 레벨교차율 그리고 미분 영교차율을 이용한 파라미터 추출 흐름도는 그림 2 와 같다.

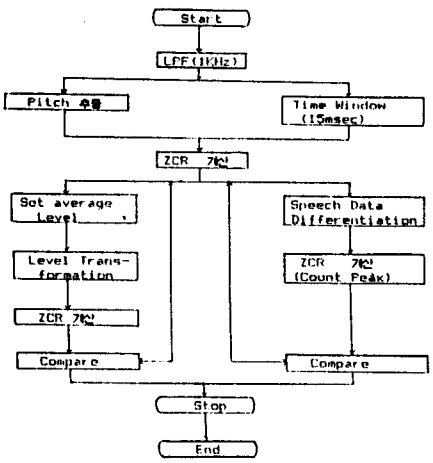


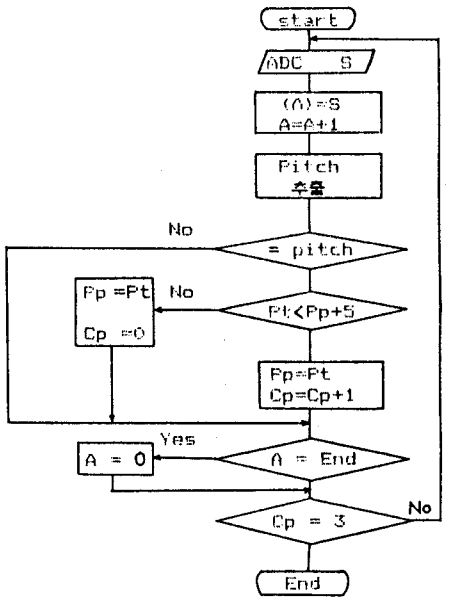
그림 2. 파라미터 추출 흐름도
Fig. 2. The flowchart of parameter extraction

III. 신호의 효율적인 저장 알고리즘

데이타를 저장하기 위하여 시작 부분과 나머지 부분으로 나누어 유성음이 나타날 때까지 시작 부분에 음성신호를 순환적으로 저장하다가 유성음이 검출되면 나머지 부분에 저장한다[5].

음성신호의 출발점을 찾기 위하여 묵음-무성음-유성음을 분류해야 한다. 묵음과 무성음을 분류한다는 것은 매우 어려우며 특히 시작점이 약한 마찰음이나 파열음이 나타날때 출발점을 찾는다는 것은 쉽지가 않다[3].

유성음은 각 음소에 따라 공진 주파수를 갖고 있어 무성음보다 더 큰 에너지를 갖고 또한 반 주기적(semiperiodic)인 특성으로 에너지가 상대적으로 작은 무성음과는 분류가 가능하다. 이러한 특성으로 인하여 한 frame(15 msec)내에서 무성음과 유성음에 대한 평균 영교차율(Zav)과 크기(Mav)는 상당한 차이가 있다. 이때 시작 버퍼를 처리하는 흐름도는 그림 3 과 같다.



Cp = Pitch Counter　　Pp = Post Pitch Period
Pt = Present Pitch Period　　S = Sampled Speech
A = Starting Buffer Address

그림 3. 데이타 저장 흐름도
Fig. 3. Flowchart of data storing

Zav와 Mav를 한 frame 단위로 하여 무성음과 유성음을 분류한다면 음성 데이타의 실시간 처리에 문제점이 생기므로 frame단위가 아닌 peak단위로한 zero-crossing interval (ZCI)의 면적(A)을 사용할 수 있다.

무성음과 유성음의 식별은 샘플된 데이타에서 영교차간의

면적을 이용하여 구할 수 있다. 즉 영교차간의 면적은 식 (5)와 같다.

$$A = \sum_{i=n1}^{n2} S(i) = \left(\frac{1}{n2-n1} \sum_{i=n1}^{n2} S(i)\right) * (n2-n1)$$

$$= Mav / Zav \qquad \text{——(5)}$$

S(i) : i번째 음성파형　　n2-n1 : 구간
Mav : 음성의 평균 크기

유성음은 제1 포만트 주파수가 무성음보다 낮기 때문에(약 3배 이상) 1/Zav는 그고 또한 높은 에너지를(약 10배 이상) 갖는다. 그리고 Mav도 무성음보다 크다. 이러한 두 인수 1/Zav 와 Mav 의 곱이 면적값 (약 30배 이상) 이다.

식 (5)를 무성음과 유성음을 분류하는데 사용한다면 frame단위가 아닌 파형의 peak 단위로 하여 쉽게 분류할 수가 있다. 이 식을 이용하여 안정한 주기 pitch를 연속적으로 3번 찾아내었다면 유성음이 검출되었다고 하자. 즉 식 (5)를 통해 검출된 pitch의 주기가 다음 조건을 연속 3번 만족한다면 유성음이 검출된 것으로 가정할 수 있다.

유성음이 검출된 다음, 데이타는 나머지 버퍼에 저장하여 필요한 파라미터를 추출 하는데 이용하며 유성음이 발견될 때까지 입력 표본값의 현재주소(present address)를 지정하기 위하여 pointer를 사용하여 음성 데이타의 word unit를 입력한 후에 버퍼의 첫번째 주소를 지정한다.

IV. 결론

음성신호의 무성음과 유성음을 분류하는데 영교차 방법이 사용된다. 또한 영교차를 임의의 레벨로 확장한 레벨 교차율과 미분 영교차율을 도입하여 음성 신호를 효율적으로 처리할 수 있다. 레벨 교차율의 기준 레벨은 음성 데이타의 평균값으로 정하고 미분 영교차율은 음성 신호의 peak수를 측정하여 모음간의 구분이 가능하게 할 수 있다.

본 논문에서는 음성 신호을 처리하기전 음성 데이타를 메모리에 저장하여야 하므로 메모리를 효율적으로 저장하는 방법에 대해 고찰했다. 즉 메모리 버퍼를 시작 버퍼와 나머지 버퍼로 나누어 시작 버퍼에서 일정한 pitch를 갖는 유성음이 3번째 pitch를 검출하면 나머지 버퍼에 그 이후의 데이타를 저장하도록 하였다. 이 방법은 Pointer를 사용 함으로써 음성을 실시간 처리하는데 유용하게 된다.

참 고 문 헌

[1] Yiu-Kei Lau and Chok-Ki Chan,"Speech Recognition Based on Zero Crossing Rate and Energy", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-33.NO.1, pp.320-323, Feb.1985.

[2] R.J.Niederjohn, "A Mathematical Formulation and Comparison of Zero-Crossing Analysis Techniques Which have been Applied to Automatic Speech Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-23, NO.4, pp.373-388, Aug.1975.

[3] L.R.Rabiner and R.W.Schafer, "digital Processing of Speech Signals", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.

[4] 이 현주, "ZCR과 LCR을 이용한 모음인식 파라미터 추출에 관한 연구", 연세대학교 석사논문 1984년.

[5] B.S.Atal and L.R.Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-24, pp.201-212, June 1976.

[6] N.C.Geckinli and Davras Yavuz, "Algorithm for Pitch Extraction Using Zero-Crossing Interval Sequence", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-25.NO.6, pp.559-564, Dec.1977.

[7] 배 명진, 이 인섭, 안 수길 , " 음성신호를 표본화할 동안 효율적인 실시간 저장기법", 한국 음향학회 학술 발표회 논문집, pp.66-70, 1986년 11월.