원 파형과 임의 반복시킨 파형의 비교에 의한 유성음의 피치검출

°배 명 진 안 수 길

천안 호서대학 전자공학과 서울대 전자공학과

The Pitch Extraction of Voiced Speech by the Comparison Between the Original and the Repeated Segmental Waveform.

Myungjin BAE Souguil ANN

Hoseo College Seoul National University

ABSTRACT

In speech signal processing, it is necessary to estimate exactly the pitch. We propose a new algorithm which uses the correlation coefficient between the original and the repeated segmental waveform in the frame as a parameter in the pitch extraction. The correlation coefficient in the frame reflects the periodic component and the transient ratio of the waveform.

1. 서 론

모든 설비가 자동화됨에 따라 음성을 통해 직접 설비를 제어및 운용하려는 음성신호처리에 관한 연구가 활발히 전개되고 있다. 음성신호처리는 필터링의 관점으로 보통 연구되고 있는데 필터의 여기신호로는 기본주파수를 갖는 성문(vocal cord)으로, 그리고 필터의 특성은 성도(vocal tract)의 물리적인 형태로 모델링하고 있다.

여기신호인 음원의 기본주파수를 검출하는 방법은 지난 이십여년 동안에 많이 제안되어져 왔지만 아직도 그 정확도는 실용화 단계에 미치지 못하고 있다. IC의 개발기법이 성숙하지 못하던 1980년 이전에는 처리시간의 단축에 주로 관심을 가졌었다. 그후 DSP칩이 실용화 되고부터는 처리의 복잡성 보다는 처리결과의 정확도에 알고리즘의 개발 목적을 두고 있다.

음성신호에서 음원정보에 해당하는 기본주파수를 정확히 검출하면 음성합성분야에서 자연성(natuallity)과 개성을 적은 데이타로 나타낼 수 있고, 음성인식분야에서는 화자(speaker)의 영향을 제거할 수 있기 때문에 인식의 정확도를 높일 수 있게된다. 또한 분석시에도 피치(pitch)에 동기시킬 수 있기 때문에 성문(vocal cord)의 영향을 제거한 parameter를 얻을 수 있어, 음성신호에서의 정확한 기본주파수를 추출하는 것이 필요하다.

음성신호에서 기본주파수를 검출하려는 연구는 지금까지 많이 진행되어 왔지만 음성신호가 유성자음으로 시작하는 구간이나, 파열음이 연결된 유성음구간이나, 비음이 끼어있는 유성음 구간에서는 검출에러가 많이 일어나고 있다. 또한 잡음이 섞인 음성신호에서의 기본주파수 검출에 대한 실용적인 결과는 아직도 얻어지지 않고 있는 실정이다.

본 논문에서는 유성음 구간을 임의 반복시켰을 때 원래의 파형과 가장 유사한 반복구간을 찾는 방법으로 시간영역에서 피치를 구하는 새로운 방법을 제안하고자 한다. 일반적인 피치추출기들이 주기성을 강조한 후에 결정논리에 의해 피치를 찾는방법을 택하고 있지만, 이것은 주기성의 강조와 검출부분이 독립적으로 이루어져서

검출에러가 누적되는 경향이 나타난다. 이에 반해 본 논문에서는 주기성의 강조와 검출과정을 동시에 수행하는 알고리즘을 제안함으로써 피치검출에 대한 파레미터로 적용함과 동시에 검출된 피치의 정확도를 알 수가 있게 된다.

2. 피치의 추출방법

음성신호는 그 발성음원에 따라 유성음, 무성음및, 묵음으로 구분지을 수 있다. 유성음의 음원은 시간영역에서 준주기적인 성질을 갖는데, 이 주기를 피치(pitch)라 한다. 피치추출은 음성파형에서 반복되는 골에서 골까지나 봉우리에서 봉우리 까지를 측정하는 것이다. 골과 봉우리 사이에는 영점을 교차하는 영교차점이 존재하므로, 반복되는 영교차구간 사이를 측정하면 또한 피치가 된다. 반복되는 영교차구간 사이에는 성도의 공명현상에 따른 포먼트들의 영향으로 몇개의 영교차구간이 더 나타날 수도 있다. 피치검출의 방법은 몇개의 영교차구간이 구해질 때 반복되는 영교차구간을 찾는 것이다. 피치검출법은 주기성강조블럭과 결정논리블록으로 나눌 수 있다.

주기성을 강조하는 방법으로는 auto-correlation function(ACF), average magnitude difference function(AMDF), parallel processing법 등이 있다. ACF와 AMDF는 다음식에 의해 주기성을 강조시키고 있다.

$$ACF(k) = \sum_{k=0}^{N} S(n)*S(n+k) \quad \text{-------(1)}$$

$$AMDF(k) = \sum_{k=0}^{N} |S(n)-S(n-k)| \quad \text{------(2)}$$

여기서 S(.)은 음성신호를 표시하고, N은 한 프레임의 길이를 나타낸다. 한프레임 동안에 상기의 식을 계산하면, 지연인자(k)가 피치와 일치하게 될때, 파형의 봉우리와 봉우리가 파형의 골은 골끼리 곱해져서 주기성이 더 강조된 파형이 된다. 따라서 결정논리에 의해 주기적인 것을 결정하면 피치가 된다.

한 프레임 안에서 주기가 거의 일정한 경우에는 ACF나 AMDF법으로 주기성을 강조시킬 수 있다. 그렇지만, 한 프레임 안에서 음소의 천이가 이루어지고 있어 주기가 변화되고 있거나, 음소의 시작및 끝 부분에서 에너지의 급작스러운 변화가 일어나고 있다면 ACF나 AMDF법으로 주기성을 강조시키기는 어렵게 된다.

이러한 문제의 한 해결방법은 다음과 같이 제안할 수 있다. 우선 검출하려는 프레임이 음소의 천이나 큰 진폭변동이 있는 지를 알수 있는 한 파레미터를 사용한다. 그러한 파레미터로는 correlation coefficient(R(.))를 사용할수 있다. 즉,

$$R(n) = Cov(S(n),S(k))/SQR[Var(S(n))*Var(S(k))]$$
$$= \{E[S(n)*S(k)]-E[S(n)]*E[S(k)]\}$$
$$/SQR\{E[S(n)^2]-E[S(n)]^2\}*\{E[S(k)^2]-E[S(k)]^2\} \quad \text{---(3)}$$

여기서 Cov(.)는 covariance를, Var(.)는 variance를, 그리고 SQR(.)는 평방근 값을 나타낸다. 계산된 상관계수 값은 음성파형 n-번째 샘플을 기준으로 k-샘플수 만큼 지연된 것이다. 따라서 한 프레임 안에서 주기와 진폭이 완전히 일치한다면 상관계수 값은 1에 가깝게 되고, 상반된다면 -1에 가깝게 된다.

이 상관계수값을 이용하면 그 프레임 구간의 주기성도 쉽게 결정할 수가 있게된다. 상관계수 값이 1에 근접하면서 최대값이 구해지는 지연인자 k-값이 피치가 된다. 실험적으로 조사해보면 유성음에서 준주기적인 프레임구간은 0.9이상이 되고, 음소의 천이구간이나 시작및 끝부분에서는 0.5이상이 되었다. 또한 그 프레임의 끝부분을 기준으로 하여 지연인자를 -k로하여 상관계수 값을 구하였을 때와 비교하면 천이구간에서는 0.2정도의 차이값을 갖게 된다.

본 연구에서 제안한 피치추출 과정은 처리과정이 간단하고 정확한 추출법이 되지만 자기상관계수를 구하는 과정이 음성파형의 전 구간마다 적용되기 때문에 처리시간이 길어진다. 따라서 처리시간을 단축시키는 알고리즘이 필요하다.

3. 처리과정의 단축

피치주기의 범위로는 2.5에서 25msec정도로 알려져 있으므로, 음성신호를 8KHz로 샘플링하였을 때는 20(Pmn)에서 200(Pmx)샘플간격 안에서 찾을 수 있다. 따라서 한 프레임의 길이(N)를 256샘플로하여 3식을 계산하면 프레임당 (Pmx-Pmn+1)*N번의 곱셈이 필요해 진다.

피치가 P-샘플간격을 갖는 유성음의 파형S(.)은 다음과 같이 나타낼 수 있다

$$S(n)=S(n-INT(n/P)*P) \quad ---(4)$$

여기서 INT(.)는 괄호속의 값에서 정수값만 취하는 함수이다. 역으로, 피치주기를 모를때는 P-를 20에서 부터 증가시키면서 그 프레임의 원래파형과 상관계수를 구한다면 원래파형의 주기와 일치할 때에 1에 근접하게 된다. 이렇게 하면 P-길이 만큼 이미 상관계수값이 계산되어 있으므로 한 프레임에 대한 곱셈의 수는 (Pmx-Pmn)^2+N번으로 단축시킬 수 있다.

음성파형이 1KHz로 대역제한 되었다면, 영교차하는 구간은 4-샘플 간격이상이된다. 상관계수가 1에 근접하는 경우는 주기가 일치하는 때이므로 파형이 영교차하는 근방에서 주로 나타나게 된다. 그리고 주기가 일치하는 경우에는 음성파형의 봉우리가 -에서 +로 또는 +에서 -로 변화하는 한 경우이므로 -에서 +로 변화는 영교차점 부근(k=-1,0,+1)에서만 상관계수를 계산한다면, 필요한 곱셈수는 프레임 마다 N+(Pmx-Pmn)^2*3/8번 보다는 작게 소요된다.

본 연구에서 제안한 피치추출 과정은 처리과정이 간단하고 정확한 추출법이 되지만 자기상관계수를 구하는 과정이 음성파형의 전구간 마다 적용되기 때문에 처리시간이 길어진다. 따라서 처리시간을 단축시키는 알고리즘이 필요하다.

4. 결 론

유성음에서 피치주기 검출은 합성시에 자연성(natuallity)과 개성을, 인식시에 화자의 영향제거를 시킬 수 있기 때문에, 정확한 추출이 필요하다. 음성신호에서 피치를 추출하려는 연구는 지금까지 많이 진행되어 왔지만, 유성자음으로 시작하는 구간이나, 파열음이 연결된 유성음구간이나, 비음이 끼어있는 유성음구간에서는 에러가 보통 발생한다.

이러한 필요성에 따라, 본 연구에서는 원래의 파형과 임의로 주기를 변형시킨 파형과 상관계수를 비교하는 방법으로 피치를 정확히 추출할 수 있는 새로운 방법을 제안하였다. 처리과정에서 정규화된 상관계수를 근거로 현재의 프레임이 천이구간인 지를 파악할 수 있으며, 주기성이 쉽게 파악되어 결정논리도 동시에 이루어진다. 또한 통계적인 문턱값(threshold level)을 취할 필요없이 원래 음성파형을 기준으로 하기 때문에 추출 후 별다른 보상이 필요없어진다.

5. 참고문헌

1) L. R. Rabiner and R. W. Schafer, Digital processing of Speech Signals Englewood Cliffs, NJ:Prentice-Hall, 1978

2) T. V. Screenivas ans P. V. S. Rao,"pitch extraction from corrupted harmonics of the power spectrum,"J. Acoust. Soc. Amer., vol. 65, pp.223-228, Jan. 1979.

3) K. K. Paliwal and P. V. S. Rao, "A synthesis-based method for pitch extraction," Speech Commun., vol. 2, pp. 37-45, May 1983.

4) C. K. Un and S. C. Yang, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," IEEE Trans. Acoust., Speech, Signal processing, vol. ASSP-25, pp. 565-572, Dec. 1977.

5) L. R. Rabiner, "On the use of autocorrelation anaysis for pitch detection," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25, pp. 24-33. Feb. 1977.

6) L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, " A comparative perfomance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp.399- 417, Oct. 1976.

7) J. A. Moorer, "the optimum comb method of pitch period anaysis of con- tinuous digitized speech," IEEE, Trans. Acoust., Speech, Sginal Proce- ssing, vol. ASSP-22, pp. 330-338, Oct. 1974.