

소음환경에 강인한 피치주기 검출 알고리즘에 관한 연구

서현수^{*}·배상범^{*}·김남호^{*}

*부경대학교 제어계측공학과

A Study on the Robust Pitch Period Detection Algorithm in Noisy Environments

Hyun-Soo Seo^{*} · Sang-Bum Bae^{*} · Nam-Ho Kim^{*}

*Dept. of Control and Instrumentation Eng., Pukyong National Univ.

E-mail : digitalshs@pknu.ac.kr

요 약

음성 신호의 피치주기 검출 알고리즘은 음성 인식, 화자 식별 등의 다양한 음성 신호 처리 분야에 적용되고 있으며, 시간영역과 주파수영역에서 많은 연구가 진행되고 있다. 피치주기 검출 알고리즘의 하나인 AMDF(average magnitude difference function)는 연산속도가 빠른 장점이 있지만, 피치주기 검출을 위한 valley점 선정에 있어서 알고리즘이 복잡해지는 문제점이 발생한다. 또한 이러한 피치주기 검출 알고리즘이 실생활에 응용되기 위해서는 다양한 환경에서 발생하는 소음으로부터 강인한 특성을 가져야 한다. 따라서, 본 논문에서는 변형된 AMDF 알고리즘을 이용하여 피치주기 검출을 위한 전체 최소 valley점 선정을 보다 용이하게 하였으며, 테스트 신호로써 지하철 등과 같은 소음환경에서의 음성신호를 사용하였다.

ABSTRACT

Pitch period detection algorithms are applied to various speech signal processing fields such as speech recognition, speaker identification, speech analysis and synthesis. Furthermore, many pitch detection algorithms of time and frequency domain have been studied until now. AMDF(average magnitude difference function), which is one of pitch period detection algorithms, chooses a time interval from the valley point to the valley point as the pitch period. AMDF has a fast computation capacity, but in selection of valley point to detect pitch period, complexity of the algorithm is increased. In order to apply pitch period detection algorithms to the real world, they have robust prosperities against generated noise in the subway environment etc. In this paper we proposed the modified AMDF algorithm which detects the global minimum valley point as the pitch period of speech signals and used speech signals of noisy environments as test signals.

키워드

pitch period, AMDF, noisy environment

1. 서 론

최근 IT기술의 급격한 발전으로 인해 음성을 이용한 다양한 시스템의 수요가 증가하고 있으며, 음성 인식, 화자 식별, 음성 분석 및 합성 등과 같은 음성 신호 처리에 관련된 다양한 연구들이 진행되고 있다. 그리고 음성 신호의 기본 주파수로

정의되는 피치주기는, 음성 신호 처리의 여러 분야에서 응용을 위해 필수적이며 기본적인 요소로 인식되고 있다. 피치는 인간의 청각에 매우 민감하게 반응하는 파라미터로써, 화자의 자연성에 큰 영향을 미치며, 이러한 피치주기를 정확히 획득함으로써 화자의 상태와 무관하게 음성신호로부터 유용한 정보를 용이하게 추출할 수 있다. 그러

므로 음성 합성과 음성 코딩에서 피치주기의 정확한 추출과 복원은 음질에 결정적인 역할을 한다. 따라서 음성 신호의 피치주기 검출을 위한 시간영역과 주파수영역에서의 많은 알고리즘이 제안되었다[1]-[3].

시간영역에서의 피치 검출 알고리즘은 파형의 주기성을 강조하여 피치를 검출하는 방법으로서 병렬처리법, ACF(autocorrelation function), AMDF 등이 있다[2]. 시간영역에서의 알고리즘은 분석을 위한 영역의 변환이 불필요하며, 합과 차 그리고 비교논리 등과 같은 간단한 연산만을 사용하게 되어 처리속도 측면에서 우수한 특성을 가진다. 그러나 음소가 천이 구간 내에 존재하여, 프레임 내에서 급격한 레벨변화가 있는 경우와 잡음이 중첩된 음성의 경우에는 피치 검출을 위한 알고리즘이 복잡해지는 단점을 가진다.

그리고 주파수영역에서의 피치 검출 알고리즘은 음성 스펙트럼의 고조파 간격을 측정하여 유성음의 기본주파수를 검출하는 방법으로서, Cepstrum법, 고조파분석법 등이 있다. 일반적으로 스펙트럼은 프레임 단위로 연산되며, 주파수대역별로 신호의 분리가 가능하므로 음소의 천이와 변동 그리고 소음과 같은 배경잡음에 대하여 그 영향을 적게 받는다. 그러나 처리과정에서 주파수영역으로의 변환과정이 필요하기 때문에 다소 복잡한 연산이 요구되며, 기본주파수의 정밀성을 높이기 위해 FFT의 포인트 수를 증가시킬 경우 처리시간이 길어지고 변화특성이 둔해지게 된다[4].

따라서 실시간시스템에서는 주파수영역으로의 변환과정을 필요로 하지 않는 시간영역에서의 알고리즘이 주로 사용된다. 특히, 적은 계산량과 정확한 피치주기 검출의 장점을 가지는 AMDF가 널리 적용되고 있으나, 피치주기 검출을 위한 valley점 선정이 용이하지 않다.

따라서, 본 논문에서는 AMDF에 회전변환을 적용한 피치주기 검출 알고리즘을 제안하였으며, 테스트 신호로써 지하철 등의 소음환경에서 획득한 음성 신호를 이용하였다. 또한 보간 알고리즘으로 SM(standard median) 필터를 적용하여 우수한 피치주기 검출 특성을 나타내도록 하였다.

II. 제안된 피치 검출 알고리즘

1. Basic-AMDF 알고리즘

음성신호 $s(n)$ 은 길이 N 을 가진 창함수 $w(n)$ 과의 연산을 통해 신호 $s_w(n)$ 이 된다. 이때, Basic-AMDF $\chi(l)$ 은 다음과 같이 정의된다.

$$\chi(l) = \sum_{n=1}^{N-l+1} |s_w(n+l-1) - s_w(n)|, \quad (1)$$

$$l = 1, 2, \dots, N$$

식 (1)로부터, $s_w(n)$ 이 유성음일 경우, $\chi(l)$ 은 아래의 그림 1과 같이 주기적인 특징을 나타내며, $\chi(l)$ 의 이러한 특징은 유성음과 무성음을 구분하거나 유성음의 피치주기를 결정하는 분야에 사용된다.

일반적으로 유성음은 그림 1에서 보는 바와 같이, Basic-AMDF 알고리즘에 의해 국부 최소 valley점이 주기적으로 A, B, C 등에서 나타나며, 전체 표본에 대해서 최초의 valley점을 피치주기로 계산한다.

따라서 그림 1에서, 국부 최소점 A와 시작점 사이의 시간 간격이 피치주기이다. 그러나 최초 국부 최소 valley점 A가 나머지 valley점 B 혹은 C보다 값이 클 경우, 전체 최소 valley점이 되지 않으므로 피치주기 검출을 위한 최초 valley점 선정에 어려움이 있다.

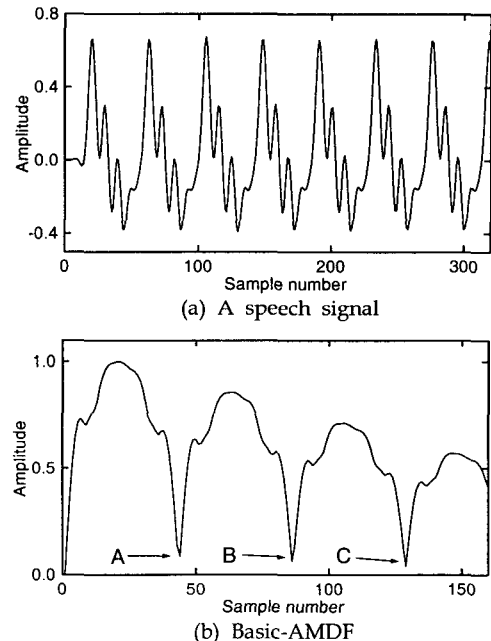
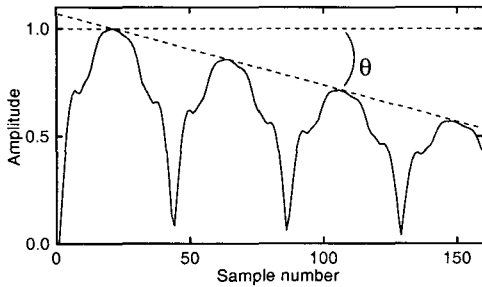


Fig. 1. Basic-AMDF in a speech signal.

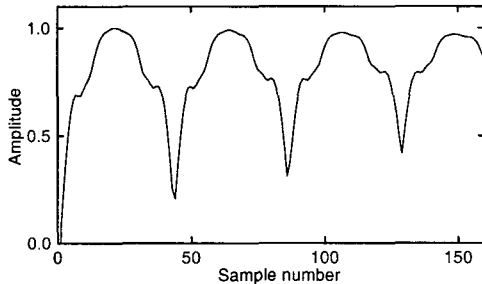
2. 변형된 AMDF

본 논문에서는 Basic-AMDF $\chi(l)$ 이 분석 프레임 내에서 크기가 감소하는 특징을 이용하여, 다음의 그림 2에서 보여지는 바와 같이, 최초 피크 지점과 N 지점을 잇는 직선에 의해 수평선과의 각 θ 를 설정한다. 그리고 이러한 각 θ 를 0°로 만들기 위해 양의 방향으로 Basic-AMDF를 회전시킨다.

다음의 식 (2)는 각 θ 를 설정하기 위해 유도된 식이다.



(a) Setting Angle θ of Modified AMDF



(b) Modified AMDF after setting angle θ

Fig. 2. Modified AMDF.

$$\theta = \arctan \frac{R_{\max}}{N} \quad (2)$$

여기서 $R_{\max} = \max\{y(l)\}$ 으로써, 프레임 내에서 Basic-AMDF의 최대값을 나타내며, N 은 $y(l)$ 의 분석을 위한 신호의 길이이다. 그리고 다음의 식 (3)은 Basic-AMDF를 위의 식에서 계산된 각 θ 만큼 회전시킨, 변형된 AMDF를 나타낸다.

$$y_M(l) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \times y(l) \quad (3)$$

식 (3)에서 제시된 변형된 AMDF는 각 θ 를 이용하여, Basic-AMDF $y(l)$ 에 회전변환을 적용함으로써 피치주기의 첫 번째 valley점에 대한 검출을 용이하게 한다. 또한, 변형된 AMDF에 다음의 식 (4)과 같이 파라미터 α 를 추가하여, 회전각에 대한 유연성을 갖도록 하였다.

$$\theta = \arctan \frac{R_{\max}}{N} + \alpha \quad (4)$$

그림 3은 식 (4)에서 제시된 파라미터 α 에 대한 영향을 나타내고 있으며, α 의 값에 따라 최초 국부 최소 valley점과 나머지 valley점 사이의 값의 차이가 더욱 커짐을 알 수 있다.

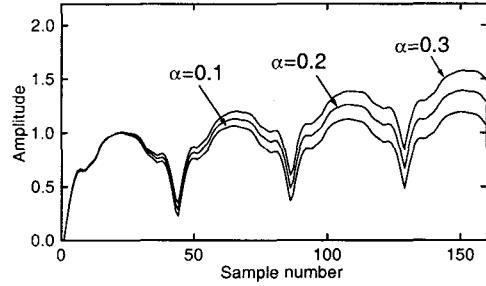


Fig. 3. Modified AMDF of several angle

3. 보간 알고리즘

유성음의 피치를 검출하기 위해서는 먼저, 유·무성음의 구분이 필요하다. 본 논문에서 제안된 피치주기 검출 알고리즘에서는 에너지 임계값을 설정하여 유성음과 무성음을 판별하였다.

또한 변형된 AMDF에 의해 각 프레임에서 검출된 피치의 오차를 최소화하기 위해, 보간법과 SM 필터를 이용하여 최종 피치주기를 결정하였다.

보간법으로써, 한 프레임의 피치주기를 인근 프레임의 피치주기의 크기와 비교하여 설정한 임계값의 범위를 벗어날 경우, 인근 프레임의 피치주기 중 큰 값을 대입하였다.

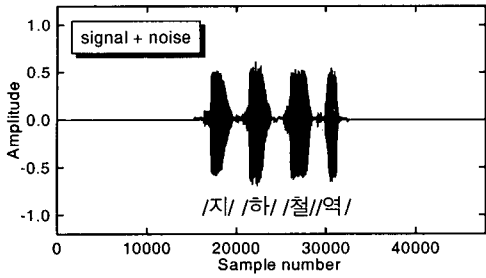
또한 필터 내의 입력값을 크기 순으로 정렬하여, 중간값을 출력값으로 선택하는 SM 필터를 다음의 식 (5)와 같이 나타내었다. 식으로부터 $y(i)$ 는 출력값이며, $x(i)$ 는 입력값, l 은 SM 필터 내부의 좌표이고, $2N+1$ 은 필터의 크기이다. SM 필터는 임펄스성 오차 및 잡음을 제거시켜주는 역할을 하는 비선형 필터이다. 따라서, 본 논문에서는 SM 필터를 이용하여 피치주기의 크기를 평활화하였으며, 그 값을 최종 출력으로 결정하였다.

$$y(i) = \text{med}\{x(i-l) : -N \leq l \leq N\} \quad (5)$$

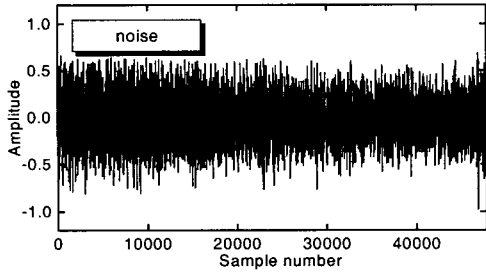
III. 시뮬레이션 및 결과

본 논문에서 제안한 피치주기 검출 알고리즘을 시뮬레이션하기 위해, IBM-PC의 사운드카드를 사용하였다. 또한 테스트 신호로 획득한 지하철 환경의 소음 및 음성 신호의 분해능은 16비트이며, 표본화 주파수는 8[kHz]로 설정하였다. 그리고 한 프레임의 길이는 10[ms](80표본)이며, 분석프레임의 길이는 4프레임(320표본)으로 하였고, 한 프레임씩 오버랩시켜 피치주기를 검출하였다.

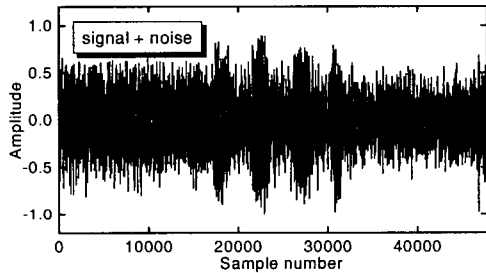
그림 4는 /지하철역/으로 발생된 음성 신호에 대한 시뮬레이션 결과를 보이고 있다. 그림으로부터 (a)는 잡음이 없는 가운데 획득한 음성 신호이며, (b)는 본 논문에서 사용된 지하철 소음이다.



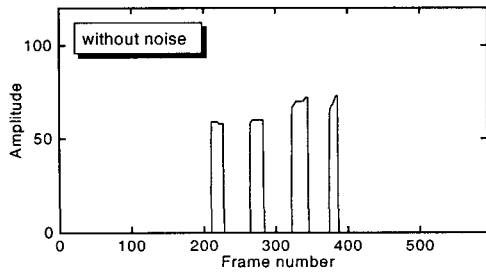
(a) A speech signal



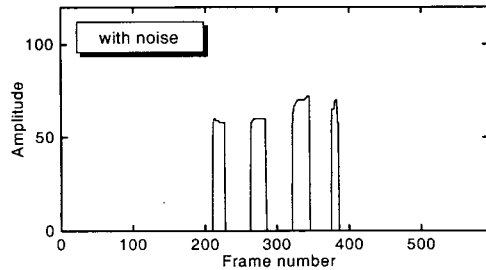
(b) A noise in the subway



(c) A test signal



(d) A pitch detection result(without noise)



(e) A pitch detection result(with noise)

Fig. 4. A simulation result.

그림 (c)는 (a)의 음성 신호와 (b)의 지하철 소음을 중첩시킨 테스트 음성이다. 그리고 그림 (d)는 본 논문에서 제안한 피치 검출 알고리즘의 결과로 나타난 각 프레임에 해당하는 피치주기이며, 이 때 사용한 변형된 AMDF의 파라미터 α 는 0.3으로 설정하였다.

그림에서 확인할 수 있듯이, 음성 신호의 획득과 변형된 AMDF를 이용한 피치주기 검출, 그리고 보간 알고리즘에 의한 평활화의 결과로부터, 제안된 알고리즘이 지하철 소음환경에서 우수하게 수행되어 음성 신호의 피치주기가 정확히 검출되었음을 알 수 있었다.

IV. 결론

본 논문에서는 소음환경에서 음성 신호의 피치주기를 검출하기 위해, 지하철 환경에서 획득한 소음과 음성 신호를 중첩시켜 시뮬레이션하였다. 또한 Basic-AMDF의 회전 변환을 이용한 변형된 AMDF 알고리즘을 제시하였으며, 에너지 임계값을 이용하여 유·무성음을 구분하였고, 보간법과 SM 필터를 이용하여 임펄스성 잡음 및 오차를 최소화하여 소음환경에 강인한 특성을 나타내도록 하였다.

시뮬레이션 결과, 소음환경에서의 음성에 대한 피치주기가 정확히 검출되었음을 확인할 수 있었다. 따라서 제안한 방법은 음성 합성, 음성 인식, 화자 식별 등과 같은 음성 신호 처리의 다양한 분야에 유용하게 적용되리라 사료된다.

참고문헌

- [1] L. R. Rabiner et al., "A Comparative performance study of several pitch detection algorithms", *IEEE Trans. Acoust., Speech, Signal Processing*, 24(5), pp. 399-417, 1976.
- [2] M. J. Ross et al., "Average magnitude difference function pitch extractor", *IEEE Trans. Acoust., Speech, Signal Processing*, 22(5), pp. 353-362, 1974.
- [3] S. Kadambe and G. F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals", *IEEE Trans. Information Theory*, 38(2), pp. 917-924, 1992.
- [4] S. Sneff, "Real Time harmonic pitch detection", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, pp. 358-365, 1978.