

전·후 처리를 이용한 3 레벨 클리핑 알고리즘의 피치검출

최승영, 유창규, 신유식, 김종교

전북대학교 전자공학과

The Pitch detection of 3 Level Clipping Algorithm using by Pre-Post Processing

Seung-Yeong Choi, Chang-Kyu Yu, You-Sik Shin, Chong-Kyo Kim

Dept. of Electronic Eng., Chonbuk National Univ.

요 약

음성신호의 특징적인 성분인 피치를 검출하는 알고리즘 중 실시간 구현이 손쉬운 3단계 클리핑 알고리즘을 PC상에서의 처리를 위하여 구현하였다. 이 알고리즘을 통하여 검출되는 피치의 안정성 및 정확성을 높이기 위해서 적용된 창함수, LPF, 클리핑 자기상관값계산, 비선형 감쇄, 동위 전처리 필터링과, 배수피치 검출 및 정정, 매디언 필터링을 사용하여 피치를 검출하였다. 또한 이 알고리즘을 이용하여 DSP의 도움을 얻지 않고 PC 상에서 음성을 분석하여 스펙트로그램, 파형, 에너지, 피치 등을 출력하는 프로그램인 Visual Analysis Tool for sounds(VAT)의 출력화면을 통하여 피치검출을 나타내었다.

1. 서론

날로 발전해가는 고도화 사회에서 컴퓨터와 인간 상호간의 통신이 절실히 요구되고 효율적인 통신을 위한 노력들이 여러 방면에서 연구가 되어 오고 있다. 특히 음성신호처리는 음성인식, 음성합성, 음성코딩으로 크게 구분할 수 있고, 인식이나 합성에서도 음의 피치에 대한 분석을 통하여 피치 정보가 이용되고 있다.

정확한 피치를 찾는다는 것은 매우 중요하며, 특히 기존의 방법을 영역별로 나누어 고려해 보면, 시간영역법, 주파수 영역법등으로 나눌 수 있다. 시간영역법은 자기상관관계법, AMDF, 병렬처리법등을 통하여 검출을 하고 있으며, 이 방법은 시간영역에서 수행되므로 영역의 변환이 불필요하고, 합, 차, 비교논리 등 간단한 연산만 필요하다. 그러나 음소가 천이구간에 걸쳐 있는 경우에는 프레임내의 레벨변화가 심하고 피치주기가 변동하기 때문에 피치검출에 어려움이 따르게 된다. 특히 잡음이 섞인 음성의 경우에는 피치검출을 위한 결정능리가 복잡해져서 검출오류가 증가하는 단점이 있다.

이 논문에서는 시간영역에서의 분석법으로서 3레벨 클리핑과 자기상관법을 이용하여 알고리즘을 PC상에서의 처리를 하기위하여 구현하였으며 피치검출법과 전체 블록도, Peak의 탐색 알고리즘을 기술하고 그 결과를 VAT화면을 이용하고 있다.

2. 피치 검출법과 전체 블록도

피치 검출법들은 일반적으로 음성파형의 주기성을 강조하고 결정논리를 적용하여 피치를 검출하게 된다. 주기성 강조는 공명현상에 나타나는 포먼트들의 영향을 제거시키고 여기원의 피치만을 강조하는데 그 목적이 있다. 따라서 성도의 영향을 제거시키면 결정능리가 간

단해진다.

피치검출의 결정논리법에는 강조된 음성파형의 기본주기를 실험적인 문턱값과 무게치를 적용하는 것이 보통이다. 그 외에도 위상이나 진폭에 따라 몇가지의 기준 패턴을 만들어 입력된 음성의 파형에서 추출한 패턴과의 유사도를 측정하여 주기성을 결정하기도 한다. 결정논리가 실험적인 문턱값에 의존하면 검출오차가 확률적으로 커지기 때문에 경우의 수에 따라서 결정문턱값을 신중히 고찰해야 하는 복잡성이 따른다. 최근에는 Fuzzy나 신경망의 결정논리를 적용한 연구 발표 논문이 계속해서 나오고 있다. 종래에 사용되는 피치검출법으로는 Parallel Processing, Autocorrelation, AMDF, Spectrum Flattener, Data Reduction, Harmonic Matching법 등이 사용되고 있다.

전체블록도를 그림1에 나타내었다. 먼저 한프레임의 데이터를 읽어들이 창함수를 씌우고 난 다음 LPF를 통과 시킨후 클리핑레벨을 결정한 후 클리핑을 수행한다. 클리핑된 데이터를 가지고 자기상관값을 계산한후 그 계산값을 취하고 피크를 탐색하면 그 탐색된 결과에서 한 프레임의 피치위치를 결정하는 것이 전체 블록도에 서 설명하고 있는 내용이다.

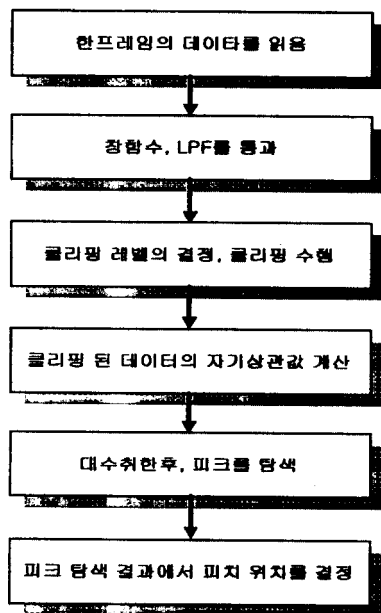


그림 1. 전체 블록도

3. Peak탐색 알고리즘

시간영역법의 피치검출법은, 1967년 Reddy에 의해서 국부 봉우리와 골을 사용한 방법이 사용되었으며, 그 이후에 Rabiner에 의한 병렬처리법이 주로 사용이 되고

있다. 이 방법들은 많은 조건을 주어 처리하기 때문에 분석시에 따르는 오차가 생겨 기본주파수 이외에 haling, doubling, tripling등이 생기는 문제점과 처리과정상의 복잡한 결정논리로 인하여 많은 계산시간이 요구된다.

따라서 처리과정이 복잡하지 않고 결정논리가 간단한 알고리즘이 필요로 하게 된다. 실제 기본 주파수를 검출하는데 걸리는 시간이 인식이나 합성을 위한 기저기술이기 때문에 실시간에 적합한 알고리즘이 요구된다.

이 논문에서는 음성 데이터가 들어오면 우선 한 프레임의 결정한 다음 그 프레임의 데이터를 읽어서 그림 2와 같은 창함수를 씌우고 다음으로 LPF를 통과시킨다. 예를들어 남성의 데이터를 가지고 실험한 결과가 그림 3과 같다. 창 함수는 식(1)과 같다.

$$w(n) = 0.5 + 0.5 \sin\left[\frac{\pi}{(N-1)} n\right] \quad (1)$$

$$0 \leq n \leq N-1$$

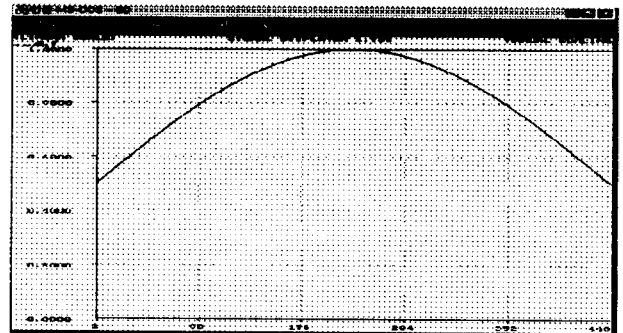


그림 2. 창함수

그림 3은 창함수를 씌운 결과와 LPF를 통과한 결과를 각각 가는 선과 굵은 선을 나타낸 것이다.

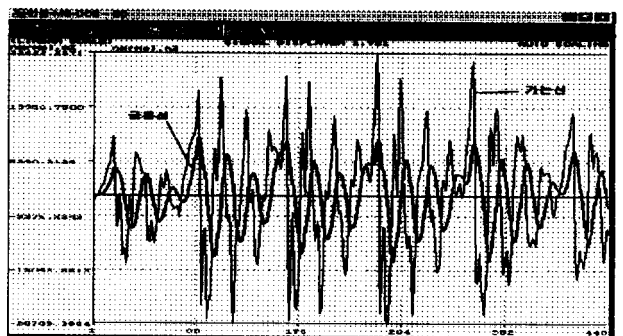


그림 3. 분석 프레임 및 LPF통과 후 (남성)

LPC를 통과시킨후에는 클리핑 레벨을 다음과 같은 순서로 결정한다.

※ 클리핑 값을 결정 방법

- ① 한 프레임을 같은 폭을 갖는 3부분으로 분할한다.
- ② 각부분에서 최대 절대값 M1, M2, M3을 구한다.
- ③ M1, M2, M3 값들중에서 가장 작은 값을 클리핑레벨로 결정한다.

$$\min(M1, M2, M3) = \text{Clip level} \quad (3)$$

- ④ 프레임의 모든값을 대상으로 절대값이 클리핑레벨의 0.8배 보다 작은 값은 0으로 자른다.
- ⑤ 클리핑된 값을 이용하여 자기 상관값을 구한다.

그림 4, 그림 5는 위에서 구한 클리핑된 값을 대수를 취한 결과이다. 계수를 치할때는 1보다 작은 값은 0으로 하였다.

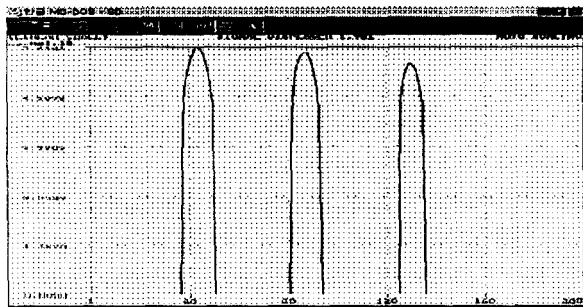


그림 4. 대수를 취한 후 상관값(여성)

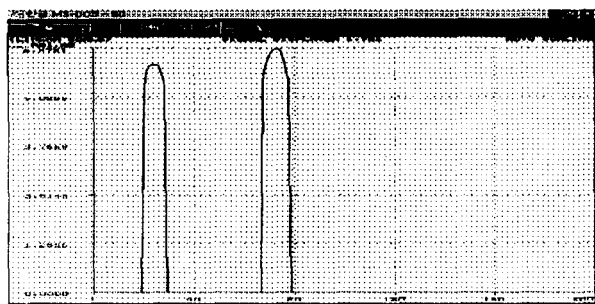


그림 5. 대수를 취한 후 상관값(남성)

이렇게 구해진 결과를 다음의 피크 탐색법을 이용하여 피치 위치를 결정하게 된다.

(1) 첫 번째 탐색법

- 가. 포인터를 피치가 존재할 수 있는 처음위치로 이동
- 나. 포인터가 탐색구간에 있으면 반복
- 다. 0이 아닌 값을 찾아서 포인터를 이동
- 라. 0이 아닌 값 들 중에서 MaxVal = 가장 큰 값,

MaxPos = 위치, Width = 폭을 추출

마. Peak 탐색법 II를 수행

바. 0인 값을 찾아서 포인터를 이동, 단계 나로 이동

사. WOfs = 폭이 가장큰 위치, POfs = 크기가 가장 큰 위치

아. WOfs == POfs 이고 매칭이 1번만 발생했으면, WOfs를 리턴

자. WOfs의 폭 > POfs의 폭+3 이면, WOfs를 리턴

차. POfs를 리턴

카. Median필터링 수행

(2) 두 번째 탐색법

MaxVal = 가장큰값, MaxPos = 큰값의위치,

Width = 폭

MatchFlag = FALSE;

for i = 0 to EndofInfo (

if(IsAnyTimes(Candid[i].Pos, MaxPos) == FALSE) continue;

if(abs(Candid[i].Width - Width) >= 2) {

if(Candid[i].Width < Width)

then MatchFlag = FALSE, break;

else MatchFlag = TRUE, break;

}

else if (Candid[i].MaxVal < MaxVal)

then MatchFlag = FALSE, break;

UpdateItem(Candid, MaxVal, MaxPos, Width);

MatchFlag = TRUE, break;

}

if (MatchFlag == FALSE)

AddItem(Candid, MaxVal, MaxPos, Width);

4. 실험 및 결과

그림 4, 그림5를 이용하여 위의 2가지 탐색법을 적용하여 계산된 피치후보값들은 다음 표 1과 같다.

순번	처음후보 위치	가장최근후보 위치	피크의 폭 (대표값)	비고
0	43	87	12	실제피크

여성화자의 피크 탐색결과

순번	처음후보 위치	가장최근후보 위치	피크의 폭 (대표값)	비고
0	43	87	12	
1	72	72	10	실제피크

남성화자의 피크 탐색결과

그림 4에서 나타난 것처럼 처음 피크가 생기는 포인트는 43정도이고 두 번째 피크는 첫번째 피크의 2배정도 떨어진 87정도에서 생긴다. 세 번째 피크는 3배정도인 125정도에서 생기는 것을 알 수 있다. 처음피크와 두 번째 피크는 폭이 거의 비슷하고 크기만 다르다. 하지만 세 번째 피크는 피크 탐색법에 어긋나므로 피치가 아님을 알 수 있다. 피크 탐색법에 의하면 우선 폭과 크기를 비교해서 폭이 넓고 크기가 크면 그 피크위치가 실제피치 위치가 된다. 그림 5에 나타난 피크 또한 피크 탐색법 알고리즘에 의해 두 번째 피크가 실제 피치가 됨을 알 수 있다.

5. 결론

이상에서와 같은 결과를 PC상에서 음성을 분석하여 스펙트로그램, 파형, 에너지, 피치 등을 출력하는 프로그램인 VAT의 출력화면을 통하여 파형과 검출된 피치만을 나타내었다. 그림 6, 그림7을 보면 남성과 여성의 출력된 피치의 모습이다. 피치 검출에 있어서 빠른 계산시간과 Peak검색에 따른 안정된 피치출력을 보여주었다. 즉 배주기, 반배주기, 피치의 감쇄에 따른 안정된 결과를 나타내주고 있다.

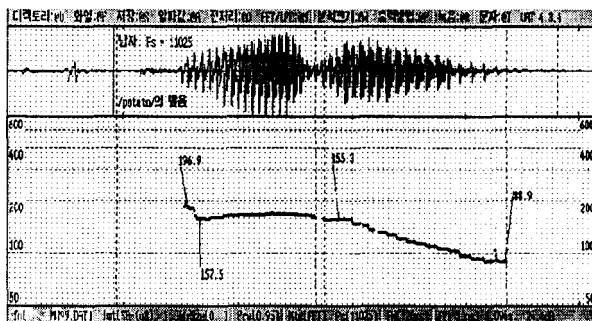


그림 6. 출력된 피치의 모습 (남성)

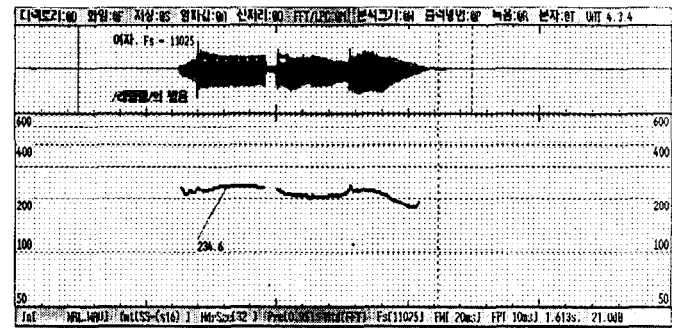


그림 7. 출력된 피치의 모습 (남성)

6. 참고 문헌

- [1] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer Verlag, New York, 1976.
- [2] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signal*, Prentice-Hall, 1978.
- [3] A. N. Ince, *Digital Speech Processing(speech coding, synthesis, and recognition)*, Kluwer Academic Publishers, 1992.
- [4] G. Bristow, *Electronic Speech Synthesis*, McGraw Hill 1984.
- [5] H. W. Strube, "Determination of the Instant of Glottal Closure from the Speech Wave," *J. Acoust. Soc. Am*, vol56, No.s, pp1625-1629 November 1974.
- [6] E. Lee, C Park M bae and S. Ann, "The High speed Pitch Extraction of Speech Signals Using the Area Comparsion Method," *KITE*, vol.22, No.2, pp101-105, 1985.
- [7] D.E Veeneman and S.L. Bement, "Automatic Glottal Inverse Filtering from Speech and Electroglottographic Signals." *IEEE Trans, Acoust, Speech, Signal Processing*, Vol. ASSP-33, No.2, pp.369-377 April 1985.