분산을 이용한 피치검출

김덕수
대구공업대학 컴퓨터정보계열

Pitch Detection Using Variance

Deok-Soo, Kim
Dept. of Computer Science, Taegu Technical College

<요 약> 음성신호는 주기성으로 유성음과 무성음으로 구분할 수 있으며 유성음은 준주기성 신호이지만 무성음은 주기성이 없다. 주기성은 음성신호를 분석하는 중요한 파라미터 중 하나이다. 본 논문에서는 시간영역에서 피치의 시작점과 주파수 영역에서 분산을 이용하여 피치 검출 알고리즘을 제안하며 864개 단독 숫자음을 이용하여 실험한 결과 99.5%의 정확도를 확인하였으며 제안된 알고리즘의 유효성을 확인하였다.

Ⅰ. 서 론

음성신호는 파형의 주기성이 있는 유성음(voiced sound)과 주기성이 없는 무성음(unvoiced sound)으로 나눌 수 있다. 이 유성음의 단위시간당 진동을 기본주파수와 기본주파수의 배수에 해당하는 주파수로 해석할 수 있으며 기본 주파수에 해당하는 시간을 피치라고 하며 음성합성, 음성인식, 화자인식 등 음성신호처리 분야에 중요한 파라미터 중의 하나이다.

피치를 검출하는 방법들이 다양하게 제안되어 있으며 시간영역법[3,5], 주파수영역법[2], 웨이브렛을 이용한 방법[4]으로 구분할 수 있다. 시간영역검출법은 파형의 주기성을 결정논리에 의하여 피치를 검출하는 방법으로, 시간 영역에서 수행되므로 주파수 영역의 변환이 불필요하고 합, 차, 비교에 의해 결정되므로 속도가 빠른 편이지만 녹음 level이나 음소의 천이 구간에 있는 경우 결정논리가 복잡해서 검출오류가 증가되는 단점이 있다. 주파수 영역의 피치 검출은 음성 스펙트럼의 고조파 간격을 측정하여 피치를 검출하는 방법으로 한 프레임 단위로 구해지므로 적용된 창함수의 길이, 종류에 따라 많은 영향을 받는다. 창함수의 길이가 긴 경우 기본 주파수의 하모닉스 봉우리가 분명치 않게 되고 짧은 경우 smearing 현상이 발생하여 왜곡된 결과가 된다.

본 논문에서는 가변 창함수를 이용하여 3개 피치의 구간을 측정하여 평균피치를 구하는 방법을 제시하며 2장에서는 피치검출 알고리즘을 제안, 3장에서는 실험 및 고찰 4장에서 결론을 맺는다.

Ⅱ. 본 론

본 논문에서는 음성신호 x(t)를 식 1과 같이 정의하고 주파수 영역에서 처리하기 위하여 식 3을 이용하여 DFT(discrete fourier transform) 처리하여 식 2와 같이 주파수 성분을 $[X_m]$으로 표시한다.[1]

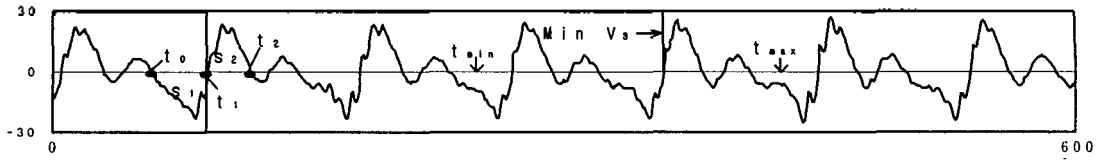
$$x(t) = [x_k] = [x_0, x_1, x_2, \cdots, x_{N-1}] \quad (1)$$

$$[X_m] = DFT[x_k] = [X_0, X_1, \cdots, X_{N/2}] \quad (2)$$

$$X_m = \left| \sum_{k=0}^{N-1} x_k e^{-j(2\pi mk/N)} \right| \quad (3)$$

$$S = X_2 + X_3 + X_4 \quad (4)$$

$$V_3 = \sum_{m=2}^{4} (m-3)^2 X_m / S \quad (5)$$



[그림 1] 음성신호 및 검출성공한 피치

[그림 2] 검출에러로 처리한 피치

식 3의 우측항은 복소수이므로 절대치를 X_m으로 하며 주파수 성분 X_m의 X_3에서 분산의 값을 표현하기 위하여 식 4와 같이 X_2, X_3, X_4의 합을 S로 정의하고 X_2, X_3, X_4의 평균값을 X_3이라고 가정하여 X_3에 대한 분산을 식 5와 같이 V_3으로 정의한다.

본 논문에서 제안하는 분산을 이용한 피치검출 알고리즘은 다음과 같다.

Algorithm average_pitch_detection
Step1 : for i = 첫프레임 to 끝프레임
　　step 10ms
Step2 : 피치의 시작점 t_1 결정
Step3 : for t_i = t_{min} to t_{max} step 영교차점
　　DFT(t_1, t_i) : 시간 t_1에서 t_i
　　까지 DFT처리를 정의
Step4 : 분산값 V_3 계산
　　next t_i
Step5 : Min V_3 for all t_i
　　next i
Step6 : Min V_3 for all 프레임

Step1: 입력된 음성데이터를 10ms 단위로 구분하여 step2~step5를 10ms 프레임 단위로 실행한다.
Step2 : 음성을 10ms 프레임 단위로 구분하여 선정된 프레임 내에서 영교차점을 이용하여 면적 최대 구간을 다음과 같이 선정한다. [그림 1]에서 면적 s_1은 t_0 ~ t_1까지 구간의 면적이고 s_2는 t_1 ~ t_2까지 구간의 면적이 된다. 음수인 경우에는 절대값을 s_1로 한다. 선정된 프레임에서 인접한 영교차점까지 면적의 최대치 구간을 구한다. 만약 최대치가 s_1 또는 s_2이라면 시작점은 t_1이며 [그림 1]에서는 시작점이 t_1인 경우이며 t_1은 선정된 프레임의 신호에서 가장 변화가 큰 시점이 된다. 최대 구간이 양수인 s_2인 경우 시작점은 t_1이 되고 만약 최대구간이 s_1인 경우 시작점은 t_1이 된다. 최대 면적의 값이 양수인 경우 구간의 시작점, 음수인 경우에는 구간의 끝점을 시작점 t_1으로 결정한다.
Step3 : [그림 1]에서 t_{min}과 t_{max} 사이에는 여러 개의 영교차점이 있으며 t_1에서 이 영교차점들까지 DFT를 각각 구한다. 음성은 일반적으로 50Hz ~ 500Hz라고 알려져 있으므로 최소피치는 2ms, 최대피치는 20ms가 된다. 본 논문에서는 3개 피치를 구하므로 t_{min}은 최소피치의 3배, t_{max}는 최대피치의 3배이므로 t_{min}과 t_{max}는 각각 6ms, 60ms가 된다. t_{min}에서 t_{max}까지는 54msec이므로 음성신호에 따라 영교차점이 여러개 될 수 있다.
Step4 : DFT 처리된 신호에서 분산값 V_3을 계산하며 V_3은 DFT 처리된 주파수 성분 X_2, X_3, X_4값에서 X_3이 평균값이라고 가정하여 X_3에 대한 분산값이

다. Step3, Step4을 t_{max}까지 반복한다.

Step5 : DFT 처리된 t_{min}에서 t_{max}까지의 각각의 V_3에서 최저치 V_3을 검출하며 최저치 V_3은 피치가 3개 되는 것을 의미한다. 최저치 V_3에 해당하는 t_i ~ t_i 시간을 3으로 나누면 해당 음성신호의 평균 피치가 된다.

Step6 : Step1 ~ Step5의 과정을 전체 프레임에 실행하고 가장 작은 V_3을 검출하여 이 음성의 피치로 결정한다.

Ⅲ. 실험 및 고찰

본 연구에서 제안된 알고리즘의 정확성을 확인하기 위하여 사용한 음성데이터는 국어공학연구소(KLE)에서 채록한 단독 숫자음 '영', '공', '일', '이', '삼', '사', '오', '육', '륙', '칠', '팔', '구' 이며 16kHz 샘플링, 16비트로 양자화 되어 있다. 실험에서는 8비트로 양자화 하였다. 이 음성 데이터는 <표 1>의 남자 39명, 여자 33명 화자들로 단독 숫자음의 종류는 12개이므로 총 864개 단독 숫자음 이다. 864개의 실험 결과를 확인하기 위하여 1개 숫자음 단위로 알고리즘에 의해 계산한 후 모니터 화면에서 [그림 1]과 같이 t_i과 최소치 V_3에 해당하는 위치를 수직선을 이용하여 확인하며 864개 음절을 반복하였다.

<표 1>의 피치검출율에서 남자화자인 경우 99.8% 여자화자인 경우 99.2%의 피치검출 성공률이며, [그림 1]은 피치검출에 성공한 경우, [그림 2]는 피치검출에서 에러로 처리한 경우이다. 864개의 단독 숫자음에서 검출에 실패한 4경우는 음성신호에 고주파가 포함된 경우이다.

<표 2>는 남여화자의 샘플수에 의한 피치분포이다. 음성신호는 피치가 유성음의 시작에서 끝점까지 일반적으로 변하지만 남성화자의 최대분포는 140샘플이며 114Hz가 되며 여성화자인 경우 최대분포는 70샘플이며 228Hz가 된다. 여성화자인 경우 130 ~ 150샘플은 12개이며 1명을 의미한다. 이 여성화자의 음성을 확인한 결과 남자음성으로 발성된다. 남성화자의 주분포는 100 ~ 179 사이이며 여성화자는 60 ~ 89 사이이므로 남성화자의 샘플수 대역이 여성화자에 비하여 넓은 것이 특징이며 90 ~ 99 샘플수에는 남여화자가 중복이 된다.

[그림 3, 4]는 음성신호 /3/, /7/에 대한 (a)는 음성신호 (b)는 피치궤적이다. [그림 3]은 남성화자가 발성한 /3/의 신호이며 (a)는 음성신호 (b)는 제안한 알고리즘에 의해 자동으로 검출한 피치궤적이며 [그림 3]의 (b)피치궤적은 약 150샘플이 중심이 되며 106Hz가 된다. [그림 4]는 여성화자가 발성한 /7/의 신호이며 [그림 3]의 (b)피치궤적은 약 85샘플이 중심이 되며 188Hz가 된다. 피치궤적은 식 5의 분산간 V_3의 임계값을 0.05로 정의하여 V_3의 값이 0.05 이상이면 무성음으로 주기성이 없으며 V_3의 값이 0.05 이하인 경우 유성음이 되며 해당하는 피치궤적은 [그림 3, 4]의 (b)가 된다.

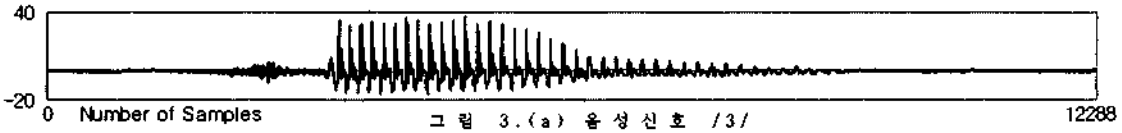
음성신호 /3/의 '사' 음성신호 /7/의 '칠'에 해당하는 무성음과 유성음의 구분이 가능하며 [그림 3~4]에서 피치의 시작점과 끝 부분도 구분 할 수 있다.

<표 1> 남여화자별 피치검출 성공률

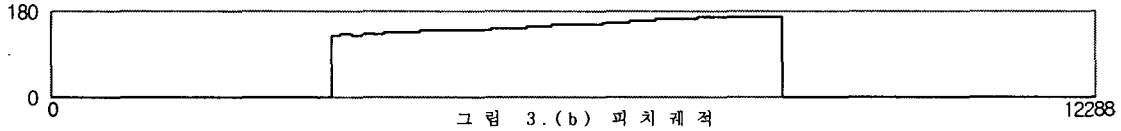
	인원	음절수	검출성공	검출성공%
남자	39	468	467	99.8
여자	33	396	393	99.2
합	72	864	860	99.5

<표 2> 남여화자의 피치분포

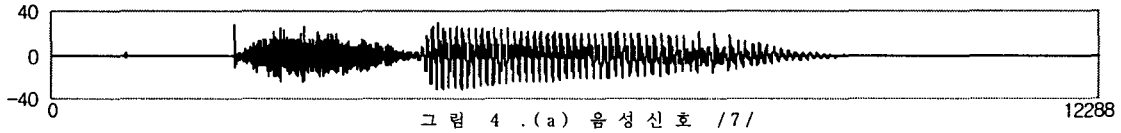
샘플수	50 ~59	60 ~69	70 ~79	80 ~89	90 ~99	100 ~109	110 ~119	120 ~129	130 ~139	140 ~149	150 ~159	160 ~169	170 ~179	180 ~189	190 ~199	합
남자	0	0	0	0	3	21	54	78	56	96	83	46	14	8	9	468
여자	1	61	188	124	10	0	0	0	1	6	5	0	0	0	0	396



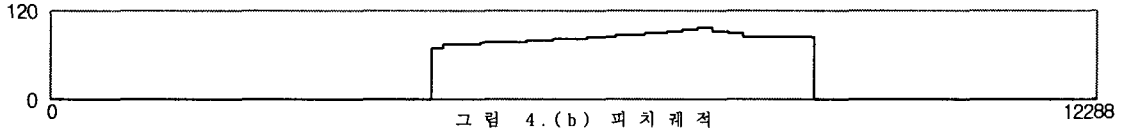
그 림 3.(a) 음 성 신 호 /3/



그 림 3.(b) 피 치 궤 적



그 림 4 .(a) 음 성 신 호 /7/



그 림 4.(b) 피 치 궤 적

감사의 글

본 연구는 2004년도 (주)퓨전소프트 연구조성비 지원에 의하여 수행되었으며, 이에 감사드립니다.

Ⅳ. 결 론

본 본문에서는 음성신호의 피치를 구하기 위하여 시간 영역에서 피치의 시작점, 주파수 영역에서 분산을 이용하여 3개 피치를 검출하는 알고리즘을 제안하였으며 주파수 영역에서 계산되므로 속도는 느리지만 정확성에 목적을 두었다. 총 864개 단독 숫자음을 이용하여 실험한 결과 99.5%의 정확도를 확인하며 제안된 알고리즘의 유효성을 확인하였으며 4개의 에러는 음성 신호에 비교적 고주파가 포함된 경우이며 이 에러는 알고리즘을 적용하기 전 LPF를 이용하면 해결된다고 판단되며 이 결과는 음성인식 시스템을 구성하는 경우 남, 여 구분에 의해 인식율이 저하되는 경우 해결할 수 있는 기술이 된다. 분산값을 이용하여 숫자음에서 'ㄱ', 'ㅅ', 'ㅊ', 'ㅍ' 초성의 구분 및 유성음의 끝점 검출이 가능하였으며 다음 연구 대상은 평균피치를 구하는 방법을 이용하여 유성음 구간 결정, 유성음에서 각 피치의 값 산출, 연속음성에 적용하는 것을 연구하고 있다.

참고문헌

[1] Samuel D. Sterns, Ruth A. David, 1988. Signal Processing Algorithms, Englewood Cliff, NJ : Prentice-Hall

[2] 강동규, 한민수. 1995. "유성음 구간에서의 피치동기식 포먼트 추출." 제 8회 신호처리 합동학술대회 논문집 제 8권 1호. 75-79

[3] 나덕수, 고장영, 배명진. 1997 " 개선된 가변대역 LPF에 의한 피치검출법" 한국통신학회 학술대회 논문집. 16권 3호. 585-588

[4] 석종원, 손영호, 배건성. 1999. "웨이브렛 변환을 이용한 피치검출." 음성과학 제 5권 제 1호. 23-33

[5] 김종국, 조왕래, 배명진. 2003. "음성 하모닉스 스펙트럼의 피크-피팅을 이용한 피치검출에 관한 연구." 음성과학 제 10권 제 2호 85-95