

음성의 시간축 처리에 의한 피치추출

Estimating Pitch of speech in the time domain

이한식

충북대학교

1. 서론

음성의 피치 추출은 음성처리의 모든 부분에 응용되는 기초기술이며, 동시에 피치를 실 시간에서 추출하는 것은 무 넓은 응용범위를 갖는 기술이다.

피치 추출을 실시간에서 실행하려고 할 때 문제로 되는 것 중의 하나는 그 피치추출의 알고리즘과 연산이 복잡할수록 고속도의 하드웨어가 필요로 하게된다. 본헌(1)에 의하면 상당한 피치 추출법에 의해 1초간의 음성 피치 추출을 NOVA 800으로 행하였을 때의 처리시간을 표 1과 같이 보고하고 있다.

표 1. 1초간의 음성 피치를 NOVA 800으로 실행하였을 때의 처리시간

구 방법	처리시간(초)
DARD	5
PPROC	7.5
AMDF	50
AUTOC	120
SIFT	250
LPC	300
CEP	400

여기서 주파수축상에서의 변환 또는 선형 예측 계수, 자기상관함수 등을 이용한 방법은 그 연산의 번잡성에 의해 처리시간이 증가하는데 비하여 음성의 시간파형의 특징으로부터 피치추출을 하는 쪽이 처리속도가 느린 것

같다. 또한 이 시간파형의 처리도 디지털적으로 처리 하는 것보다 아날로그적으로 처리(가령 디지털 필터 보타 아날로그 필터)하면 처리속도의 향상을 예상 할 수 있다.

본 연구에서는 이와같은 점에 고려, 저렴한 고 범용 마이크로 컴퓨터에 의해 실시간 피치추출을 실행하는 시스템 즉 Automatic triggering Method (ATM 로 약기함)을 이형 컴퓨터시뮬레이션에 의한 결과를 발표한다.

2. 본론

(1) 시스템구성 및 동작원리

ATM의 블록을 만들어 입력신호는 통과역을 50 - 200 Hz 의 대역필터(BPF)을 통과시킴으로써 필요한 고조파, 잡음등이 제거된다. BPF를 통과한 신호는 경의(부외) 마이크로호출드회로는 일단 회로에서 샘플값을 받으면 출력전압은 강제적으로 입력전압까지 증가(상승)하도록 하였다. 이들의 마이크로호출드회로의 출력은 두개의 저항에 의해 나누어져 분압된 출력이 다음단의 두 비교회로 I, II에 V_+ , V_- 로서 주어진단. I는 V_+ 를 드레시호출드 레벨로서 입력신호는 이 값을 웃돌면 다음단의 RS-FF를 셋한다. II는 V_- 를 드레시호출드 레벨로서 입력신호가 이 값을 밑돌면 RS-FF를 리셋한다. RS-FF의 출력 QQ의 상승에지에서 미분회로에 의해 푸이 좁은 펄스가 만들어져 경부의 마이크로호출드회로에 각각 샘플 펄스로서 주어진다. 이들의 동작 타임 차트를 그림 1에 나타낸다.

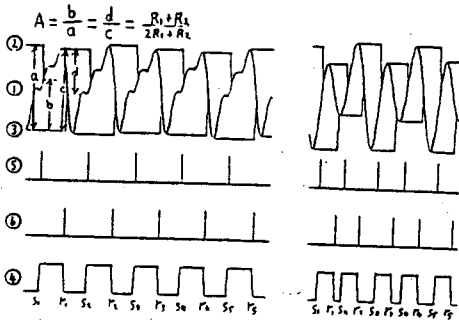


그림 1. 파형처리 부동 타임차드



그림 2. 강제 샘플동작

또 입력파형에 과도적인 큰 진폭부분이 있을 경우 타운저 포락선의 급격한 하강이 있을 경우 이 과도에 의하면 RS-FF의 출력은 H 또는 L을 계속시켜 나갈 우려가 있다. 이를 방지하기 위해 뒷 단에서 소프트웨어 플루트로 그림 2와 같이 RS-FF의 출력이 시간 T 이상 H 또는 L을 지속시켰을 경우 강제적으로 샘플 펄스를 발생하도록 한다. 이 같은 예측 가능한 최장피치(25ms)로 하면 좋다.

(2) 소프트웨어 플루

앞 단에서 파형처리 블록의 RS-FF 출력은 마이크로 컴퓨터를 사용한 소프트웨어 플루에 들어간다. 이 소프트웨어 플루에서는 RS-FF 출력 Q의 가장 최근 5개의 상승에지 시각을 S_1, S_2, S_3, S_4, S_5 로, 하강에지 시각을 r_1, r_2, r_3, r_4, r_5 로 각각 기억시켜둔다. 여기서 피치의 후보로 되는 것은 S_1-S_j 또는 r_1-r_j 이다. (그림 1 참조) .면

저 최종적인 피치후보를 선택하기 위해 S_1-S_j r_1-r_j 를 표 2와 같은 4 개의 그룹으로 나뉘워진다. 동일 그룹내에 있는 두 요소 $t_{mn}, t_{mn'}$ 의 차가 ϵ 이내 즉 $|t_{mn} - t_{mn'}| \leq \epsilon$ 때 이 두 요소는 일치된 것으로 부인될 것으로 한다.

최종적인 피치후보는 일치하는 요소를 가장 많이 가진 요소로서 선택된다. 이 최종적인 피치후보가 갖는 일치요소의 개수 C 가 적을 때에는 RS-FF 출력은 규칙적으로 H, L 를 되풀이 된 것이 아니며 입력파형은 피치를 갖지않는 무성음 또는 무음이었을 가능성이 강하다. 따라서 이 C 가 어느값 V 이상일 때는 최종피치 후보를 피치로 출력하며 V 미만일 때는 무성 또는 무음으로 간주하며 그 해당을 출력한다. 이상의 실제 예를 보이기 위해 $\epsilon = 100\mu s, V = 5$ 에서 표 2 의 각 요소가 표 3과 같이되었다면 각각의 일치하는 요소의 개수는 표 4와 같이된다.

표 2, S_1-S_j, r_1-r_j 의 그룹분류

group1	group2	group	group
group1	group2	group3	group4
$t_{11}=s_2-s_1$	$t_{21}=s_3-s_1$	$t_{31}=s_4-s_1$	$t_{41}=s_5-s_1$
$t_{12}=s_3-s_2$	$t_{22}=s_4-s_2$	$t_{32}=s_5-s_2$	$t_{42}=r_5-r_1$
$t_{13}=s_4-s_3$	$t_{23}=s_5-s_3$	$t_{33}=r_4-r_1$	$t_{43}=r_5-r_2$
$t_{14}=s_5-s_4$	$t_{24}=r_3-r_1$	$t_{34}=r_4-r_2$	$t_{44}=r_5-r_3$
$t_{15}=r_2-r_1$	$t_{25}=r_4-r_2$	$t_{35}=r_5-r_3$	
$t_{16}=r_3-r_2$	$t_{26}=r_5-r_3$		
$t_{17}=r_4-r_3$			
$t_{18}=r_5-r_4$			

표 3, 표 2의 실제예

group1	group2	group3	group4
7.5ms	15.1ms	22.8ms	30.2ms
7.6	15.3	22.9	30.1
7.7	15.1	22.6	
7.6	15.0	22.6	
7.5	15.1		
7.5	15.1		
7.6			
7.5			

표 4, 표3 의 각 요소에 일치하는 요소 개수

수	group1	group2	group3	group4
	7	5	2	2
	8	1	2	2
	4	5	2	
	8	5	2	
	7	5		
	7	5		
	8			
	7			

예를들면 t_{11} 와 일치하는 요소는 t_{11} 자신과 t_{12} , t_{14} , t_{15} , t_{16} , t_{17} , 7 개다. 표 4에서 최종 피치 후보는 $t_{12} = t_{14} = t_{17} = 7.5ms$ 가된다. $C=825=V$ 이므로 7.5ms가 피치로서 출력된다. 이상 설명에서 알 수 있지만 그루프 1, 2, 3, 4, 5 의 각 요소가 같이 일치하는 요소의 개수는 각각 8, 6, 4, 2 보다 크게되지는 않는다. 따라서 $V=5$ 로 하면 그루프 3, 4 의 요소는 피치로서 출력될 수는 절대로 없기 때문에 이들의 그루프에 대한 계산을 할 필요는 없다. 값이 작아져서 이 점에 유의하여 불필요한 계산을 생략할 수 있다.

3. 결론

이상 시스템에 의해 컴퓨터시뮬레이션을 실행하여 본 결과를 그림 3, 4 에 나타낸다. 샘플로 사용한 음성은 "대한전기공학회" 와 "고슴도치" 의 두 종류를 남성음성으로 각각 ATM 로 구한 피치로 전술한 파라메터 A, ξ, τ 의 값은 각각 $A=0.8, \xi=0.3ms, \tau=5$ 이다. 이상 본 연구에서 ATM 은

첫째, 국히 간단한 하드웨어 구성으로 저렴한 비용으로 실현가능한 실시간 피치추출 시스템이며,

둘째, 사용된 컴퓨터처리의 알고리즘은 기본적인 정수의 합, 차, 대소비교 만으로 구성되어 있으며 고속으로 처리가능하며 또한 이에 필요한 소프트웨어양도 적으며 현재 시중에 흔한 한컴 마이크로컴퓨터로도 실현된다.

대한 전기 공 학 회

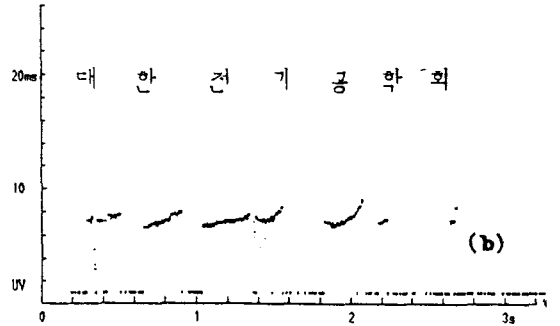
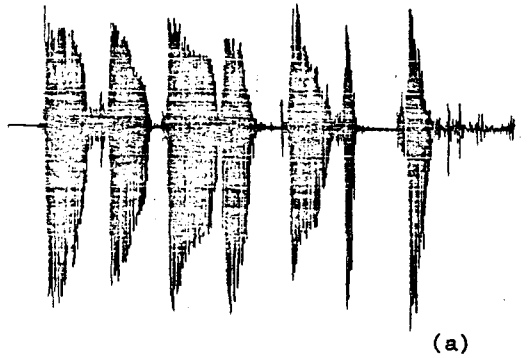


그림 3 "대한 전기 공 학 회" 의 시간파형(a) ATM 결과(b) 고 슝 도 치

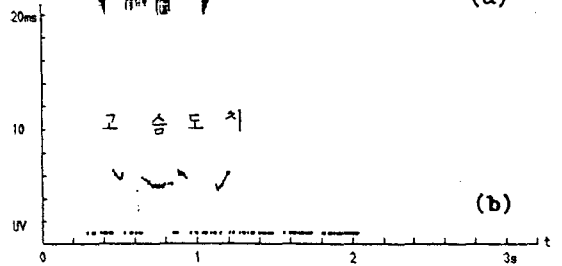


그림 4 "고슴도치" 의 시간파형(a) ATM 결과(b)

참 고 문 헌

(1) L.R.Rabiner, M.J.Cheng, A.E.rosenberg and C.A.Mcgonagal, "A Comparative Performance Study of Several Pitch Detection Algorithms" IEEE Trans. Acoust., Speech, Signal Processing, Vol.ASSP-24, pp.399-418, Oct.1976.
 (2) J.D.Markel, A.H.Gray, Jr. Linear prediction of speech. Springer-Verlag Berlin Heidelberg new york 1980.