

음성신호의 조파구조를 리용한 피치주파수추출의 한가지 방법

박성윤, 림현수

본문에서는 음성인증에서 중요한 기술의 하나인 피치주파수추출의 한가지 방법을 연구하였다.

선행연구[1]에서는 음성신호의 자기상관함수의 루적평균을 취하여 그것의 최소값을 얻어 피치주파수를 추출하는 방법(YIN)을 제기하였다.

선행연구[2]에서는 케프스트람결수의 극대점을 잡는 추출방법과 음성신호의 조파구조를 리용하여 주파수영역에서 반복적인 주파수들의 진폭을 곱한 값의 최대값을 찾는 HPS (Harmonic Product Spectrum)방법을 결합한 피치주파수추출방법을 취급하였다.

선행연구[3]에서는 음성신호의 연속웨블레트결수들의 척도에 따르는 평균을 계산하고 그것의 극대점을 얻어 피치주파수를 추출하는 방법을 주었다.

선행연구[1, 2]의 방법은 계산량이 작은 우점이 있지만 음성신호의 조파구조만 고려하고 반조파구조를 고려하지 못한것으로 하여 반피치 및 배피치오류률이 높은 결함을 가지고있으며 선행연구[3]의 방법은 정확성이 높으나 계산량이 많은 결함을 가지고있다.

본문에서는 조파구조와 반조파구조의 진폭비를 리용하여 정확한 피치주파수를 고속으로 추출하는 알고리즘을 연구하였다.

1. 반조파구조와 조파구조의 진폭비(SHR)

$A(f)$ 를 음성신호의 주파수영역에서의 진폭함수라고 하자.

f_0 과 f_{\max} 를 $A(f)$ 의 기본주파수와 최대주파수라고 하자. 이제 조파구조의 진폭합을 다음과 같이 표시하자.

$$SH = \sum_{n=1}^N A(nf_0) \quad (1)$$

여기서 N 은 조파구조의 최대개수이다.

$f > f_{\max}$ 이면 $A(f)=0$ 으로 한다. 피치주파수를 찾는 구간을 $[F0_{\min}, F0_{\max}]$ 라고 하면 $N = \text{floor}(f_{\max} / F0_{\min})$ 로 된다. 실천에서는 계산량을 줄이기 위하여 한부분만을 리용한다.

이때 $f_{\max} = 450\text{Hz}$ 로 놓는다.

한편 반조파구조의 진폭합은 다음과 같이 정의된다.

$$SS = \sum_{n=1}^N A((n-1/2)f_0) \quad (2)$$

SHR는 SS를 SH로 나누어 얻어진다.

$$SHR = SS / SH \quad (3)$$

2. SHR에 의한 피치주파수추출알고리즘

SH 를 계산하려면 주파수령역을 따라 주파수의 배수만큼씩 이동해가야 한다. 이것을 표시하기 위하여 다음의 함수를 받아들인다.

$$SUMA(f)_{even} = \sum_{n=1}^{2N} A(2nf) \quad (4)$$

$f > f_{\max}$ 일 때 다음의 사실을 얻는다.

$$SUMA(f_0/2)_{even} = SH \quad (5)$$

$$SUMA(f_0/4)_{even} = SH + SS \quad (6)$$

이와 유사하게 주파수령역을 따라 주파수의 $(2n-1)/2$ 배만큼씩 이동해가면 다음의 사실을 얻는다.

$$SUMA(f)_{odd} = \sum_{n=1}^{2N} A((2n-1)f) \quad (7)$$

$$SUMA(f_0/2)_{odd} = SS \quad (8)$$

$$SUMA(f_0/4)_{odd} = \Delta \quad (9)$$

여기서 Δ 는 $(1/4+n)f_0$ 을 나타낸다.

새로운 함수를 하나 정의하자.

$$SUBA(f) = SUMA(f)_{even} - SUMA(f)_{odd} \quad (10)$$

식 (7)–(9)로부터 다음의 결과를 쉽게 알 수 있다.

$$SUBA(f_0/2) = SH - SS \quad (11)$$

$$SUBA(f_0/4) = SH + SS - \Delta \quad (12)$$

만일 음성이 규칙적인 음성이라면 즉 반조파구조의 영향이 충분히 약하여 $SS \approx 0$ 이라면 $SUBA(\cdot)$ 의 최대값은 $f_0/2$ 으로 될 것이다. 한편 반조파구조가 실질적인 영향을 준다면 최대값은 대체로 $f_0/4$ 으로 될 것이다. 이때 $\Delta \approx 0$ 이고 두번째 최대값은 $f_0/2$ 으로 된다.

식 (11), (12)를 리용하여 SHR 를 대략적으로 계산할 수 있다.

$$\frac{SUBA(f_0/4) - SUBA(f_0/2)}{SUBA(f_0/4) + SUBA(f_0/2)} = \frac{SS - \Delta/2}{SH - \Delta/2} \quad (13)$$

최대값을 찾을 때에는 우선 대역적 최대값을 선택하고 그 위치를 f_1 로 한다. 다음 구간 $[1.9375f_1, 2.0625f_1]$ 에서의 최대값을 선택하고 f_2 로 한다.

그러나 다음의 경우에는 다르게 처리한다.

① 임의의 f 에 대하여 $SUBA(f) \leq 0$ 일 때 이 프레임구간은 무성구간이라고 본다.

② $1.9375f_1 > F_{0\max}$ 일 때에는 f_1 이 피치주파수로 된다.

③ $SUBA(f_1) > 0$ 이고 $SUBA(f_2) \leq 0$ 일 때에는 f_1 의 피치주파수로 된다.

2개의 최대점을 찾은 다음 식 (13)으로부터 SHR 를 계산할 수 있다.

$$SHR = \frac{SUBA(f_1) - SUBA(f_2)}{SUBA(f_1) + SUBA(f_2)} \quad (14)$$

위의 식으로 계산한 SHR 값이 턱값 1보다 작으면 그 프레임은 조파구조에 따른다고

본다. 이때 f_2 가 선택되고 피치주파수값은 $2f_2$ 로 된다. 한편 SHR 값이 턱값 2보다 큰 경우에는 f_1 이 선택되어 피치주파수값은 $2f_2$ 로 된다. 실험적으로 확인한데 의하면 2개의 턱값은 0.2와 0.4이다.

3. 성능평가실험

논문에서는 피치주파수검출의 정확성을 평가하기 위하여 5명의 남성과 5명의 여성의 음성자료가 들어있는 자료기지를 선정하였다. 이 자료기지에 대하여 선행연구에서 언급한 3개의 피치주파수추출방법과 논문에서 연구된 방법의 성능을 비교하였다.

실험결과는 다음과 같다.(표 1, 2)

표 1. 남성음성자료에 대한 피치주파수추출방법비교

피치추출방법	반피치 및 배피치오율/%					평균
	남성 1	남성 2	남성 3	남성 4	남성 5	
방법[1]	4.49	11.36	2.74	2.59	1.59	4.35
방법[2]	4.36	9.32	2.30	2.96	1.59	3.90
방법[3]	3.32	8.78	1.27	1.66	1.08	3.22
론문의 방법	2.9	7.49	0.41	1.55	1.27	3.08

표 2. 여성음성자료에 대한 피치주파수추출방법비교

피치추출방법	반피치 및 배피치오율/%					평균
	여성 1	여성 2	여성 3	여성 4	여성 5	
방법[1]	11.23	6.15	6.62	7.15	2.74	6.78
방법[2]	6.31	3.21	4.98	4.66	1.08	3.99
방법[3]	6.02	2.78	3.24	4.26	1.13	3.48
론문의 방법	4.22	2.63	1.66	3.61	0.59	2.74

론문의 방법에 의한 피치주파수추출은 선행연구[3]에서보다 3배이상 빠르다.

참 고 문 헌

- [1] D. Zhiyao; Computer Audition, 21, 7, 10, 2017.
- [2] H. Ding, L. Yanping; IEEE, 48, 8, 2, 2017.
- [3] M. K. Reddy, K. S. Rao; IEEE, 50, 5, 1, 2016.

주체107(2018)년 9월 8일 원고접수

A Pitch Determination Method using Harmonic of Speech Signal

Pak Song Yun, Rim Hyon Su

In this paper, we study a pitch determination method which is one of the most important parts of speaker recognition. We find out pitch candidates and compute SHR and determinate correct pitch among them.

Key words: pitch, harmonic, sub-harmonic