포먼트 주파수 추출 알고리즘들의 성능 비교평가 연구

손 성 용, 김 상 진, 김 영 민, 한 민 수
한국정보통신대학원대학교

A Comparative Study on Formant Frequency Extraction Performances

Sungyung Son, Sang-Jin Kim, YoungMin Kim, Minsoo Hahn
Information Community University
E-mail : {thill, sangjin, ymkim, mshahn}@icu.ac.kr

Abstract

In this paper, we compared formant frequency extraction algorithms with various conditions, and show their performances. The formant frequency is the resonance frequency which is decided by the vocal tract characteristics. It is related with phonemes, or characteristics of the physical condition of the vocal track. Since the speech signal is influenced by both the sound source and the vocal tract, it is difficult to calculate the exact formant frequencies. Many studies on the formant frequency extraction had been executed already. Besides, any new formant frequency extraction algorithm is hardly found recently.

Ⅰ. 서 론

음성의 특징을 잘 나타내는 요소 중 하나는 성도(vocal tract)의 특성이다. 성도는 성문으로부터 입술까지 일종의 파이프 모양으로 이루어진 구조이다. 음성은 이 파이프의 모양에 따라 여러 종류로 변하게 된다. 즉 화자의 성별, 나이 및 신체적 특성, 발성하고자 하는 음소에 따라 달라지게 된다. 이런 원리는 파이프의 길이와 굵기에 따라 다른 소리가 생성되는 파이프 오르간에서 쉽게 이해할 수 있다. 이러한 성도의 구조가 알려져 있다면 성도의 공진 주파수는 계산될 수 있다. 하지만 음성 신호는 음원과 성도, 두 가지의 영향을 모두 받기 때문에 정확히 계산되어지기 어렵다. 예를 들어, 만약 음원의 스펙트럼이 성도의 공진 주파수 중 하나와 근접한 영점을 가지고 있다면 각각의 포먼트 및 대역폭 등을 정의하기 어렵다. 포먼트 추출에 관한 연구는 이미 오래 전에 활발히 이루어졌으며, 최근까지도 새로운 추출방법이 제시된 예는 찾아보기 어렵다. 본 연구의 목적은 이미 소개되어 있는 포먼트 추출 알고리즘을 구현하고 그들을 성능을 비교하여 가장 좋은 결과를 보이는 방법을 찾아보는데 있다.

Ⅱ. 포먼트 추출 알고리즘

2.1 Peak-Picking Method

Peak-Picking method는 분석구간의 스펙트럼 상에서 peak를 찾음으로써 포만트의 위치를 구하는 방법이다[1]. 본 연구에서는 LPC, LPC-Cepstrum, FFT-Cepstrum을 몇 가지 창 함수에 대하여 Peak-Picking 방법을 적용해 보았다.

2.1.1 LPC / LPC-Cepstrum / FFT-Cepstrum

선형 예측 계수, LPC는 음성 분석 방법의 가장 기본이 되는 것 중의 하나로써 널리 이용되어 왔다. 시간 t 에서의 음성샘플을 $x(t)$라고 하자. 이때, 선형 예측 분석법에서는 현재의 음성 샘플을 이전 예측값과 실제값 사이의 차이를 p개의 샘플로부터 예측한다. 이

때, 예측값과 실제값 사이의 차이를 $e(t)$ 라고 하면 아래의 식(1)이 성립한다.

$$x(t)=\sum_{i=1}^{p}a_i x(t-i)+e(t) \tag{1}$$

여기서 오차 $e(t)$의 값을 최소화하는 p개의 a_i 값들을 선형예측계수(LPC)라고 한다. 이렇게 얻어진 LPC 계수를 식(2)를 이용하여 LPC cepstrum을 구할 수 있다[2].

$$c(n)=a_n+\sum_{i=1}^{n-1}\left(\frac{i}{n}\right)c(i)a_{n-i} \tag{2}$$

주파수영역의 음성샘플을 $X(w)$이라고 할 때 FFT cepstrum은 아래의 식(3)에서와 같이 $X(w)$의 역 푸리에변환으로 구해진다.

$$c(n)=F^{-1}\{\log|X(w)|\} \tag{3}$$

2.1.2 다양한 창 함수 적용

Bartlett/Blackman/Hanning/Hamming/Rectangular와 같은 창 함수들이 프레임에 적용되었다.

2.1.3 Pre-emphasis

입술의 방사효과를 제거하기 위해 적용하는 전처리 과정의 하나인 pre-emphasis의 계수를 0.0, 0.95, 0.97로 변화, 적용하였다.

2.2 LPC Root-Finding Method

음성신호의 포만트는 극점의 집단에 의해 특징 지워진다는 사실에 근거하여 극점을 구함으로써 포만트의 위치를 추정 할 수 있다[3].

LPC을 구하는 방법은 여러 가지가 있으나 일반적으로 autocorrelation을 이용한 방법을 많이 사용하며, 보다 안정된 LPC를 구하기 위한 방법으로 Burg's LPC 알고리즘이 제안되기도 했다.

2.2.1 Autocorrelation LPC

Autocorrelation방법은 식 (4)을 만족하는 a_k를 구하는 방법이다.

$$\sum_{k=1}^{p}a_k R_n(|1-k|)=R_n(i)\ ,\quad 1\le i\le p \tag{4}$$

이를 만족하는 a_k는 다음과 같은 Durbin 알고리즘을 통하여 구할 수 있다.

$$\begin{aligned}
E^{(0)} &= R(0)\\
k_i &= \left(R(i)-\sum_{j=1}^{i-1}a_j^{(i-1)}R(i-j)\right)/E^{(i-1)},\ 1\le i\le p\\
a_i^{(i)} &= k_i\\
a_j^{(i)} &= a_j^{(i-1)}-k_i a_{i-j}^{(i-1)}\ ,\ 1\le j\le i-1\\
E^{(i)} &= (1-k_i^2)E^{(i-1)}\\
a_j &= a_j^{(p)}\ ,\ 1\le j\le p
\end{aligned} \tag{5}$$

2.2.2 Burg's LPC

Lattice structure를 이용하여 standard direct-form FIR filter가 구현되면 식(6)를 이용하여 우리는 직접적으로 reflection coefficients를 계산할 수가 있다.

$$k_i=\frac{\sum_{m=0}^{N-1}f_{i-1}(m)g_{i-1}(m-1)}{\left\{\left(\sum_{m=0}^{N-1}f_{i-1}(m)^2\right)\left(\sum_{m=0}^{N-1}f_{i-1}(m-1)^2\right)\right\}^2} \tag{6}$$

이 필터가 안정하기 위해서는 reflection coefficients는 1보다 작아야 한다. 여기서 reflection coefficients의 공식을 식 (7)로 변환해 줌으로써 보다 강한 안정된 조건에서 LPC 값을 구할 수가 있다[4].

$$k_i=\frac{\sum_{m=0}^{N-1}f_{i-1}(m)g_{i-1}(m-1)}{\frac{1}{2}\left\{\sum_{m=0}^{N-1}f_{i-1}(m)^2+\sum_{m=0}^{N-1}f_{i-1}(m-1)^2\right\}} \tag{7}$$

2.2.3 LPC Root-Solving Algorithm

선형 예측 다항식의 근은 n차의 LPC 계수를 이용하여 Laguerre's method로 근을 구한다. Laguerre's method 알고리즘은 다음 식과 같다[5].

$$\begin{aligned}
P_n &= (x-x_1)(x-x_2)\ldots(x-x_n)\\
\ln|P_n(x)| &= \ln|x-x_1|+\ln|x-x_2|+\ldots+\ln|x-x_n|\\
P_n'(x) &= (x-x_2)\ldots(x-x_n)+(x-x_1)\ldots(x-x_n)+\ldots\\
&= P_n(x)\left(\frac{1}{x-x_1}+\ldots+\frac{1}{x-x_n}\right)\\
\frac{d\ln|P_n(x)|}{dx} &= \frac{1}{x-x_1}+\frac{1}{x-x_2}+\ldots+\frac{1}{x-x_n}\\
&= \frac{P_n'(x)}{P_n(x)}\equiv G(x)
\end{aligned} \tag{8}$$

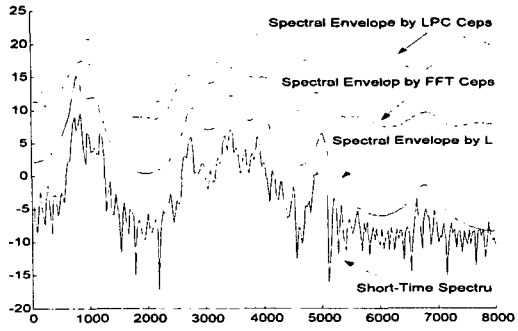

그림 1. LPC/LPC-Cep/FFT-Cep 스펙트럼 포락선 비교

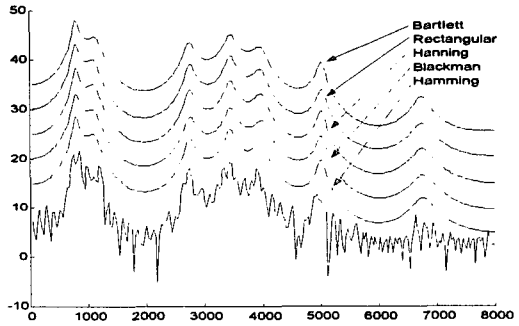

그림 2. 창 함수에 따른 LPC 스펙트럼 포락선

$$\frac{d^2 \ln|P_n(x)|}{dx^2} = \frac{1}{(x-x_1)^2} + \frac{1}{(x-x_2)^2} + \ldots + \frac{1}{(x-x_n)^2} = \left[\frac{P_n'(x)}{P_n(x)}\right]^2 - \frac{P_n''(x)}{P_n(x)} \equiv H(x)$$

여기서 $a = x - x_1$, $b = x - x_i$, $i = 2,3,\ldots,n$ 라 가정하면,

$$G = \frac{1}{a} + \frac{n-1}{b}, \quad H = \frac{1}{a^2} + \frac{n-1}{b^2} \tag{9}$$

이제 식 (9)를 정리하면 다음을 얻는다.

$$a = \frac{n}{\max[G \pm \sqrt{(n-1)(nH - G^2)}]} \tag{10}$$

초기값 x를 위 식들에 대입하여 a를 구하고 $x - a$ 값으로 다시 대입, 충분히 작은 a를 구할 때까지 반복한다. 이때 x값이 원하는 해가 된다.

2.2.4 LPC 근을 이용한 포만트 위치와 대역폭 추출

식(11)을 이용하면 위에서 구한 극점과 샘플링 주기 T_s을 이용하여 포먼트의 위치와 대역폭을 구할 수 있다[6].

$$F_k = (1/2\pi T_s) Im(\ln z_k) \tag{11a}$$

$$B_k = (1/\pi T_s) Re(1/\ln z_k) \tag{11b}$$

Ⅲ. 실험 및 결과

3.1 실험 조건

20대 남성이 발성한 모음/아/에 대해 16 kHz로 샘플링하고 16 bit로 저장하여 실험에 사용하였으며, LPC 차수는 18을 사용하였다.

3.2 Peak-Picking Method

그림 1을 통하여 LPC, LPC-cepstrum, 및 FFT-cepstrum을 비교해 보면 LPC가 포먼트 구조를 보다 잘 표현함을 알 수 있다. Peak-Picking 방법을 통하여 구한 표 1을 통하여 이를 재차 확인할 수 있다.

그림 2는 다양한 창 함수를 사용하였을 때 스펙트럼 포락선을 보여주며, 표 2를 통하여 포먼트 추출 결과를 비교할 수 있다. Hanning창을 사용한 경우 가장 좋은 결과를 얻을 수 있었다.

Pre-emphasis의 영향은 그림 3과 표 3을 통해서 확인할 수 있다. 0.97을 계수로 사용한 경우 가장 좋은 결과를 보임을 알 수 있다.

표 1. 20대 남성, 모음 /a/에 대한 포만트 추출결과

/아/ 모음	Ref	FFT-Ce		LPC		LPC-Cep	
		Est.	Error	Est.	Error	Est.	Error
1st	757	297	478	**775**	**-18**	992	-235
2nd	1080	1798	-718	**1054**	**26**	3317	-2237
3rd	2723	2108	615	**2728**	**-5**	4030	-1307
4th	3416	3038	378	**3410**	**6**	5425	-2009
5th	3962	3844	118	**3937**	**25**	7471	-3509

표 2. 창 함수별 포만트 예측 결과

/아/ 모음	Ref	Rect		Bart		Hamm		Hann		Black	
		Est.	Err	Est.	Err	Est.	Err	Est.	Err	Est.	Err
1st	757	775	-18	775	-18	775	-18	**775**	**-18**	775	-18
2nd	1080	1054	26	1085	-5	1054	26	**1085**	**-5**	1054	26
3rd	2723	2728	-5	2728	-5	2728	-5	**2728**	**-5**	2728	-5
4th	3416	3410	6	3410	6	3410	6	**3410**	**6**	3410	6
5th	3962	3906	56	3906	56	3937	25	**3937**	**25**	3937	25

표 3. Pre-Emphasis 변화에 따른 포먼트 추출 비교

/아/ 모음	Ref	alpha = 0		alpha = 0.95		alpha = 0.97	
		Est.	Err	Est.	Err	Est.	Err
1st	757	806	-49	775	-18	**775**	**-18**
2nd	1080	2759	-1679	1054	26	**1085**	**-5**
3rd	2723	3472	-749	2728	-5	**2728**	**-5**
4th	3416	3937	-521	3410	6	**3410**	**6**
5th	3962	5022	-1060	3937	25	**3937**	**25**

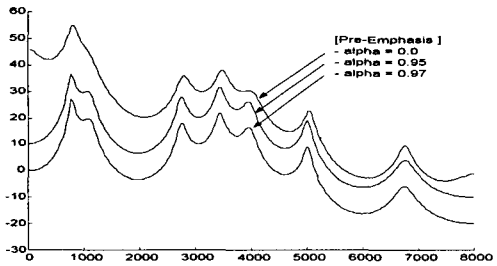

그림 3. Pre-Emphasis 변화에 따른 스펙트럼 포락선

3.3 LPC Root-Solving Method

앞서 Peak-Picking 방법에서 가장 좋은 결과를 보였던, Hanning 창 함수에 0.97 pre-emphasis의 조합을 적용 후 18차 LPC를 구했다.

Autocorrelation LPC와 Burg LPC에 대해 Root-Solving을 통한 포먼트 주파수 추출 결과는 표 4와 같다. Root-Solving 방법이 보다 정확한 1,2차 포먼트 주파수를 추출함을 알 수 있다.

Ⅳ. 결 론

포먼트 주파수의 추출을 위해서 가장 간단한 방법인 Peak-Picking 방법과 일반적인 LPC Root-Solving 방법을 구현해 보았다.

Peak-Picking 방법으로 포만트를 추출하기 위해서 성도의 적절한 모델링, 적절한 창 함수의 선택이 요구된다. LPC, LPC-Cep, 및 FFT-Cep에 대하여 다양한 창 함수, 즉 Rectangular, Bartlett, Hanning, Hamming, Blackman 창 함수를 Pre-emphasis의 변화와 함께 적용하였다. LPC 스펙트럼에 Hanning, Hamming, Blackman 창 함수를 사용했을 경우 가장 좋은 결과를 보였으며, 입술 방사 효과 제거를 위해 0.97의 계수를 가지는 선형 함수로 필터링을 하였을 때, 제1, 2, 3, 4, 5 포만트는 각각 18Hz, 5Hz, 5Hz, 6Hz, 25Hz의 오차를 보였다.

일반적으로 많이 사용되는 Autocorrelation을 이용한 LPC와 보다 안정적이라고 알려진 Burg's LPC에 대하여 Root-Solving 알고리즘을 이용하여 포먼트를 추출하였다. Autocorrelation LPC와 Burg's LPC를 이용하여 추출된 포먼트가 대체적으로 비슷한 결과를 보이며, Peak-Picking 방법보다 Root-Solving을 사용한 경우에 좀 더 정확한 1,2차 포먼트를 구할 수 있었다.

포먼트 추출시 주의할 점은, 두개 이상의 포만트가 서로 가까운 위치에 존재할 때 마치 하나의 포만트로 예측되는 점을 들 수 있으며 이와 같은 경우에도 정확한 포만트 예측을 위해서는 좀 더 보완이 이루어져야 하겠다. 한 문장에 대한 포먼트 주파수 궤적 추출의 구현을 진행 중이다.

표 4. LPC Root-Solving Method 결과 비교

/아/ 모음	Ref.	Peak-Picking		AutoCorr.LPC		Burg's LPC	
		Est.	Error	Est.	Error	Est.	Error
1st	757	775	-18	748	9	749	8
2nd	1080	1085	-5	1078	2	1078	2
3rd	2723	2728	-5	2713	10	2713	10
4th	3416	3410	6	3385	31	3385	31
5th	3962	3937	25	3899	63	3899	63

참고문헌

[1] J.L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer-Verlag, 1972.

[2] B.S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Am., vol.55, no.6, pp.1304-1312, 1974.

[3] H. Mizuno and M. Abe, "Voice conversion algorithm based on piecewise linear conversion rule of formant frequency and spectrum tilt," *Speech Communication*, Vol.16, pp.153-164, 1995.

[4] J.D. Markel and A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, 1976.

[5] W.H. Press, et al., *Numerical Recipes in C*, Cambridge University Press, 1988.

[6] B.S. Atal and S.L. Hanauer "Speech analysis & Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Am.*, Vol.50, pp.637-665, Aug. 1971.