

RLSL 적응선형예측필터를 이용한 형성음 및 조음운동 궤적 추정에 관한 연구

○ 김 동준*, 송 영수*, 윤 태성**, 박 상희*

* 연세대학교 전기공학과 ** 창원대학교 전기공학과

A Study on Estimation of Formant and Articulatory Motion using
RLSL Adaptive Linear Prediction Filter

○ Dong Jun Kim*, Young Soo Song*, Tae Sung Yoon**, Sang Hui Park*
* Dept. of Electrical Engineering, Yonsei Univ.
** Dept. of Electrical Engineering, Changwon Univ.

ABSTRACT

In this study, the extractions of formant and articulatory motion trajectories from Korean diphthongs are performed by using the RLSL adaptive linear prediction filter. This enables us to extract spectrum transition of speech signal accurately.

This study shows that the RLSL algorithm is superior to the Levinson algorithm, specially in transition part of speech.

1. 서론

음성에서 형성음 주파수(formant frequency)는 가장 중요한 매개변수중의 하나이다. 형성음 주파수를 추출하는 방법은 일반적으로 특정길이의 음성신호를 블록 단위로 계수를 추출하고 스펙트럼을 추출하게 된다. 이러한 방법은 신호의 특성이 변화하는 부분에서는 정확한 분석이 어렵다.

본 연구에서는 이중 모음 사이의 통계적 특징 변화가 급격한 곳에서도 정확한 해석을 하기 위하여 신호에 대한 적응속도가 빠른 RLSL(Recursive Least Square Lattice)적응 알고리즘을 사용하여 선형 예측 필터를 구성하고, 대수 스펙트럼을 구한 후, 그로부터 형성음 주파수를 추출하고자 한다. 또한 추출된 형성음 주파수를 이용하여 조음기관 앞 부분의 조음운동을 추정하여 결과를 고찰하고자 한다. 그리고 객관적인 평가를 위하여 Levinson 알고리즘에 의한 결과를 비교·고찰한다.

2. RLSL 적응 알고리즘을 이용한 선형 예측 필터

그림 1은 적응 알고리즘을 적용한 선형 예측 필터를 나타낸다. 대표적인 적응 알고리즘인 LMS 알고리즘을 적용한 선형 예측 필터는 식 (1)과 같다.

$$\xi = E[en^2] = A^T R A \quad (1)$$

여기서, A는 $[1 \ a_1 \ a_2 \ a_3 \ \dots \ a_M]^T$ 인 필터 계수이며, M은 필터의 차수이다.

n번째 계수에 대한 시간 n에서의 변화율은 다음과 같다.

$$\Delta a_m(n) = -2 \mu e_n y_{n-m} \quad (2)$$

여기서, μ 는 적절하게 선택되는 상수값이다.

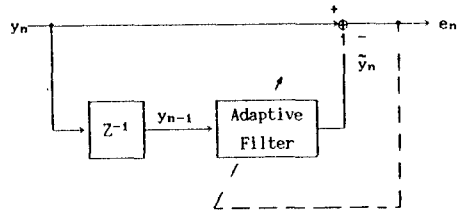


그림 1 적응 선형 예측 필터

LMS 알고리즘은 계산이 간단하고 예측 계수가 수치적으로 안정성을 가지고 있으나 최적의 수치에 이르기까지 적응속도가 낮기 때문에 음성 신호 및 빠른 천이 부분이 포함된 신호에 대해서는 적합하지 않다. 그러므로 이중 모음 사이의 천이 부분에 대한 정확한 분석을 위해서는 보다 빠른 적응성을 갖는 알고리즘이 필요하다.

이를 위하여 본 연구에서는 RLS(Recursive Least Square) 알고리즘을 사용하여 lattice구조의 선형 예측 필터를 구성하고 음성 신호의 천이 부분에 적용하여 정확한 분석을 할 수 있도록 한다.

일반적인 RLS 알고리즘을 lattice 구조의 필터로 구성하는 것은 다음과 같은 장점을 갖는다.

1. 전향(forward)예측 오차나 후향(backward)예측 오차가 동시에 효율적으로 간단히 구해질 수 있다.
2. 각 lattice 단 마다 서로 'decoupled'의 효과가 있다. 즉, 전·후향 예측 오차와 입력 신호간에 서로 orthogonal한 특성이 있다.
3. 구조 자체가 모듈화되어 있기 때문에 차수의 증가시 단 순히 한 단씩만을 추가하면 되므로 진단에서의 과정에 대해 서로 무관하다.
4. 구조의 단순성 때문에 응용분야에 따른 VLSI 기술의 적용이 쉽다.
5. 음성발생 모델에 일치되는 특성을 가지고 있다.

이와 같은 lattice 구조의 장점과 RLS 알고리즘의 빠른 수렴성의 결과로 우수한 선형 예측 결과를 얻게 된다. lattice 구조로 선형 예측 필터를 구성하면 그림 2와 같다.

표 1은 double/direct RLSL 알고리즘을 요약한 것으로 autocorrelation이나 가중 오차 계급 합을 계산하지 않고도 posteriori/priori 오차를 사용하여 빠른 적응성과 정확한 수치를 산출해 내는 장점이 있다[1].

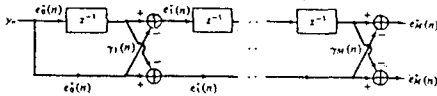


그림 2 Lattice 선형 예측 필터

표 1 Double/direct RLSL 알고리즘

1.	$e_{0a} = \tilde{e}_{0a} - \gamma_{0b} \tilde{e}_{0b}$ $e_{0b} = \tilde{e}_{0b} - \gamma_{0a} \tilde{e}_{0a}$
2.	$\gamma_{1a} = \gamma_{0a} + e_{0b} \frac{\tilde{e}_{1a}}{\tilde{E}_{1a}}$ $\gamma_{1b} = \gamma_{0b} + e_{0a} \frac{\tilde{e}_{1b}}{\tilde{E}_{1b}}$
3.	$e_{1a} = \tilde{e}_{1a} - \gamma_{1b} \tilde{e}_{1b}$ $e_{1b} = \tilde{e}_{1b} - \gamma_{1a} \tilde{e}_{1a}$
4.	$E_{1a} = E_{0a} - e_{1a} e_{0a}$ $E_{1b} = E_{0b} - e_{1b} e_{0b}$

3. 형성음 및 조음운동 추정

3.1 대수 스펙트럼 추출

lattice 선형 예측 필터에서 구하여진 반사 계수는 식 (3)의 Levinson recursion에 의하여 선형 예측 계수로 전환되어 스펙트럼이 구해진다.

$$\begin{bmatrix} 1 \\ a_{m+1,1} \\ a_{m+1,2} \\ a_{m+1,3} \\ \vdots \\ a_{m+1,m} \\ a_{m+1,m+1} \end{bmatrix} = \begin{bmatrix} a_{m,1} \\ a_{m,2} \\ a_{m,3} \\ a_{m,4} \\ \vdots \\ a_{m,m} \\ 0 \end{bmatrix} - \gamma_{m+1} \begin{bmatrix} 0 \\ a_{m,m} \\ a_{m,m-1} \\ a_{m,m-2} \\ \vdots \\ a_{m,1} \\ 1 \end{bmatrix} \quad (3)$$

여기서, γ_m 은 반사 계수, $a_{m,n}$ 은 선형 예측 계수, m 은 선형 예측 차수를 나타낸다.

윗 식에서 구하여진 선형 예측 계수를 이용하여 임의의 순간 n 에서의 스펙트럼 $S_n(\omega)$ 을 구하면

$$S_n(\omega) = \frac{1}{|1 + \sum_{i=0}^m a_i(n) e^{-j\omega i}|^2} \quad (4)$$

이며, 대수 스펙트럼(log spectrum) $L_n(\omega)$ 은 식 (5)에 의하여 얻게 된다.

$$L_n(\omega) = 20 \log_{10} \{ S_n(\omega) \} \quad (5)$$

적용 알고리즘에서는 각 계수가 입력신호에 대해 매 순간마다 계속적으로 적용되므로 원하는 순간에 스펙트럼을 추출해 낼 수 있다.

3.2 형성음 주파수 추출

형성음 주파수는 성문을 거친 공기가 조음기관의 형태에 따라 성도내에서 공진되어 발생하는 주파수이다.

본 연구에서는 선형 예측 분석에 의해 구해진 계수로부터 대수 스펙트럼을 구하고 그로부터 첨두치를 검출하여 형성음 주파수를 구한다. 그러나 형성음이 뚜렷한 첨두치가 아닌 번곡점으로 나타나는 경우, 일반적인 첨두치 검출 방법에 의해 형성음을 구하면 검출되지 않을 수도 있다. 따라서 1차, 2차 미분을 이용한 첨두치 추출방법을 사용하여 형성음 주파수를 찾아내기로 한다[2].

3.3 조음 운동 추정

음성 신호의 반사 계수 또는 PARCOR(partial correlation) 계수는 인간의 성도 모델과 일치한다. 따라서 음성 분석 및 성도에 대한 고찰을 위하여 lattice구조로 필터를 구성하는 것은 매우 적합하다고 볼 수 있다.

추출된 1, 2, 3차 형성음 주파수를 이용하여 조음기간 앞단의 운동 거리를 추정한다[3].

양 입술이 발음할 때 움직이는 거리는 식 (6)으로 표현된다. 식의 계수들은 Harshman등의 PARAFAC 분석 결과를 한국어 모음에 맞게 수정한 실험치이다[3][7].

$$D = aF_2 + bF_2F_3 + c(F_1/F_2) + d \quad (6)$$

여기서, F_1, F_2, F_3 는 1, 2, 3차 형성음 주파수이며,

- $a = 0.3 \times 10^{-3}$
- $b = -0.343 \times 10^{-6}$
- $c = 4.143$
- $d = -0.174$

4. 실험 및 결과 고찰

4.1 전체 시스템 구성

전 과정에 대한 블록선도는 그림 2와 같다.

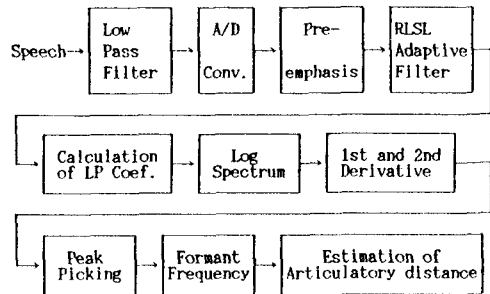


그림 2 음성 특징 추출 과정 블록선도

Pre-emphasis 과정 ($1-\mu z^{-1}$)에서 μ 를 0.95로 고정시켰다. 선형 예측 필터의 차수 P 는 15차로 선정하였다. RLSL 선형 예측 필터내의 forgetting factor λ 는 실험적인 결과로 $\lambda = 0.985$ 일 때 여러가지 실험 목적에 잘 부합되어 이를 선택하였다.

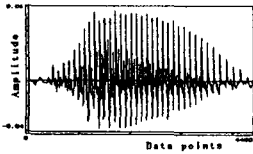
4.2 스펙트럼 추정을 위한 순간 검출법

블록 처리 알고리즘에서 그 윈도우를 대표하는 선형 예측 계수를 얻는 방법과는 달리 적응 알고리즘에서는 입력이 들어오는 때 순간마다 계수들이 보정되므로 어느 순간에서나 적절한 스펙트럼을 구할 수 있다. 그러나 본 연구에서는 가급적 최소의 오차를 갖는 스펙트럼을 검출하기 위한 방법을 제시한다.

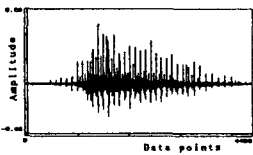
그림 3은 이중 모음 /야/에 대한 예측 오차 및 반사 계수의 합을 나타낸다. 그림 (b)에서 예측 오차는 음성 신호와 같은 주기로 큰 값을 갖게 된다. 그림 (c)는 입력 신호에 대한 각 lattice에서의 전향 반사 계수(forward reflection coefficients)들의 합을 나타낸 것으로서 아래 식에 의해 구해진 값이다.

$$\text{Sum} = \sum_{l=0}^P [\gamma(l) + 1] \quad (7)$$

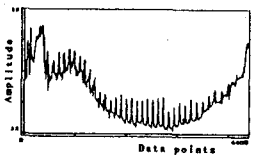
여기서, $\gamma(n)$ 은 전향 반사 계수이며, P는 선형 예측 필터의 차수이다.



(a) 이중 모음 /야/의 음성 신호



(b) 필터 마지막 단계에서의 예측 오차



(c) 전향 반사 계수의 합

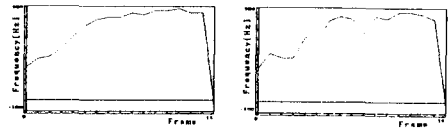
그림 3 이중모음 /야/에 대한 예측오차 및 반사계수의 합

식 (7)에서 $-1 < \gamma < 1$ 이므로 γ 의 부호에 의한 Sum의 감소를 막기 위해 1을 더한다. 그림 (c)에서 볼 수 있듯이 첨두치는 그림 (b)의 선형 예측 오차가 크게 되는 순간과 일치하는 것을 알 수 있다. 따라서 첨두치들을 피해서 계수를 구하면 정확한 선형 예측 계수들을 얻을 수 있다. 이 방법을 이용하면 윈도우에 의해 일정한 시간간격으로 스펙트럼을 구할 때 발생할 수 있는 부정확성을 해결할 수 있다.

실험에 의하면 첨두치가 있는 부분으로부터 대략 50 데이타 시퀀스(5[msec])정도 더 진행된 순간에서 안정된 평균값에 이르는 것을 볼 수 있었다.

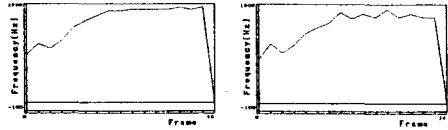
그림 4는 이중 모음 /와/에 대하여 본 연구에서 제안한 방법에 의한 형성음 주파수 추출 성능을 나타낸 것이다.

그림 4에서 -1, -2, -3은 각각 1, 2, 3차 형성음 주파수를 의미한다. 그림 (a)는 안정된 값들을 이루는 반면, 그림 (b)의 경우 불안정한 변화가 있는 곳이 나타난다.



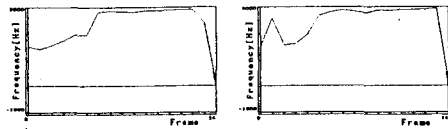
(a-1)

(b-1)



(a-2)

(b-2)



(a-3)

(b-3)

(a) 제안된 방법으로 구해진 /와/의 형성음 주파수

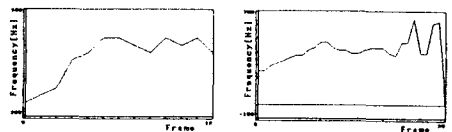
(b) 첨두치가 포함될 경우에 /와/의 형성음 주파수

그림 4 제안된 방법의 형성음 주파수 추출성능 비교

4.3 적응 알고리즘과 블록처리 알고리즘의 결과 비교

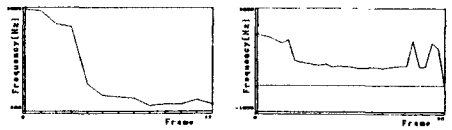
RLSL 적응 알고리즘은 매 순간마다 통계적 변화가 나타나는 음성 신호에 대하여 빠르게 적응하며, 수렴 속도가 낮은 여러 적응 알고리즘들의 단점이 극복되어 정확한 결과를 얻을 수 있다.

비교하려는 블록화 처리 방법인 Levinson 알고리즘의 선형 예측 필터의 차수는 15로 하였고, 입력 신호를 25.6 [msec] 길이의 Hamming 윈도우를 사용하여 50%중첩처리하였다. 그림 5는 이중 모음 /요/에 대한 결과 비교를 나타낸다.



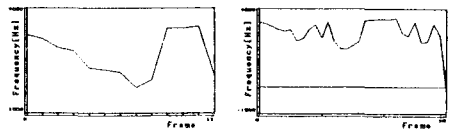
(a-1)

(b-1)



(a-2)

(b-2)



(a-3)

(b-3)

(a) 적응 알고리즘에 의한 형성음 주파수

(b) Levinson 알고리즘에 의한 형성음 주파수

그림 5 이중 모음 /요/의 형성음 주파수

형성음 주파수의 천이 과정에서 나타난 것과 같이 RLSL 적용 알고리즘으로 추출된 경우, 이중 모음 사이의 천이 부분에서도 두 모음의 특징을 정확히 반영하고 있으나, Levinson 알고리즘을 이용한 경우는 불안정한 변화들이 나타남을 알 수 있다.

4.4 조음운동 추정 결과

1, 2, 3차 형성음 주파수를 이용하여 입술의 조음 운동 추정에 적용하였다. 그림 6은 조음운동의 비교를 나타낸 것이다. 그림에서 세로축에 표시된 운동 거리의 범위는 그래픽을 위하여 적절하게 조정된 상대수치이다. 그리고 가로축은 형성음 주파수가 추출된 프레임을 표시하고 있다. 왼쪽의 그림은 RLSL 적용 알고리즘으로부터 구해진 결과이고, 오른쪽의 그림은 Levinson 알고리즘의 결과이다.

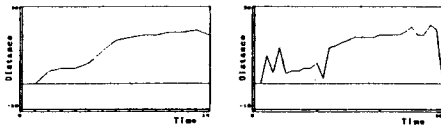


그림 6 이중 모음 /유/에 대한 조음 거리

그림에서 본 연구에서의 RLSL 적용 알고리즘을 이용한 결과가 Levinson 알고리즘의 결과보다 더 안정된 조음 운동을 나타냄을 알 수 있다. 이는 급격한 변화가 일어나지 않는 인간의 발음기관을 고려해 볼 때 정확한 결과라고 볼 수 있다.

5. 결론

본 연구에서는 RLSL 적용 선형 예측 필터를 구성하고 한국어 이중 모음에 대하여 천이 부분에서 정확한 형성음 주파수를 추출하고자 하였다. 또한 이를 이용하여 조음 운동 추정에 적용하였다.

얻어진 결과는 다음과 같다. 첫째, RLSL 적용 알고리즘에 의해 추출된 형성음 주파수가 이중 모음의 천이 부분에서 Levinson 알고리즘에 의한 결과보다 안정되고, 정확하게 변화한다. 둘째, lattice로 구성된 RLS 알고리즘이 일반적인 알고리즘보다 빠르고 정확한 결과를 얻을 수 있었다. 셋째, 반사 계수의 합을 이용하여 선형 예측 오차가 작은 순간을 검출하는 방법을 제안하여 정확한 스펙트럼을 구할 수 있었다. 넷째, 1, 2차 미분법을 사용하여 두 형성음 주파수가 겹쳐진 부분에서도 정확한 추출을 하였다. 다섯째, 형성음 주파수를 이용하여 입술의 조음운동을 추정하여 본 결과, 이중 모음의 천이 부분에서 RLSL 적용 알고리즘에 의한 조음운동은 안정되고, 연속적인 변화를 하였으나, Levinson 알고리즘의 경우는 급격한 변화가 일어나는 곳이 있었다.

참고 문헌

[1] S. Saito and K. Nakata : Fundamentals of Speech Signal Processing, Academic Press, 1985.
 [2] 송 영수, 안 상필, 이 재혁, 윤 태성, 박 상희, "LPC spectrum의 1, 2차 미분을 이용한 Formant 추출," 신 호처리 합동학술대회, 1990.9.
 [3] 김동준, 이재혁, 윤태성, 박상희, "조음도를 이용한 발음 혼련 기기의 개발," 대한전기학회논문집, Vol.41, No.2, pp.209~216, 1992.2.

[4] S. L. Marple : Digital Spectral Analysis, Prentice-Hall, 1987.
 [5] L. J. Griffiths : "Rapid measurement of digital instantaneous frequency", IEEE Trans. ASSP, Vol. ASSP-23, No.2, pp.207~221, April 1975.
 [6] K. S. Nathan, Y. T. Lee and H. F. Silverman : "A time varying analysis method for rapid transitions in speech", IEEE Trans. Signal Processing, Vol.39, No.4, pp.815~824, April 1991.
 [7] P. Ladefoged, R. Harshman and L. Rice : "Generating vocal tract shapes from formant frequencies", J. Acoust. Soc. Am., Vol.64(4), Oct. 1978.
 [8] S. Haykin : Adaptive Filter Theory, 2nd ed., Prentice-Hall, 1991.
 [9] J. Orfanidis : Optimum Signal Processing, McGraw Hill, 1988.
 [10] B. Widrow and S. D. Stearns : Adaptive Signal Processing, Prentice-Hall, 1985.