

음성천이구간에서의 성도 파라메타 시변추정에 관한 연구

Time-varying Estimation of Vocal Track Parameters During the Speech Transition Regions

최 홍 섭*
(Hong-Sub Choi*)

요 약

음성의 천이구간에서의 특징 파라메타를 찾아내기 위하여 본 논문에서는 AR모형을 사용하여 적응적으로 성문폐쇄구간을 찾은 후, 이를 제외한 구간에서 성도 파라메타를 추정함으로써 음원의 피치바이어스 영향을 제거하는 SSRLS (Sample Selective RLS) 방법을 제안한다. 성능을 비교하기 위하여 합성음과 실제음에 대하여 포먼트 추정실험을 했으며, 실험결과 제안된 방법이 WRLS 보다 우수함을 알 수 있었다.

ABSTRACT

In this paper, sample selective RLS(SSRLS) method is proposed, which aims to eliminate the influence of pitch bias. Its basic concepts are as follows. First it extracts the open glottis interval by using the residual signals, then estimates the formant values from the selected speech samples excluding above open glottis interval. This method has some analogy with the SSLPC in the basic concept, but processes the speech data in recursive sample by sample way. To assess the performance of the SSRLS, the simulation is conducted upon the synthetic and real speech. From these results, we find more usefulness of the proposed method than the conventional ones.

Key words: recursive estimation, pitch bias, sample selective RLS estimator

I. 서 론

음성신호처리에서 성도를 파라메타로 모델링하여 이것으로부터 포먼트(formant)와 그것들의 대역폭(bandwidth)을 추정하는 것은 음성인식과 합성에서 매우 중요하다. 성도파라메타 모델링의 대표적인 것이 선형예측방법(LPC)이다. 그러나 선형예측방법은 구간단위(frame-based) 처리방법이므로 음성의 빠른 천이구간에 적용하는 데는 적절하지 못하다. 이러한 문제점은 보통 순차적인(recursive or sequential) 추정기법을 통하여 해결하고 있는데 크게 WRLS(weighted recursive least square)와 Kalman filter를 이용한 것으로 구분된다.

그러나 구간분석이나 순차추정의 어떤 방법을 이용하더라도, 음성신호의 분석에서 우선적으로 고려해야하는 사항은 음원의 임펄스열에 의해서 성도파라메타 추정오차를 발생하는 피치 바이어스 문제다. 이에 대한 해결 방안으로 구간분석에서는 성문(glottis)의 폐쇄구간에 해당

하는 음성데이터에 대해서만 선형예측을 하는 방법과 음성음의 음원신호를 ϵ -혼합정규분포(ϵ -contaminated Gaussian mixture distribution)로 가정하고 분석처리하는 강인 LPC 방법등이 있다. 순차방식으로 접근한 방법들의 경우에도 잔차신호(residual signal)가 상대적으로 크지 않은 표본만을 가지고 처리하기 때문에 물리적 개념에 있어서 구간분석방법인 강인 LPC나 다중펄스(multi-pulse) LPC와 비슷하다. 따라서 이들은 성문폐쇄구간 개념을 이용하지 않으므로, 엄밀히 말하면 음원신호의 물리적 특성을 부적절하게 이용하는 방법인 것이다.

따라서 본 논문에서는 성도를 AR(auto-regressive) 모델로 표시한 후 적응적으로 성문폐쇄구간을 찾아서, 순차적으로 성도파라메타를 추정하는 방법을 제안한다. 이 방법은 LS(least square) 개념이 기본이고 성문폐쇄구간을 선택하여 처리하기 때문에 표본선택 순차최소차승(SSRLS: sample-selective recursive least square)이라 명명된다. 이 방법은 성도파라메타를 시변으로 표시하여 음성의 천이구간에 대해서도 특징파라메타를 적절히 찾아낼 수 있으며, 또한 성문폐쇄구간을 추정하여 분리처리함으로써 피치 바이어스를 줄일 수 있다.

제안된 방법의 우수성을 보이기 위해서 음원신호를 필

*대전대학교 전자공학과
접수일자: 1997년 1월 8일

스열이나 백색잡음으로 가정된 WRLS 방법과 보의실험을 통해 비교한다. 여기서는 피치 바이어스의 영향을 증명하기 위해 두 방법의 포만트추정 결과를 합성음과 실제음에 대해서 실험해 본다. 실험결과는 제안된 방법이 WRLS 방법보다 우수하다는 것을 보여준다.

II. 표본선택 순차최소자승법

우리는 입술의 방사특성을 고려하여 사전강화(pre-emphasis) 처리한 음성신호가 전극모델에 의하여 다음과 같이 생성된다고 가정한다.

$$s(n) = - \sum_{i=1}^p a_i(n)s(n-i) + u(n), \tag{1}$$

여기서 $s(n)$ 은 n 번째 시간의 음성신호의 표본을, $u(n)$ 은 음원신호를, p 는 전극모델의 차수를, $a_i(n)$ 은 시변성을 반영한 전극모델 파라메타를 각각 나타낸다. 만약 음원신호가 임펄스열이나 백색잡음이라면 기존의 WRLS 방법이 좋은 성능을 보인다. 그러나 실제의 음원신호는 무성음의 경우에는 백색잡음으로 여겨질 수 있지만 유성음의 경우에는 피치를 주기로 갖는 임펄스열로만 보기가 어렵다. 실제로 유성음의 음원신호는 성문이 열린 구간에서는 그것이 존재하고 성문이 닫힌 구간에서는 불규칙잡음으로 간주될 수 있다. 따라서 WRLS 방법은 유성음 구간 처리시에 추정된 파라메타가 성문이 열렸을 때의 음원신호에 의한 영향인 피치 바이어스를 겪게 된다.

1) 피치바이어스 제거를 위한 표본선택 순차최소자승법의 개념

WRLS가 피치 바이어스를 겪게 되는 것은 최소자승법을 식(1)에 적용할 때, 유성음에서 성문이 열린 구간을 포함했기 때문이다. 그러므로 성문이 열린 구간을 판정하여, 이것을 최소자승법에 반영한다면 이 영향을 줄일 수 있다. 이것은 구간처리 방식에 있어서는 표본선택 최소자승법인 표본선택 선형예측(SSLPC)방법[7]과 같은 개념이다. 식(1)에서 음성신호의 예측치를

$$\hat{s}(n) = - \sum_{i=1}^p \hat{a}_i(n)s(n-i) \tag{2}$$

라 한다. 식(2)에서 $\hat{a}_i(n)$ 은 $a_i(n)$ 의 추정치를 나타낸다. 따라서 잔차신호(residual signal)는

$$\begin{aligned} r(n) &= s(n) - \hat{s}(n) \\ &= \sum_{i=1}^p \hat{a}_i(n)s(n-i) \end{aligned} \tag{3}$$

이고, 여기서 $\hat{a}_0(n) = 1$ 이다. 우리는 잔차신호의 누적자승오차를 구할 때, 피치 바이어스를 줄이기 위해 성문이 열린 구간을 제외한 후 그것을 구한다. 편의상 시간구간을

$(1, n)$ 으로 생각하고 $a_i(n)$ 의 순차해(recursive solution)를 구하기 위해 이 구간에서는 성도파라메타가 $a_i(n)$ 값으로서 일정하다고 본다. 이때 누적자승오차 E_n 은

$$E_n = \sum_{k=1}^n \lambda^{n-k} \Pi(k) r^2(k) \tag{4}$$

이 된다. 식(4)에서 λ 는 소멸인자(forggetting factor) 이고

$$\Pi(k) = \begin{cases} 0, & k = \text{성문 열린 시간} \\ 1, & k = \text{그외 시간} \end{cases} \tag{5}$$

이다. 식(4)를 행렬 형태로 고쳐 쓰면,

$$E_n = \{Z(n) + H(n)\hat{\Theta}(n)\}^T G(n) \{Z(n) + H(n)\hat{\Theta}(n)\} \tag{6}$$

이 되고, 여기서 위 첨자 T 는 행렬의 전치(transpose)를 나타내고,

$$Z(n) = [s(n) s(n-1) \dots s(p+1)]^T, \tag{7a}$$

$$H(n) = [\phi(n) \phi(n-1) \dots \phi(p+1)]^T, \tag{7b}$$

$$\phi(n) = [s(n-1) s(n-2) \dots s(n-p)]^T, \tag{7c}$$

$$\hat{\Theta}(n) = [\hat{a}_1(n) \hat{a}_2(n) \dots \hat{a}_p(n)]^T, \tag{7d}$$

그리고

$$G(n) = \text{diag}\{\Pi(n) \lambda \Pi(n-1) \dots \lambda^{n-p} \Pi(p+1)\} \tag{7e}$$

이다. 식(4)와 식(6)을 보면 이것들은 보통의 누적자승오차에 가중치 $\Pi(k)$ 와 소멸인자 λ 를 반영한 것이다. 따라서 E_n 을 최소화하는 $\hat{\Theta}(n)$ 과 이것의 시간변화에 따른 순차해를 구하면 [11, pp.56-58][12],

$$\hat{\Theta}(n) = \hat{\Theta}(n-1) - \Pi(n) P(n) \phi(n) v(n), \tag{8}$$

여기서

$$P(n) = \frac{1}{\lambda} \left[P(n-1) - \frac{\Pi(n) P(n-1) \phi(n) \phi^T(n) P(n-1)}{\lambda + \Pi(n) \phi^T(n) P(n-1) \phi(n)} \right], \tag{9}$$

$$v(n) = s(n) + \phi^T(n) \hat{\Theta}(n-1) \tag{10}$$

이고, 식(10)은 예측오차(prediction error)를 나타낸다. 식(8)과 식(9)는 $\Pi(n) = 1$ 일 때는 소멸인자를 가진 WRLS방법이 되고, $\Pi(n) = 0$ 일 때는 식(8)은 $\hat{\Theta}(n) = \hat{\Theta}(n-1)$ 이 된다. 이것을 다시 말하면, 성문이 열리지 않은 시간에는 WRLS 방법으로 $\hat{\Theta}(n)$ 이 구해지나 성문이 열린 시간에는 $\hat{\Theta}(n)$ 이 변화하지 않는다는 것을 말해준다. 따라서 우리가 $\Pi(n)$ 즉, 성문의 개폐시간을 추정할 수 있다면 식(8)-식(10)에 의해 $\hat{\Theta}(n)$ 을 구할 수 있게 된다.

2) 예측오차를 이용한 성문개폐시간 추정

성문개폐시간의 추정은 식(3)의 잔차신호와 식(10)의 예측오차를 이용하여 이루어진다. 식(1)에서 $s(n)$ 은 입술 방사특성을 원래 음성신호로부터 제거하기 위해 사전강화(pre-emphasis)된 것이다. 이런 $s(n)$ 을 가지고 식(8)-식(10)에 의해 일단 $\hat{\Theta}(n-1)$ 이 제대로 찾아졌다고 가정한다. 만약 n 시간에서 성문 열린 구간이 시작된다면 n 시간에서의 예측 오차의 절대치는 비교적 커질 것이다. 따라서 이때의 예측오차의 절대치는 $n-1$ 시간까지 나온 잔차신호들의 표준편차보다는 커진다. 잔차신호의 표준편차는 구간표준편차(interval standard deviation)로 추정한다. 이것은

$$Q(n) = \frac{1}{L} \sqrt{\sum_{i=1}^L r^2(n-i)} \quad (11)$$

로서 구하고, 여기서 L 은 피치주기보다 약간 크게 선택되는 상수이다. 그러면 적당한 양의 상수 c 를 사용하여 다음 조건식 $cQ(n) \leq |v(n)|$ 을 만족하는지를 판단하면, n 시간이 성문열린구간에 있는지를 판정할 수 있다. 보통 천이구간에서 예측오차는 성문이 닫힌 시간에서도 커지는 경향이 있다. 만약 식(11) 대신에 전체구간에서 구한 표준편차를 사용하여 성문개폐시간을 추정한다면, 천이구간에서 성문이 닫힌 시간을 열린 시간으로 잘못 추정하는 경우가 많아진다. 따라서 식(11)을 사용하면 적응적으로 추정할 수 있게 된다.

한편 식(1)에 WRLS를 적용하면, 성문이 열린 구간의 끝 시간에서 예측오차가 피크(peak)를 띠게 된다고 알려져 있다[11]. 또 성문이 한번 열리면 어느 정도 지속되는데, 여기서는 이 지속시간이 T_0 로 일정하다고 가정한다. 이런 가정은 물리적으로 어느 정도 타당성이 있으며, 좋은 결과를 보여줬다[7]. 위의 사항들을 종합하면 성문개폐시간 추정방법은 표 1과 같이 얻을 수 있다.

표 1. 성문개폐시간 추정방법

- (1) $v(n)$ 과 $Q(n)$ 을 WRLS방법으로 계산한다.
- (2) $cQ(n) > |v(n)|$ 이면 $\Pi(n) = 1$. 그리고 $n = n + 1$ 로 놓고 순서 (1)을 반복하고, 그렇지 않으면 순서 (3)로 간다.
- (3) WRLS방법으로 $v(k)$ 을, $n + 1 \leq k \leq T_0 + n - 1$, 구한다. k 를 $T_0 + n - 1$ 부터 감소 시키면서 $|v(k)| \geq |v(n)|$ 을 만족하는 최초의 k 를 찾고 이것을 m 이라 놓는다.
- (4) $\Pi(k) = 0$, 성문(glottis) 열린구간, $m - T_0 + 1 \leq k \leq m$ 로 정한다.
- (5) $n = m + 1$ 로 하고 순서 (1)로 간다.

3) 성도파라메타와 성문개폐시간의 동시추정 알고리즘
순차방식으로 성도파라메타를 찾으려면 식(8)-식(10)과 표 1을 결합하여야 한다. 표 1을 보면, 현재 시간을 n 이라 했을 때 미래의 음성신호 ($T_0 - 1$)개가 필요하다는 것을 알 수 있다. 이것은 $\Pi(n)$ 과 $\hat{\Theta}(n)$ 이 ($T_0 - 1$)만큼 지연된다는 것을 말해준다. 또 표 1에서 $m = n$ 이면, $n - T_0 + 1 \leq k$

$\leq n$ 구간에서 $\Pi(n) = 0$ 이 됨을 알 수 있다. 따라서 이때는 식(8)에 의해 $\hat{\Theta}(k) = \hat{\Theta}(n - T_0)$ 가, $n - T_0 + 1 \leq k \leq n$, 된다. 한편, $n = 1, 2, 3, \dots$ 에서 식(8)-식(10)을 적용할 때, 초기값으로서

$$P(0) = b^2 I \quad (12a)$$

$$\hat{\Theta}(0) = 0 \quad (12b)$$

을 사용하는데[13], 여기서 I 는 단위행렬(identity matrix), b 는 매우 큰 상수이다. 또 식(11)을 보면 구간표준편차를 구할 때 잔차신호가 필요한데 이들의 초기값으로는

$$r(k) = d \{ rand(k) - 0.5 \}, \quad 1 - L \leq k \leq 0 \quad (13)$$

로 한다. 식(13)에서 $rand(\cdot)$ 는 0과 1 사이의 값을 불규칙하게 발생하는 함수이고 d 는 매우 큰 상수이다. 초기의 $1 \leq n \leq L$ 에서는 WRLS를 그대로 따름으로서 제대로 된 구간표준편차가 나오도록 한다. 위의 사항을 종합하면, 표 2와 같이 제안된 방법을 얻는다.

표 2. 표본선택 순차최소자승법의 알고리즘

- (1) $P(0) = b^2 I$, $\hat{\Theta}(0) = 0$,
 $r(k) = d \{ rand(k) - 0.5 \}$, $1 - L \leq k \leq 0$,
 $n = 1$
- (2) 구간표준편차 $Q(n)$ 을 먼저 계산한다. WRLS방법으로 $v(n)$, $\hat{\Theta}(n)$, $P(n)$ 을 계산.
 $n \leq L$ 이면 $n = n + 1$ 로 놓고 순서 (2) 반복
- (3) $cQ(n) > |v(n)|$ 이면 $\Pi(n) = 1$, 성문(glottis) 닫힌구간,
그리고 $n = n + 1$ 로 놓고 순서 (2)로 간다.
- (4) 표 1의 순서 (3) 수행
- (5) 표 1의 순서 (4) 수행
 $\hat{\Theta}(k) = \hat{\Theta}(m - T_0)$,
 $P(k) = \lambda^{-k - (m - T_0)} P(m - T_0)$, $m - T_0 + 1 \leq k \leq m$
- (6) $n = m + 1$, 순서 (2)로 간다.

III. 실험결과

제안된 방법에는 사전에 선택해야 할 상수들이 있다. 이 상수들의 값을 정하기 위해 먼저 합성음에 대한 실험 결과로부터 최적의 상수값을 결정한다. 이 실험과정에서 제안된 방법이 WRLS 방법보다 우수하다는 것이 나타난다. 또 실제음에 대한 실험에서도 제안된 방법이 우수함을 알 수 있다.

합성음으로 한국어의 대표 모음인 '아 어 오 우 으 이'를 사용했다. 표 3은 이들의 포먼트와 포먼트 대역폭을 나타낸다.

합성음은 직렬 포먼트 합성방식으로 만들었으며 이때 사용한 성문파로는 성문파의 성질을 최대한 갖기 위하여 임펄스열이 아닌 LF 모델[15]을 사용했다. 이 모델에 의

표 3. 합성음의 포먼트(f_1 - f_5)와 대역폭(b_1 - b_5) (단위: Hz)

모음	f_1/b_1	f_2/b_2	f_3/b_3	f_4/b_4	f_5/b_5
아	783/67	1165/68	2733/84	3557/107	3889/87
어	579/59	816/57	2756/45	3587/88	3945/100
오	423/19	663/27	2777/58	3451/105	3843/111
우	303/42	676/88	2405/358	3389/49	4550/287
으	353/14	1037/52	2532/143	3431/57	3700/74
이	265/21	2347/45	3110/109	3493/65	4137/146

한 미분된 성분파가 그림 1에 보이며, 이때 $t_c=16$ (samples)로 고정시켰으며, 피치는 표준화주파수 10kHz에 대해서 200Hz로 했다.

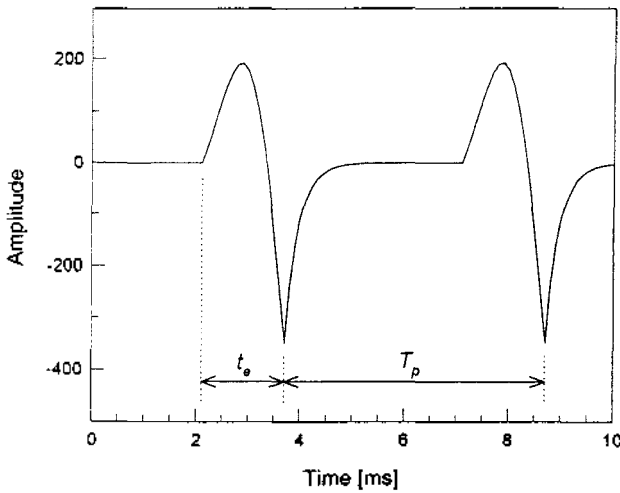


그림 1. LF모델에 의한 미분된 성분파의 파형

표 4는 표 3에 의해서 만들어진 합성음에 대해 상수(c , T_0)의 여러 가지 값들을 가지고 제안된 방법을 적용했을 때의 포먼트 추정오차를 나타낸다. 포먼트 추정오차는 식(14)로 계산한다.

$$E_f = \sum_{i=1}^6 \sum_{j=1}^5 |f_{ij} - \hat{f}_{ij}| / f_{ij} \quad (14)$$

표 4. 상수 (c , T_0)에 대한 합성음의 포먼트추정오차 (단위: %)
($p=12$ 그리고 $\lambda=0.9922$)

T_0	c	1.0	1.5	2.0	3.0	7.0	7.5
1	28.69	11.71	5.36	2.82	4.58	4.58	4.58
4	2.40	3.81	3.80	3.80	4.58	4.58	4.58
7	1.64	2.25	2.25	2.25	4.58	4.58	4.58
10	1.09	2.10	2.10	2.10	4.58	4.58	4.58
13	0.98	1.65	1.65	1.65	4.58	4.58	4.58
16	0.59	0.59	0.59	0.59	4.57	4.58	4.58
19	0.63	0.63	0.63	0.63	4.57	4.58	4.58
22	0.90	0.91	0.91	0.91	4.58	4.58	4.58
25	1.49	1.53	1.54	1.54	4.58	4.58	4.58
28	2.06	2.11	2.15	2.15	4.58	4.58	4.58
31	2.40	2.52	2.59	2.59	4.58	4.58	4.58

여기서 f_{ij} 는 i 번째 모음의 j 번째 포먼트를 나타낸다. 표 4에서 $c \geq 7.5$ 일 때가 WRLS와 동일한 결과이다.

위의 실험결과에서 보면 제안된 방법이 $1.0 \leq c \leq 3.0$ 와 $13 \leq T_0 \leq 25$ 와 같은 상수범위에서 WRLS 방법보다 월등히 좋은 결과를 보여주고 있다. 표 5, 표 6, 그리고 표 7는 각각 차수, 피치, 그리고 소멸인자 변화에 따른 포먼트 추정오차를 나타낸다. 마찬가지로 제안된 방법이 WRLS 방법보다 우수하다는 것을 알 수 있다. 일반적인 순차최소자승(RLS)방법에 의한 추정기의 유효창길이(N_{eff})와 소멸인자(λ) 사이의 관계는 식(15)와 같이 표현된다[2].

$$N_{eff} = 1/(1 - \lambda) \quad (15)$$

표 5. 차수에 대한 합성음의 포먼트추정오차 (단위: %)

($c=2.0$, $T_0=16$, 그리고 $\lambda=0.9922$)

차수	10	12	14	16	18	20
WRLS	29.3	26.4	25.0	25.2	22.2	22.0
SSRLS	13.8	3.2	3.5	3.7	3.5	3.1

표 6. 피치에 대한 합성음의 포먼트추정오차 (단위: %)

($c=2.0$, $T_0=16$, $p=12$, 그리고 $\lambda=0.9922$)

피치(Hz)	100	125	150	175	200	225	250	275	300	325	350
WRLS	10.3	12.3	13.0	19.1	26.4	22.7	25.0	24.1	21.2	29.5	29.2
SSRLS	1.7	1.2	1.0	1.6	3.2	3.2	2.8	2.9	4.0	8.0	11.8

표 7. 유효창길이에 대한 합성음의 포먼트추정오차 (단위: %)

($c=2.0$, $T_0=16$, 그리고 $p=12$)

창길이 (samples)	30	60	90	120	150	180	210	240	270
WRLS	27.6	27.1	26.8	26.5	26.0	25.7	25.5	25.3	25.0
SSRLS	0.0	0.2	1.4	2.8	4.0	4.9	5.5	6.1	6.5

제안된 방법을 천이구간에 대하여 실험하기 위해 두개의 모음을 연결하였다. 두 모음을 연결할 때 천이구간에서 포먼트들은 한 피치구간에서는 일정하다고 가정하였다. 따라서 앞 모음의 j 번째 포먼트가 뒤 모음의 j 번째 포먼트로 천이할 때, 그것은 피치주기를 똑바로 갖는 계단 모양으로 천이하게 된다. 이와같은 합성음의 포먼트 구조는 그림 2에 보인다.

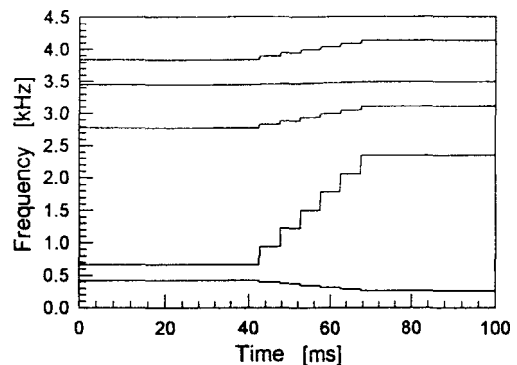


그림 2. 합성음의 포먼트 구조

그림 3은 이렇게 합성된 /오이/음의 천이구간이 포함된 일부분을 나타낸다. 그림 4는 제안된 방법과 WRLS 방법에 의하여 찾은 포만트궤적을 보여준다. 포만트는 추정기의 차수가 12차의 경우, 최고로 6개의 포만트가 생기는데 이를 대역폭의 크기가 작은 순서대로 선택한다. 그림 4에서 볼 수 있듯이 제안된 방법(SSRLS)에 의하여 찾아진 포만트가 WRLS에 의한 것들보다 합성시에 사용된 포만트에 보다 많은 부분에서 일치하고 있음을 알 수 있다. 즉, 피치주기 동안 일정하게 유지하는 합성음의 포만트 구조를 WRLS 보다 충실히 추적하고 있음을 그림에서 보여주고 있다.

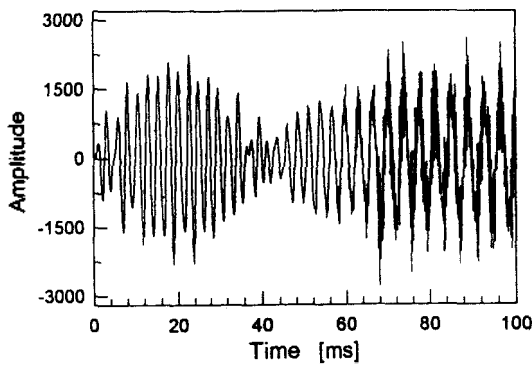
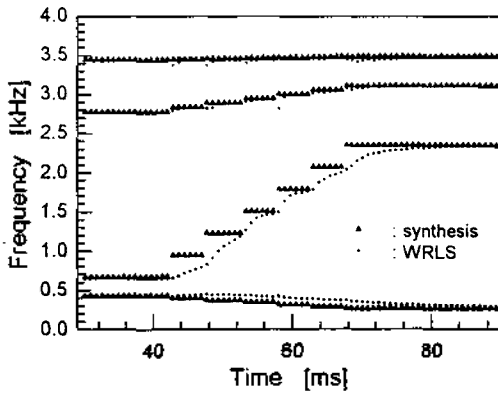
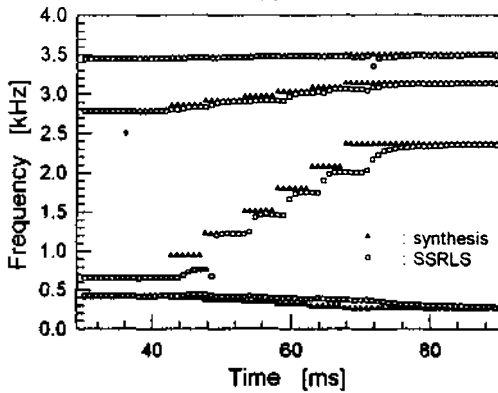


그림 3. 합성음/오이/의 천이구간에서의 파형



(a)

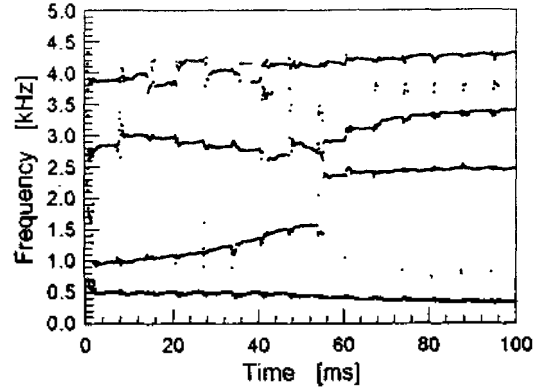


(b)

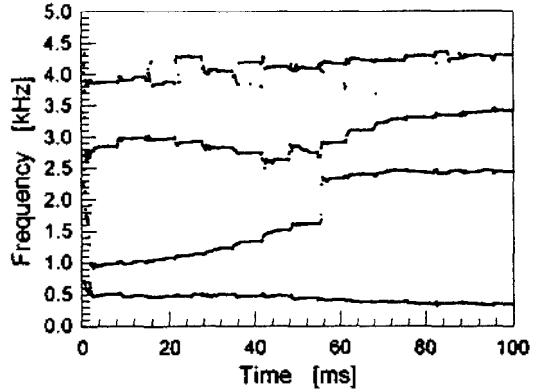
그림 4. 합성음에서 추정된 포만트궤적의 비교

(a)WRLS의 포만트궤적 (b)SSRLS의 포만트궤적

실제음으로 20대 남성화자가 발음한 /오이/음의 천이구간의 일부분을 사용하였다. 표본화주파수는 10kHz이고, 12bit A/D 변환기를 사용하여 음성데이터를 얻었다. 그림 5(a), (b)는 제안된 방법과 WRLS방법으로 찾은 포만트의 궤적을 보여준다.



(a)



(b)

그림 5. 실제음에 대한 SSRLS와 WRLS의 성능비교

(a)WRLS의 포만트궤적 (b)SSRLS의 포만트궤적

실험에 사용한 소멸인자의 상수값은 $\lambda = 0.9667$, 즉 유효장길이는 30샘플로 정하므로써 피치의 추정기에 대한 영향을 증대시킨 후, 성능비교를 하였다. 실험에 이용한 다른 상수로는 $c = 2.0$, $T_0 = 16$, 그리고 $p = 12$ 이다. 여기서 제안된 방법의 포만트가 WRLS방법의 포만트보다 정상구간과 천이구간에서 더욱 평탄하고, 깨끗하게 보인다. WRLS로 구한 포만트궤적은 피치의 영향으로 특히 3kHz와 4kHz 사이의 3차 및 4차 포만트와 1kHz 부근에서 포만트의 값이 주기적으로 줄어드는 것을 볼 수 있다. 따라서 WRLS 방법에 의해서 구한 포만트궤적은 피치의 영향을 많이 받고 있지만, 제안된 방법에 의한 포만트궤적은 피치의 영향이 많이 제거된 것을 확인할 수 있다. 위와 같은 현상은 여러 다른 실제음에 대해서도 비슷하게 나타나고 있어서 제안된 표본선택 순차최소자승방법이 기존의 순차추정방법보다 우수한 성능을 갖고 있음을 알 수 있다.

IV. 결 론


본 논문에서는 표본선택 순차최소자승법(SSRLS)을 제안하였다. 이 방법은 성도를 AR 모델로 표시하고, 음원 신호의 영향을 줄이기 위해 먼저 적응적으로 성분의 개폐시간을 추정한다. 다음 성분 열린 구간을 음성의 분석 구간에서 제외한 후, 순차적으로 성도파라미터를 추정하는 방법이다. 제안된 방법에는 사전에 선택되어야 할 여러 가지 상수가 있다. 합성음을 가지고 식(14)로 정의한 포먼트 추정오차를 척도로 하여 WRLS 방법과 비교하면서 이 상수들의 영향을 살펴보았다. 추정기의 성능에 중요한 상수로는 구간표준편차의 비례상수인 c 와 성분 열린 구간의 길이인 T_0 이다. 모의실험에서 이들을 변화시키면서 최적의 상수값을 찾아 선택하면 우리는 합성음에 대해 SSRLS가 WRLS보다 매우 우수하다는 것을 알 수 있었다. 또 선택된 최적의 상수값들을 가지고 실제음에 대해서 SSRLS를 적용해본 결과, 포먼트 궤적이 WRLS보다 대체로 매끄럽게 나왔고 추정된 포먼트들이 피치 바이어스에 의한 영향을 적게 받는 것으로 검증되었다. 따라서 제안된 방법은 음성의 순차분석에 앞으로 많이 이용되리라 기대된다.

참 고 문 헌

1. J. Makhoul, "Linear prediction: a tutorial review," *IEEE Proc.* vol. 63, pp. 561-580, June 1973.
2. Y. T. Ting, D. G. Childers, "Speech analysis using the weighted recursive least squares algorithm with a variable forgetting factor", *Proc. ICASSP*, 1990, pp. 389-392.
3. Y. S. Cho, S. B. Kim, and F. J. Power, "Time-varying spectral estimation using AR models with variable forgetting factors", *IEEE Trans. Signal Processing*, vol. SP-39, pp. 1422-1426, June, 1991.
4. J. D. Gibson, J. L. Melsa and S. K. Jones, "Digital speech analysis using sequential estimation techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 362-369, Aug. 1975.
5. G. A. Mack and V. K. Jain, "Speech parameter estimation by time-weighted error Kalman filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1300-1303, Oct. 1983.
6. S. Parthasarathy and D. W. Tufts, "Excitation-synchronous modeling of voiced speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1241-1249, Sep. 1987.
7. Y. Miyoshi, K. Yamato, R. Mizoguchi, M. Yanagida, and O. Kakusho, "Analysis of speech signals of short pitch period by a sample-selective linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1233-1240, Sep. 1987.
8. C. H. Lee, "On robust linear prediction of speech," *IEEE Trans. ASSP*, vol. 36, pp. 642-650, May 1988.

9. Y. Miyanaga, N. Miki and N. Nagai, "Adaptive identification of a time-varying ARMA speech model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 423-433, June, 1986.
10. Y. Miyanaga, N. Miki, N. Nagai and K. Hatori, "A speech analysis algorithm which eliminates influence of pitch using the model reference adaptive system," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 88-96, Feb. 1982.
11. S. Singal and B. S. Atal, "Amplitude optimization and pitch prediction in multipulse coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 317-327, March 1989.
12. L. Ljung and T. Soderstrom, *Theory and practice of recursive identification*, MIT press, 1987.
13. J. M. Mendel, *Lessons in digital estimation theory*, Prectice-Hall, Inc., 1987.
14. D. Y. Wong, J. D. Markel, and A. H. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 350-355, August 1979.
15. G. Fant, J. Liljencrants and Q. G. Lin, "A four-parameter model of glottal flow," *STL-QPSR 4/1985*, pp. 1-13, 1985.

▲ 최 홍 섭(Choi Hong-Sub) 1957년 10월 3일생



1985년 2월: 서울대학교 전자공학과 (공학사)
 1987년 2월: 서울대학교 전자공학과 (공학석사)
 1988년 3월~1993년 2월: 서울대학교 공학연구소 조교
 1994년 8월: 서울대학교 전자공학과 (공학박사)

1995년 3월~현재: 대전대학교 전자공학과 전임강사