

음성신호에 적용한 AR모델상에서의 강인 순차형 추정기

허창원⁰⁾, 이윤복¹⁾, 김외철²⁾, 하판봉³⁾
상원대학교 전자공학과¹⁾Robust Sequential Estimator Using AR Models Applied to
Speech SignalChang-Won Heo⁰⁾, Yun-Pog Lee¹⁾, Oe-Cheul Kim²⁾, Pan-Bong Ha³⁾
Dept. of Electronic Eng., Changwon National University¹⁾

요 약

본 논문에서 다루는 두 가지의 큰 내용은 블럭처리 차원에서 많이 연구되고 이용되어 오던 추정 연구를 순차처리로 확산시키는 것이고, 나머지 하나는 피치 바이어스에 대해 강인한 추정을 할과 동시에 비정상(non-stationary)신호의 특징도 간과하지않은 않는 순차형 추정기를 제안한다.

1. 서 론

성도시스템을 모델링하기 위해 음성신호의 파라미터를 분석하게 된다. 이 때 음성신호에 포함되어 있는 피치신호의 영향이 그대로 분석된 결과에 남아 있게 된다.

이러한 영향을 제거하기 위해 블럭처리방법으로 SSLP(Sample Selective Linear Prediction)[1], RBLP (Robust LP)[2] 등이 제안되었다. 그러나 정상상태가 아닌 폐쇄음, 마찰음의 시작 및 자음과 모음의 천이 부분에서의 급격한 파라미터 변동은 블럭처리방법의 맹점인 프레임 간의 파라미터 불연속성을 더 부각시킨다.

그래서 제안하는 추정기는 순차적인 처리방법으로 널리 사용되고 있는 WRLS(Weighted Recursive Least Squares)추정기를 개선해서 음성신호의 피치 바이어스 영향과 시변성을 동시에 극복하는 강인 순차형 추정기이다.

제안된 순차형 강인 추정기의 성능을 평가하기 위해 합성음에 적용해서 주어진 포먼트에 대한 추정오차를 WRLS방법과 비교해 보았고, 실제음성에서 변화가 심한 부분에 대해 각 강인 추정기의 강인성을 비교해 보았다.

2. 기본적인 WRLS

성도시스템을 모델링하는 일반적인 AR모델의 식은

$$s(n) = -\sum_{i=1}^p a_i(n)s(n-i) + GU(n) \quad (1)$$

이다[4]. 즉 시간 n 에서의 음성신호 $s(n)$ 은 이전의 p 개 sample들의 선형결합으로 모델링될 수 있다. 여기서 G 는 이득인자, $u(n)$ 은 시간 n 에서의 여기신호이다. WRLS에서 예측계수는 누적자승오차 $E(n)$ 를 최소화하는 값들로 이루어진다.

$$E(n) = \sum_{k=1}^n \lambda^{n-k} |e(k)|^2 \quad (2)$$

$$\begin{aligned} e(k) &= s(k) - \hat{s}(k) \\ &= s(k) + \sum_{i=1}^p \hat{a}_i(k) s(k-i) \\ &= s(k) + \hat{\mathbf{a}}_i^T(k) \mathbf{S}(k) \end{aligned} \quad (3)$$

여기서

$$\begin{aligned} \hat{\mathbf{a}}_i(k) &= [\hat{a}_1(k) \hat{a}_2(k) \cdots \hat{a}_p(k)]^T \\ \mathbf{S}(k) &= [s(k-1) s(k-2) \cdots s(k-p)]^T \end{aligned}$$

이다[3].

망각계수 λ 는 0과 1사이의 값으로서 RLS계산시에 시간적으로 어느 데이터에 큰 비중을 두어 계산할지를 결정한다. 이것으로써 서서히 변화하는 파라미터를 추적해 나갈 수 있다. 그래서 망각계수는 비정상적인 데이터의 파라미터 추정에 있어 매우 중요한 요소이다. 비정상적인 추정에 있어서 데이터의 유효창이 작아짐으로써 파라미터를 효율적으로 구할 수 있게 된다.

WRLS추정기에서는 covariance 매트릭스의 역행렬과 파라미터 벡터에 대해 순차적인 계산이 이루어

진다[3].

$$P(n) = \frac{1}{\lambda} [P(n-1) - K(n) S^T(n) P(n-1)] \quad (4)$$

$$a(n) = a(n-1) + K(n) e(n) \quad (5)$$

여기서

$$K(n) = \frac{P(n-1) S(n)}{\lambda + S^T(n) P(n-1) S(n)} \quad (6)$$

이다.

3. 강인 순차형 추정

음성신호에서 파라미터 값을 예측시키는 주된 요소가 여기신호이다. 이런 여기신호에 의한 영향으로 음성의 산호부분에서 주기적으로 그리고 순간적으로 큰 폭의 신호가 나타난다. 이것을 피치라 한다. 서서히 변화되거나 모음부분에서처럼 대체로 일정하던 파라미터 값이 이러한 피치 때문에 순간적으로 크게 바뀐다. 그래서 이 때 예측오차값이 커지게 되는데 이렇게 큰 예측오차가 순차적인 파라미터의 계산에서 그대로 반영되기 때문에 파라미터 값도 순간적으로 큰 폭의 변화를 보이게 된다. 그래서 강인 순차 추정에서는 이렇게 큰 예측오차에 의한 영향을 둔화시킴으로써 강인한 추정을 하게 된다.

강인 순차형 추정기는 식(2)의 누적자승오차에 가중함수 $W(n)$ 을 포함한 다음의 식을 최소화하는 파라미터를 구해낸다.

$$E_{n,h}(n) = \sum_{k=1}^n \lambda^{n-k} W(k) |e(k)|^2 \quad (7)$$

이 식의 미분에 의한 계산으로부터 식(6)은 다음과 같이 된다[7].

$$K_{n,h}(n) = \frac{W(n) P(n-1) S(n)}{\lambda + W(n) S^T(n) P(n-1) S(n)} \quad (8)$$

이 식에서 어떤 가중함수를 사용할 것인지에 대해 본 논문에서는 세 가지 가중함수의 사용예를 보였다.

그리고 가중함수가 갖는 의미를 쉽게 보이기 위해 $\Psi(e(n))$ 를 다음과 같이 정의한다.

$$\Psi(e(n)) = W(n) \cdot e(n) \quad (9)$$

강인 순차형 추정에서 예측오차의 임시적인 분포를 계산하기 위해 확장예측오차를 정의하게 된다[5].

$$Q(n) = \frac{1}{M} \sum_{i=0}^{M-1} e^2(n-i) \quad (10)$$

그리고 예측오차의 구간표준편차는 다음과 같다.

$$\sigma(n) = \sqrt{Q(n)} \quad (11)$$

i) Hard-Rejection 가중함수

가장 쉽게 생각할 수 있는 형태의 가중함수이다. 예측오차가 어떤 기준치 이상 되면 파라미터의 계산에서 제외시킨다. 모음과 같이 파라미터의 값에 변화가 완만한 음성신호의 강인추정에 매우 강력하다. 이것은 SSLP에서 이용되는 함수이다.

$$W_H(n) = \begin{cases} 1 & |e(n)| \leq \sigma(n) \cdot T_H \\ 0 & |e(n)| > \sigma(n) \cdot T_H \end{cases} \quad (12)$$

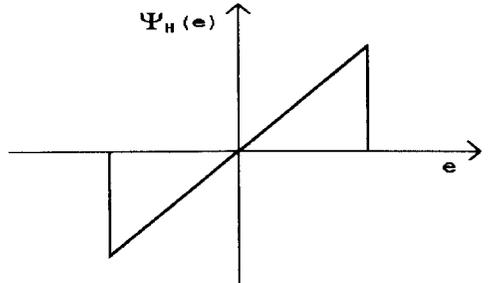


그림 1. Hard-Rejection 가중함수의 $\Psi(e(n))$ 함수

ii) Huber 가중함수

예측오차가 어떤 기준치 이상 되면 그 이상의 예측오차의 차이를 무시한다. Hard-Rejection함수에 비해 큰 예측오차의 존재를 어느 정도 인정하게 된다. 그래서 변화가 큰 음성부분에서의 효과적인 강인 추정에 적합하다.

$$W_{Hh}(n) = \begin{cases} 1 & |e(n)| \leq \sigma(n) \cdot T_{Hh} \\ \frac{\sigma(n) \cdot T_{Hh}}{|e(n)|} & |e(n)| > \sigma(n) \cdot T_{Hh} \end{cases} \quad (13)$$

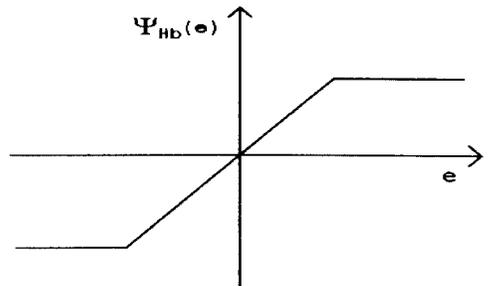


그림 2. Huber 가중함수의 $\Psi(e(n))$ 함수

iii) Cauchy 가중함수

앞의 두 가중함수의 특징을 반반씩 가진 가중함수이다. 음성의 정상상태와 비정상상태에 대해 적당한 선에서의 전실한 추정에 이용될 수 있다. 다음과 같은 가중함수를 갖는다.

$$W_c(n) = \frac{1}{1 + \left(\frac{e(n)}{\sigma(n) \cdot T_c}\right)^2} \quad (14)$$

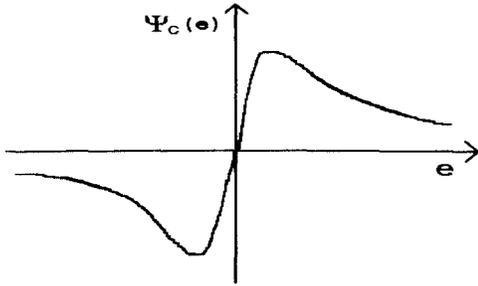


그림 3. Cauchy 가중함수의 $\Psi(e(n))$ 함수

4. 실험 및 고찰

실험은 강인추정의 충실도를 측정하기 위해 합성음에 대해서 한 번 실행했고, 강인추정의 강인성의 연관 관측치를 보기 위해 실제음성에 대해 한 번 실시했다.

첫번째 실험에 이용한 합성음은 “아이”라고 발음한 실제음성을 분석해서 그 포먼트로써 재합성한 것이다. 피치주파수는 200Hz, 성도가 열린 구간은 2 msec로 주어 5 msec단위로 합성했다. 각 추정방법을 평가하기 위한 포먼트 추정오차는 다음과 같다[6].

$$E_f = \frac{1}{M \cdot N} \sum_{f=1}^N \sum_{m=1}^M \left[\frac{|F_f(t) - \hat{F}_f(t)|}{F_f(t)} \right] \quad (15)$$

각 추정기의 $\lambda=0.98$, $p=10$ 로 했다. 강인추정기에서 식(10)의 확장예측오차를 구하기 위한 $M=25$ 로 했다. 실험 결과를 표 1과 그림 4, 5, 6, 7, 8에 나타내었다.

표 1. 각 순차 추정기의 추정오차

	WRLS	Hard-Rejection ($T_H=1.25$)	Huber ($T_{H0}=0.75$)	Cauchy ($T_C=0.75$)
예측오차 (%)	4.3	2.7	3.6	3.0

두번째 실험은 20대 남성화자가 발음한 “구”라는 음성의 자음에서 모음으로 천이되는 부분을 각각의 추정기로 파라미터를 구해 보았다. 피치라고 생각되는 부분을 확대해서 나타낸 그림이 그림 11이다. 강인성의 대소관계는

WRLS < Huber < Cauchy < Hard-Rejection
이다.

5. 결론

본 논문에서는 음성신호처리의 각종 응용분야의 기반이 되는 신호의 파라미터 분석에 대해 다루었다. 신호의 순차 처리는 Real-Time을 위해 꼭 거쳐야만 하는 관문이다. 순차 처리에서는 차례로 입력되는 하나 하나의 샘플 데이터가 다음에 입력되는 값과 연관되어 중요한 의미를 갖게 되고, 이 때문에 외란의 영향에 민감할 수 밖에 없다. 견실한 분석에 치중하게 되면 급변하는 파라미터의 변화를 놓치게 되고, 반대로 변하는 파라미터에 집중하면 강인성을 손해보게 된다. 이런 이율배반적인 상황에서 더욱더 효율적인 추정기를 구현하려면 처리하고자 하는 음성 신호 자체에 대한 예측할 수 있는 사전지식들을 충분히 활용할 필요가 있다고 본다.

참고문헌

[1] 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.
 [2] P. J. Huber, *Robust Statistics*, John Wiley & Sons, 1981.
 [3] J. R. Deller, Jr., J. G. Proakis and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.
 [4] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
 [5] Y. S. Cho, S. B. Kim and E. J. Powers, "Time-Varying Spectral Estimation Using AR Models with Variable Forgetting Factors," IEEE Trans. Signal Processing, vol. 39, pp. 1422-1426, Jun. 1991.
 [6] H. S. Choi, S. C. Bang and S. G. Ann, "A Robust Sequential Parameter Estimation for Time-Varying Speech Signal Analysis," in Proc. ICASSP-94, pp. 457-460, Apr. 1994.
 [7] L. Ljung and T. Soderstrom, *Theory and practice of recursive identification*, MIT press, 1987.

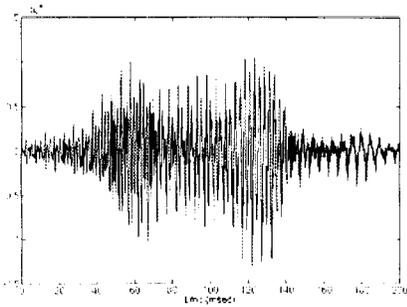


그림 4. 합성음 “아이”의 파형

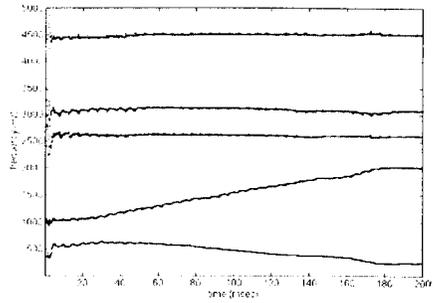


그림 8. Cauchy 추정기로 구한 포먼트 궤적

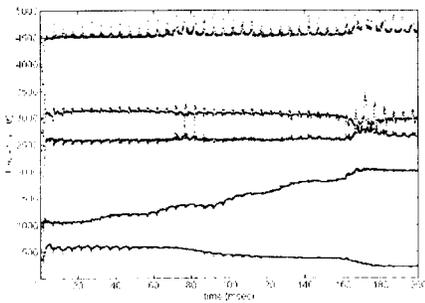


그림 5. WRLS로 구한 포먼트 궤적

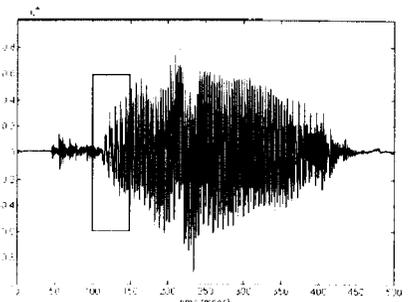


그림 9. 숫자음 “구”의 실이부분

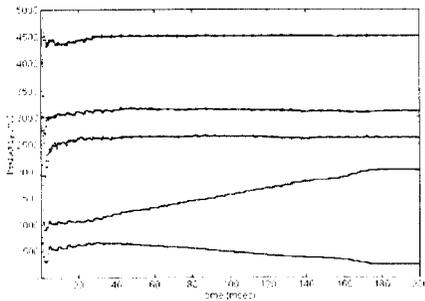


그림 6. Hard-Rejection 추정기로 구한 포먼트 궤적

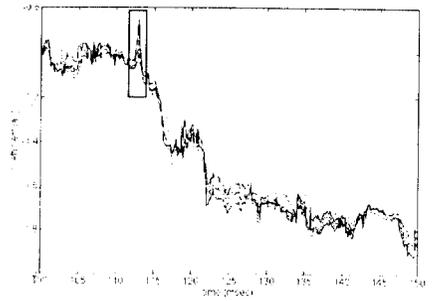


그림 10. “구”의 앞부분 신호의 파라미터(a_1)의 궤적

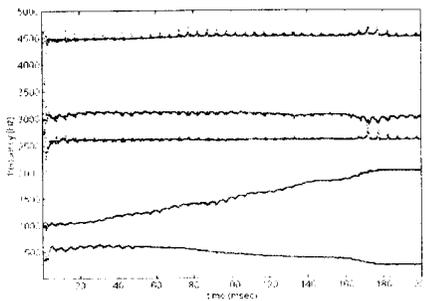


그림 7. Huber 추정기로 구한 포먼트 궤적

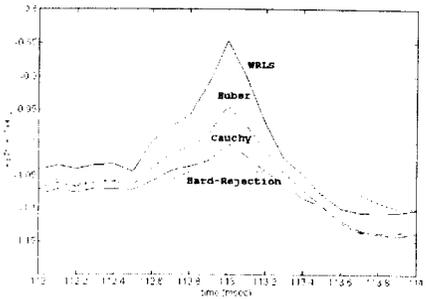


그림 11. 각 추정기의 강인성 비교(그림 8의 확대)