

견실 순차 특이치분해를 이용한 음원추정

홍성훈*, 최홍섭*, 방승찬*, 이정철*, 이용주**, 안수길*

*: 서울대학교 전자공학과 통신이론 연구실

** : 한국전자통신연구소 자동통역 연구실

Voice Source Estimation Using Robust Sequential SVD

SungHoon Hong*, HongSub Choi*, SeungChan Bang*, JungChul Lee*, YongJu Lee*, Souguil Ann*

*: Communication Theory Lab., Dept. of Electronics Engineering, Seoul National University

** : Automatic Translation Lab., Electronics Telecommunication Research Institute

X 본 논문은 1993년도 한국전자통신연구소 위탁과제의 지원으로 이루어졌습니다.

요약

본 논문에서는 변화가 심한 음원파형을 추정하는 새로운 순차처리 알고리즘을 제안한다. 먼저, 1) 기존의 순차처리 분석법중 대표적인 분석법인 RLS(recursive least square)의 문제점들을 검토하고, 2) 이를 개선하기 위해서 관측행렬(observation matrix)을 최적차수의 SVD(reduced-rank singular value decomposition)로 재구성하고, 3) 이에 견실개념(robustness concept)을 적용해서 최적의 성도변수(vocal tract parameter)를 찾아내고 역필터를 적용해서 음원(voice source)을 효과적으로 구분해낸다. 본 논문에서 제안된 방법으로 음원을 추정할 경우, 변화가 심한 음원파형을 잘 추정할 수 있으며, 음원의 특성을 구분해낸 성도 파라미터도 효과적으로 추정할 수 있다. 본 연구내용은 음성합성에서 자연성 개선 및 개인성 구현을 위해서 필수적이며, 다양한 형태의 음성을 표현하기 위해 사용되어질 수 있다. 또한, 음성코딩, 화자인식, 음성인식에서도 사용되어질 수 있다.

1. 서론

현재 대부분의 음성분석 및 합성은 음원(voice source)이라는 입력이 성도(vocal tract)라는 필터를 통과한 출력으로 음성파형을 간주하는 선형생성이론에 기초를 두고 있다 [9]. 이와 같은 음성분석에서는 합성음에서의 자연성과 개인성을 구현하는데 있어서 음원은 특히 중요하다. 일반적으로 유성음의 경우에는 폐에서 발생된 공기의 압력이 성문에서의 떨림에 의해 준주기적인 성문파로 변하여 성도에 입력되는 과정을 거쳐 유성이 생성된다. 그리고, 무성음의

경우 성도는 열린채 공기의 흐름을 일정지점에서 억제하여 일종의 음향잡음을 음원으로 발생시킨다. 파열음의 경우 일정기간동안 폐쇄점에서 공기의 흐름을 차단한 후 압력을 급격히 높였다가 성도로 음원을 내보내게 된다. 이러한 음원의 발생방식은 사람에 따라 달라지고 음원의 자체적인 변화패턴이 자연성에 영향을 주게 되므로 음원을 부정확하게 모델링하면 합성음의 자연성과 개인성이 떨어진다는 것을 예측할 수 있을 것이다 [10].

그러나, 음원의 특성에 대한 부정확한 추정으로 인해서 기존의 음성분석 및 합성은 다양한 실제음의 특성을 구현하기 어려운 문제점을 안고 있다. 이러한 문제를 해결하기 위하여 여러가지 추정법들을 제안하여 사용하고 있지만, 여전히 합성음의 자연성과 개인성이 실제음에 비해 떨어지는 실정이다.

음성과 음원의 추정 방법은 한 번에 처리하는 데이터의 양에 따라 크게 블록처리 방법과 순차처리 또는 sample by sample 방법으로 나눌 수 있다 [11].

블록처리 방법은 선형 예측 방법(LPC) [9] 처럼 분석 데이터를 블록 단위로 처리한 후 계수를 추정하는 방법이며, 순차처리 방법은 매 샘플마다 계수를 추정하는 방법으로 대표적인 예들로는 RLS(recursive least square) [3],[6], 칼만 필터링(Kalman filtering) [2] 등이 있다. 블록처리 방법은 안정된 계수의 추정과 계산량이 상대적으로 적다는 장점이 있는 반면 시변의 음성신호, 특히 음절간의 천이구간에서 계수 추정이 어렵게 되는 단점이 있다. 따라서 이러한 시변의 음성신호에서 생기는 문제점을 극복하는 방법으로서 순차처리 방법에 대해 많은 연구가 진행되고 있다.

그러나 LPC에서는 유성음의 음원신호를 임펄스열로 모델링하게 되므로 음원의 피치 주기가 음성 파라미터 추정에 영향을 미치게 된다. 이러한 영향을 피

견실 순차 특이치분해를 이용한 음원추정

치 바이어스라 부르며, 특히 여성이나 어린이의 음성
과 같이 피치 주기가 짧은 경우에 심각한 영향을 받
는다. 따라서 피치 바이어스 효과를 줄인 정확한 성
도의 필터 계수를 구하기 위해서 한 피치 주기 내에
서 선형 예측 방법을 적용하는 Pitch synchronous
LPC 방법과 한 피치 주기중 성문(glottis)의 폐쇄구간
에 해당되는 음성 데이터만을 선택하여 분석하는
SSLP (sample selective linear prediction) 방법 [5]
이 제안되었다. 한편 순차처리 방법의 경우에서도 피
치의 영향을 줄이기 위해 한 피치 구간 내의 성문폐
쇄 구간에서만 추정을 하는 방법 [4] 과, 미리 음원신
호를 추정된 후 이를 보상하는 방법 등이 제안되었
다. 그러나 위의 방법은 구간추출과 음원추정등에 많
은 계산량이 요구되는 단점이 있다.

본 논문에서는 변화가 심한 음원파형을 추정하는
새로운 순차처리 알고리즘을 제안한다. 먼저, 기존의
순차처리 분석법중 대표적인 분석법인 RLS(recursive
least square)의 문제점들을 검토하고, 이를 개선하기
위해서 관측행렬(observation matrix)을 최적차수의
SVD(singular value decomposition) [7],[8] 로 재구
성한다. 이에 견실개념(robustness concept) [12] 을
적용해서 최적의 성도 파라미터를 찾아내고 역필터를
적용해서 음원을 효과적으로 구분해내는 견실 순차
특이치분해(robust sequential SVD)를 이용한 음원
추정법을 제안한다.

II. 순차 특이치분해(sequential SVD)를 이용한
음원추정(voice source estimation)

지금까지 블록단위의 음원추정방법은 음성신호를
짧은 구간에서는 stationary process로 볼 수 있다는
가정에서, 매 분석구간 마다 그 구간의 특징 파라메
터를 하나의 대표값으로 추출하게 되므로 시간에 따
라 변화하는 음성신호에서 시변 계수를 추출하는대는
적절하지 않다. 따라서 순차처리 방법의 대표적인 알
고리즘으로 사용되고 있는 RLS를 음성 분석에 이용
하고 있다.

p 를 성도 필터의 차수, t 를 샘플수라고 할때, 입
력되는 음성신호의 벡터를 z 라고 하고, 추정된 계수
의 벡터를 a 라고 하면, RLS의 알고리즘은 표 1 과
같다 [6].

표 1 RLS(recursive least square) 알고리즘

```
Initialize :  $A^{-1}(-1) = aI$ ,  $a(-1) = [0 \dots 0]^T$ 
For  $t = 0, 1, 2, \dots$ 
    Input :  $z(t)$ ,  $x(t)$   $0 \leq \lambda \leq 1$ 
     $k(t) = A^{-1}(t-1)z(t)$ 
```

$$a(t) = \lambda + z^T(t)k(t)$$

$$k^*(t) = k(t)/a(t)$$

$$A^{-1}(t) = \lambda^{-1} [A^{-1}(t-1) - k^*(t)k^T(t)]$$

$$\varepsilon(t) = x(t) - z^T(t)a(t-1)$$

$$a(t) = a(t-1) + k^*(t)\varepsilon(t)$$

$$e(t) = x(t) - z^T(t)a(t)$$

end loop

그러나, RLS 알고리즘은 수치적으로 불안정하고
음의 고유치(negative eigenvalue)를 가질 수 있다.
따라서, 일반적으로 순차처리의 추정식은 잡음이 존
재할 경우 안정도(stability)와 정확도(accuracy)면에서
뛰어나지 못하다. 이러한 문제를 해결하기 위하여
역공분산행렬(inverse covariance matrix) A^{-1} 을
SVD로 재구성한다. 역공분산행렬의 SVD는 고유치
를 양수화하며 안정도를 효율적으로 개선할 수 있다
[7],[8].

추정직전의 역공분산행렬 $A^{-1}(t-1)$ 을 직교행렬
들(orthogonal matrices) U, V 와 대각행렬(diagonal
matrix) Σ 의 곱으로 표현하면,

$$A^{-1}(t-1) = U(t-1)\Sigma(t-1)V^T(t-1) \quad (1)$$

여기서 U, Σ, V 의 행렬들을 dominant subspace와
subdominant subspace들로 구분할 수 있다.
subdominant subspace들은 잡음과 음원의 영향을 포
함하고 있다고 가정하고, 이들을 제거할 수 있다. 즉,
식 (2)에서 r 차의 dominant subspace들만으로 역공
분산행렬을 근사화시킬 수 있다.

$$A^{-1}(t-1) = [U(r)|U(p-r)] \begin{bmatrix} \Sigma(r) & 0 \\ 0 & \Sigma(p-r) \end{bmatrix} \begin{bmatrix} V^T(r) \\ V^T(p-r) \end{bmatrix} \quad (2)$$

식 (2)에서 reduced-order r 은 평균자승오차
(mean-squared error)로부터 계산할 수 있으며, 식
(3)과 같다 [8].

$$r^* < - \min_r \{ z^T(P(A) - P(r))z + (2r-p)\sigma^2 \} \quad (3)$$

위에서 제시한 순차 특이치분해(sequential SVD)
를 RLS에 적용하면, 최적의 최소자승값(least square

solution) \hat{a} 를 제공하며, 음원 및 잡음의 영향도 어느 정도 구분해낼 수 있다. 또한, 양의 고유치를 보장할 수 있으며, 안정도(stability)도 개선할 수 있다.

III. 견실 순차 특이치분해(robust sequential SVD)를 이용한 음원추정

II 장에서 제안한 순차 특이치분해는 기존의 RLS에 비해서 음원의 영향을 효과적으로 구분해낼 수 있다. 하지만, 여전히 음원의 영향을 완전하게 구분해내지는 못한다. 음원의 영향을 효과적으로 구분해내기 위해서 일반적으로 많이 사용하는 방법은 closed-phase 개념을 적용하는 것이다. 즉, 음원의 영향이 거의 나타나지 않는 구간을 구분해내고 이 구역에서만 성도 파라미터를 추정하고 이의 역필터로 음원을 추정한다. 그러나, closed-phase 개념은 몇가지 문제점들을 가지고 있다. 즉, 음원의 영향이 나타나지 않는 구간을 구분해내는 것이 어렵다. 특히, 여성 및 어린이 목소리나 음질의 전이구간 등에서는 특히 어렵다.

따라서 본 논문에서는 가중함수를 이용하여 mixture Gaussian model 확률 분포를 갖는 잡음에 대해서 사용하는 robust 개념을 적용하였다 [12]. 위의 RLS 알고리즘으로부터 다음과 같이 robustness theorem을 세운다.

Theorem 1. Robustness Theorem

$$a(t) = a(t-1) + k^*(t)\psi(\varepsilon(t))$$

여기서, $\varepsilon(t) = x(t) - z^T(t)a(t-1)$

$\psi(\varepsilon(t))$ 를 적절히 선택, 조절하면 아래와 같이 robustness bound δ 이내로 조절된다.

$$E(x(t) - \hat{x}(t))(x(t) - \hat{x}(t)) \leq \delta$$

이 알고리즘은 강인 신호처리에 사용되고 있는 가중함수를 사용하여 정해 놓은 문턱값을 넘는 큰 값을 갖는 잔차 신호에 대해서는 상대적으로 낮은 가중치를 부과하는 방법이다. 따라서, 성도 및 음원 추정 알고리즘은 표 2 와 같다.

표 2. 견실 순차 특이치분해 알고리즘

Initialize : $A^{-1}(-1) = \sigma I, a(-1) = [0 \dots 0]^T$
 For $t=0, 1, 2, \dots$
 Input : $z(t), x(t) \quad 0 \leq \lambda \leq 1$

$$[U, \Sigma, V^T] = SVD(A^{-1}(t-1))$$

$$r^* = \min_r \{ z^T(P(A) - P(r))z + (2r-p)\sigma^2 \}$$

$$A^{-1}(t-1) = U_{p \times r} \Sigma_{r \times r}^{-1} V_{r \times p}^T$$

$$k(t) = A^{-1}(t-1)z(t)$$

$$a(t) = \lambda + z^T(t)k(t)$$

$$k^*(t) = k(t)/a(t)$$

$$A^{-1}(t) = \lambda^{-1} [A^{-1}(t-1) \cdot k^*(t)k^T(t)]$$

$$\varepsilon(t) = x(t) - z^T(t)a(t-1)$$

Compute $\psi(\varepsilon(t))$

$$a(t) = a(t-1) + k^*(t)\psi(\varepsilon(t))$$

$$e(t) = x(t) - z^T(t)a(t)$$

end loop

여기서 알고리즘의 강인특성은 위의 가중함수 ψ 의 선택에 따라 크게 영향을 받게 되는데 일반적으로 가중함수는 입력신호 $\varepsilon(t)$ 의 확률분포에 의해서 함수의 형태가 결정된다. 여기서 사용한 함수는 Huber 함수라 부르며, 아래의 형태를 갖는다.

$$\psi_\varepsilon(t) = \begin{cases} t & |t| \leq K \\ K \cdot \text{sgn}(t) & |t| > K \end{cases} \quad (4)$$

이때 문턱값 K 는 혼합계수 ε 의 함수이다. 즉, 입력신호의 non Gaussian 정도에 의해 상수로 결정된다. 그러나 음성신호에서는 혼합계수 ε 값을 알 수 없을 뿐만 아니라 오차신호의 분산값 R_ε 를 일정하게 정할 수 없으므로 그에 대한 대안으로 다음과 같이 가중함수의 문턱값 K 를 시간에 따라 적응 변화시켜서 사용할 것을 제안하였다.

$$K(n) = c \hat{\sigma}_v = c \left[\frac{1}{M} \sum_{i=1}^M v^2(n-i) \right]^{\frac{1}{2}} \quad (5)$$

여기서 $\hat{\sigma}_v$ 은 일정구간의 잔차신호 $\varepsilon(t)$ 에 대한 구간표준편차이며, M 은 구간표준편차를 계산하는 창 길이이다. 일반적으로 음성의 피치주기는 최소한 20 샘플보다는 크다고 보기 때문에 구간표준편차 $\hat{\sigma}_v$ 를 구할때 평균을 취하는 구간의 길이는 20 샘플로 정했다. 또한 비례상수 c 의 값은 실험을 통하여 2.0-2.5 내의 값으로 사용하였다.

IV. 실험 및 검토

먼저 기존의 순차처리분석의 예로 RLS와 Kalman을 이용해서 얻은 음원신호 즉, 오차신호를 검토한다. 20대의 남성화자를 대상으로 한국어 '아'음을 표본화율(sampling rate), 10KHz, 16 bit로 녹음했다. 차수는 20차로 했으며, 가중치는 0.9로 놓은 후 분석했다. 그 결과는 그림 1 과 같다. 그림 1의 결과에서 보면 기존의 순차처리분석으로는 음원의 특성을 정확히 추정해 낼 수 없고, 모델 오차가 심하게 나타남을 알 수 있다. 이러한 문제점을 개선하기 위해서 성문닫힘구간(glottal closure region)에서만 RLS 변수와 Kalman 추정기의 변수들을 개선시킨 순차표본선택최소자승(sample selective RLS) [11] 과 closed-phase Kalman 추정기 [2] 의 결과를 그림 2에 나타내었다. 그림 1의 결과보다는 많이 개선되었지만 여전히 음원의 특성을 추정해내지 못하고 있음을 알 수 있다.

그림 3에 본문에서 제안한 견실 순차 특이치분해를 이용한 음원추정의 결과를 나타내었다. 20차의 차수로 추정을 했다. 실험결과 15차에서 16차정도에서 reduced-order가 결정되었고, 유원파형을 효과적으로 추정해냄을 알 수 있다.

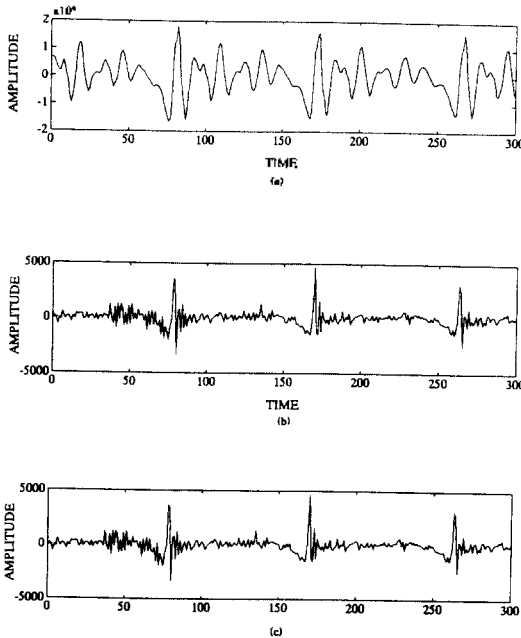


그림 1 RLS와 Kalman 추정기를 이용한 음원추정 (수평축은 0.1 ms 단위의 시간, 수직축은 amplitude)
 (a) 실제음 '아' (b) RLS의 음원추정
 (c) Kalman 추정기의 음원추정

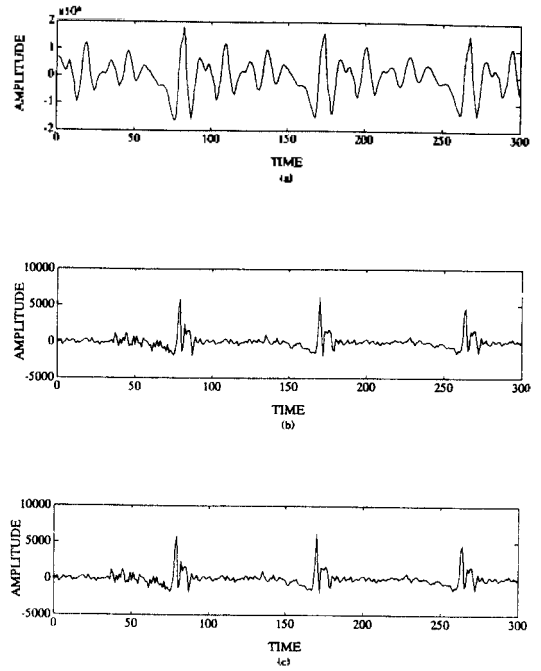


그림 2 순차표본선택 최소자승법과 closed-phase Kalman 추정기를 이용한 음원추정 (수평축은 0.1 ms 단위의 시간, 수직축은 amplitude)
 (a) 실제음 '아' (b) RSSLS의 음원추정
 (c) closed-phase Kalman 추정기의 음원추정

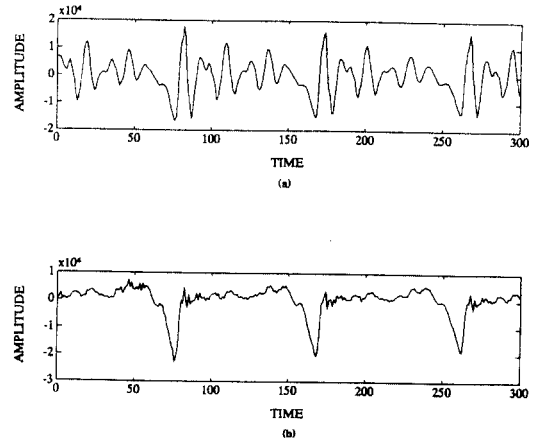


그림 3 견실 순차 특이치분해를 이용한 음원추정 (수평축은 0.1 ms 단위의 시간, 수직축은 amplitude)
 (a) 실제음 '아' (b) 추정된 음원신호

V. 결론

본 논문에서는 변화가 심한 음원파형을 추정하는 견실 순차 특이치분해 알고리즘을 제안했다. 먼저, 1) RLS(recursive least square)의 문제점들을 검토하고, 2) 이를 개선하기 위해서 관측행렬(observation matrix)을 최적차수의 SVD(reduced-rank singular value decomposition)로 재구성하고, 3) 이에 견실개념(robustness concept)을 적용해서 최적의 성도변수(vocal tract parameter)를 찾아내고 역필터를 적용해서 음원(voice source)을 효과적으로 구분해내었다. 본 논문에서 제안된 방법으로 음원을 추정할 경우, 변화가 심한 음원파형을 잘 추정할 수 있으며, 음원의 특성을 구분해낸 성도 파라미터도 효과적으로 추정할 수 있다. 특히, 음원추정이 어려웠던 여성과 어린이의 목소리나 음절의 천이구간 등에서의 문제점들을 개선할 수 있을 것이다. 또한, 제한한 음원모델을 사용할 경우 성문파의 다양한 형태를 표현할 수 있다. 따라서 합성기 등에서 음원의 특성을 좀 더 다양하게 표현할 수 있을 것이다.

본 논문의 연구내용은 음성합성분야에서의 자연성 개선 및 개인성 구현을 위해 쓰일 수 있으며, 음성 코딩에서의 음원모델에 적용가능하다. 또한 화자인식을 위한 원천기술로도 필요할 것이다.

참고문헌

- [1] D.Y.Wong, J.D.Markel and A.H.Gray, "Least Square Glottal Inverse Filtering from the Acoustic Speech Waveform", IEEE Trans. on ASSP, Vol. ASSP-27, No. 4, pp 350-355, August, 1979
- [2] S.Crisafulli, J.D.Mills and R.R.Bitmead, "Kalman Filtering Techniques in Speech Coding", Proceedings of ICASSP, Vol. 1, pp 77-80, 1992
- [3] Y.T.Ting and D.G.Childers, "Speech Analysis Using the Weighted Recursive Least Square Algorithm with a Variable Forgetting Factor", Proceedings of ICASSP, S7.5, pp 389-392, 1990
- [4] J.D.Gibson, J.L.Melsa, and S.T.Jones, "Digital Speech Analysis Using Sequential Estimation Techniques", IEEE Trans. on ASSP, Vol. ASSP-23, No. 4, pp 362-369, August, 1975
- [5] 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. on ASSP, Vol. ASSP-35, No. 9, September, 1987
- [6] P.Strobach, *Linear Prediction Theory*, Springer-Verlag Berlin Heidelberg, 1990

1993년도 한국음향학회 학술논문발표회 논문집(제 12권 1(a)호)

- [7] F.Deprette (editor), *SVD and Signal Processing : Algorithms, Applications and Architectures*, North Holland, 1988
- [8] R.Vaccaro (editor), *SVD and Signal Processing, II : Algorithms, Analysis, and Applications*, Elsevier Science Publishers B. V., 1991
- [9] J.D.Markel and A.H.Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, 1976
- [10] 홍성훈, 이정철, 안수길, "유성음의 성문파 추정", 제 9 회 음성통신 및 신호처리 워크샵 논문집, pp 59-73, 1992
- [11] 방승찬, 최홍섭, 이정철, 홍성훈, 이용주, 안수길, "순차표본선택 최소자승법을 이용한 음성분석", 제 10 회 음성통신 및 신호처리 워크샵 논문집, pp 128-132, 1993
- [12] C.J.Masrelez and R.D.Martin, "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE Trans. on Automatic Control, Vol. AC-22, No. 3, June, 1977