퍼지 및 지능시스템학회 논문지 2007, Vol. 17, No. 7, pp. 957-963

최소 제곱 서포트 벡터 회귀 기반 비선형 자귀회귀 방법을 이용한 지속 모음 모델링

Sustained Vowel Modeling using Nonlinear Autoregressive Method based on Least Squares-Support Vector Regression

장승진* · 김효민* · 박영철** · 최홍식*** · 윤영로*

Seung-Jin Jang* · Hyo-Min, Kim* · Young-Choel Park** · Hong-Shik Choi*** · Young-Ro Yoon*

* 연세대학교 의공학과

** 연세대학교 컴퓨터통신공학부

*** 연세대학교 이비인후과

요 약

본 연구에서는 비선형 지속 모음 모델링을 위한 최소 제곱 서포트 벡터 회귀 기반 비선형 자귀회귀 방법을 소개하고 분석하였다. 비주기적인 파형 특성을 갖는 양성 후두 질환자 43명의 지속 모음을 대상으로 한 실험에서 제안된 비선형 합성기는 거의 완벽하게 혼란한 지속 모음을 생성하고 선형 예측 코딩은 할 수 없는 주파수 변동과 같은 자연스러운 음의 특성 또한 보존할 수 있었다. 하지만 일부 모음의 합성 결과 실제 원음과 다른 차이점을 보였다. 이러한 결과들은 단일 밴드 모델이 음의 고주파 성분을 조정, 분해 못하기 때문에 발생한 것이라 가정된다. 그러므로 웨이블릿 필터 뱅크를 이용한 멀티 밴드 모델을 단일 밴드 모델과 대치하여 실험을 수행한 결과 향상된 안정성을 보였다. 결과적으로 최소 제곱 서포트 벡터 회귀 기반 비선형 자귀회귀 방법은 성공적으로 원음에 가까운 합성음을 생성할 수 있다는 것을 확인 할 수 있었다.

키워드 : 지속 모음 모델링, 최소제곱법 서포트 벡트 회귀, 비선형 자귀회귀 모델, 웨이블릿, 멀티밴드

Abstract

In this paper, Nonlinear Autoregressive (NAR) method based on Least Square-Support Vector Regression (LS-SVR) is introduced and tested for nonlinear sustained vowel modeling. In the database of total 43 sustained vowel of Benign Vocal Fold Lesions having aperiodic waveform, this nonlinear synthesizer near perfectly reproduced chaotic sustained vowels, and also conserved the naturalness of sound such as jitter, compared to Linear Predictive Coding does not keep these naturalness. However, the results of some phonation are quite different from the original sounds. These results are assumed that single-band model can not afford to control and decompose the high frequency components. Therefore multi-band model with wavelet filterbank is adopted for substituting single band model. As a results, multi-band model results in improved stability. Finally, nonlinear sustained vowel modeling using NAR based on LS-SVR can successfully reconstruct synthesized sounds nearly similar to original voiced sounds.

Key Words : Sustained Vowel Modeling, Least Squares-Support Vector Regression(LS-SVR), Nonlinear Autoregressive Model(NAR), Wavelet, Multi-band

1. 서 론

음성질환 연구 분야에 있어 음성 조음 모델링의 이론적 배경 및 원리를 기반으로 지속 모음을 모델링하여 음성의 이상 유무 및 질환의 종류 등을 검사하려는 연구들이 진행되고 있다[1-5].

큰 범주로 구분해서 음성 조음 모델 개발 방법은 조음기관 모델링과 음향적 모델링 2가지 방법이 존재한다. 조음기관 모델링 접근법은 성대의 형상과 운동을 가능한 물리적으로 세세하게 표현하는 것을 목표로 하여, 실제 음성과 비슷한 출력을 만드는 것을 기본 원리로 한다. 조음기관 모델링은 단순히 조음기관을 제어함으로써 만족할만한 수준의 음성 신호에 대한 재생산성을 갖는다는 장점을 갖고 있지만, 성대의 3차원적인 정보를 구하기 위하여 조음기관의 움직임에 대한 자세한 분석을 필요로 한다. 하지만, 각 조음기관의 정보는 획득하기에 어려운 구조적 특성들이 존재하며, 종종 침습적인 측정방법들이 요구되는 단점을 갖는다[6]. 반면에 음향

접수일자 : 2007년 10월 20일
완료일자 : 2007년 12월 3일
감사의 글 : 본 연구는 보건복지부에서 지원하는 바이오산업화기술개발 사업의 일환으로 추진되고 있는 특정센터연구지원/의료기기(과제고유번호: A020602) 분야의 재택 건강관리 시스템 연구센터 사업의 지원에 의하여 이루어진 것임. (교신저자: 윤영로)

적 모델링 방법의 접근 방법은 시간영역 또는 주파수 영역과 같은 음성파형으로부터 직접 취득된 정보를 활용하여 음성을 재생할 수 있다. 음향적 모델링은 분석에 있어 단지 음성파형만이 필요하므로 마이크로폰을 이용해서 쉽게 수집할 수 있다는 장점이 있다. 또한 지각적으로 실제와 근접한 음성 합성 및 그와 관련된 모델링 분석 이벤트들을 위한 정확한 음성 파형 또는 스펙트럼의 일치가 필요하지 않다는 점도 장점으로 작용한다.

음성 코딩, 음성 합성과 같은 음성모델링 기술에서 가장 일반적인 음향적 분석 방법은 선형 예측 코딩(Linear Predictive Coding : LPC) 방법이 있다. 음성신호 분석에서 LPC의 장점은 적용의 용이성, 단일해, 합성시 성대필터와 음성소스의 완벽한 분리 그리고 성대의 손실 없는 acoustic tube 모델링을 기반으로 한다는 것들이 존재한다[7].

그러나 LPC 방법은 몇 가지 치명적인 단점을 가지고 있다. 첫 째, 무성음의 경우 성대함수에 0의 값을 포함하기 때문에 minimum phase방식의 all-pole LPC 모델에서는 만족할 수 없는 모델링 성능을 보여준다. LPC 모델에서, 영점은 높은 예측 차수의 극점의 집합들로 근사화되고 이러한 이유로 스펙트럼 상에서 영점 근처에 존재하는 스펙트럼 피크를 검출하기 어렵기 때문에, LPC 모델에 의한 무성음의 합성은 어렵다고 여겨진다[8]. 두 번째로 LPC 모델 변수들은 합성을 위해 사용되는 모델의 출력에서 얻어지는 실질적인 에러보다는 평균 제곱 예측 에러를 최소화하기 위한 구조로 설계되어 있기 때문에 합성을 위한 용도에 최적화된 모델이 아니라는 단점이 존재한다. 세 번째로 음성 신호를 분석하는데 있어서 음원과 음원 여과기를 분리할 수 없는 설계 구조로 인해, LPC 필터모델은 음원, 성대, 입술파열음의 영향력이 조합된 구조로 되어 있다. 그리하여 선형 예측을 통한 음성 합성의 질은 원 음성보다 낮아지게 된다. 마지막으로 선형적인 모델링 특성으로 인해 Jitter, Shimmer와 같은 음성 신호의 비선형적인 요소를 반영하지 못하는 제약요소가 있다. 이러한 문제점으로 인해 다른 대안적인 조음 모델링의 요구가 부각되고 있는 추세이다.

본 연구에서는 LPC 모델의 한계점을 극복할 수 있는 서포트 벡터 머신(Support Vector Machine: SVM)기반 비선형 외인성 자귀회귀 모델을 이용한 비선형 조음 모델링 방법을 소개하고 구현하여 실제 비주기적이고 복잡한 형태를 갖는 음성질환자의 음성신호에 적용하여 무질서하고 변동이 심한 지속 모음을 위한 모델링 방법을 제시하고자 한다.

2. LS-SVR 기반 NAR 모델

Vapnik(1985, 1988)에 의해 제안된 SVM은 부분해와 같은 오류를 발생할 수 있는 신경망 이론과 같은 머신 러닝과는 달리 구조적으로 통계적인 오류를 최소화하게끔 설계되어져 있다[9]. 즉, 다른 머신 러닝 방법들과 비교했을 때, 다른 방법들이 종종 불안전한 성능(원하는 출력과 상이한 경우)을 보여주는 반면, SVM은 일반적으로 만족할만한 성능을 보여준다. 이러한 이유는, SVM 계산식에는 일반화 오류를 제거하기 위한 정규화된 수식이 포함되어 있기 때문이다. 또한 SVM은 부분해가 존재하지 않으며, 적은 변수만을 사용하여 수식을 전개하기 때문에 실제 구현하는데 있어 용이한 장점을 갖는다.

최소제곱 서포트 벡터 회귀(Least Squares-Support Vector Regression: LS-SVR)을 이용한 비선형 자귀 회귀(Nonlinear Autoregressive Regression: NAR) 모델을 설계할 때, 일단 음성 신호를 재구성된 임의의 고차원 공간으로 변환하게 되면, 비선형 모델링의 생성은 LS-SVR을 사용하여 시스템을 예측하는 함수로 바뀌게 된다. 이것을 수식적으로 해석해 보면, 먼저 LS-SVR 시스템을 모델링하고자 할 때, $\{x_k,y_k\}_{i=1}^{N} \subset \mathbf{R}^d \times \mathbf{R}$ 으로 정의되는 트레이닝 데이터 입력 값들과 출력 값이 있다고 가정한다. 이 때 입력 $x_1,...,x_N$에 대한 $f:\mathbf{R}^d \rightarrow \mathbf{R}$ 특성을 갖는 함수 f와 상관관계가 없는 랜덤 에러 $e_1,...,e_N$들을 가지고 다음과 같은 회귀 모델 식 (1)을 정의할 수 있다.

$$y_i = f(x_i) + e_i \tag{1}$$

단, 랜덤 에러들은 $E[e_i]=0, E[e_i^2]=\sigma_e^2 < \infty$와 같은 특성을 만족한다고 가정한다. LS-SVR은 비선형 함수 f를 추정하기 위하여 식 (2)와 같은 수식를 이용하여 모델링을 한다.

$$f(x) = w^T \Phi(x) + b \tag{2}$$

기저 함수 Φ는 $\Phi(x):\mathbf{R}^d \rightarrow \mathbf{R}^{n_f}$ 특성을 갖는, 즉 잠재적으로 무한 차원에 가까운 특성 맵핑을 의미하며, w는 가중치 벡터와 b는 바이어스를 의미한다. 식 (2)를 기저함수를 이용한 표현으로 수식을 바꾸면, 출력 데이터 y_k가 식 (3)으로 정의될 때, 정규화된 최소 제곱법의 손실 함수는 식 (4)와 같이 w,b,e를 최소화하는 식으로 표현 할 수 있다.

$$y_k = w^T \Phi(x_k) + b + e_k \ , \ k=1,...,N \tag{3}$$

$$\min_{w,b,e} \mathfrak{J}(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{n} e_k^2 \tag{4}$$

이 때 정규화 상수로 작용하는 파라미터 γ를 조정하여 실제 해에 가까운 데이터 적합을 얻을 수 있게 된다. 이러한 최적화 수행 방법은 ridge regression으로 알려져 있는 방법을 이용하여 수행된다[10]. 이후 제한된 최적화 문제를 풀기위하여 라그랑제를 이용하면 식 (5)와 같이 식 (4)를 변경할 수 있고, 이 때 a_k는 라그랑제 승수가 된다.

$$\begin{aligned} \mathfrak{L}(w,b,e:\alpha) &= \mathfrak{J}(w,e) \\ &- \sum_{k=1}^{N} \alpha_k \{ w^T \Phi(x_k) + b + e_k - y_k \} \end{aligned} \tag{5}$$

결론적으로 최적화를 위한 조건들은 식 (6-9)와 같이 정의하여 각각의 파라미터 값을 풀어 가면 해를 구할 수 있다.

$$\frac{\delta \mathfrak{L}}{\delta w} = 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k \Phi(x_k) \tag{6}$$

$$\frac{\delta \mathfrak{L}}{\delta b} = 0 \rightarrow \sum_{k=1}^{N} \alpha_k = 0 \tag{7}$$

$$\frac{\delta \mathfrak{L}}{\delta e_k} = 0 \rightarrow \alpha_k = \gamma e_k, \ k=1,...,N \tag{8}$$

$$\frac{\delta \mathfrak{L}}{\delta \alpha_k} = 0 \rightarrow y_k = w^T \Phi(x_k) + b + e_k, k=1,...,N \tag{9}$$

위의 파라미터들을 구하는 수식들에 대한 결과를 행렬 형태로 변환하여 요약하면 식 (10)과 같이 간략하게 표현할 수 있다.

$$\begin{vmatrix} 0 & 1_N^T \\ 1_N & \Omega+\gamma^{-1}I_N \end{vmatrix} \begin{vmatrix} b \\ \alpha \end{vmatrix} = \begin{vmatrix} 0 \\ y \end{vmatrix} \quad (10)$$

이 때 $y=[y_1 \cdots y_N]^T$, $1_N=[1 \cdots 1]^T$, $\alpha=[\alpha_1 \cdots \alpha_N]$, $\Omega_{kl}=\Phi(x_k)^T\Phi(x_l)=K(x_k,x_l)$ 으로 정의되며 K는 커널 트럭으로 불리는 양한정 행렬(커널)의 특성을 갖는다. 최종적으로 시스템에 적용할 수 있는 LS-SVR 모델의 수식은 새롭게 입력된 데이터 x_φ에 대하여 식 (11)과 같이 정의할 수 있으며, $\theta=(b,\alpha)$으로 정의되는 해를 의미한다.

$$\hat{f}(x_\varphi:\theta)=\sum_{k=1}^{N}\alpha_k K(x_\varphi,x_k)+b \quad (11)$$

3. 실험 방법

3.1 실험 데이터

음성파형은 연속해서 다른 음소를 발성할 때 주파수 특성이 끊임없이 변화하며, 일반적으로 이러한 특성을 평가하기 위해서는 매개 변수의 적정한 비율로 끊임없이 개선을 할 수 있는 파형의 시변 모델이 필요하다. 전형적으로, 단시간 분석이 사용되었고, 음성파형은 약 50ms의 중첩된 파형 간격으로 나누어 각 파형 간격에서 새로운 파형 모델의 파라미터 집합으로 계산하였다. 본 연구에서는 연세대학교 영동세브란스 병원 음성언어의학연구소에서 2003년 6월부터 2007년 3월 사이에 검진하여 유사 주기적이고 불규칙적인 음성을 주로 발성하는 양성 후두질환자(Benign Vocal Fold Lesions: BVFL)로 판명된 43명의 지속 모음에 대하여 조음 모델링을 설계하고 분석하였다. 특히 음성질환은 후두에 구조적으로 외형적인 변형이 발생한 용종, 낭종, 결절 음성질환자의 음성만을 대상으로 실험하였다. 실험 데이터의 대상자는 성별에 상관없이 나이 분포는 18~71세이며, 평균 나이는 43.7세이고, 음성 data는 잡음이 없는 환경에서 컴퓨터 화된 음성녹음 장비를 (Kay Elemetrics CSL) 이용하여 음성신호와 성문파형신호(Electroglottography: EGG)를 동시에 수집하였다. 모든 음성신호와 EGG data는 22㎑로 샘플링 되어졌고 16bit 분해능을 갖고 있다. 실험 프로토콜에 의해 성별을 구분하였으며, 음소의 발성 차이점으로 인해 지속 모음의 주파수 특성이 다르다는 점을 감안하여 최소 2-3초간 유지되는 지속 모음(/a/, /e/, /i/, /o/, /u/)별로 샘플을 구분하여 분석하였다.

3.2 최적화 파라미터 선택

커널 함수는 feature 공간에서 내적에 상응하는 함수이다. 커널함수는 Φ기저 함수를 이용하여 입력 데이터의 차원보다 높은 차원의 feature 공간으로 트레이닝 데이터를 맵핑시키고, 분류를 위해 마진폭을 최대한으로 갖는 분리된 hyper-plane을 설계한다. 비록 Mercer의 이론[11]에 따라 준양한정 대칭행렬 (semi-positive definite symmetric) 함수를 만족하는 많은 커널 함수들이 존재하며, 또한 커널 함수의 선택에 따라 분류 성능이 상이해질지라도, 일반적인 경우 원형 기준 함수(Radial Basis Function)를 기반으로 한 커널 함수는 합리적인 커널 함수로 간주될 수 있다. 즉, 여러 개의 파라미터를 갖는 보다 좋은 성능을 보이는 많은 수의 기저 함수들이 존재할지라도, 이러한 기저 함수들이 보다 과대 적합(over-fitting)하게 될 확률을 높이게 되고, 따라서 일반화 오류를 높이게 할 수도 있다. 그러므로 본 연구에서는 식(12)과 같이 다른 커널 함수 선택 없이 (σ^2, γ)두 개의 파라미터를 갖는 원형 기준 함수를 LS-SVR의 커널함수로 적용하였다. σ^2는 SVM에서 사용되는 가중치 벡터의 상한 값을 제어하기 위한 loss 상수이고, γ는 원형 기준 함수의 반경을 결정하는 파라미터이다.

$$K(x_i,x_j)=exp(-\gamma\|x_i-x_j\|^2) \quad ,\gamma>0 \quad (12)$$

하지만, 일반적으로 SVM 분류기의 성능이 다른 머신 러닝 방법들 보다 뛰어나고 일반화 오류가 최적화되게끔 설계할지라도[12], (σ^2, γ) 파라미터들은 원형 기준 함수 커널 모델에서 이용되는 유일한 파라미터이며, 또한 잘못된 (σ^2, γ) 값을 선택할 경우 과대 적합한 트레이닝 모델을 산출하게 되므로 보다 정확한 트레이닝 정확도를 구하기 위하여 좋은 (σ^2, γ)계수 값을 구하는 것은 매우 중요한 일이다. 그리하여, 본 연구에서는 최적의 (σ^2, γ)값을 선정하기 위한 목적으로 learning set을 동일한 사이즈를 갖는 v(=5)개의 파티션으로 나눈 후, v-1개의 subsets들에 의해 훈련된 분류기를 가지고 남아 있는 하나의 subset에 테스트하여 가장 좋은 성능을 보이는 모델을 선택하는 v-fold cross-validation 방법을 이용하였다. 이 방법을 사용함으로써 과대 적합 문제를 방지하고, 트레이닝에 포함되지 않은 데이터의 분류에 대한 성능을 좀 더 향상 시킬 수 있게 된다. 휴리스틱한 방법을 사용할 수밖에 없는 SVM 분류기의 (σ^2, γ)파라미터 선택은 [그림 1]과 같이 Grid search을 이용하여 제한된 범위에서 최적의 파라미터 (σ^2, γ) 값을 선택하였다.

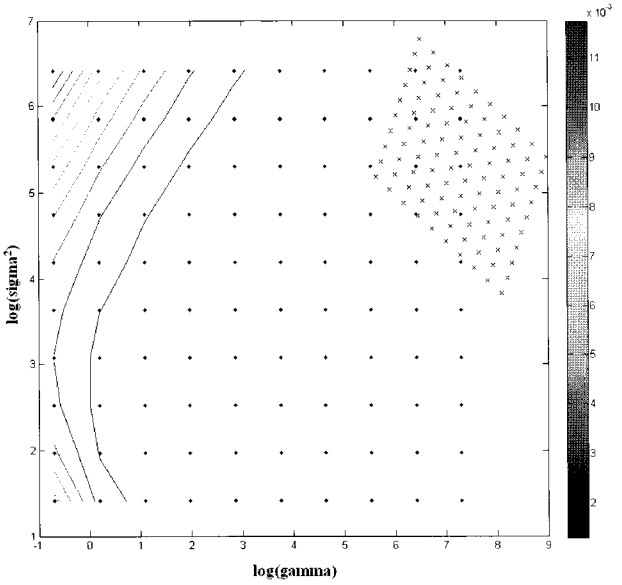


그림 1. 남성의 /i/ 모음을 위한 최적화된 σ^2, γ 검출

표 1. 성별과 발음에 따른 최적화 파라미터 σ^2, γ 값 산출 결과

sex	Phonation	σ^2		γ	
		Mean	S.D.	Mean	S.D.
male	/a/	17.31	2.26	791.40	21.91
	/e/	22.08	3.37	802.07	18.04
	/i/	26.08	4.70	814.64	24.99
	/o/	23.53	4.08	814.69	27.61
	/u/	28.41	4.18	810.09	21.25
female	/a/	27.04	4.81	827.66	21.86
	/e/	28.18	4.63	829.35	21.16
	/i/	37.65	4.19	828.65	27.75
	/o/	30.49	4.19	837.39	35.61
	/u/	30.19	4.84	841.39	37.00

4. 실험 및 성능 분석

4.1 LS-SVR기반 NAR 스키마 구조

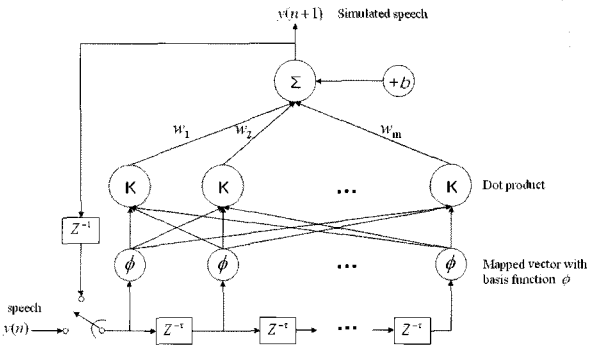


그림 2. LS-SVR 기반 NAR 스키마

그림 2와 같이 입력 데이터들을 원형 기준 함수로 맵핑시켜서 고차원 공간으로 변형시킨 후 각 변형된 입력 데이터 간의 내적을 취한 값들을 적분하고 바이어스 값을 더하여 시뮬레이션된 음성을 구해낸다. 모음이 발성되는 임의의 구간에서 동일 지속 모음의 50ms 구간에 대하여 훈련시켜서 표 1과 같이 최적 파라미터를 설계하였고, 이후 50ms 샘플 데이터를 지연시켜서 다음에 발생하는 지속 모음을 예측하였다.

4.2 Single-Band 모델

그림 3과 같이 LS-SVR을 사용한 NAR모델은 성공적으로 유성음 /a/를 합성해내는 것을 볼 수 있다. 하지만 일부의 발음의 경우 그림 4의 /i/ 모음에서 살펴볼 수 있듯이 음성신호의 재구성에 어려움이 있음을 알 수 있다. 재구성된 음성은 특히 고주파 성분을 발생하는 신호로 인해 차이가 발생함을 알 수 있다. 이러한 원인으로 발성 /i/, /o/, /u/ 모음의 경우와 같이 고주파수 성분이 많은 음성에서 주로 이러한 현상이 발생하는 것을 발견할 수 있었다. 이러한 값의 정도를 분석하기 위하여 실제 원음과 시뮬레이션된 지속 모음간의 mse를 분석한 결과 표 2와 같은 결과를 구할 수 있었다. 또한 전반적으로 남성보다 여성의 지속 모음에 대한 mse 결과가 더 높게 산출되는 것으로 보아 높은 피치를 갖는 음성에 대한 예측이 더 어려움을 알 수 있었다. 이러한 문제점을 해결하기 위하여 시간상의 음성신호를 다차원의 주파수로 분해하여 신호를 분석하는 multi-band 모델을 이용하는 것을 제안하였다.

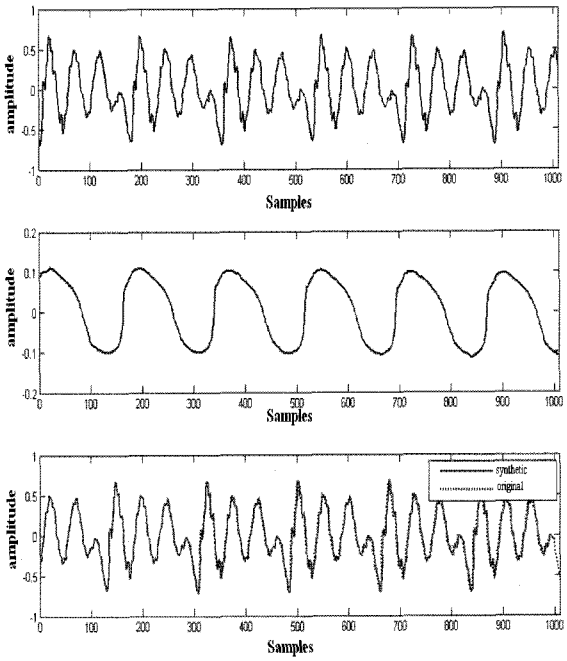


그림 3. 합성된 음성 vs 실제 음성신호((time delay = 50 샘플): 남성 모음 /a/에 대한 (top) 실제 음성신호, (middle) EGG 신호, (bottom) 합성 + 실제 음성 신호

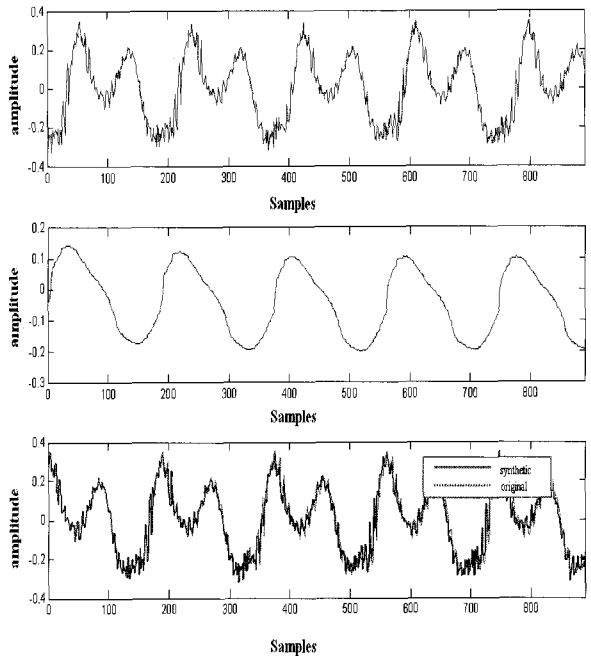


그림 4. 합성된 음성 vs 실제 음성신호((time delay = 50 샘플): 남성 모음 /i/에 대한 (top) 실제 음성신호, (middle) EGG 신호, (bottom) 합성 + 실제 음성 신호

표 2. 성별과 발음에 따른 실제 원음과 시뮬레이션된 지속 모음의 평균 제곱 에러 결과

sex	Phonation	Mean Square Error	
		Mean	S.D.
male	/a/	1.26×10^{-2}	0.72×10^{-2}
	/e/	1.28×10^{-2}	0.93×10^{-2}
	/i/	3.69×10^{-2}	1.04×10^{-2}
	/o/	4.17×10^{-2}	1.84×10^{-2}
	/u/	4.62×10^{-2}	1.47×10^{-2}
female	/a/	3.83×10^{-2}	1.37×10^{-2}
	/e/	4.23×10^{-2}	1.24×10^{-2}
	/i/	7.11×10^{-2}	3.69×10^{-2}
	/o/	6.51×10^{-2}	2.87×10^{-2}
	/u/	6.06×10^{-2}	3.15×10^{-2}

4.3 Multi-Band 모델

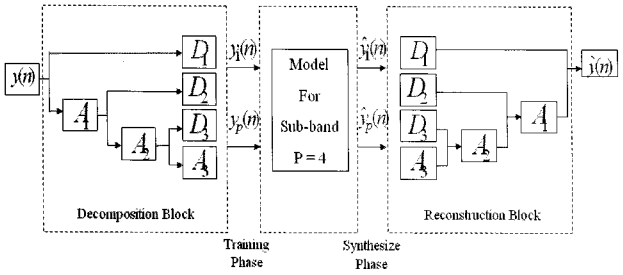


그림 5. 웨이블릿 필터를 이용한 Multi-band NAR 모델

Multi-band 모델은 선형 적응형 필터에서 잘 알려진 모델이다. 이 모델의 Sub-Band 기술은 전대역 접근법에서 기대되는 몇 가지 장점을 가지고 있다[13-15]. 첫 번째 특징으로, 이것은 전처리 과정 전에서 신호의 decimation으로 의해 계산의 효율성을 가진다. 두 번째 입력신호의 분리에 따라 최소자승 적용 알고리즘에서 보다 좋은 수렴 성능을 나타내게 된다[16].

Multi-band 신호 처리의 중요한 주제로 필터뱅크 선택이 여겨진다. 그림 5에서 볼 수 있듯이, 본 연구에서는 웨이블릿 필터뱅크를 이용하여 다중 대역에 대한 음성을 분해한 후 각각에 분해시킨 신호를 LS-SVR의 입력 신호에 인가하여 예측한 후 각각의 예측된 신호를 다시 재구성하여 합성된 신호를 만들었다. 웨이블릿은 3차 미세(detail) 신호까지 분해하였다. 이러한 결과, 그림 4에서 제대로 합성하지 못한 /i/ 지속 모음을 Multi-Band를 이용하여 다시 합성한 결과 그림 6과 같이 성공적으로 합성할 수 있었으며, 고주파 대역이 많이 존재하는 모음뿐만 아니라 /a/, /e/와 같은 주로 저주파 성분을 갖는 지속모음에 대해서도 성공적으로 유성음을 합성해내는 것을 알 수 있다. 이러한 결과는 표 3 mse 분석 결과에서 확연하게 살펴 볼 수 있었으며 그림 7에서와 같이 스펙트로그램 분석결과에서 고주파 대역의 일부 영역을 제외하고는 비슷한 패턴을 보임을 알 수 있었다.

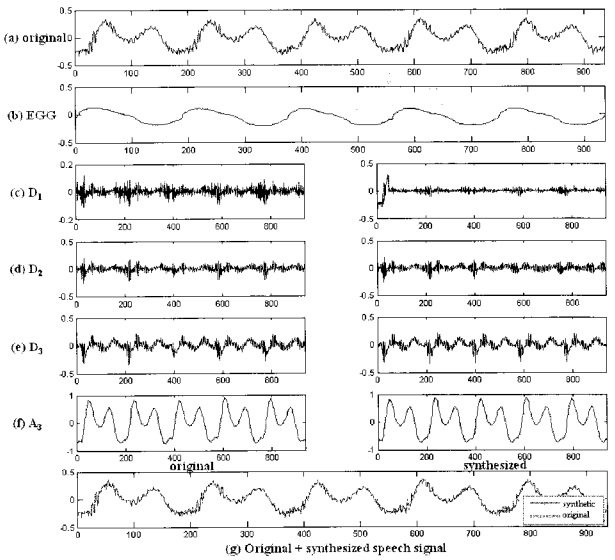


그림 6. 합성 신호 vs 실제 신호 비교 (time delay = 50 샘플): 남성의 /i/ 모음에 대한 (a) 실제 신호, (b) EGG 신호, (c-f) 각각의 D1-3(합성 신호), A1-A3 sub-band (실제 신호), (f) 합성 + 실제 음성 신호

표 3. 성별과 발음에 따른 실제 원음과 시뮬레이션된 지속 모음의 평균 제곱 에러 결과

sex	Phonation	Mean Square Error	
		Mean	S.D.
male	/a/	0.41×10^{-2}	0.14×10^{-2}
	/e/	0.33×10^{-2}	0.13×10^{-2}
	/i/	0.86×10^{-2}	1.15×10^{-2}
	/o/	1.07×10^{-2}	0.97×10^{-2}
	/u/	1.05×10^{-2}	0.89×10^{-2}
female	/a/	0.93×10^{-2}	0.24×10^{-2}
	/e/	1.18×10^{-2}	0.19×10^{-2}
	/i/	1.21×10^{-2}	1.36×10^{-2}
	/o/	1.42×10^{-2}	1.40×10^{-2}
	/u/	1.05×10^{-2}	1.48×10^{-2}

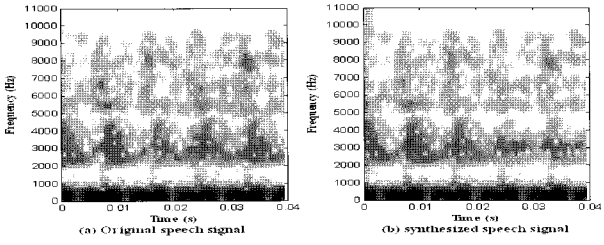


그림 7. 남성 /i/ ([그림 4]의 신호)신호에 대한 실제 음성 신호와 합성 신호의 스펙트로그램 비교

4.4 주파수 변동(Jitter) 성능 분석

음성질환자의 언어 장애(Dysphonia) 연구 분야에서도 피치 추정을 이용한 주파수 변동(Jitter)은 음성 장애의 정도를 평가할 수 있는 중요한 파라미터로 활용되고 있다. 식(13)으로 정의되는 Jitter(%) 값이 원신호와 시뮬레이션된 신호 사이에서 얼마나 유사한지에 대해 Welch's rank sum 검정을 기반으로 분석한 결과 <표 4>에서와 같이 통계적으로 원신

호와의 사이에서 별 다른 차이가 없음을 알 수 있었다.

$$jitter(\%) = \frac{\frac{1}{n-1}\sum_{i=n-1}^{1}|F_{i+1}-F_i|}{\frac{1}{n}\sum_{i=1}^{n}F_i}, \quad F_i : ith\ F_0 \qquad (13)$$

표 4. 원신호와 시뮬레이션된 지속 모음의 shimmer 분석

sex	Phonation	synthesized		original		p
		Mean	S.D.	Mean	S.D.	
male	/a/	3.76	1.21	3.71	1.54	.721
	/e/	4.78	0.34	4.15	0.84	.747
	/i/	3.86	0.37	3.48	1.11	.106
	/o/	4.66	0.90	4.54	0.83	.932
	/u/	4.78	0.94	4.71	0.78	.801
female	/a/	4.83	0.91	4.47	1.22	.574
	/e/	5.74	2.37	4.99	1.60	.321
	/i/	4.89	1.31	4.38	1.77	.557
	/o/	5.49	1.55	5.74	1.22	.676
	/u/	5.61	1.38	5.08	1.44	.724

5. 결 론

본 연구에서 제시된 LS-SVR 기반의 NAR 모델을 가지고 무질서하고 불규칙적인 지속 모음을 모델링하고 예측하였다. 이러한 비선형 합성기는 LPC 기반의 선형 모델이 예측하기 어려운 비선형 신호 예측에 대해서 거의 완벽하게 원음을 재생할 수 있었으며, 주파수 변동과 같은 특성 또한 원신호의 값과 유사하게 생성함을 알 수 있었다. 다만 고주파 성분이 존재하는 /i/, /o/, /u/와 같은 일부 발성에서는 원 신호와 유사한 결과를 산출하지 못하였으며, 이를 해결하기 위하여 웨이블릿 필터를 이용한 multi-band 기반의 비선형 자기회귀 모델을 이용하여 다시 모델링을 재구성한 결과 성별, 발음에 구분 없이 향상된 결과를 나타내었다. 결론적으로 본 연구에서 제시된 최소 제곱 서포트 벡터 회귀 기반 비선형 자귀회귀 방법을 이용한 지속 모음 모델링은 양성후두질환과 같은 혼잡한 지속 모음 신호의 예측 및 모델링에 우수한 결과를 산출할 수 있음을 보였다.

향후 과제로는 설계된 모델링의 결과를 향상시키기 위하여 고주파 대역의 성분에 대한 입력 데이터 차수와 원음과 시뮬레이션 신호와의 mse 관계를 분석하는 것이 요구되며, 신호를 multi-band 필터로 분해, 재구성할 때 발생할 수 있는 에러를 줄이는 것에 대한 연구가 필요하다.

참 고 문 헌

[1] Giovanni A, Robert D, Estubier N, Teston B: Objective evaluation of dysphonia: Preliminary results of a device allowing simultaneous acoustics and aerodynamics measurements. Folia, Phon. Logop.

[2] Banci G, Monini S, Falaschi A, Sario N: Vocal fold disorder evaluation by digital speech analysis, J.Phonetics,1986,vol.14,pp.495-499.

[3] Gavidia-Ceballos L, Hansen L: Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection., IEEETr.on BiomedicalEng., 1996, vol.43,pp.373-383.

[4] Laver J, Hiller S, Mackenzie J, Rooney E: An acoustic screening system for the detection of laryngeal pathology. J.Phonetics,vol.14,pp.517-524.

[5] J.C. Principe, A. Rathie, J.M. Kuo, Prediction of chaotic time series with neural networks and the issue of dynamic modeling, Int.J.BifurcationChaos, 1992, vol.2,pp.989 - 996.

[6] C.S. Blackburn, Articulatory Methods for Speech Production and Recognition, PhD Thesis, Cambridge University Engineering Department, 1996.

[7] Rabiner L. and Juang B. H., Fundamentals of speech recognition, Prentence Hall, NJ, 1993.

[8] Klatt, D, Review of text-to-speech conversion for english, J.ofAcoustSocofAm.,1987,vol.82,pp.737-793.

[9] V. Vapnik, The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.

[10] Golub, G.H. and C.F. Van Loan, Matrix Computations. John Hopkins University Press, 1989.

[11] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, Philos. Trans. Roy. Soc. London 1909.

[12] B. Schlkopf, A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2001.

[13] H. Yasukawa, Signal restoration of broad band speech using nonlinear processing, Proceedings of EUSIPCO'96, Trieste, Italy, Sept. 1996.

[14] R.E. Crochiere, L.R. Rabiner, Multirate Digital Signal Processing, Prentice-Hall, Englewood CliLs, NJ, 1983.

[15] N.J. Fleige, Multirate Digital Signal Processing (Multirate systems, Filter Banks, Wavelet), Wiley, New York, 1994.

[16] M.R. Petraglia, S.K. Mitra, Performance analysis of adaptive filter structures based on subband decomposition, Proceedings of the IEEE International Symposium on Circuit and Systems, Chicago, IL, 1993, pp. 60 - 63.

저 자 소 개

장승진(Seung-Jin Jang)
2000년 : 연세대 의용전자공학과 학부 졸업
2002년 : 연세대 의공학과 석사 졸업
2007년 : 연세대 의공학과 박사 졸업
현재 연세대 의공학과 박사후과정

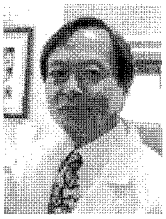
관심분야 : 생체/음성 신호처리, 의료정보, 패턴 인식, 머신 러닝
Phone : 033-760-2809
Fax : 033-763-1953
E-mail : highnoon@yonsei.ac.kr

김효민(Hyo-Min Kim)
2007년 : 연세대 의공학과 학부 졸업.
2007년~연세대 의공학과 석사과정

관심분야 : 영상처리, 음성신호처리, 패턴인식
Phone : 010-9906-3376
Fax : 033-763-1953
E-mail : hyominkim@yonsei.ac.kr

박영철(Young-Choel Park)
1986년 : 연세대 전자공학과 학부 졸업.
1988년 : 연세대 전자공학과 석사 졸업.
1993년 : 연세대 전자공학과 박사 졸업.
현재 연세대 컴퓨터정보통신공학부 부교수

관심분야 : 음성신호처리
Phone : 033-760-2744
Fax : 033-763-4323
E-mail : young00@yonsei.ac.kr



최홍식(Hong-Shik Choi)
1978년 : 연세대 의과대학 졸업.
1981년 : 연세대 대학원 의학석사 졸업
1986년 : 연세대 대학원 의학박사 졸업
현재 연세대학교 음성언어의학연구소 소장

관심분야 : 후두생리, 인공후두, 후두이식
Phone : 02-3497-3460
Fax : 02-3463-4750
E-mail : hschoi@yumc.yonsei.ac.kr

윤영로(Young-Ro Yoon)
1981년 : 연세대 전자공학과 학부 졸업.
1985년 : California State University 전기공학 석사 졸업
1991년 : Purdue University 전기공학 박사 졸업
현재 연세대 의공학부 교수

관심분야 : 의료정보, 원격진료, 생체신호처리

Phone : 033-760-2440
Fax : 033-763-1953
E-mail : yoon@yonsei.ac.kr