# JOINT SOURCE-FILTER MODELING USING FLEXIBLE BASIS FUNCTIONS

*Daryush D. Mehta*[*] *Daniel Rudoy*[*] *Patrick J. Wolfe*[*†]

[*]Statistics and Information Sci. Laboratory
Harvard University
{dmehta, rudoy, patrick}@seas.harvard.edu

[†]Speech & Hearing Biosci. & Technology
Harvard-MIT Div. of Health Sci. & Technology

## ABSTRACT

Improving on recent work on joint source-filter analysis of speech waveforms, we explore improvements to an autoregressive model with exogenous inputs represented by flexible basis functions. Following a brief review of the maximum likelihood estimators of the model parameters, the Cramér-Rao bounds are derived to provide evidence for the challenging nature of estimating source and filter characteristics with overlapping spectra. Wavelet expansion of the exogenous inputs is employed, and the selection of an appropriate subset of wavelets is described as an online, signal-adaptive approach. Results from synthesized and real vowel analysis illustrate the promise of iterative wavelet shrinkage using soft and hard thresholding and an alternative regularization method.

***Index Terms—*** Glottal flow, harmonics-to-noise ratio, linear prediction, spectral estimation, wavelet regression

## 1. INTRODUCTION

In classical source-filter representations of acoustic speech waveforms, autoregressive (AR) models parameterize the filter transfer function of the vocal tract with a white Gaussian noise term modeling the glottal source waveform and radiation characteristics. To address the model mismatch due to the presence of quasi-periodic source waveforms during speech [1], our group previously incorporated the ARX model to include a time-varying eXogenous component $\mu[n]$ to explicitly capture the quasi-periodic nature of the glottal airflow derivative [2]. In this parameterization, $\mu[n]$ is modeled by a basis function expansion that exhibits flexible properties, which preclude the need to estimate pitch periods and/or glottal closure instants during speech analysis.

First, the current work derives the Cramér-Rao bounds of the resultant estimators of the ARX model parameters and explores sensitivity issues dependent on the spectral content of the source and filter. Second, as a wavelet expansion is proposed for the exogenous variable, we address the issue of how to adaptively select the appropriate wavelet subspace to represent source waveforms in synthesized and real speech examples. Wavelet thresholding and $\ell_1$-penalty approaches are compared through experiments on vowel waveforms with constant and time-varying fundamental frequency.

## 2. MODEL FORMULATION

Modifying the classical AR($p$) speech model for length-$N$ signal $x[n]$, here we briefly describe the ARX($p$) model of [2] that adds an *exogenous* input $\mu[n]$:

$$\text{ARX}(p): \quad x[n] = \sum_{i=1}^{p} a_i x[n-i] + \mu[n] + \sigma w[n]. \tag{1}$$

The discrete-time difference equation of (1) comprises an all-pole, linear time-invariant system driven by a Gaussian process $w[n]$ with constant variance $\sigma^2$ and a *time-varying mean* $\mu[n]$, which we use to capture the quasi-periodic nature of the glottal airflow derivative. Specifically, the sequence $\mu[n]$ is defined as

$$\mu[n] \triangleq \sum_{k=1}^{r} \beta_k g_k[n], \tag{2}$$

where the $r$ basis functions $\{g_1[n], g_2[n], \ldots, g_r[n]\}$ are specified before any data are observed.

The ARX($p$) model specified by (1) and (2) reduces to the classical AR($p$) process when $\beta_k = 0$ for $1 \leq k \leq r$. The application of the ARX($p$) model to speech source estimation has appeared in other literatures. Its study was initiated in [3–6] from a theoretical perspective (i.e., estimation and asymptotic analysis) and as an application to econometrics. It has since been used in control theory [7] and signal processing [8]. As an alternative to the representation of (2), one may also assume a linear [9] or nonlinear [10] parameterization for the source waveform $\mu[n]$.

A variety of natural choices for the basis functions exists, including a small set of low-order polynomials [8]. If $\mu[n]$ were periodic, a Fourier series could be used; pitch period estimates would be required, however, leading to robustness issues regarding natural pitch variation and irregular phonation due to the *global* support of the Fourier basis functions. The selection of *time-localized* functions aids in addressing these issues; thus, the use of wavelets as flexible basis functions first proposed in [2] is explored further in this paper.

## 3. PARAMETER ESTIMATION

### 3.1. Maximum Likelihood

The ARX($p$) model of (1) and (2) is specified by a vector of AR coefficients $\boldsymbol{a} \triangleq \begin{pmatrix} a_1 & a_2 & \cdots & a_p \end{pmatrix}^T \in \mathbb{R}^{p\times 1}$, a vector of expansion coefficients $\boldsymbol{\beta} \triangleq \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \end{pmatrix}^T \in \mathbb{R}^{r\times 1}$, and the noise variance $\sigma^2$. The maximum likelihood (ML) estimator of the model parameters $\boldsymbol{\theta} \triangleq (\boldsymbol{a}^T, \boldsymbol{\beta}^T, \sigma^2)^T$ is the least-squares solution to the following linear regression problem:

$$\boldsymbol{x}_{N-p} = \boldsymbol{X}\boldsymbol{a} + \boldsymbol{G}\boldsymbol{\beta} + \sigma \boldsymbol{w}_{N-p} = \begin{pmatrix} \boldsymbol{X} \mid \boldsymbol{G} \end{pmatrix} \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{\beta} \end{pmatrix} + \sigma \boldsymbol{w}_{N-p}, \tag{3}$$

where $\boldsymbol{w}_{N-p} \triangleq \begin{pmatrix} w[p] & w[p+1] & \cdots & w[N-1] \end{pmatrix}^T \in \mathbb{R}^{(N-p)\times 1}$, and matrices $\boldsymbol{X} \in \mathbb{R}^{(N-p)\times p}$ and $\boldsymbol{G} \in \mathbb{R}^{(N-p)\times r}$ are defined by

$$\boldsymbol{X} \triangleq \begin{pmatrix} x[p-1] & \cdots & x[0] \\ x[p] & \cdots & x[1] \\ \vdots & \ddots & \vdots \\ x[N-2] & \cdots & x[N-p+1] \end{pmatrix}, \; \boldsymbol{G} \triangleq \begin{pmatrix} g_1[p] & \cdots & g_r[p] \\ g_1[p+1] & \cdots & g_r[p+1] \\ \vdots & \ddots & \vdots \\ g_1[N-1] & \cdots & g_r[N-1] \end{pmatrix}.$$

Following prior derivations [2], the least-squares estimates of $\boldsymbol{a}$, $\boldsymbol{\beta}$, and $\sigma^2$ are

$$\begin{pmatrix} \widehat{\boldsymbol{a}} \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{G} \\ \boldsymbol{G}^T\boldsymbol{X} & \boldsymbol{G}^T\boldsymbol{G} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}^T \\ \boldsymbol{G}^T \end{pmatrix} \boldsymbol{x}_{N-p}, \tag{4a}$$

$$\widehat{\sigma}^2 = \frac{1}{N-p}\|\boldsymbol{x}_{N-p} - \boldsymbol{X}\widehat{\boldsymbol{\alpha}} - \boldsymbol{G}\widehat{\boldsymbol{\beta}}\|_2^2. \tag{4b}$$

Note that if all the basis functions $g_k$ were equal to 0 everywhere, then the model of (1) reduces to the classical AR model, and the least-squares estimators of (4) reduces to the standard covariance method of linear prediction.

### 3.2. Cramér-Rao Bounds

The Cramér-Rao lower bounds on the variance of estimators (4a) and (4b) provide insight into their performance on signals representative of speech. The Fisher information $\boldsymbol{I}$ for the ARX($p$) model (1) has been derived in [8] as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\begin{pmatrix} \widetilde{\boldsymbol{R}} + \boldsymbol{M}^T\boldsymbol{M} & \boldsymbol{M}^T\boldsymbol{G} & \boldsymbol{0}_{p\times 1} \\ \boldsymbol{G}^T\boldsymbol{M} & \boldsymbol{G}^T\boldsymbol{G} & \boldsymbol{0}_{r\times 1} \\ \boldsymbol{0}_{1\times p} & \boldsymbol{0}_{1\times r} & (N-p)/(2\sigma^2) \end{pmatrix}, \tag{5}$$

where $\widetilde{\boldsymbol{R}} = (N-p)\boldsymbol{R}$ and $\boldsymbol{R}$ is the symmetric Toeplitz matrix constructed from the autocorrelation sequence $(r_{xx}[0], r_{xx}[1], \ldots, r_{xx}[p-1])$.

To define matrix $\boldsymbol{M} \in \mathbb{R}^{(N-p)\times p}$, let $m[n] \triangleq \mathbb{E}(x[n])$ and suppose that the first $p$ means $m[0], m[1], \ldots, m[p-1]$ are known. Since $w[n]$ is a zero-mean process, it follows from (1) that $m[n]$ can be recursively computed for any $p \le n \le N-1$ as $m[n] = \sum_{i=1}^{p} a_i m[n-i] + \mu[n]$. For each $1 \le i \le p$, let the vector $\boldsymbol{m}_i \in \mathbb{R}^{(N-p)\times 1}$ be defined via

$$\boldsymbol{m}_i \triangleq \begin{pmatrix} m[p-i] & m[p+1-i] & \cdots & m[N-1-i] \end{pmatrix}^T,$$

and define the matrix $\boldsymbol{M}$ as

$$\boldsymbol{M} \triangleq \begin{pmatrix} \boldsymbol{m}_p & \boldsymbol{m}_{p-1} & \cdots & \boldsymbol{m}_1 \end{pmatrix}.$$

The Cramér-Rao lower bounds of interest (in the sense of matrix inequalities) are thus given by

$$\mathrm{Cov}(\widehat{\boldsymbol{a}}) \ge \frac{1}{\sigma^2}\left(\widetilde{\boldsymbol{R}} + \boldsymbol{M}^T\boldsymbol{M} - \boldsymbol{M}^T\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{M}\right)^{-1}, \tag{6}$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) \ge \frac{1}{\sigma^2}\left(\boldsymbol{G}^T\left(\boldsymbol{I}_{N-p} + \boldsymbol{M}\widetilde{\boldsymbol{R}}^{-1}\boldsymbol{M}^T\right)^{-1}\boldsymbol{G}\right)^{-1}, \tag{7}$$

$$\mathrm{Var}(\widehat{\sigma}^2) \ge \frac{2\sigma^4}{N-p}.$$

Now consider (6) and (7) when entries of $\boldsymbol{M}$ are large. This may occur when the spectral energy of the columns of $\boldsymbol{G}$ (and, consequently, the mean signal $\mu[n] = \boldsymbol{G}\boldsymbol{\beta}$) and the all-pole spectrum associated with $\boldsymbol{a}$ concurrently take large values over the same set of frequencies. In this case, (6) and (7) dictate that the trace of the CRLB for the AR coefficients $\widehat{\boldsymbol{a}}$ *decreases*, whereas the trace of the CRLB of the expansion coefficients $\widehat{\boldsymbol{\beta}}$ *increases*.

We illustrate this somewhat counterintuitive result using $N = 250$ observations of the following ARX(2) process:

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + \beta_1 \cos[2\pi n\omega/f_s] + w[n]. \tag{8}$$

In the first experiment, we set the variance of $w[n]$ to 1, $a_1$ and $a_2$ such that the corresponding second-order resonator has a center frequency of 2 kHz and bandwidth of 51 Hz, and $f_s = 16$ kHz signal. Consider the CRLBs for $\widehat{\boldsymbol{a}}$ and $\widehat{\beta}_1$ as a function of $\omega$. The largest spectral overlap of the exogenous sinusoid and the autoregressive filter should occur when $\omega$ is around 2 kHz, verified by the left three panels of Figure 1.

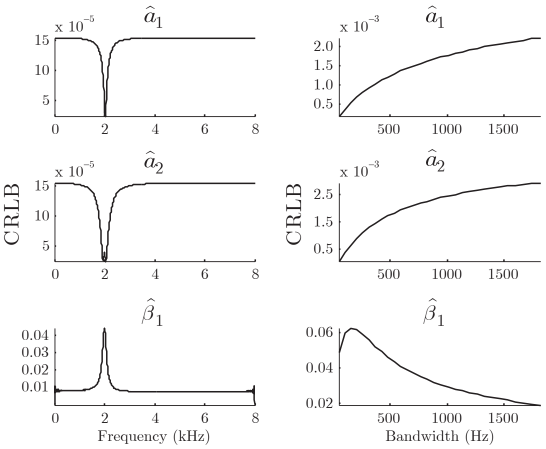

**Fig. 1**. Square root of the Cramér-Rao lower bound (CRLB) for estimates $\widehat{a}_1$, $\widehat{a}_2$, and $\widehat{\beta}_1$ in the ARX(2) model of (8). Left: CRLBs over varying frequencies of the exogenous sinusoid. Right: CRLBs over varying bandwidths of the AR filter.

In the second experiment, we fix $\omega$ at 2 kHz and vary $a_1$ and $a_2$ such that the center frequency of the associated second-order digital resonator is fixed *also* at 2 kHz but with bandwidth varied. As the bandwidth approaches 0, the frequency response of the all-pole filter sharpens around 2 kHz. In fact, when the bandwidth is equal to 0, the AR(2) model converges to the exogenous sinusoid exactly. The right three panels of Figure 1 confirm that, as the bandwidth decreases, the entries of the matrix $\boldsymbol{M}$ decrease, resulting in the predicted increases and decreases in the CRLB for $\widehat{\beta}_1$ and $\widehat{\boldsymbol{a}}$, respectively.

This suggests that the coefficients of $\mu[n]$–e.g., the sinusoid in (8)–in the ARX model may be more difficult to estimate when the spectral content of $\mu[n]$ significantly overlaps with that of the autoregressive filter. This is not too surprising from the point of view of deconvolution. For instance, in the context of speech processing with $\mu[n]$ representing a glottal source waveform, this result implies that it is more difficult to estimate the spectral characteristics of the source if they match those of the filter, as is often the case with high-pitched speakers [11].

## 4. SUBSPACE SELECTION

To apply wavelets, it may seem natural to use all wavelets in a basis supported on $N-p$ samples (assuming that $N-p$ is a power of 2). In this case, $\boldsymbol{G}$ would be a full rank matrix ($r = N-p$), and the columns of $\boldsymbol{X}$ would lie in the span of the columns of $\boldsymbol{G}$, rendering an ill-defined matrix inverse in estimator (4). One approach to avoid this problem is to select a subspace of dimension $r << N-p$ to model the time-varying mean. In this section, we explore three algorithms that provide such wavelet shrinkage.

Given a time series of observations, an appropriate subset of a wavelet basis for $\mathbb{R}^{N-p}$ is selected *online* in a signal-adaptive manner:

$$
\begin{aligned}
\mu[n] = & \sum_{j=0}^{L} \sum_{k=0}^{2^j-1} \phi_{j,k} 2^{-j/2} \phi[2^{-j}n-k] + \\
& \sum_{j=L+1}^{M} \sum_{k=0}^{2^j-1} \psi_{j,k} 2^{-j/2} \psi[2^{-j}n-k],
\end{aligned} \tag{9}
$$

where $\psi[n]$ is the mother wavelet, $\psi_{j,k}$ are the wavelet function coefficients, $\phi[n]$ is the low-pass scaling function, $\phi_{j,k}$ are the scaling function coefficients, and $L$ is an integer such that $2^L$ denotes the number of scaling function coefficients in the decomposition.

To begin, we assume that $N-p$ is a power of two to admit dyadic sampling, and let columns of a matrix $\boldsymbol{W} \in \mathbb{R}^{(N-p)\times(N-p)}$ contain the elements of some wavelet basis for $\mathbb{R}^{N-p}$. Thus, $\boldsymbol{W}$ is the orthonormal discrete wavelet transform (DWT) matrix so that $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$. We define the vector $\boldsymbol{x}_w \triangleq \boldsymbol{W}^T\boldsymbol{x}_{N-p}$ and the matrix $\boldsymbol{X}_w \triangleq \boldsymbol{W}^T\boldsymbol{X}$ as the DWTs of the data $\boldsymbol{x}_{N-p}$ and the design matrix $\boldsymbol{X}$, respectively. Rewriting the normal equations of (3) [12] in the wavelet domain yields the estimators

$$
\widehat{\boldsymbol{a}} = \left(\boldsymbol{X}_w^T\boldsymbol{X}_w\right)^{-1} \boldsymbol{X}_w^T(\boldsymbol{x}_w - \widehat{\boldsymbol{\beta}}), \tag{10a}
$$

$$
\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{W}^T\boldsymbol{W}\right)^{-1} \boldsymbol{W}^T\left(\boldsymbol{x}_{N-p} - \boldsymbol{X}\widehat{\boldsymbol{a}}\right) = \left(\boldsymbol{x}_w - \boldsymbol{X}_w\widehat{\boldsymbol{a}}\right), \tag{10b}
$$

which, in conjunction with wavelet thresholding, provide a natural *iterative* estimation approach described in Algorithm 1.

To fully specify Algorithm 1, recall that $L$ is an integer such that $2^L$ denotes the number of scaling function coefficients in the wavelet decomposition. Thus there are $2^L$ "coarse" (low-pass) coefficients associated with translations and dilations of the scaling function and $(N-p-2^L)$ "detail" (high-pass) coefficients associated with translations and dilations of the mother wavelet. Accordingly, we partition the coefficient vector $\boldsymbol{\beta}$ via

$$
\boldsymbol{\beta} = \begin{pmatrix}\boldsymbol{\beta}_c^T & \boldsymbol{\beta}_d^T\end{pmatrix}^T,
$$

with coarse coefficients $\boldsymbol{\beta}_c$ and detail coefficients $\boldsymbol{\beta}_d$.

At the $i$th iteration of Algorithm 1, the estimate of $\mu[n]$ based on $\widehat{\boldsymbol{\beta}}^{(i-1)}$ is subtracted from the waveform, and the covariance method of linear prediction is used to estimate $\boldsymbol{a}$ from the residual.[1] Next the waveform is inverse filtered using a moving average filter with coefficients $\widehat{\boldsymbol{a}}^{(i)}$, and the discrete wavelet transform of the residual is obtained. A nonlinear thresholding step is then applied to the resultant detail coefficients $\widehat{\boldsymbol{\beta}}_d^{(i)}$ in order to regularize the overall solution. Thresholding is achieved through a hard or soft thresholding rule with threshold $\lambda_1$, using an estimate of $\widehat{\sigma}$ advocated by [13]. Hard thresholding in Algorithm 1 explicitly promotes sparsity in the estimate of $\boldsymbol{\beta}$. The algorithm stops after the convergence of AR coefficients or after a specified number of iterations. Other thresholding rules and stopping criteria are possible [13]. A follow-up step selects the subspace matrix $G$ from $W$ by taking the top-$r$ $\beta$ coefficients to produce a smoothed estimate of $\mu[n]$ by joint estimation using (4).

Alternatively, we can penalize the detail coefficients directly using an $\ell_1$-norm criterion, leading to the following convex optimization problem:

$$
\left(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{\beta}}\right) = \underset{(\boldsymbol{a},\boldsymbol{\beta})}{\operatorname{argmin}} \left(\|\boldsymbol{x}_w - \boldsymbol{X}_w\boldsymbol{a} - \boldsymbol{\beta}\|_2 + \lambda_2\|\boldsymbol{\beta}_d\|_1\right), \tag{11}
$$

where $\lambda_2$ is an appropriately-chosen threshold. Efficient solutions to (11) can be found using, e.g., CVX [14].

**Algorithm 1** Subspace Selection via Iterative Shrinkage

```
• Initialization: Set tolerance level $\epsilon_0$, iteration counter $i = 1$, number of coarse levels $L$, and initial estimate of the coefficients $\widehat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}_{(N-p)\times 1}$
• While $\epsilon > \epsilon_0$ and $i < i_0$
    – Update estimates $\widehat{\boldsymbol{a}}$ and $\widehat{\boldsymbol{\beta}}$ using (10)
          $\widehat{\boldsymbol{a}}^{(i)} = \left(\boldsymbol{X}_w^T\boldsymbol{X}_w\right)^{-1}\boldsymbol{X}_w^T\left(\boldsymbol{x}_w - \widehat{\boldsymbol{\beta}}^{(i-1)}\right)$
          $\widehat{\boldsymbol{\beta}}^{(i)} \triangleq \left(\widehat{\boldsymbol{\beta}}_c^{(i)T} \quad \widehat{\boldsymbol{\beta}}_d^{(i)T}\right)^T = \boldsymbol{x}_w - \boldsymbol{X}_w\widehat{\boldsymbol{a}}^{(i)}$
    – Calculate threshold $\lambda_1$
        ∗ Calculate $\widehat{\sigma}$ as median absolute deviation of finest-resolution wavelet coefficients, divided by 0.6745 [13]
        ∗ Set $\lambda_1$ to $\sqrt{2\widehat{\sigma}^2\log(N-p-2^L)}$
    – Thresholding: for all $1 \le j \le N-p-2^L$
          Hard: $\widehat{\boldsymbol{\beta}}_d^{(i)}(j) = \begin{cases}\widehat{\boldsymbol{\beta}}_d^{(i)}(j) & \text{if} \quad |\widehat{\boldsymbol{\beta}}_d^{(i)}(j)| > \lambda_1 \\ 0 & \text{if} \quad |\widehat{\boldsymbol{\beta}}_d^{(i)}(j)| \le \lambda_1\end{cases}$
          Soft: $\widehat{\boldsymbol{\beta}}_d^{(i)}(j) = \operatorname{sgn}\left(\widehat{\boldsymbol{\beta}}_d^{(i)}(j)\right)\max\left(\left|\widehat{\boldsymbol{\beta}}_d^{(i)}(j)\right| - \lambda_1, 0\right)$
    – Compute change in AR coefficient vector and increment: $i = i + 1$
          $\epsilon = \frac{1}{N-p}\left\|\widehat{\boldsymbol{a}}^{(i)} - \widehat{\boldsymbol{a}}^{(i-1)}\right\|_2^2$
• Return $\widehat{\boldsymbol{a}}^{(i)}$, $\widehat{\boldsymbol{\beta}}^{(i)}$, and $\widehat{\sigma}$
```

## 5. EXPERIMENTS: SIGNAL-ADAPTIVE SUBSPACE SELECTION

Here we illustrate the algorithms on a synthesized and real vowel. The synthesized waveform ($f_s = 16$ kHz, $N = 518$, $p = 6$ so that $N-p$ is a power of 2) is generated using the linear source-filter model with a Rosenberg pulse derivative as source and 10 dB source SNR. Constant-pitch and time-varying pitch contours are simulated for the phoneme /i/. The real vowel is the phoneme /ae/ produced by an adult male exhibiting glottalized voice quality.

First, we apply Algorithm 1 to *iteratively* estimate $\mu[n]$ and the AR spectrum. As a follow-up step, the wavelets associated with the 64 largest-magnitude inferred coefficients are taken as columns of matrix $\boldsymbol{W}$, and the conditional maximum likelihood estimators of (4) are used to *jointly* fit the basis function coefficients and AR coefficients. Finally, the $\ell_1$-regularization approach of (11) is applied. For all methods, we employ Daubechies 6 wavelets with $L = 0$ and $p = 6$. Results are evaluated using the root-mean-square-error of estimates of $\mu[n]$ and the log-spectral distance ($d_{\text{LS}}$) of estimates of the AR spectrum [2]. For comparison, the residual of the covariance method is applied with $p = 8$ and pre-emphasis filter.

Figure 2A and B show that Algorithm 1 with hard thresholding is able to accurately estimate both $\mu[n]$ and the AR spectrum in the synthesized constant-pitch condition, improving upon the non-white residual of the covariance method. Figure 2C and D show the potential of hard thresholding to handle a synthesized time-varying

[1] Thus the estimate of $\boldsymbol{a}$ during the first iteration is equivalent to that obtained by the covariance method.

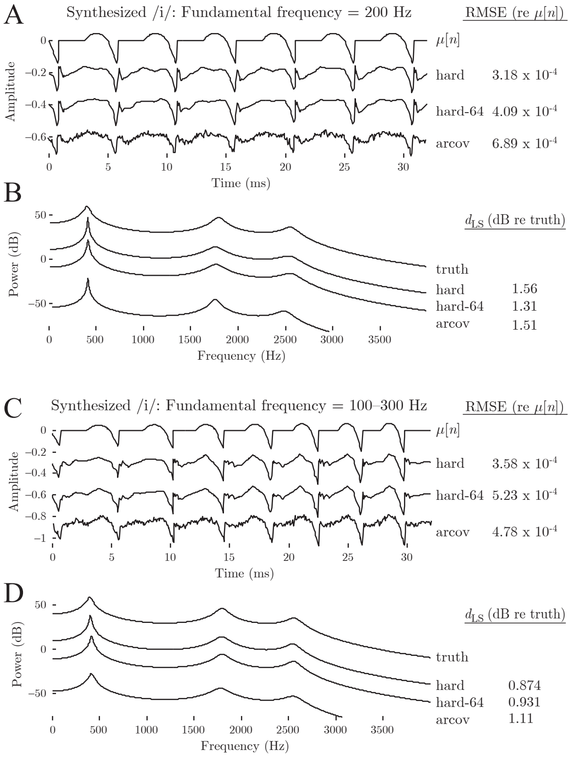

**Fig. 2**. Subspace selection via iterative wavelet shrinkage and hard thresholding on synthesized /i/ with constant (A, B) and time-varying (C, D) pitch contours. Covariance method (arcov) residual and filter estimates shown. Waveforms offset for visualization.

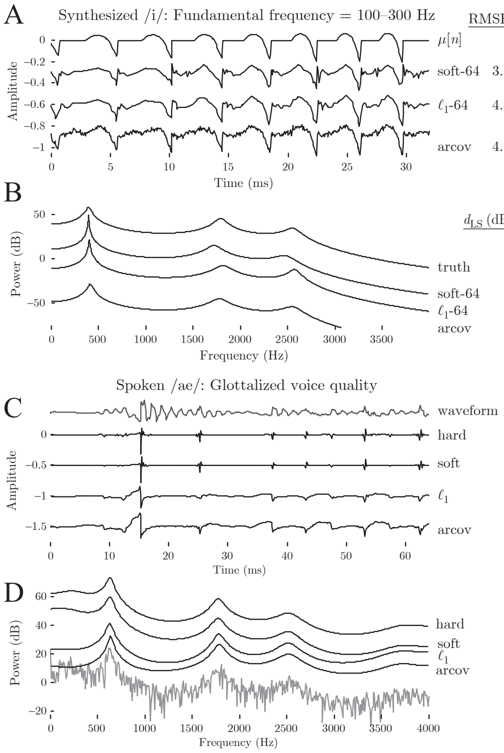

**Fig. 3**. Subspace selection via iterative wavelet shrinkage with soft thresholding and $\ell_1$ regularization on synthesized /i/ with time-varying fundamental frequency (A, B). Algorithm performance on real vowel /ae/ (C, D). Covariance method (arcov) residual and filter estimates shown. Waveforms offset for visualization.

pitch contour. Soft thresholding and $\ell_1$-regularization outputs are displayed in Figure 3A and B, providing comparable performance as additional approaches. Finally, Figure 3C and D show the outputs of Algorithm 1 (hard and soft thresholding) and $\ell_1$ regularization applied to the spoken vowel. Here, the vowel's irregular pitch periods do not affect the applicability of the ARX methods, whose estimates of the source and AR coefficients absorb the non-white components of the covariance method residual.

This initial evaluation of sustained vowels is a natural first step recognizing the need for accurate clinical voice assessment where voice quality is often acquired in a controlled environment. The applied estimators show promise for obtaining source-related information immune to pitch irregularities, warranting further investigation. Level-dependent thresholding or penalization of wavelet coefficients may further improve performance of the presented algorithms.

## 6. REFERENCES

[1] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, pp. 411–423, 1991.

[2] M. A. Berezina, D. Rudoy, and P. J. Wolfe, "Autoregressive modeling of voiced speech," in *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process.*, 2010, pp. 5042–5045.

[3] T. W. Anderson and H. Rubin, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Stat.*, vol. 20, pp. 46–63, 1949.

[4] T. W. Anderson and H. Rubin, "The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Stat.*, vol. 21, pp. 570–582, 1950.

[5] J. Durbin, "Estimation of parameters in time-series regression models," *J. Roy. Stat. Soc. Ser. B*, vol. 22, pp. 139–153, 1960.

[6] T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley and Sons, 1971.

[7] L. Ljung, *System Identification*, Upper Saddle River, NJ: Prentice-Hall, 1999.

[8] D. Sengupta and S. M. Kay, "Parameter estimation and GLRT detection in colored non-Gaussian autoregressive processes," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, pp. 1661–1675, 1990.

[9] P. Milenkovic, "Glottal inverse filtering by joint estimation of an AR system with a linear input model," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, pp. 28–42, 1986.

[10] D. Vincent, O. Rosec, and T. Chonavel, "A new method for speech synthesis and transformation based on an ARX-LF source-filter decomposition and HNM modeling," in *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process.*, 2007, pp. 525–528.

[11] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, 2002.

[12] P. Speckman, "Kernel smoothing in partial linear models," *J. Roy. Stat. Soc. Ser. B*, vol. 50, pp. 413–436, 1988.

[13] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming (web page and software)," June 2009. [Online]. Available: `http://stanford.edu/~boyd/cvx`.