RECURSIVE LINEAR PREDICTION USING OBE IDENTIFICATION WITH AUTOMATIC BOUND ESTIMATION

J. R. Deller, Jr. *T. M. Lin* *M. Nayeri*

Michigan State University
Department of Electrical Engineering / 260 EB
Speech Processing and Adaptive Signal Processing Labs
East Lansing, MI 48824-1226 USA

ABSTRACT

Application of set-membership (SM) identification to real-time speech processing is made possible by the *optimal bounding ellipsoid algorithm with automatic bound estimation* (OBE-ABE) that blindly deduces model-input bounds. To date, lack of any tenable approach to estimating bounds in speech models has rendered these interesting new SM methods impractical. OBE-ABE is consistently convergent, offers significant computational advantages, and provides a set of feasible solutions in finite time.

1. INTRODUCTION

Optimal bounding ellipsoid (OBE) identification algorithms (e.g., [4, 5, 8]) have strong potential for application to speech-processing problems involving linear prediction (LP) and other parametric models. With respect to the entrenched LP analysis methods (e.g., [6]), OBE identifiers offer superior adaptation, improved accuracy, efficient use of innovation in the data, improved computational efficiency, robustness to measurement noise, robustness to deviation from the assumed input model, a set of feasible solutions rather than a single point estimate, and the ability to compute the solution recursively in time without block processing or windows (e.g., [5]). In spite of the potential benefits of OBE identification, however, a significant practical impediment concerning input modeling has precluded widespread application of the methods to speech processing. This paper introduces a new class of OBE algorithms that include a major breakthrough for practical application, and demonstrates the benefits of the new OBE-based technique in speech modeling.

In the LP formulation, a stationary frame of speech $\{y_n\}$ is modeled as

$$y_n = \sum_{i=1}^{m} a_i y_{n-i} + v_{n*} = \theta_*^T \mathbf{x}_n + v_{n*} \tag{1}$$

in which $\theta_*^T = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}$ is the unknown parameter vector, and the conceptual input, $\{v_{n*}\}$, is stationary white noise in the unvoiced case, and a unit sample train of appropriate pitch for voiced speech (e.g., [6]). In fact, for satisfactory performance of conventional LP identifiers, it is only necessary that the input sequence have *spectral* properties similar to the assumed models.

On the contrary, OBE methods make no assumption about correlation properties of the input. Rather, OBE algorithms are based on the premise that a sequence of energy bounds, $\{\gamma_n\}$, is known *a priori* such that $v_{n*}^2 \le \gamma_n$ for each n. Specific time-domain information (often simpler and easier to pose in signal-processing problems) takes the place of restrictive correlation assumptions about the driving sequence.

Unfortunately, failure to prescribe accurate bounds in OBE processing is potentially catastrophic. Underestimated bounds may cause divergence to a parameter vector that is not even capable of generating the observed data (outside the *feasible* set), while overestimated bounds may cause the estimate to "freeze" at a biased estimate. In either case, improper bounds imply statistical inconsistency.

The *OBE algorithm with automatic bound estimation* (OBE-ABE) is the first SM method to solve the difficult problem of blindly estimating speech-model input bounds[1]. OBE-ABE removes the practical roadblock to LP modeling of speech. The parameter estimator inherent in OBE-ABE converges consistently under conditions on $\{v_{n*}\}$ expected to be met in speech signals, and additionally provides the customary OBE set of feasible solutions in finite time. This work significantly advances speech results in [2, 3] where an OBE method was applied with heuristic bound estimation.

2. BENEFITS OF OBE PROCESSING

Before proceeding with formal developments, we elaborate upon the significance of OBE-ABE as an LP identifier for speech, and upon the scope of this paper in that regard.

Fundamentally, OBE-ABE is a recursive-in-time identifier, offering the algorithmic convenience of temporal recursion, avoiding the inelegant and hard-to-analyze process of batch processing of overlapping frames. Neglecting window effects, the RLS algorithm (e.g., [11]) has always been available to speech processors as a (framewise) theoretical equivalent to batch-LP solutions, but one with inadequate finite-time convergence, poor adaptation, and $\mathcal{O}(m^2)$ computational complexity compared with $\mathcal{O}(m)$ for batch-LP methods (e.g., [6]). OBE-ABE, on the other hand, is an RLS-like set of recursions which, in its suboptimal variant [9], is of $\mathcal{O}(m)$ complexity, and which converges to an excellent solution in time intervals typical of speech frames. (This convergence behavior is also superior to conventional OBE algorithms with *ad hoc* methods for bound estimation [2, 3].) With an enhancement similar to the selective-forgetting approach used with conventional OBE algorithms [5], OBE-ABE can be made to quickly adapt to changing dynamics in the speech [9]. With temporal recursion, the adaptation

[1] The "ABE" enhancement can be incorporated into *any* OBE algorithm [4]. Here we employ the OBE version known as *set-membership–stochastic approximation* (SM-SA).

occurs continuously rather than in a "forced" and "quantized" manner due to block processing.

Further, OBE-ABE (and all OBE algorithms) incorporates only data that are sufficiently innovative in a well-defined sense. Typically, only 10% of the data are actually used in the estimate. This fact, combined with the tracking capabilities noted above, offer interesting possibilities for efficient speech coding and compression. For the same reason, empirical quality of the OBE-ABE-based LP solution is often superior to that obtained from batch processing.

OBE-ABE solutions are more robust to measurement noise and other practical deviations from the presumed production model. The measurement noise issue is discussed in [9]. To the extent that the linear model (1) is accurate, the "true" $\{v_{n*}\}$ is more likely to be a noisy, "phase-scattered" version of such a pulse train in the voiced case to account for model errors and nonminimum-phase effects [6]. In either the voiced or unvoiced case, the speech signal might also be better-modeled by driving the system with an alternative input sequence such as a multi-pulse excitation [1]. One of the benefits of the OBE methods is that unbiased identification is not rigidly dependent upon the spectral properties (whiteness) of $\{v_{n*}\}$. In fact, OBE, and in particular OBE-ABE, algorithms operate according to a very different principle which requires sufficient visitation of the input to its amplitude extrema to effect convergence. Even a correlated input model is acceptable with certain restrictions. The convergence properties of OBE-ABE are rigorously understood with respect to conditions on $\{v_{n*}\}$ [9]. The adherence of $\{v_{n*}\}$ to the prescribed model form is of significantly less importance to good estimation performance than with batch LP methods.

These benefits will be more fully developed and discussed in future publications. The purpose of this paper is to introduce the OBE-ABE algorithm and to demonstrate its elementary application to speech modeling.

3. OBE-ABE ALGORITHM

With knowledge of energy bounds $v_{n*}^2 \leq \gamma_n$, it can be demonstrated that $\theta_* \in \Omega_n$ where $\Omega_n \subseteq \mathrm{R}^m$ is a hyperellipsoidal set based on the observations at time n, $\Omega_n \stackrel{\text{def}}{=} \left\{\theta | \left(\theta - \theta_n\right)^T \mathbf{P}_n^{-1} \left(\theta - \theta_n\right) < \kappa_n\right\}$ where the ellipsoid center, θ_n, and the defining matrix $\mathbf{P}_n$ are computed recursively using recursions (2)–(6) in Table 1. These recursions comprise a general OBE algorithm. The ellipsoid center θ_n is used as an estimator of the parameters θ_* at each n. For details see, for example, [4, 5, 8].

In LP identification of speech via OBE estimation, the choice of a proper bounding sequence $\{\gamma_n\}$ is both critical and difficult. Since precise bounds are unknown, and since underbounding is riskier, practitioners are conservative in applying bounds, "erring" on the side of overbounding. A conservative sequence of bounds $\{\gamma_n\}$ will assure a meaningful feasibility set of parameters Ω_n at each n in the sense that $\theta_* \in \Omega_n$. However, achieving an ultimately small set Ω_n depends explicitly on a set of very tight bounds [10]. This means that the estimator may be imprecise, even asymptotically, if the bounds are too "loose." Formally, it has been shown that if a *persistency of excitation* (PE) [11] condition holds, then the estimator of any OBE algorithm converges to a finite neighborhood of the true parameter vector θ_* [10]. However, consistency of the estimator (reduction of Ω_n to a point) results can be proved only when exact bounds (i.e., overbounds not exceeding the true bounds by more than an arbitrarily small $\epsilon > 0$) are employed for infinitely many samples. Thus, an OBE estimator is not guaranteed to be consistent when conservative bounds are used. OBE-ABE relaxes the requirement for precise *a priori* bounds.

OBE-ABE augments the conventional OBE iterations with a recursion for the estimation of the bound γ_n. The ABE step prevents potentially catastrophic bound violation by converging to the true (but unknown) bound from above, and results in faster convergence speed, as well as improved robustness to measurement noise [9].The ABE recursion is shown in Table 1 which summarizes OBE-ABE (quantities ϵ, N, M, and index J are described presently).

The novelty in OBE-ABE is based on recent work on the convergence behavior of general OBE algorithms [10]. OBE-ABE itself is guaranteed to converge consistently (either with probability one, or in the weaker probability sense) given a PE condition and one of several distribution conditions on $\{v_{n*}\}$ assuring sufficient amplitude visitation to the input bounds [9]. Therefore, in OBE-ABE, the usual OBE requirement of exact knowledge of the bound sequence $\{\gamma_n\}$, and the conventional model assumption of input "whiteness," are replaced by the much less demanding requirement of characterizing the "tail probabilities" of the random variables v_{n*} through the small numbers ϵ and δ. Roughly speaking, δ is the probability with which $\{v_{n*}\}$ visits ϵ-neighborhoods of the bounds. Practically, ϵ is simply taken as a small number as described below. The proofs of these convergence results are based on the notion that OBE will cease to update asymptotically if the bounds are overestimated. Therefore, if at some time n, the bounds are overestimated, there is guaranteed to be a future interval $\mathcal{I}$ of length N over which no updating takes place. Such an interval indicates the need to lower the bound which is done using the data at time $J = \arg\max_{n \in \mathcal{I}} \varepsilon_n^2$. If M denotes the number of data points available on the frame, then for a given interval size N, the number of bound updates on the frame is less than M/N. As $M \to \infty$ and $\epsilon \to 0$, the set of feasible solutions for the model will approach a single point with probability greater than $\left[1 - (1-\delta)^N\right]^{M/N}$ [9]. In practice, since M is finite, ϵ is first chosen to be a small positive number. Given this ϵ, a δ is obtained. We then choose a sufficiently large M and $N(= \sqrt{M}$, for example) such that the probability $[1 - (1-\delta)^N]^{\frac{M}{N}}$ is close to unity. These choices are illustrated in the experiments in Section 4.

4. LP ANALYSIS USING OBE-ABE

In this section we briefly demonstrate the application of OBE-ABE to stationary frames of speech.

Speech is a very dynamic signal. Even a short frame (e.g., 256 points) of speech can be regarded as quasi-stationary at best. Hence, any recursive algorithm must converge satisfactorily within short time frames. With the OBE-ABE algorithm, we obtain below comparable results to those obtained using conventional batch LP methods (autocorrelation method [6]) with all the advantages of OBE processing, but without the difficulties encountered in *ad hoc* attempts to apply OBE [2, 3].

A critical point is that, in the reported and similar experiments, optimal OBE-ABE selects only 12% of the data for updating the estimator on average in voiced cases and 8%

I. Initialization: 1. Set $\theta_0 = \mathbf{0}$, $\kappa_0 = \mu$, and $\mathbf{P}_0 = \frac{1}{\mu^2}\mathbf{I}$, where μ is a small number, typically 10^{-3}. **2.** Set $\gamma_0 =$ any overestimated bound. **3.** Choose ϵ (small positive number), N and M. (See discussion in example applications.)

II. Recursion: For $n = 1 : M$

If $c_n < 0$, execute recursions (2)–(6). (Note: c_n, α_n, and β_n are described in III below.)

$$G_n = \mathbf{x}_n \mathbf{P}_{n-1} \mathbf{x}_n \tag{2}$$

$$\varepsilon_n = y_n - \theta_{n-1}^T \mathbf{x}_n \tag{3}$$

$$\mathbf{P}_n = \frac{1}{\alpha_n}[\mathbf{P}_{n-1} - \frac{\beta_n \mathbf{P}_{n-1}\mathbf{x}_n\mathbf{x}_n^T\mathbf{P}_{n-1}}{\alpha_n + \beta_n G_n}] \tag{4}$$

$$\theta_n = \theta_{n-1} + \beta_n \mathbf{P}_n \mathbf{x}_n \varepsilon_n \tag{5}$$

$$\kappa_n = \alpha_n \kappa_{n-1} + \beta_n \gamma_n - \frac{\alpha_n \beta_n \varepsilon_n^2}{\alpha_n + \beta_n G_n}. \tag{6}$$

Otherwise, if a time interval $\mathcal{I}$ of length N over which $c_n \geq 0$ is found, set

$$\gamma_n = \{\gamma_{n-1} - d_J \text{ if } d_J > 0; \quad \gamma_{n-1} \text{ otherwise}\}, \tag{7}$$

$$\text{where } d_J = \kappa_{J-1} G_J / m - \epsilon(2\sqrt{\gamma_{n-1}} - \epsilon) \text{ and } J = \arg\max_{n \in \mathcal{I}} \varepsilon_n^2. \tag{8}$$

III. Optimization details: $\{\alpha_n\}$ and $\{\beta_n\}$ are positive weighting sequences determined by the particular OBE algorithm employed. In almost every OBE algorithm, the weights are chosen to minimize the "size" of Ω_n at each n. When such optimal weights do not exist, the updating need not take place. In the SM-SA algorithm (used in this paper), the volume of the ellipsoid (proportional to $\det \kappa_n \mathbf{P}_n$) is minimized at each iteration by letting $\beta_n = \lambda_n$, $\alpha_n = 1 - \lambda_n$, and seeking the optimal λ_n in light of the current measurements. If $c_n < 0$ [see (9) below], the optimal weight is given by the unique positive root of the quadratic equation in λ with coefficients (in descending power order)

$$\begin{aligned} a_n &= m\gamma_n - m\varepsilon_n^2 + mG_n^2\gamma_n - 2mG_n\gamma_n - \kappa_{n-1}G_n + \kappa_{n-1}G_n^2 + G_n\gamma_n - G_n^2\gamma_n - \varepsilon_n^2 G_n \\ b_n &= 2m\varepsilon_n^2 - 2m\gamma_n + 2mG_n\gamma_n + 2\kappa_{n-1}G_n - \kappa_{n-1}G_n^2 - G_n\gamma_n + \varepsilon_n^2 G_n \\ c_n &= m\gamma_n - m\varepsilon_n^2 - \kappa_{n-1}G_n. \end{aligned} \tag{9}$$

If $c_n \geq 0$, the optimal weight is zero.

Table 1. The OBE algorithm with automatic bound estimation (OBE-ABE).

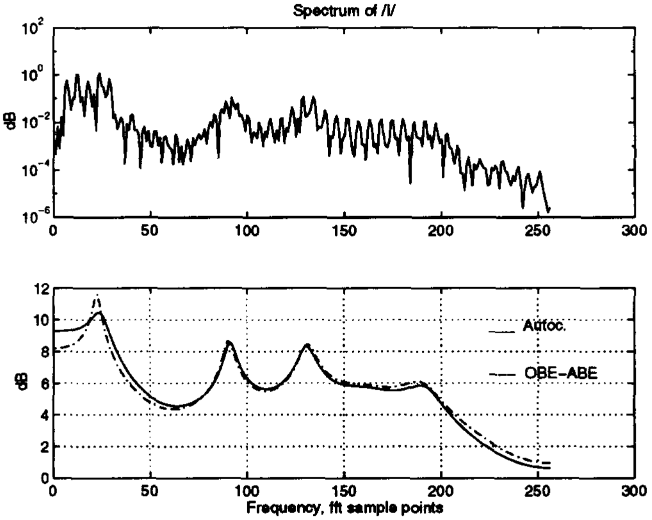

Figure 1. Spectra of voiced /ɪ/ phoneme from an utterance of "six" in the TIMIT database. A 256-point frame (rectangular window) is used in each case. Upper graph: FFT spectrum. Lower graph: OBE-ABE spectrum (dashed curve), and LP autocorrelation method spectrum (solid curve).

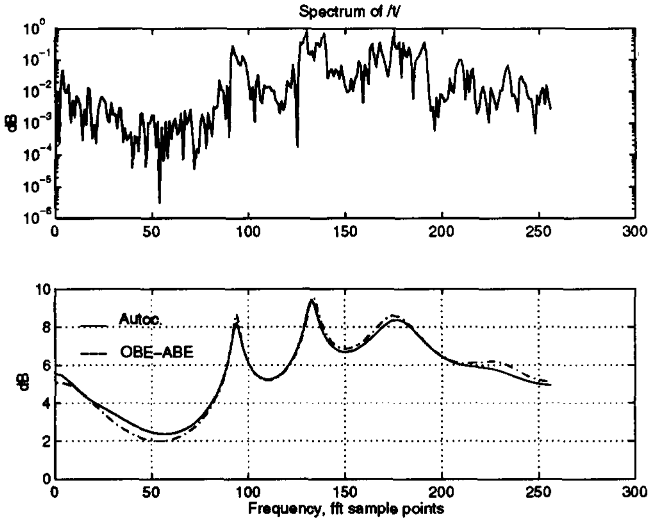

Figure 2. Results similar to those in Fig. 1 for the unvoiced /t/ phoneme.

for unvoiced. For suboptimal OBE-ABE, these results are 7% and 12%, respectively, so that the process is of $\mathcal{O}(m)$ complexity (see [9]). Earlier attempts to apply conventional OBE algorithms resulted in approximately 30% data selection in both cases [2, 3]. Further, in applying the OBE-ABE algorithm, the speech frames (voiced or unvoiced) are unwindowed and have a consistent length of 256 points. The LP model size ($m = 14$) also remains the same for both voiced and unvoiced cases. In the earlier work with conventional OBE, it was found necessary to increase the window size and decrease the model order for unvoiced speech (*c.f.* [3]). This makes the application of OBE-ABE straightforward, without requiring *a priori* knowledge of the voiced / unvoiced status of the frame.

As typical examples of stationary-frame analysis, we show results for the vowel /ɪ/ (voiced phoneme, from utterance "six") in Fig. 1 and for the unvoiced plosive /t/ (from "eight") in Fig. 2. The speech data are taken from the TIMIT database [7]. The upper portions of the figures show the spectra of the speech frames themselves based on an FFT. The lower portions show the smoothed spectra based on LP parameters obtained from the autocorrelation method [6] (solid line), and the OBE-ABE algorithm (dashed line). The OBE-ABE algorithm produces similar spectra to those of conventional batch method while using only a small fraction ($\sim$ 10%) of the data. Further, the analysis produces a feasible set of solutions (with 100% statistical confidence) that might be useful in certain applications. Results of processing the same two data frames with suboptimal OBE-ABE are similar.

In Sections 1 and 2, we have enumerated a number of benefits of OBE-ABE processing of speech, but the space available for these simple experiments with stationary speech does not permit extensive illustration of these advantages. These issues will be the subject of future publications. Notably, a modification of OBE-ABE to provide *adaptive* estimates to track speech parameters is described in [9].

REFERENCES

[1] B.S. Atal and J.R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," *Proc. ICASSP '82*, Paris, pp. 614-617, 1982.

[2] J.R. Deller, "Applications of OBE Algorithms to Speech Processing," Ch. 29 in M. Milanese *et al.* (Editors), *Bounding Approaches to System Identification*, London: Plenum, 1996.

[3] J.R. Deller and T.C. Luk, "LP analysis of speech based on set-membership theory," *Computer Speech and Language*, vol. 3, pp. 301-327, Oct. 1989.

[4] J.R. Deller, M. Nayeri and M.S. Liu, "Unifying the landmark developments in optimal bounding ellipsoid identification," *Int. J. Automatic Control and Signal Processing*, vol. 8, pp. 48-63, Jan.-Feb. 1994.

[5] J.R. Deller, M. Nayeri and S.F. Odeh, "Least square identification with error bounds for real-time signal processing and control," *Proc. IEEE*, vol. 81, pp. 813-849, June 1993.

[6] J.R. Deller, J.G. Proakis and J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, New York: Macmillan, 1993.

[7] W.M. Fisher, G.R. Doddington and K.M. Goudie, "The DARPA speech ... database," *Proc. DARPA Speech Recog. Workshop*, pp. 93-99, 1986.

[8] E. Fogel and Y.-F. Huang, "On the value of information in system identification – Bounded noise case," *Automatica*, vol. 18, pp. 229-238, 1982.

[9] T.M. Lin, *Optimal Bounding Ellipsoid Algorithms with Automatic Bound Estimation* (Ph.D. dissertation), Michigan State U., 1996.

[10] M. Nayeri, M.S. Liu and J.R. Deller, Jr. "Do interpretable OBE algorithms converge? Parts I & II," *Proc. 10th IFAC Symp. System Identification*, Copenhagen, vol. 3, pp. 389-400, July 1994.

[11] T. Söderström and P. Stoica, *System Identification*, Englew'd Cliffs, New Jersey: Prentice-Hall, 1989.