GRAPH LINEAR PREDICTION RESULTS IN SMALLER ERROR THAN STANDARD LINEAR PREDICTION

Arun Venkitaraman, Saikat Chatterjee, and Peter Händel

Department of Signal Processing, ACCESS Linnaeus Centre,
KTH Royal Institute of Technology Stockholm
arunv@kth.se, sach@kth.se, ph@kth.se

ABSTRACT

Linear prediction is a popular strategy employed in the analysis and representation of signals. In this paper, we propose a new linear prediction approach by considering the standard linear prediction in the context of graph signal processing, which has gained significant attention recently. We view the signal to be defined on the nodes of a graph with an adjacency matrix constructed using the coefficients of the standard linear predictor (SLP). We prove theoretically that the graph based linear prediction approach results in an equal or better performance compared with the SLP in terms of the prediction gain. We illustrate the proposed concepts by application to real speech signals.

Index Terms— Linear prediction, Graph signal processing, autoregressive model.

1. INTRODUCTION

Many naturally occurring and synthetic-made signals exhibit a high degree of correlation over consecutive samples. For simplicity and compactness of representation, it is a typical practice to model such a signal as an outcome of an autoregressive (AR) process. A signal x generated by a L^{th} order AR model is expressed as:

$$x(n) = \sum_{i=1}^{L} a_i x(n-i) + e(n), \quad n = 1, 2, \cdots, N \qquad (1)$$

where $x(n)$ denotes the signal sample at the n^{th} instant, a_i, the i^{th} AR coefficient, and $e(n)$, the additive white noise. For brevity and ease of description, we use the terms 'instant' or 'time' to refer to the independent variable in this paper, though the analysis does not assume time-series data. The linear estimate or linear predictor (LP) of $x(n)$ from its past P samples is given by [1]

$$\hat{x}(n) = \sum_{i=1}^{P} \hat{a}_i x(n-i), \quad n = 1, 2, \cdots, N \qquad (2)$$

The authors would like to acknowledge the support received from the Swedish Research Council.

where $\hat{a}_i$s denote the predictor coefficients obtained by minimization of the squared prediction error $\xi(\mathbf{x}, \mathbf{a}) = \sum_n \left(x(n) - \hat{x}(n)\right)^2$, $\mathbf{x} = [x(1), x(2), \cdots, x(N)]^T \in \mathbb{R}^N$, and $\mathbf{x}^T$ denotes the transpose of $\mathbf{x}$. Approaches that incorporate statistics of the data explicitly often involve minimization of the expected value of ξ [2, 3]. We shall, however, consider the more general data-driven version of the LP obtained by minimization of ξ directly over the parameter vector $\mathbf{a}$ and signal $\mathbf{x}$.

Linear prediction finds applications in a wide range of domains such as geophysics, speech coding, image coding, and neurophysics, among others. We refer to the article by Makhoul [1] for an exhaustive review of the techniques and applications of linear prediction. Linear prediction with time-varying predictor coefficients have also been proposed for analysis of nonstationary signals [4, 5, 6]. LP models that incorporate dimensionality and sparsity constraints have also been considered [7, 8]. Recently, Sandryhaila and Moura introduced the notion of linear prediction for signals defined over connected graphs and demonstrated its application in the analysis of sensor network data [9]. Motivated by this idea, we propose a specific graph LP approach for analysis of standard AR signals (signals represented as in (1)) using a graph adjacency matrix constructed from the coefficients obtained from the standard linear predictor. We prove theoretically that the proposed graph linear predictor results in smaller prediction error than the standard predictor. We validate the theory by application of the proposed approach to speech signals.

1.1. Related work

Signal processing on graphs has gained considerable interest from the research community recently. The framework deals with the analysis of signals defined over the nodes of a connected graph, with particular focus on defining and extending many of the foundational discrete signal processing concepts such as shift-invariant filtering, convolution, Fourier transform, modulation analysis, wavelet transforms, to non-Euclidean geometries [10, 11, 12, 13, 14]. Such an analysis seems particularly well-suited to the present scenario where the applications involving large connected datasets is on the

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

increase.

The techniques for signal processing on graphs may be grouped under two different approaches. The first approach is based on the graph Laplacian, whose spectrum forms the building block for the definition of spectral operations such as the Fourier transform and in modulation analysis. Shuman et al. present a detailed review of this approach in [15]. The second approach takes the graph adjacency matrix as its foundational unit, using it to define operations such as graph shift, graph Fourier transform, frequency ordering, and is based on the concepts of algebraic signal processing [16]. Sandryhaila and Moura discuss this approach in great depth in their recent article [17]. We follow the second approach as advocated in [9, 17].

2. PRELIMINARIES

2.1. Graph signal processing

We briefly review some of the basic building blocks of signal processing on graphs. We denote the graph by $G = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ and $\mathbf{A} \in \mathbb{R}^{N\times N}$ denote the vertices/nodes and the weighted adjacency matrix, respectively. The i^{th} graph-shift of a graph signal $\mathbf{x} \in \mathbb{R}^N$ is defined as [9, 17]:

$$\tilde{\mathbf{x}}_i = \mathbf{A}^i\,\mathbf{x}.$$

The Q^{th} order graph linear predictor of $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$, is defined as linear combination of Q graph-shifts of $\mathbf{x}$, that is,

$$\hat{\mathbf{x}} = \sum_{i=1}^{Q} h_i \tilde{\mathbf{x}}_i = \sum_{i=1}^{Q} h_i \mathbf{A}^i\,\mathbf{x} = \mathbf{B}\,\mathbf{h},$$

where $\mathbf{B} = [\mathbf{b_1}\mathbf{b}_2\cdots\mathbf{b}_Q]$, $\mathbf{b_i} = \tilde{\mathbf{x}_i} = \mathbf{A}^i\mathbf{x}$ and $\mathbf{h} = [h_1, h_2, \cdots, h_Q]^T \in \mathbb{R}^N$. The optimal predictor coefficients are obtained by minimizing squared error $\xi_g(\mathbf{x}, \mathbf{h}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 = \|\mathbf{x} - \mathbf{B}\,\mathbf{h}\|_2^2$. The solution is obtained as

$$\begin{aligned}\mathbf{h}_* &= \arg\min_{\mathbf{h}} \xi_g(\mathbf{x}, \mathbf{h}) = \arg\min_{\mathbf{h}} \|\mathbf{x} - \mathbf{B}\,\mathbf{h}\|_2^2 \\ &= \mathbf{B}^\dagger\mathbf{x} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x}. \end{aligned} \quad (3)$$

Typically, Q is much smaller than N.

2.2. Standard linear prediction

The squared error for the standard LP is given by

$$\xi(\mathbf{x}, \mathbf{a}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2,$$

where $\hat{x}_n$ obtained using (2). Let us define the matrix $\mathbf{X} \in \mathbb{R}^{N\times P}$ as follows:

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ x(1) & 0 & 0 & \dots & 0 \\ x(2) & x(1) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(N-1) & x(N-3) & x(N-3) & \dots & x(N-P) \end{bmatrix}$$

Then, the prediction $\hat{\mathbf{x}}$ can be written as $\hat{\mathbf{x}} = \mathbf{X}\mathbf{a}$, where $\mathbf{a} = [a_1, a_2, \cdots, a_P]^T$, or equivalently,

$$\xi(\mathbf{x}, \mathbf{a}) = \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_2^2,$$

and the optimal predictor coefficients are obtained as:

$$\mathbf{a}_* = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\,\mathbf{a}\|_2^2 = \mathbf{X}^\dagger\mathbf{x} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x} \quad (4)$$

Let us denote the optimal predictor by $\mathbf{x}_* = \mathbf{X}\,\mathbf{a}_*$. The corresponding minimum squared error is given by $\xi(\mathbf{x}_*, \mathbf{a}_*)$. In general, $\mathbf{e} = \mathbf{x} - \mathbf{x}_* \neq \mathbf{0}$ and $\xi(\mathbf{x}_*, \mathbf{a}_*) > 0$. In order that the system is identifiable and stable we make the following assumption:

Assumption 1. *$x_1 \neq 0$ and $a_{1*} \neq 0$.*

3. GRAPH LINEAR PREDICTION

Let us now consider $\mathbf{x}$ as a signal over a given graph $G = (\mathcal{V}, \mathbf{A})$. By applying (3) on $\mathbf{x}$, we get that the optimal GLP coefficients are given by:

$$\mathbf{h}_* = \arg\min_{\mathbf{h}} \|\mathbf{x} - \mathbf{B}\,\mathbf{h}\|_2^2 = \mathbf{B}^\dagger\mathbf{x}, \quad (5)$$

and the minimum squared error is given by $\xi_g(\mathbf{x}_{g*}, \mathbf{h}_*)$. The optimal graph prediction is given by $\mathbf{x}_{g*} = \mathbf{B}\,\mathbf{h}_*$. In general it is reasonable to assume that $\mathbf{e}_g = \mathbf{x} - \mathbf{x}_{g*} \neq \mathbf{0}$, that is, the predicted values do not exactly coincide with the original signal.

Goal: To design an adjacency matrix $\mathbf{A}$ such that $\xi_g(\mathbf{x}_{g*}, \mathbf{h}_*) \leq \xi(\mathbf{x}_*, \mathbf{a}_*)$.

Consider the following adjacency matrix constructed using the optimal P^{th} order SLP coefficient vector $\mathbf{a}_*$:

$$\mathbf{A}_* = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ a_{1*} & 0 & \dots & 0 & 0 & \dots & 0 \\ a_{2*} & a_{1*} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ a_{P*} & a_{P-1*} & \dots & a_{1*} & 0 & \dots & 0 \\ 0 & a_{P*} & & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 \\ 0 & \dots & a_{P*} & a_{P-1*} & \dots & a_{1*} & 0 \end{bmatrix} \quad (6)$$

We note that $\mathbf{x}_* = \mathbf{X}\mathbf{a}_* = \mathbf{A}_*\mathbf{x}$, and that by construction, the maximum rank of $\mathbf{A}_*$ is $N-1$ and hence, $\mathbf{x} \notin \text{span}(\mathbf{A}_*)$ ($\mathbf{x}$ being an arbitrary N-dimensional vector). We also note that for the adjacency matrix $\mathbf{A}_*$, the corresponding $\mathbf{B}$ matrix has columns given by $\mathbf{b}_k = \mathbf{A}_*^k\mathbf{x}$ for $1 \leq k \leq Q$, and $\mathbf{b}_1 = \mathbf{A}_*\mathbf{x} = \mathbf{x}_*$.

Lemma 1. *For the adjacency matrix $\mathbf{A}_*$ constructed as in (6), $\mathbf{B}$ has full column rank.*

Proof. Let us denote $\mathbf{b}_k = [b_{1,k}, b_{2,k}, \ldots, b_{N,k}]^T$. Then, we have $\mathbf{b}_1 = \mathbf{A}_*\mathbf{x} = [0, b_{2,1}, \ldots, b_{N1}]^\top$ and $\mathbf{b}_2 = \mathbf{A}_*^2\mathbf{x} = \mathbf{A}_*\mathbf{b}_1 = [0, 0, b_{3,2} \ldots, b_{N,2}]^\top$. In general, we have that the first k elements of $\mathbf{b}_k$ are equal to zero, and therefore, $\mathbf{B}$ is expressible as follows:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ b_{2,1} & 0 & 0 & \ldots & 0 & 0 \\ b_{3,1} & b_{3,2} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & b_{Q,Q-1} & 0 \\ b_{Q,1} & \vdots & \vdots & \vdots & & b_{Q+1,Q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{N,1} & b_{N,2} & \vdots & \vdots & \vdots & b_{N,Q} \end{bmatrix}, \quad (7)$$

where $b_{2,1} = a_{1*}x_1$, $b_{3,2} = a_{1*}^2 x_1$, and finally $b_{Q+1,Q} = a_{1*}^Q x_1$. Using Assumption 1, we have that $b_{2,1} \neq 0$, $b_{3,2} \neq 0$, leading to $b_{Q+1,Q} \neq 0$, which in turn shows that $\mathbf{B}$ has full column rank. □

We next show that the error from the GLP of $\mathbf{x}$ using $\mathbf{A}_*$ and $\mathbf{h}_*$ is always bounded from above by the error from the SLP, which is formally stated in the following theorem.

Theorem 1. *For the adjacency matrix $\mathbf{A}_*$, we have*

$$\xi_g(\mathbf{x}_{g*}, \mathbf{h}_*) \leq \xi(\mathbf{x}_*, \mathbf{a}_*),$$

that is, GLP error energy is smaller than that of SLP.

Proof. $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_Q] = [\mathbf{x}_*, \mathbf{b}_2, \ldots, \mathbf{b}_Q]$. As $\mathbf{B}$ has full column rank, by the least squares principle we have that $\xi_g(\mathbf{x}_{g*}, \mathbf{h}_*) = \|\mathbf{x} - \mathbf{B}\mathbf{h}_*\|^2 = \|\mathbf{x} - [\mathbf{x}_*, \mathbf{b}_2, \ldots, \mathbf{b}_Q]\mathbf{h}_*\|^2 \leq \|\mathbf{x} - \mathbf{x}_*\|^2 = \xi(\mathbf{x}_*, \mathbf{a}_*)$. □

The inclusion of Q graph filter taps gives additional degrees of freedom in the minimization of ξ_g using same information as in the SLP, namely, the signal $\mathbf{x}$. Thus, the GLP may be viewed as an approximation to a higher order SLP, particularly for order since the first filter tap has already been optimized using the SLP). In this formulation, we implicitly assume that the underlying AR order L is typically much larger than P and Q.

4. EXPERIMENTS

We apply the proposed GLP to speech samples taken from the CMU Arctic database [18] for different values of P and Q. We compare the performance with SLP of order P, with GLP of varying graph orders Q. The sampling rate of the speech samples is 16 kHz. We use 10ms window frames for analysis. For each frame, we compute the prediction gain defined as $10 \log \left(\frac{P_x}{P_{x-\hat{x}}} \right)$, where $P_x = \|\mathbf{x}\|_2^2$ and $P_{x-\hat{x}} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$.

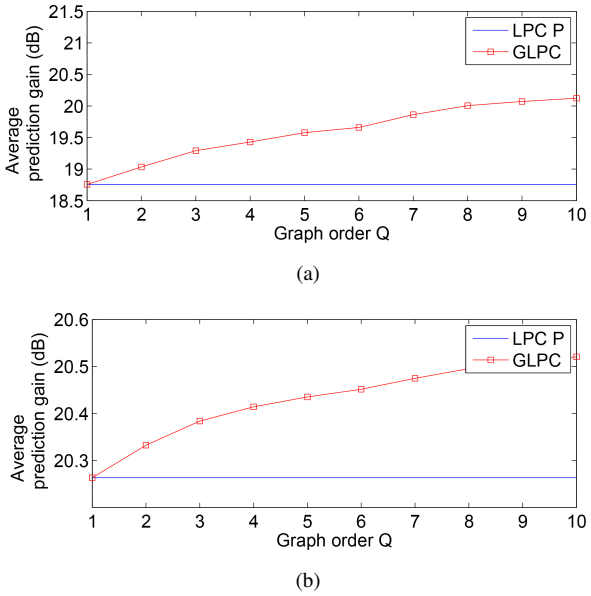


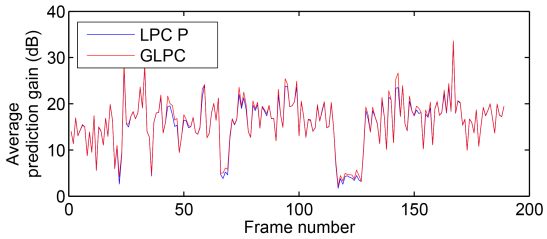
Fig. 1. Average prediction gain for different graph orders (a) $P = 3$, and (b) $P = 8$.

In Figure 1, we show the prediction gain as a function of the graph order Q, for $P = 3$ and $P = 8$, corresponding to a male speech sample. The gain is obtained by averaging over all the signal frames. We observe that the GLP performs consistently better than the P^{th} order SLP as expected. We also observe that the increase in prediction gain from the GLP reduces from $P = 3$ to $P = 8$. This is so because the coefficients of higher order SLP capture most of the signal information and performance is not expected to improve greatly with additional degrees of freedom obtained from the GLP.

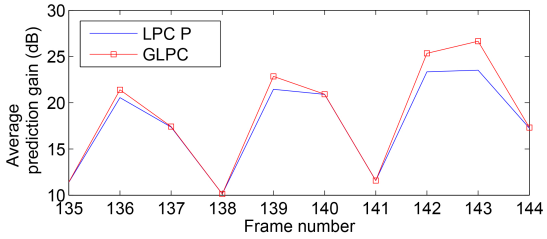
In Figures 2 and 3, we show the predictor gains for different frames of the signal considered in Figure 1, for particular values of P and Q. We also show the zoomed-in plots around a frame range where the performance difference can be seen easily. As before, the increase in prediction gain reduces as we go from $P = 3$ to $P = 8$. Similar performance curves are obtained for a large number of speech samples, for both male and female speakers. We show here a particular example to illustrate the general trend. As we have discussed in the earlier sections, our approach is valid for general AR models and the choice of examples from speech data is for the purposes of illustration only.

5. CONCLUSIONS

We proposed a graph linear predictor for the analysis of signals generated from AR models. The analysis was carried out using a graph adjacency matrix constructed from the coefficients of standard linear predictor. We proved theoretically that the proposed approach results in an equal or better perfor-

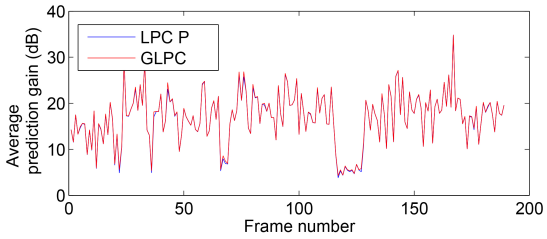


(a)

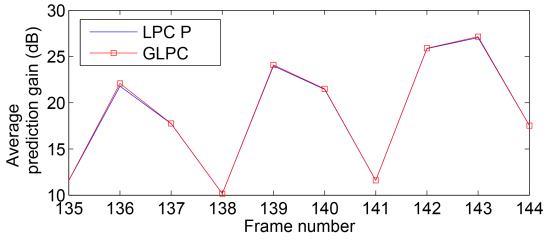


(b)

Fig. 2. Prediction gain for $P = 3$ and $Q = 3$ (a) Complete signal, and (b) Zoomed-in plot of (a).



(a)



(b)

Fig. 3. Prediction gain for $P = 8$ and $Q = 3$ (a) Complete signal, and (b) Zoomed-in plot of (a).

mance in comparison with the standard linear predictor, while using the same prediction coefficients. The graph predictor so obtained may be viewed as approximating the performance of higher order standard linear predictor. Experiments were performed on speech data to demonstrate the validity of the proposed theory.

The proposed approach is a specific case of graph linear prediction for AR models. The choice of the graph order and the prediction order both decide the performance of the graph predictor. A large Q results in better fit, at the cost of fitting the noise as well. Further analysis is necessary in order to ascertain the nature of the trade-off between P and Q in the prediction performance. It would also be interesting to investigate if similar philosophy could be applied to more general signal models. We intend to explore along these lines in the near future.

REFERENCES

[1] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, April 1975.

[2] P. P. Vaidyanathan, *The theory of linear prediction*, Morgan & Claypool, 2008.

[3] L.R. Rabiner and R.W. Schafer, "Introduction to digital speech processing," *Foundations and Trends in Signal Processing*, vol. 1, no. 1, pp. 1–194, 2007.

[4] D. Rudoy, T.F. Quatieri, and P.J. Wolfe, "Time-varying autoregressions in speech: Detection theory and applications," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 977–989, May 2011.

[5] K. Schnell and A. Lacroix, "Time-varying linear prediction for speech analysis and synthesis," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 3941–3944.

[6] M. Hall, A. V. Oppenheim, and A. Willsky, "Time-varying parametric modeling of speech," in *IEEE Conference on Decision and Control including the 16th Symposium on Adaptive Processes and A Special Symposium on Fuzzy Set Theory*, Dec 1977, pp. 1085–1091.

[7] D. Giacobello, M.G. Christensen, M.N. Murthi, S.H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1644–1657, July 2012.

[8] D. Giacobello, M.G. Christensen, T.L. Jensen, M.N. Murthi, S.H. Jensen, and M. Moonen, "Stable 1-norm error minimization based linear predictors for speech modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 5, pp. 912–922, May 2014.

[9] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, April 2013.

[10] W. Raymond, H. Choi, R. G. Baraniuk, and D. Veronique, "Distributed wavelet transform for irregular sensor network grids," in *IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, Jul. 2005, pp. 1196–1201.

[11] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 53 – 94, 2006, Special Issue: Diffusion Maps and Wavelets.

[12] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129 – 150, 2011.

[13] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*, March 2012, pp. 3921–3924.

[14] S. K. Narang and A. Ortega, "Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4673–4685, 2013.

[15] D. I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[16] M. Puschel and J.M.F. Moura, "Algebraic signal processing theory: Foundation and 1-d time," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3572–3585, Aug 2008.

[17] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, Sept 2014.

[18] J. Kominek, A. W. Black, and V. Ver, "Cmu arctic databases for speech synthesis," Tech. Rep., 2003.