SIGNAL MODELING WITH NON-UNIFORM TOPOLOGY LATTICE FILTERS

Sacha KRSTULOVIĆ

IDIAP C.P. 592 - CH-1920 Martigny - Switzerland
sacha@idiap.ch

Frédéric BIMBOT

IRISA - Campus Beaulieu, 35042 Rennes - France
bimbot@irisa.fr

ABSTRACT

This article presents a new class of constrained and specialized Auto-Regressive (AR) processes. They are derived from lattice filters where some reflection coefficients are forced to zero at a priori locations. Optimizing the filter topology allows to build parametric spectral models that have a greater number of poles than the number of parameters needed to describe their location. These NUT (Non-Uniform Topology) models are assessed by evaluating the reduction of modeling error with respect to conventional AR models.

1. INTRODUCTION

Lattice filters are a well-known signal analysis and coding tool. Their parameters, the reflection coefficients, have a good robustness to noise and quantization effects [1]. These filters also present a formal analogy with the process of wave propagation into lossless discrete acoustic tube models (possibly used as vocal tract models) [2]. But they don't incorporate any other a priori knowledge about the process they represent. For instance, it is classically implied that the individual portions forming a discretized tube all have a unit length, whereas it may be more accurate to represent a priori knowledge about unequally spaced tube interfaces.

By generalizing the lattice formalism to the case of tube portions with any length, this article defines a class of processes, called Non-Uniform Topology (NUT) lattice processes, that represent a constrained case of Auto-Regressive (AR) filtering. Section 2 is dedicated to the description of their formalism and general properties. Section 3 deals with the estimation of their parameters for signal analysis. Section 4 exposes experimental results that assess the spectral modeling accuracy of this new model.

2. NON-UNIFORM TOPOLOGY LATTICES

2.1. Basic principle

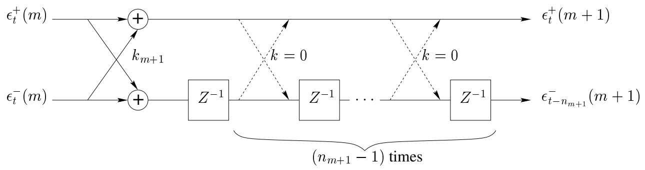
The transfer function $H(z) = \frac{1}{A_M(z)} = \frac{1}{\sum_{i=0}^{M} a_i z^{-i}}$ associated with an M^{th} order Auto-Regressive (AR) signal model can be built recursively by application of the following matrix recursion [2]:

$$\begin{aligned} \begin{bmatrix} A_{m+1}(z) \\ B_{m+1}(z) \end{bmatrix} &= \begin{bmatrix} 1 & k_{m+1} \\ k_{m+1}z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} A_m(z) \\ B_m(z) \end{bmatrix} \\ \begin{bmatrix} A_0(z) \\ B_0(z) \end{bmatrix} &= \begin{bmatrix} 1 \\ -z^{-1} \end{bmatrix} \end{aligned} \qquad (1)$$

where:

- $A_m(z)$ is the transfer function of an m^{th} order forward predictor, modeling the current sample as a linear combination of $m-1$ past samples;
- $B_m(z)$ is the transfer function of an m^{th} order backward predictor, modeling the m^{th} past sample as a linear combination of $m-1$ future samples;
- k_{m+1} is the reflection coefficient allowing to grow the predictors from order (m) to order $(m+1)$.

Suppose that from step m of this recursion, a known number $(n_{m+1}-1)$ of the reflection coefficients following k_{m+1} are fixed to zero[1]. The equivalent lattice flow chart looks like:



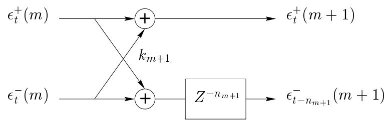
and the corresponding portion of the matrix recursion becomes:

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}}_{(n_{m+1}-1) \text{ times}} \begin{bmatrix} 1 & k_{m+1} \\ k_{m+1}z^{-1} & z^{-1} \end{bmatrix} \qquad (2)$$

which can be compacted into a single matrix:

$$\begin{bmatrix} 1 & k_{m+1} \\ k_{m+1}z^{-n_{m+1}} & z^{-n_{m+1}} \end{bmatrix} \qquad (3)$$

This matrix describes an inverse filtering cell of the form:



Hence, if the recursion steps are re-numbered from 1 to the number of non-zero reflection coefficients (i.e. if the steps with null reflection coefficients are ignored in the indexing), the whole matrix recursion can be rewritten as:

$$\begin{aligned} \begin{bmatrix} A_{m+1}(z) \\ B_{m+1}(z) \end{bmatrix} &= \begin{bmatrix} 1 & k_{m+1} \\ k_{m+1}z^{-n_{m+1}} & z^{-n_{m+1}} \end{bmatrix} \begin{bmatrix} A_m(z) \\ B_m(z) \end{bmatrix} \\ \begin{bmatrix} A_0(z) \\ B_0(z) \end{bmatrix} &= \begin{bmatrix} 1 \\ -z^{-n_0} \end{bmatrix} \end{aligned} \qquad (4)$$

where the delays $z^{-n_{m+1}}$ can be of any order greater than or equal to 1 for each step of the recursion.

This work is supported by the Swiss National Science Foundation, grant nr. 20-55.634.98 for the ARTIST II project.

[1] In an acoustic tube model, this would correspond to connecting n_{m+1} elementary tube portions that have an equal cross-section [3].

Method	Error criterion	Estimator for k_{p+1}
Forward	$\xi^2(p+1) = \sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p+1)^2$	$k_{p+1} = \frac{-\sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p)\,\epsilon_{t-n_p}^-(p)}{\sum_{t=\Sigma_p+1}^{N} \left(\epsilon_t^+(p)\right)^2}$
Backward	$\xi^2(p+1) = \sum_{t=\Sigma_p+1}^{N} \epsilon_t^-(p+1)^2$	$k_{p+1} = \frac{-\sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p)\,\epsilon_{t-n_p}^-(p)}{\sum_{t=\Sigma_p+1}^{N} \left(\epsilon_{t-n_p}^-(p)\right)^2}$
Burg	$\xi^2(p+1) = \frac{1}{2}\left\{\sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p+1)^2 + \sum_{t=\Sigma_p+1}^{N} \epsilon_t^-(p+1)^2\right\}$	$k_{p+1} = \frac{-2\sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p)\,\epsilon_{t-n_p}^-(p)}{\sum_{t=\Sigma_p+1}^{N} \left(\epsilon_t^+(p)\right)^2 + \sum_{t=\Sigma_p+1}^{N} \left(\epsilon_{t-n_p}^-(p)\right)^2}$
Itakura-Saito		$k_{p+1} = \frac{-\sum_{t=\Sigma_p+1}^{N} \epsilon_t^+(p)\,\epsilon_{t-n_p}^-(p)}{\sqrt{\sum_{t=\Sigma_p+1}^{N} \left(\epsilon_t^+(p)\right)^2 \sum_{t=\Sigma_p+1}^{N} \left(\epsilon_{t-n_p}^-(p)\right)^2}}$

Table 1. Various estimators for the reflection coefficients of inverse NUT lattice filters, where Σ_p denotes the sum of all the delays from order 1 to order p.

2.2. Constrained Linear Prediction

Generally speaking, all the relations that describe the mathematics of standard lattice filters are still valid in the framework of NUT lattices : they will only undergo formal modifications due to the inclusion of zero-values at particular places. For instance, the transfer function $A_M(z)$ of the forward-error filter remains a polynom in z^{-1}. Similarly, the backward predictor $B_m(z)$ can be deduced from the forward predictor $A_m(z)$, using the expression :

$$B_m(z) = -z^{-\sum_{i=0}^{m} n_i} A_m(1/z) \tag{5}$$

The forward predictor's growth can thus still be formalized as :

$$A_{m+1}(z) = A_m(z) - k_{m+1} B_m(1/z) \tag{6}$$

Nevertheless, the inclusion of the a priori null values introduces interesting structural constraints to the Linear Prediction modeling method.

Some of these constraints appear when computing the prediction coefficients $a_i^{(m)}$ from the reflection coefficients k_i. This can be done through the classical Levinson procedure [1], but including the a priori null k_i values at the relevant iterations. This procedure is described by :

$$
\begin{aligned}
E^{(0)} &= R_0 && (7)\\
k_{m+1} &= -\left[R_{m+1} + \sum_{i=1}^{m} a_i^{(m)} R_{m+1-i}\right] / E^{(m)} && (8)\\
a_{m+1}^{(m+1)} &= k_{m+1} && (9)\\
a_i^{(m+1)} &= a_i^{(m)} + k_{m+1}\, a_{m+1-i}^{(m)}\,;\ \ i = 1, \cdots, m && (10)\\
E^{(m+1)} &= \left(1 - k_{m+1}^2\right) E^{(m)} && (11)
\end{aligned}
$$

where R_m are the values of the autocorrelation function. Forcing $k_{m+1} = 0$ at step $(m+1)$ has the following effects :

- from equations (9) and (10), it simply means that the predictor has not changed between step (m) and step $(m+1)$;
- from equation (11), it means that the energy of the prediction error stays the same;
- from equation (8), setting $k_{m+1} = 0$ induces :

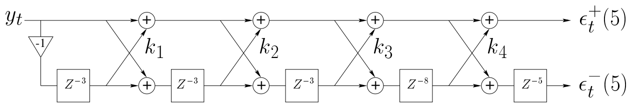
$$R_{m+1} = \sum_{i=1}^{m} a_i^{(m)} R_{m+1-i} \tag{12}$$

The last effect deserves particular attention, because it represents a way of constraining the autocorrelation function : the value for R_{m+1} is turned into a linear combination of the previously considered autocorrelation coefficients. This can be put in parallel with the fact that correlation between the forward and backward prediction errors is created only for some particular lags, i.e. those where the reflections coefficients are not constantly null. Consequently, the corresponding power spectral density contains some "genuine" energy peaks together with peaks resulting from harmonic combinations. Spectral modeling with a NUT lattice is therefore more specialized than modeling with an unconstrained Auto-Regressive production models, since it accounts precisely for frequency combinations that comply with the underlying generative model.

From equation (4), one can also remark that the global order of $A_M(z)$ is equal to the sum of the various delays $n_{m,\ m=0,\cdots,M-1}$. In the classical case, where $n_m = 1\ \forall m$, the global order is equal to the number of reflection coefficients. Conversely, in the NUT lattice case, the global order can be greater than the number of unconstrained reflection coefficients.

As a matter of fact, the reflection coefficients represent some intrinsic degrees of freedom (DoFs) for the equivalent linear predictor. Constraining some of them to be zero-valued amounts to reducing the intrinsic number of DoFs without changing the global order. Hence, the corresponding spectral model contains a number of poles greater than the number of parameters needed to describe their location. Alternately, a signal sample can be predicted from an increased portion of its past if the number of DoFs is kept fixed while the global order is grown.

In the following, the various lattice configurations will be identified by strings starting with the number of delay blocks expressed over the number of spanned unit delays, and followed by the enumeration of their lengths. An example would be : [5/22:3x3,8,5.], which reads : "a NUT lattice with 5 cells spanning 22 unit-delays, and which has three 3^{rd} order delays, one 8^{th} order delay and one 5^{th} order delay"[2]. The corresponding flow chart would look like :



[2]This would be equivalent to a lossless acoustic tube model made of 5 unequal-length sections distributed over 22 unit sections. See [4] for more details about the equivalence between non-uniform tubes and lattice filters.

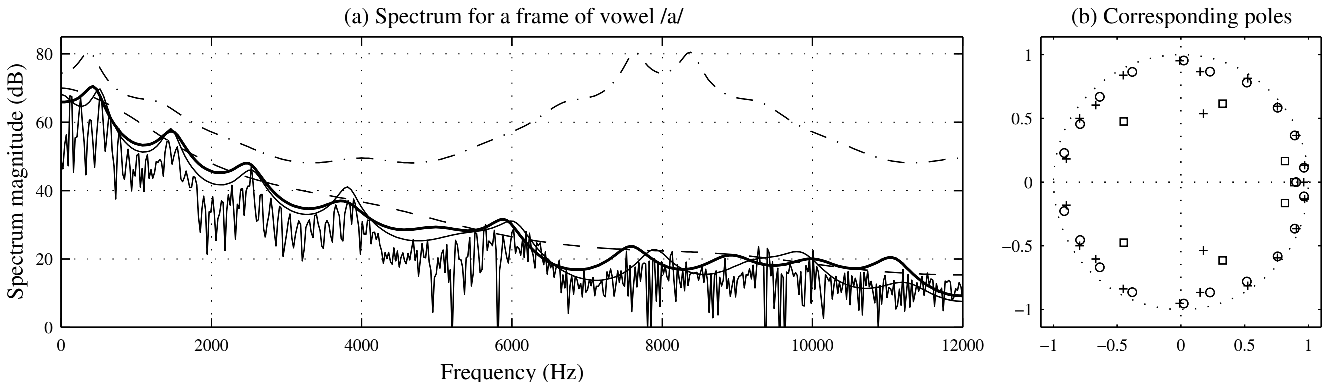


Fig. 1. *(a)* Spectrum of a frame of vowel /a/ @ 24kHz. Superimposed over the FFT: 23^{rd} order LPC [24/24:24x1.] (continuous line); optimal NUT lattice w/ 8DoFs [8/24:1,1,1,5,4,4,5,3.] (bold line) ; 7^{th} order LPC [8/8:8x1.] (dashed); "random" NUT lattice w/ 8DoFs [8/24:8x3.] (dash-dot). - *(b)* Corresponding poles for [8/8] (□), optimal [8/24] (o) and [24/24] (+).

3. MODELING METHOD

3.1. First optimization level: constrained estimators

The non-null k_i coefficients of the NUT lattices can be estimated by accounting for the non-uniform delays into the classical [5] lattice-based estimators. The result is given in table 1. The modified forward, backward and Burg estimators can be derived analytically by differentiating the error criterion ξ^2 and equating the result to zero. The modified Itakura estimator is the geometric mean of the reflection coefficients found by the forward and backward methods. Following the remarks made in section 2.2, it can be observed that imposing non-uniform delays modifies the lag and the summation boundaries considered into the partial correlation measures that define the k_i coefficients.

The stability of the constrained filters is preserved since forcing some reflection coefficients to zero respects the general stability condition for a lattice filter [5], namely $|k_i| < 1 \ \forall i$ (every k_i should have a value between -1 and 1). Furthermore, it can be easily verified that the modified Burg and Itakura estimators always generate values that lie between -1 and 1.

3.2. Second optimization level: optimal filter topology

Various repartitions of delays lead to different inverse filtering performances in terms of a higher or lower residual error ξ^2 for a signal frame. It is therefore interesting to find the best performing topology given a number of degrees of freedom to be distributed over a given global order, i.e. to find the best match in the set of NUT production processes that respect the two specifications.

To search for the best configuration, all the filters in the set are generated and systematically used to inverse-filter a test frame. The one bringing the least residual error is regarded as the best topology. Figure 2 shows that this search plays a significant role in the accuracy of the model. Random configurations (dark bars) perform significantly worse than optimal ones (light bars).

Further constraints, such as a minimum delay order, can be imposed to the production process to make the number of tested filters more tractable (at the price of a reduced modeling accuracy [3]). For instance, in the case of an [8/32] constraint, 2'629'575 filters have to be tested. Imposing the minimum delay to be no shorter than 2 units reduces this number to 245'157 filters.

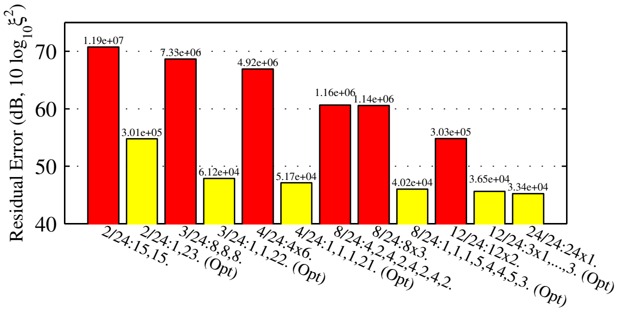


Fig. 2. Comparison of Log Residual Errors for various NUT lattice models (for a frame of vowel /a/ @ 24kHz). The original value of ξ^2 is indicated above the bars.

4. EXPERIMENTAL RESULTS

The results presented in this paper have been computed from signal frames extracted from a test sentence in French, spoken by a male speaker, recorded in a very quiet environment and sampled at 32kHz. When needed, the frames have been down-sampled to 24kHz or 8kHz using the polyphase method. Throughout the experiments, the modified Burg estimator and corresponding error criterion have been used.

Dependency to the signal - Table 2 gives the optimal filter configurations found for frames of various vowels with a [13/24] constraint. The configurations are naturally frame-dependent. As a follow-up, it would be interesting to check the stability of the topology optimization scheme across vowel classes.

Spectral shapes - Spectral shapes are computed from NUT lattices by evaluating the corresponding constrained All-Pole transfer function over the unit circle. The spectral shape obtained for a frame of /a/ is shown in figure 1(a). Again, it is clear that the topology optimization stage helps minimizing the spectral distortion induced by the reduction of the number of DoFs. In the optimal case, this distortion stays acceptable for low frequencies, in the sense that the first formants are reasonably well captured. This is confirmed by inspection of the pole locations given in figure 1(b), and has been observed for all the studied vowels.

Filter accuracy versus number of DoFs - Figure 3 compares the residual error in regular lattices and NUT lattices of type [DoFs/24] as a function of the number of DoFs. It shows that opti-

Vowel	Optimal 12 DoFs NUT lattice
/a/	13/24:3x1,4,2x1,3,2,2x1,4,1,3.
/&/	13/24:4x1,2,1,2x2,1,2x4,3,1.
/i/	13/24:10x1,2,11,1.
/o/	13/24:6x1,6,3,1,5,3x1.
/y/	13/24:3x1,2,1,3,1,2,4,1,2x3,1.

$\frac{\xi^2(M)}{\xi^2(0)}$ (dB)	Opt./a/	Opt./&/	Opt./i/	Opt./o/	Opt./y/
/a/	**-53.51**	-52.79	-52.82	-52.79	-52.66
/&/	-62.57	**-64.94**	-63.88	-62.53	-63.01
/i/	-53.23	-53.84	**-54.79**	-53.35	-52.77
/o/	-62.01	-63.27	-63.12	**-63.50**	-63.06
/y/	-55.22	-55.94	-57.37	-55.98	**-57.81**

Table 2. *Top table*: Optimal NUT lattice configurations found for frames of different vowels @ 24kHz, constraint [13/24]. (Phoneme labels are given in Worldbet notation.) - *Bottom table*: relative residual error when applying the optimal configurations to every vowel.

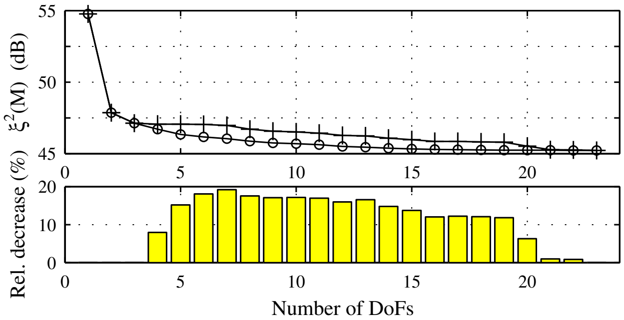


Fig. 3. Comparison of the residual error $\xi^2(M)$ for NUT lattices (o) and classical lattices (+) in function of the number of degrees of freedom $M = 1 \cdots 23$ (for a frame of vowel /a/ @ 24kHz).

mal NUT filters produce a lower residual error than unconstrained filters with and equal number of DoFs. The observed error reductions typically range from a few percent to about 45% in the case of vowel /&/. Figure 4 shows the decrease of the residual error for the reverse experiment, namely keeping a fixed number of DoFs (or parameters) and augmenting the global order of the lattice. The "flat" portions of the curve represent zones where only the last delay's order is increased, which does not change the forward error filter's transfer function (as seen in section 2.2).

5. POTENTIAL APPLICATIONS

Speech coding - The experimental results suggest that with NUT lattices, part of the signal coding task is transfered from the coefficient values to the filter structure. A coding scheme exploiting this model would replace p conventional reflection coefficients (or log area ratios [1]) with $d < p$ coefficients distributed over a p^{th} global order, plus a codeword to index the optimal filter topology. The quality compromise found by adjusting these specifications (in addition to a classical coefficient quantization system) may allow to reach a better coding quality at a lower bit rate than unspecialized AR models.

A related research track would consist in learning the NUT filters configurations on speech segments that span more than one

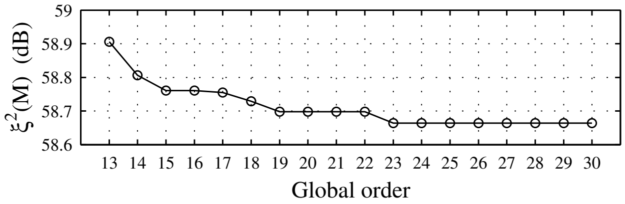


Fig. 4. Residual error decrease versus increase in global order (for a frame of vowel /a/ @ 8kHz), when going from [13/13] (usual 12^{th} order LPC) to [13/30] (greater global order).

frame, and finding the ones that produce the least *mean* residual error. This would allow to build NUT lattices that are specialized to the modeling of classes of signals instead of being specialized to one particular frame [3], and would represent a sort of "macro quantization" of filter structures. Hence, indexing of the non-uniform configurations would use a lower number of bits.

Adequation with articulatory modeling - As pointed out in the introduction, the NUT lattice idea originally arose from the study of the analogy between lattice filtering and acoustic filtering in lossless tubes [2, 4]. While the purpose of the present article was to describe and explore NUT lattices from a pure signal processing point of view, further experiments are needed to determine whether the second optimization layer is able to capture actual acoustic phenomena (e.g., nodes of stationary sound waves, or speaker-specific relative formant positions).

Speech enhancement - Finally, using NUT filters trained on clean speech data for the parameterization of noisy speech may allow to increase the robustness of feature extraction schemes, because the filters would hopefully have retained some structure related to speech production.

6. CONCLUSION

We have presented a constrained parametric spectral model able to model more poles with fewer parameters. Results show that with the same number of degrees of freedom, this model is more accurate than a classical unconstrained All-Pole model. Potential applications are numerous.

7. REFERENCES

[1] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-23, pp. 309–321, June 1975.

[2] H. Wakita, "Direct estimation of the vocal-tract shape by inverse filtering of acoustic speech waveforms," *IEEE Transactions on Audio and Electroacoustics*, pp. 417–427, October 1973.

[3] Sacha Krstulović and Frédéric Bimbot, "Inverse lattice filtering of speech with adapted non-uniform delays," in *Proc. ICSLP 2000*, 2000.

[4] Sacha Krstulović, "Acoustico-articulatory inversion of unequal-length tube models through lattice inverse filtering," IDIAP-RR 16, IDIAP, 1998.

[5] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-25, no. 5, pp. 423–428, October 1977.