INVERSE LATTICE FILTERING OF SPEECH WITH ADAPTED NON-UNIFORM DELAYS

Sacha KRSTULOVIC 1 Frédéric BIMBÔT 2

1 IDIAP C.P. 592 - CH-1920 Martigny - Switzerland - sacha@idiap.ch
2 IRISA - Campus Beaulieu, 35042 Rennes, France - bimbôt@irisa.fr

ABSTRACT

A particular form of constraint is incorporated to Linear Prediction lattice filter models in the form of unequal-length delays. This constraint amounts to reducing the number of intrinsic degrees of freedom defined by the reflection coefficients without modifying the LPC order of the corresponding transfer function. It can be optimized by a simple exhaustive search scheme. Preliminary results show that the prediction error is slightly decreased with respect to a conventional predictor using the same number of reflection coefficients.

1. INTRODUCTION

It is well known [MG76, WAK73] that the process of AR filtering is equivalent, under certain hypothesis, to the acoustic filtering in discrete lossless tubes. While this is traditionally established for tubes with discrete sections of even unitary lengths, an extension concerning the case of tubes with a non-uniform repetition of lengths has been formulated recently [Krs08]. Non-uniform lengths, or equivalently non-uniform delays in a lattice filter, form a particular kind of production constraint.

This article presents a first series of experiments aiming at observing the modifications of inverse filtering performance induced by the incorporation of this constraint into a classical Auto-Regressive modeling framework. After having reviewed the theory of non-uniform lattice filters and their estimation, we will describe a method to optimize the tube configuration in order to minimize the inverse filtering residual error. Experimental results pertaining to the application of the method to various signals will then be given and discussed.

2. REDUCTION OF DEGREES OF FREEDOM IN AN AR MODEL

2.1. AR models and acoustic tubes

The transfer function $A(z) = \frac{1}{D_m(z)}$ of a lossless tube discretized in $M$ unequally lengthy individual sections can be computed in a digital processing framework by application of the following matrix recursion [Krs08]:

$$
\begin{bmatrix}
D_m^0(z) \\
D_m^1(z)
\end{bmatrix} = 
\begin{bmatrix}
1 & k_{m+1} \\
k_{m+1}z^{-m+1} & 1
\end{bmatrix}
\begin{bmatrix}
D_m^0(z) \\
D_m^1(z)
\end{bmatrix}
$$

where:

- $D_m^0(z)$ is the inverse transfer function for the forward traveling sound wave after connection of the $m$th tube section ($D_m^0(z)$ being the transfer function for the backward traveling wave)
- $k_{m+1}$ is the reflection coefficient between section $(m)$ and section $(m+1)$, having areas $S_m$ and $S_{m+1}$ respectively. It is defined as $k_{m+1} = \frac{Z_m - Z_{m+1}}{Z_m + Z_{m+1}}$.
- $n_m$ is the delay order of section $m$, with variable $z$ defined as $z = \exp(j\omega\Delta_{m+1})$. $\Delta_{m+1}$ is the time necessary for sound to travel along one unit of length of a tube section. The unit length is defined as the greatest common divisor of the sections’ lengths. The definition of the corresponding $\Delta_{m+1}$ time constant is where the connection between physical measures in a tube and dimensionless signal processing operates [Krs08].

Equation (1) corresponds to a lattice filtering structure where each matrix

$$
\begin{bmatrix}
1 & k_{m+1}z^{-m+1} \\
k_{m+1}z^{-m+1} & 1
\end{bmatrix}
$$

(2)

corresponds to an inverse filtering cell of the form:

$$
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
\begin{array}{c}
k_{m+1}
\end{array}
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
$$

(see figure 1 for a complete filter). When the delays $n_{m+1}$ are non-unit, these matrices can be expanded as:

$$
\begin{bmatrix}
1 & 0 & \ldots & 0 & k_{m+1}z^{-m+1}
\end{bmatrix}
$$

(3)

i.e., in the lattice form:

$$
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
\begin{array}{c}
k_{m+1}
\end{array}
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
$$

(4)

$$
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
$$

(5)

i.e., in the lattice form:

$$
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
\begin{array}{c}
k_{m+1}
\end{array}
\begin{array}{c}
\text{\$e^\text{"u}$} (m) \\
\text{\$e^\text{"u}$} (m+1)
\end{array}
$$

(6)
Figure 1: Inverse lattice filter accounting for a particular tube configuration, with 30 units sections / 7 degrees of freedom. In this example, 22 reflection coefficients are fixed to zero, and 7 are “free”.

Hence, the unequal length delays topology is equivalent to constraining some reflection coefficients to zero in a filter with a given number of unit-length delays.

It can be verified that the function $D_k^{j+}(z)$ resulting from the application of recursion (1) is always a polynomial in $z^{-1}$. Hence, $A(z) = \frac{\tilde{D}_m}{D_0}$ is an Auto-Regressive transfer function. In the classical unit-delays case, the order of the transfer function is equal to the number of reflection coefficients used to build it. Conversely, in the constrained case, the order of $A(z)$ is the sum of the non-unit delays $n_m$. This order is equal to the number of unit sections in the corresponding tube model, and it is greater than the number of reflection coefficients used to parameterize the function since the number of free reflection coefficients is only equal to the number of unequal-length sections in the tube. Hence, using non-uniform delays amounts to imposing an intrinsic number of degrees of freedom (DoFs) in an LPC model with a given order. Alternatively, it allows to predict a signal sample from an increased portion of its past while keeping the number of model parameters fixed.

The stability of the constrained filter is preserved since forcing some reflection coefficients $k_i$ to zero respects the general stability condition for a lattice filter [MC76], namely $|k_i| < 1 \forall i$ (every $k_i$ should have a value between $-1$ and 1).

2.2. Modeling with non-uniform filters

The parameters $k_i$ of the constrained filters can be estimated by expressing the constraints into Burg’s estimator [Mak77]. Denoting by $\Sigma_p$ the sum of all the delays from order 1 to order ($p$), Burg’s mean squared error criterion can be expressed as:

$$\xi^2(p+1) = \frac{1}{2} \left\{ \sum_{t=p+1}^{N} \epsilon^2_f(t) + \sum_{t=p+1}^{N} \epsilon^2_s(t-1) \right\}$$  \hspace{1cm} (4)

where $\epsilon^2_f(p)$ (resp. $\epsilon^2_s(p)$) is the inverse filtering residual at the output of the forward (resp. backward) branch of the lattice filter. Minimizing this error criterion through differentiating and equating to zero gives:

$$k_{p+1} = \frac{-2 \sum_{t=p+1}^{N} \epsilon^2_f(t) \epsilon^2_s(t-1)}{\sum_{t=p+1}^{N} \epsilon^2_f(t) + \sum_{t=p+1}^{N} \epsilon^2_s(t-1)}$$  \hspace{1cm} (5)

It can be easily verified that this estimator always generates values that lie between -1 and 1, and hence always produces stable filters.

2.3. Detail of the analysis method

Our modeling method is similar to the classical frame-based analysis, but using the modified estimator:

1. **pre-emphasize the input speech signal.** This classical step is performed to compensate for the effects introduced by the glottal waveform shape and the radiation effect at the lips, in the hope that the estimated parameters will better capture some vocal tract properties.

2. **estimate the filter coefficients every 10ms by computing reflection coefficients $k_i$**, using expression (5) with 25ms observation windows. The extracted reflection coefficients can be further transformed into log area ratios or area functions.

3. **inverse-filter the signal and compute the Mean Squared prediction Error (MSE)** pertaining to the input data and the tested model:

$$MSE = \frac{1}{2N} \sum_{t=1}^{N} \left( \epsilon^2_f(M) + \epsilon^2_s(M) \right)$$  \hspace{1cm} (6)

2.4. Optimization

Various constraints lead to different filtering performances in terms of a higher or lower MSE for a test signal. Hence, it is interesting to find the best performing topology given a number of degrees of freedom to be distributed over a given total LPC order.

To search for this best configuration, all the lattice filter topologies respecting a given number of DoFs are generated and systematically used to inverse-filter a test sentence. The one bringing the least MSE is regarded as the best topology.

The various configurations are identified by strings of the form, e.g., [5/223c3,8,5]. This example reads: “5 unequal-length sections distributed on a 22 unit sections model, with three 3rd order delayers, one 8th order delay and one 5th order delay”.

Mean Squared Error

3. EXPERIMENTAL RESULTS

3.1 Importance of the Optimization

The optimization procedure is crucial for achieving the best possible performance of the model. The optimization process involves adjusting the model's parameters to minimize the Mean Squared Error (MSE) between the predicted and actual values. This is done through various algorithms such as gradient descent, where the model's parameters are iteratively updated to reduce the error.

The optimization process is typically performed using a variety of techniques, including but not limited to, stochastic gradient descent, and mini-batch gradient descent. The choice of optimization algorithm depends on the specific characteristics of the dataset and the model being used.

The significance of optimization in machine learning models cannot be overstated. A well-optimized model not only performs better in terms of accuracy but also generalizes better to unseen data, which is crucial in real-world applications.

In this section, we will discuss the importance of the optimization procedure, its impact on the model's performance, and some best practices for optimizing models effectively.
3.4. Dependence upon the signal

Non-speech test signals, scaled to various means and standard deviation values, have yielded different length repartitions (table 2). This confirms that the optimally constrained filters are somehow specialized to the nature of the signals used during optimization.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Best configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>pulse train, 100Hz</td>
<td>6/32:1,3,2,2,2,2,2,2,2</td>
</tr>
<tr>
<td>sine wave, 100Hz</td>
<td>6/32:2,2,15,2,2,2,2,2,2</td>
</tr>
<tr>
<td>white noise</td>
<td>6/32:2,2,2,2,2,2,2,2,2,2,18</td>
</tr>
</tbody>
</table>

Table 2: Best configurations for non-speech test signals with an 7 DoFs /32nd order LPC /minimum length = 2 constraint.

4. CONCLUSION

This article has presented a constrained parametric signal analysis scheme derived from Linear Prediction put in parallel with non-uniform acoustic tube models. This scheme allows to reduce the number of required LPC modeling parameters while keeping the related increase of prediction error to its lowest level. Equivalently, it allows to predict a signal sample from a longer portion of its past with fewer parameters than the usual Auto-Regressive models. This constitutes a preliminary study of a scheme which has a wide variety of potential applications, in particular speech recognition or speech coding.

ACKNOWLEDGMENTS

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

5. REFERENCES


