INTERPOLATION OF LONG GAPS IN AUDIO SIGNALS USING THE WARPED BURG'S METHOD

Paulo A. A. Esquef and Vesa Välimäki

Lab. of Acoustics and Audio Signal Processing
Helsinki University of Technology
paulo.esquef@hut.fi

Kari Roth and Ismo Kauppinen

Lab. of Optics and Spectroscopy
University of Turku/Department of Physics
kari.roth@utu.fi

ABSTRACT

This paper addresses the reconstruction of missing samples in audio signals via model-based interpolation schemes. We demonstrate through examples that employing a frequency-warped version of Burg's method is advantageous for interpolation of long duration signal gaps. Our experiments show that using frequency-warping to focus modeling on low frequencies allows reducing the order of the autoregressive models without degrading the quality of the reconstructed signal. Thus a better balance between qualitative performance and computational complexity can be achieved.

1. INTRODUCTION

Reconstruction of missing samples in audio and speech signals is needed in many real-life digital signal processing applications. For instance, signal drop-outs can be caused by transmission errors through digital channels [1]. Moreover, signal losses may occur due to the presence of corrupting spurious noises such as clicks, pops, and crackles, which are associated with the reproduction of old disk records [2]. Reconstruction or interpolation of discrete-time signals has been approached through various means. For a survey on audio interpolation methods see [3, 4].

In this paper, we tackle the problem of signal reconstruction in long gaps of missing samples via interpolation methods based on autoregressive (AR) modeling. These methods are usually suitable for interpolating relatively short gaps only. For long gaps AR-based interpolation schemes perform poorly, especially toward the middle of the gap [5]. This happens due to the least squares (LS) minimization of the prediction error, used to solve for the unknown samples. The LS criterion yields an excitation whose variance is too low. As a result, towards the middle of the gap, the reconstructed signal tends to vanish, since the unknown samples are obtained via a linear combination of already predicted ones.

Perhaps the simplest solution for the problem just described is to increase the order of the AR model. Alternatively, the problem can be alleviated by imposing a lower limit to the prediction error variance within the minimization procedure [6, 7]. A third option consists in employing two separate AR models, one for the fragment that precedes the gap and another for the segment that succeeds the gap [8].

The same rationale of employing two different AR models for interpolation of long gaps has been presented in [9, 10]. In this method the missing samples in the gap are recovered by means of AR-based signal extrapolations. For appropriate reconstruction of typical audio signals accurate AR models are needed. This usually implies employing high-order models.

However, for signals with non-uniformly distributed energy in frequency, as commonly found in audio programs, signal modeling in a warped frequency scale allows for model-order reduction [11]. As regards the model estimation task, using Burg's method is a convenient choice, since it guarantees an always stable model, which is essential for the extrapolation problem.

In this paper, we present a frequency-warped version of Burg's method and show the benefits of its use in interpolation of long gaps. Model estimation and filtering in the warped domain are computationally more demanding than conventional procedures. Even so, the possibility of model-order reduction can compensate for the additional costs. Hence, a better tradeoff between computational complexity and qualitative performance can be achieved.

This paper is organized as follows. Section 2 reviews the conventional Burg's method and presents its frequency-warped version. Section 3 reviews the interpolation method and elaborates further on the impact of using warped AR models within it. Section 4 describes the experiments that were conducted and compares the performance of the warped-based and conventional interpolation method. Conclusions are drawn in Section 5.

2. WARPED BURG'S METHOD

An autoregressive process y_n is defined by [4]

$$y_n = -\sum_{m=1}^{p} a_m y_{n-p} + e_n, \tag{1}$$

where p is the model order, a_m are the model coefficients, and e_n is the excitation signal. The latter can be also seen as prediction error sequence, which results from inverse filtering the process through its model. Minimizing the prediction error energy with respect to the model coefficients allows solving for them.

In the conventional Burg's method, a modular lattice structure is used to compute the forward and backward prediction errors. The model parameters are retrieved from the reflection coefficients, which are estimated through the minimization of the sum of prediction errors [12].

In the warped version of Burg's method, the modified lattice filter shown in Figure 1 is used instead. For this filter the forward and backward prediction errors are, respectively, given by

$$\begin{aligned} f_n^{(l)} &= f_n^{(l-1)} + \tilde{k}_l \tilde{b}_n^{(l)} \\ b_n^{(l)} &= \tilde{b}_n^{(l-1)} + \tilde{k}_l f_n^{(l-1)} \end{aligned} \qquad n = l, l+1, \ldots, N-1, \tag{2}$$

where $\tilde{k}_l$ are the reflection coefficients of the stage l. The output of

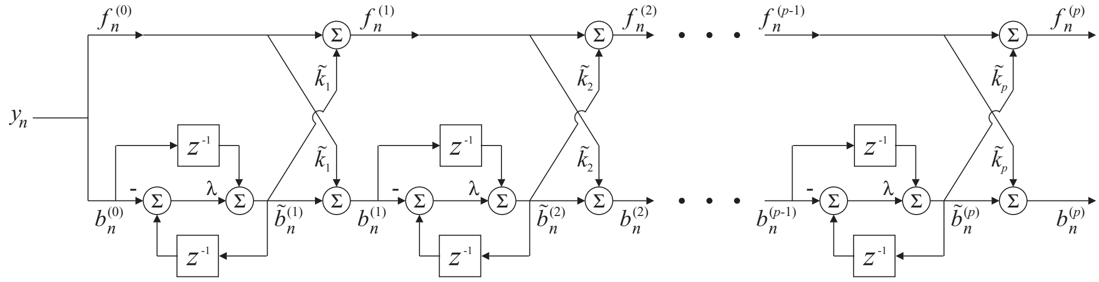

Figure 1: Warped prediction-error filter. The delay elements of the lattice structure have been replaced by first-order allpass filters.

the allpass elements at each stage l is computed via the recursion

$$\tilde{b}_n^{(l)} = b_{n-1}^{(l-1)} - \lambda \left[b_n^{(l-1)} - \tilde{b}_{n-1}^{(l)} \right] \quad \text{for} \quad n = l, l+1, \ldots, N-1, \tag{3}$$

initialized with $\tilde{b}_{l-1}^{(l)} = 0$, being λ the warping factor.

After obtaining $f_n^{(l)}$ and $\tilde{b}_n^{(l)}$ the estimation of the reflection coefficients and the further retrieval of the AR parameters is carried out in the same way as in the conventional Burg's method. The value of the reflection coefficient $\tilde{k}_l$ that minimizes the prediction error is given by

$$\tilde{k}_l = \frac{-2 \sum_{n=l}^{N-1} f_n^{(l-1)} \tilde{b}_n^{(l)}}{\sum_{n=l}^{N-1} \left(f_n^{(l-1)} \right)^2 + \left(\tilde{b}_n^{(l)} \right)^2}. \tag{4}$$

The warped AR-coefficients $\tilde{a}_m$ can be calculated from the reflection coefficients $\tilde{k}_l$ by the ordinary Levinson-Durbin recursion. See [12] and [13] for more details.

3. INTERPOLATION SCHEME

The interpolation method used in this work follows basically the scheme presented in [10]. In such method the missing signal information within a given gap is recovered by means of AR-based signal extrapolation. Here, the method is modified to cope with the use of frequency-warped AR models.

The steps of the proposed warped-based interpolation method are summarized below and illustrated in Figure 2.

1. Estimate an AR model for the segment that precedes the gap using the warped Burg's method (see Section 2), with an appropriate value for the warping parameter λ.
2. Forward extrapolate the signal across the gap by exciting the frequency warped AR model with a zero-padded excitation. The implementation of the warped extrapolation is described in Section 3.2.
3. Repeat the previous two items for the signal segment that succeeds the gap. In this case, a backward signal extrapolation is used instead.
4. Cross-fade the two extrapolated sequences in order to obtain the reconstructed signal.

The window used in the cross-fading is defined by [10]

$$w(n) = \begin{cases} 1 - \frac{1}{2}\left(2u(n)\right)^\alpha, & u(n) \le \frac{1}{2} \\ \frac{1}{2}\left(2 - 2u(n)\right)^\alpha, & u(n) > \frac{1}{2} \end{cases}, \tag{5}$$

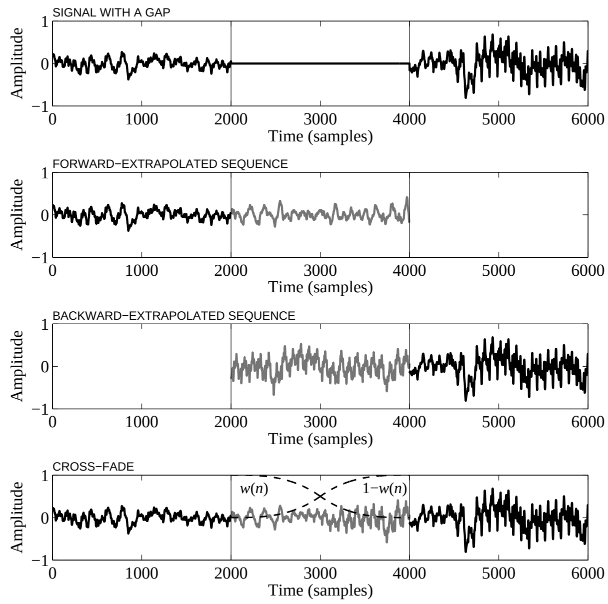

Figure 2: Steps of the interpolation scheme. The reconstructed sequence is obtained by summing the windowed ($\alpha = 3$) forward- and backward-extrapolated sequences within the gap (delimited between the vertical lines).

where $u(n) = (n - n_\mathrm{s})/(n_\mathrm{e} - n_\mathrm{s})$, with n_s and n_e being, respectively, the indices of the beginning and end of the gap. The steepness of the window's roll-off is adjusted via parameter α[1]. For instance, a linear slope down is attained with $\alpha = 1$ whereas a step-like transition is obtainable with $\alpha \to \infty$. Cross-fading is carried out by multiplying the forward-extrapolated sequence by $w(n)$ and the backward-extrapolated sequence by $1 - w(n)$.

3.1. Model Estimation

The length of the fragments (immediately before and after the gap) upon which the AR models are estimated should comply with the usual assumptions on the stationarity of the signal involved. Typically, wide sense stationarity for audio signals is acceptable within short-term segments whose duration may range from 20 to 50 ms. These limits restrict also the maximum order of the AR model that can be attributed to a given segment. As a rule of thumb, the length (in samples) of the signal upon which the model is es-

[1] $\alpha = 3$ is employed in all simulations of this work.

timated should be at least twice as long as the model order. In frequency-warped AR modeling using longer segments is recommendable.

3.2. Signal Extrapolation

The forward extrapolation can be simply carried out by inverse filtering the modeled fragment, appending as many zeros to the resulting excitation as the length (in samples) of the gap to fulfill, and exciting the model with this zero-padded excitation. The inverse filtering and the following resynthesis procedure have to be performed using warped filters [11]. Similarly, the backward extrapolation can be straightforwardly accomplished by appropriate time-reversals of the segments involved in the task.

Signal extrapolation is performed via all-pole synthesis filters in both the conventional and the warped-based interpolation methods. For the latter a warped IIR filter structure has to be employed.

3.2.1. Implementation of the Warped Filters

Frequency-warped filter structures can be designed by replacing the unit delays with first-order allpass filters. In z-domain the transfer function of an all-pole warped filter is given by

$$H(z) = \frac{1}{\sum_{m=1}^{p} \tilde{a}_m \{D(z)\}^m}, \tag{6}$$

where

$$D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \quad \text{with } -1 < \lambda < 1, \tag{7}$$

is the transfer function of the first-order allpass filter.

Replacing the unit delays with first-order allpass filters renders warped IIR filters unrealizable. This is due to the presence of delay-free loops in the filter structure [11, 14]. Realizable warped IIR filters can be achieved by moving the taps into the allpass filter elements, as seen in the structure depicted in Figure 3 [14]. This resource introduces an additional tap and also requires re-computing the tap coefficients. The new σ_l coefficients can be obtained from the warped AR-coefficients $\tilde{a}_m$ via the following recursion

$$\begin{aligned} \sigma_{p+1} &= \lambda \frac{\tilde{a}_p}{\tilde{a}_0} \\ \sigma_l &= \lambda \frac{\tilde{a}_{l-1}}{\tilde{a}_0} + \left(1 - \lambda^2\right) \sum_{m=l}^{p} (-\lambda)^{m-l} \frac{\tilde{a}_m}{\tilde{a}_0} \\ &= \frac{\tilde{a}_l}{\tilde{a}_0} + \lambda \left[\frac{\tilde{a}_{l-1}}{\tilde{a}_0} - \sigma_{l+1} \right] \quad \text{for } l = p, p-1, \ldots, 2 \\ \sigma_1 &= \sum_{m=1}^{p} (-\lambda)^{m-1} \frac{\tilde{a}_m}{\tilde{a}_0}. \end{aligned} \tag{8}$$

The gain factor χ is

$$\chi = -\sum_{m=1}^{p} (-\lambda)^m \frac{\tilde{a}_m}{\tilde{a}_0}. \tag{9}$$

The inverse filtering scheme that yields the excitations for the synthesis filters is carried out via warped FIR filters. However, implementing these filters is a straightforward task, since replacing the unit delays with allpass filter elements does not imply delay-free loops in the filter structure. For more details see [11] and [14].

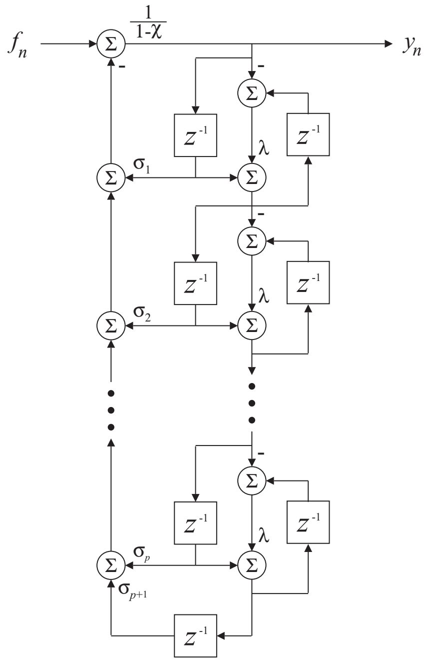

Figure 3: Warped IIR filter. The unit-delay elements of the IIR filter structure have been replaced by allpass filters and the taps have been placed so that there are no delay-free loops in the structure.

3.3. Computational Complexity

The computational complexity of the warped-based interpolation method is of the same order of magnitude as that of the conventional scheme. For a given processing setup, i.e., model order, number of samples used to compute the model, and length of the gap to interpolate, the warped-based method is about 77% more expensive than the conventional procedure. Alone, the estimation of the warped AR model accounts for 33% of the cost increase. The remaining extra costs come from employing warped filters to carry out signal extrapolation. Thus, if one wants to compare both interpolation methods at same cost levels, a fair choice is to set the orders of the AR models used in the conventional method approximately 1.77 times higher than in the warped-based scheme.

3.4. Case Study

We first demonstrate the benefits of using the warped-based interpolation method via a case study that features a low-pitch piano tone (sampled at 44.1 kHz). A gap of 2000 samples is artificially created in the signal and reconstruction is carried out via the proposed interpolation method under different parameter setups. From the results, which are summarized in Figure 4, we observe a satisfactory signal reconstruction when using conventional AR models ($\lambda = 0.0$) of order $p = 1000$. Conversely, the vanishing effect becomes clear when the order is reduced to $p = 100$. However, using warped AR models of this very same order may yield better results. For example, by setting $\lambda = 0.8$, which focus the modeling on the low frequencies, the energy of the reconstructed signal at low frequencies is preserved. This is preferable perceptually over a mute.

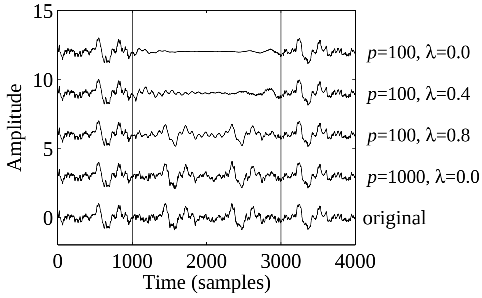

Figure 4: Reconstruction performance under several parameter setups. The gap of 2000 samples lies between the vertical lines.

4. EXPERIMENTAL RESULTS

In this section, the performance of the interpolation methods is assessed by means of objective measures. In order to do so we work on a set of four test signals that covers different music styles. Moreover, each signal is artificially degraded, allowing a reference for comparisons against different restored versions. In the sequel we describe the chosen test signals, the degradation mechanism, and the restoration procedure.

4.1. Test Signals

All test signals used in this work are monaural signals sampled at 44.1 kHz. A brief description of the test signals is given below. Most of the signals lasts about 15 s, except for the signal **Piano**, which lasts about 3 s only.

1. **Piano:** An isolated low-pitch piano tone.
2. **Classic:** An excerpt of classical orchestral music.
3. **Pop:** A fragment of Finnish pop music.
4. **Singing:** Singing *a capella* (Tom's diner sung by S. Vega).

4.2. Degradation Mechanism

The degradation procedure consisted in creating gaps (zeroing) of 2000 samples in the signal at periodical instants. The beginnings of successive gaps are placed 50000 samples apart from each other. This gives more than enough samples before and after the gap to estimate the AR models. Moreover, by forcing periodical gaps each signal will be corrupted with several gaps. These gaps are likely to occur in regions with different time-frequency characteristics. Thus, the interpolation method can be more thoroughly evaluated.

As regards the length of the gaps, the choice for 2000 samples is arbitrary. However, it is supported by the commonly accepted assumptions on the wide sense stationarity within short-term (20 to 50 ms) segments of audio signals [4]. We wanted to have gaps that would be hard to fulfill through a reconstruction method based on a single AR model. Therefore, the length of gap is set near the upper limit at which short-term stationarity is still considered valid.

4.3. Restoration Setups

Signal reconstruction of the missing samples was performed via the interpolation method described in Section 3. Both the conventional Burg's method and its warped version were employed for the model estimation task. In all cases, the length of the frames upon which the model is estimated was fixed to 2000 samples, which correspond to about 45 ms @ 44.1 kHz sample-rate. Moreover, the length of the gaps is fixed to 2000 samples. This leaves us with only two processing parameters to deal with: the order of the AR models p and the value of the warping factor λ.

Several processing parameter setups were tested for restoring the degraded test signals. More specifically, for the warped-based interpolation method, the orders of the AR models were restricted to $p \in \{100,\ 325,\ 565\}$. For each model order, the value of the warping factor was set to $\lambda \in \{0.0,\ 0.2,\ 0.4,\ 0.5,\ 0.6,\ 0.7, 0.8,\ 0.9\}$. In addition, signal restoration was carried out via the conventional method using model orders $p \in \{177,\ 575,\ 1000\}$. These values of p assure that the computational cost of the conventional method is leveled with that of the warped-based scheme for $p \in \{100,\ 325,\ 565\}$, respectively (see Section 3.3).

4.4. Objective Evaluation

Objective evaluation of the quality of the restored signals was conducted by computing their signal-to-noise ratio (SNR) and the Perceptual Audio Quality Measure (PAQM) [15, 16] with respect to the reference test signals.

4.4.1. Signal-to-Noise Ratio

Although perceptually naive, the SNR is the most trivial option to assess the performance of the interpolation method. In this case, the reconstruction error (original $-$ reconstructed) is taken as the noise component. Here, we measured the SNR only within the reconstructed portions. The obtained results are depicted in Figure 5.

Except for the results associated with the signal **Singing**, we verify that, for the conventional method, increasing the model order leads to reconstructed signals with higher SNRs. Moreover, for low model orders the warped-based scheme yields a better balance between cost and performance. For example, the SNRs attained by the warped-based method with model order $p = 100$ and $0.6 \leq \lambda \leq 0.9$ surpass those of the (as costly) conventional scheme with $p = 177$. Similarly, the setups $p = 325$ and $0.4 \leq \lambda \leq 0.7$ are more advantageous than the equivalently expensive conventional method with order $p = 575$.

However, the warped method with $p = 565$ performs, in general, poorer than the equivalently costly conventional algorithm with order $p = 1000$. The sole exception occurs for the signal **Pop** with $p = 565$ and $\lambda = 0.6$, whose SNR is the highest.

As for the effect of the value of λ on the performance, there seems to be an intriguing dependence on the model order, as observed from the curves shown in Figure 5. Note that the values of λ that yield the highest SNRs tend to decrease as the order of the warped model increases.

Finally, a the strange SNRs of the restored versions of the signal **Singing** may reside on the intrinsic nature of the fragment, which contains several short pauses (or silent regions). For instance, a critical situation would be a gap starting right after the end of an utterance followed by a pause. In this case, the interpolation algorithm will unnecessarily prolong the segment. Of course, such mis-processing decreases the SNR of the restored version. Moreover, the higher the model order, the more prolonged the segment will be, and consequently, the lower the final SNR. However, inadvertently prolonging some segments may be harmless in perceptual terms. This motivates the use of objective measures that take psychoacoustic phenomena into consideration.

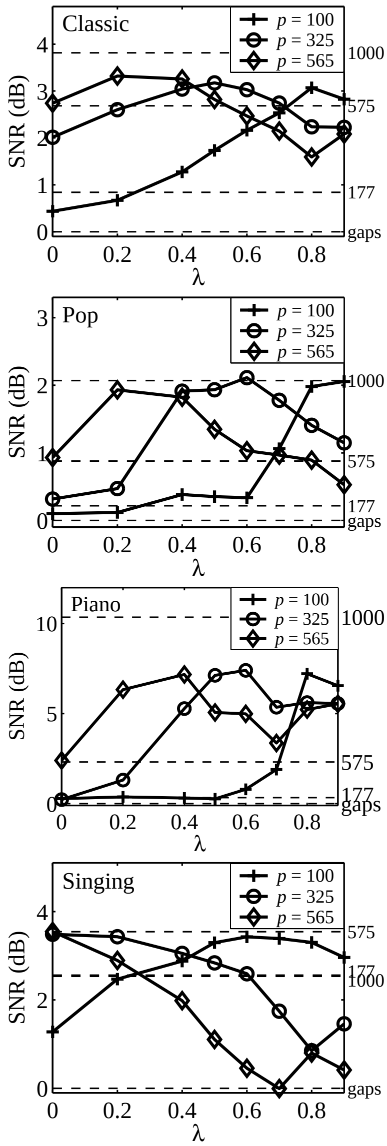

Figure 5: Average SNR of the interpolated signals under several parameter setups. Solid lines are the SNR of the restored signals as function of λ, for the indicated model orders. Dashed lines show the SNR of the signals when restored via the conventional method with the indicated model orders (at the end of the line on the right). The SNR of the corrupted signal is also shown (gaps).

4.4.2. Perceptual Audio Quality Measure

A more realistic way of evaluating the performance of the interpolation method is through the Perceptual Audio Quality Measure (PAQM) of the restored signals. The PAQM is obtained by comparing the uncorrupted (reference) and a restored (processed) version of the test signal. The procedure consists in segmenting both signals in short frames and computing the internal ear representation for each signal frame. Then, a cognitive model compares these representations and outputs the PAQM, which indicates an overall dissimilarity index. Thus, the closer to zero the PAQM, the more similar to the reference the processed signal is.

The PAQMs associated with the restored versions of the test signals were measured. Differently from the average SNR, which is measured only within the restored portions of the signal, the PAQM is computed along the whole extension of the signal. The results are plotted in Figure 6.

The PAQM curves corroborate some conclusions drawn from the SNR measures. It is clear that, at low model orders, there are values of λ for which the warped-based interpolation method is more advantageous. More specifically, choosing $p = 100$ and $0.4 \leq \lambda \leq 0.7$ yields better results than the conventional method with $p = 177$ (that costs the same). It is also worth noting that the performance gains indicated by the PAQMs are less significant than those shown in the SNRs.

Similarly, the warped-based scheme under setups $p = 325$ and $0.5 \leq \lambda \leq 0.8$ is more advantageous than the equivalently expensive conventional method with order $p = 575$. The situation is different, however, when employing higher model orders. For instance, the performance of the warped-based scheme for $p = 565$ is at best equivalent to that of the conventional method with $p = 1000$, e.g., signals **Classic** and **Pop** with $\lambda = 0.5$.

In agreement with the SNR curves, the optimum values of λ for which the lowest PAQMs are achieved tend to decrease as the model order increases. On the other hand, the PAQMs associated with the signal **Singing** are more coherent than the corresponding SNRs. According to the PAQMs, the performance improves as the model order increases.

4.5. Discussion

Both the SNR and the PAQM results indicate that the warped-based interpolation method is advantageous when adopting low-order models. A remaining question concerns the perceptual salience of the achieved gains in performance. In other words, one may wonder whether the mutes are still heard in the restored signals, despite of their higher SNRs or lower PAQMs. Answering this question would require conducting extensive listening tests.

5. CONCLUSIONS

This paper presented a model-based signal interpolation scheme that employs AR models computed via a frequency-warped version of Burg's method. The warped-based method is computationally more expensive than the conventional scheme. For the same model order the increase in the number of floating-point operations lies around 77%.

The interpolation method was applied to signal reconstruction within long gaps of missing samples. In particular, the effect of the model order and warping parameter on the results was verified. Moreover, the interpolation performance of both the conventional and warped-based methods was evaluated. In order to assure fair comparisons the model orders were adjusted as to equalize their computational load.

Experiments were conducted on a set of four test signals with distinct sonic characteristics. Performance assessment was carried out by means of objective measures taken over the restored versions of the test signals. More specifically, the signal-to-noise ratio (SNR) and the Perceptual Audio Quality Measure (PAQM) were employed. In both cases the obtained results show that, in general and for lower model orders, the warped-based scheme is more advantageous. However, for high model orders (within the range

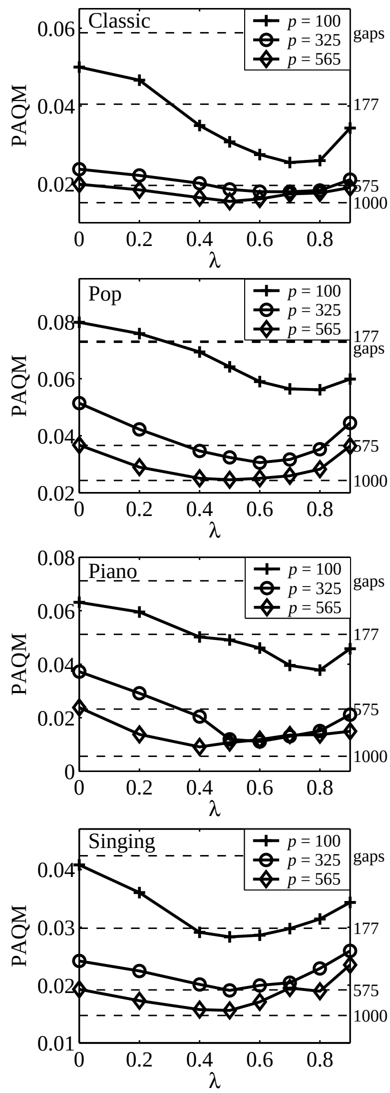

Figure 6: PAQMs of the restored signals processed under several parameter setups. Solid lines are the PAQM of the restored signals as function of λ, for the indicated model orders. Dashed lines are the PAQM of the signals restored via the conventional method with the indicated model orders (at the end of the line on the right). The PAQM of the corrupted signal is also shown (gaps).

evaluated), the warped-based method performs at best as good as the conventional scheme.

Furthermore, the results indicate that the value of the warping parameter that optimizes the interpolation performance varies with the adopted model order. More specifically, it tends to decrease as the model order increases. This indicates that for high-order models frequency warping no longer helps to improve the performance.

6. ACKNOWLEDGMENTS

The work of Paulo A. A. Esquef has been supported by the Brazilian National Council for Scientific and Technological Development (CNPq-Brazil). The authors wish to thank Dr. Luiz W. P. Biscainho for providing the code used in the PAQM computation.

7. REFERENCES

[1] N. S. Jayant and S.W. Christensen, "Effects of Packet Losses in Waveform Coded Speech and Improvements Due to an Odd-Even Sample-Interpolation Procedure," *IEEE Trans. Communications*, vol. CAM-29, pp. 101–109, Feb. 1981.

[2] P. Wilson, "Record Contamination: Causes and Cure," *J. Audio Eng. Soc.*, vol. 13, no. 2, pp. 166–176, Apr. 1965.

[3] R. N. J. Veldhuis, *Restoration of Lost Samples in Digital Signals*, Prentice-Hall, 1990.

[4] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration — A Statistical Model Based Approach*, chapter 5, Springer-Verlag, London, UK, 1998.

[5] A. J. E. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive Interpolation of Discrete-Time Signals That Can Be Modeled as Autoregressive Processes," *IEEE Trans. ASSP*, vol. ASSP-34, no. 2, pp. 317–330, Apr. 1986.

[6] P. J. W. Rayner and S. J. Godsill, "The Detection and Correction of Artefacts in Degraded Gramophone Recordings," in *Proc. IEEE ASSP WASPAA'91*, Oct. 1991, pp. 151–152.

[7] M. Niedźwiecki, "Statistical Reconstruction of Multivariate Time Series," *IEEE Trans. Signal Processing*, vol. 41, no. 1, pp. 451–457, Jan. 1993.

[8] W. Etter, "Restoration of a Discrete-Time Signal Segment by Interpolation Based on the Left-Sided and Right-Sided Autoregressive Parameters," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1124–1135, May 1996.

[9] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A Method for Long Extrapolation of Audio Signals," *J. Audio Eng. Soc.*, vol. 49, no. 12, pp. 1167–1180, Dec. 2001.

[10] I. Kauppinen and J. Kauppinen, "Reconstruction Method for Missing or Damaged Long Portions in Audio Signal," *J. Audio Eng. Soc.*, vol. 50, no. 7/8, pp. 594–602, July/Aug. 2002.

[11] A. Härmä, M. Karjalainen, V. Välimäki, L. Savioja, U. Laine, and J. Huopaniemi, "Frequency-Warped Signal Processing for Audio Applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, Nov. 2000.

[12] M. H. Hayes, *Statistical Signal Processing and Modeling*, chapter 6, John Wiley & Sons, Inc., 1996.

[13] K. Roth, I. Kauppinen, P. Esquef, and V. Välimäki, "Frequency Warped Burg's Method," in *Proc. IEEE WASPAA'03*, Mohonk, 2003.

[14] A. Härmä, "Implementation of Frequency-Warped Recursive Filters," *Signal Processing*, vol. 80, pp. 543–548, 2000.

[15] J. G. Beerends and J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation," *J. Audio Eng. Soc.*, vol. 40, no. 12, pp. 963–978, Dec. 1992.

[16] J. G. Beerends, "Audio Quality Determination Based on Perceptual Measurements Techniques," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 1. Kluwer Academic Publishers, 1998.