

INTERPOLATION OF MISSING DATA VALUES FOR AUDIO SIGNAL RESTORATION USING A GABOR REGRESSION MODEL

Patrick J. Wolfe

Harvard University
Division of Engineering and Applied Sciences
33 Oxford Street
Cambridge, MA 02138-2901
USA

Simon J. Godsill

University of Cambridge
Department of Engineering
Trumpington Street
Cambridge CB2 1PZ
UK

ABSTRACT

In this paper we present a method to address the (inherently ill-posed) problem of missing data interpolation over repeated short gaps in audio signals. By formulating the problem in terms of a Gabor regression model, we show that it is possible to leverage information from the surrounding time-frequency plane in order to obtain an interpolation in keeping with the qualities of the signal under consideration. As an exploratory investigation of this technique’s potential, we consider two example restoration scenarios in which over one third of the data values in total are missing.

1. INTRODUCTION

We present here a method for missing data interpolation in audio time series. Much related work has been done in the areas of audio signal interpolation and extrapolation; for an overview we refer the reader to [1, 2, 3, 4, 5] and references therein. However, it is well known that schemes such as autoregressive modeling can suffer from audible distortion in situations involving voiced speech or music extracts [6, Chapter 5]. Additionally, these and other methods require the tuning of parameters such as model order and block length, and can sometimes lead to reconstructions which are overly smooth in comparison to typical audio signals [6, Chapter 5]. By contrast, we outline a method based on the principles of a Gabor regression model [7, 8] as a potential way to avoid these shortcomings.

2. SIGNAL MODEL

The model considered here stems from earlier work in which the signal under consideration is decomposed according to the principles of Gabor analysis over finite cyclic groups [7, 8]. Simply put, this method is a formalization of the tried-and-true overlap-add method commonly used for audio signal analysis and synthesis (see, e.g., [9] for a detailed exploration of this relationship).

To this end, we recall that the standard practice for modification of an audio times series vector $\mathbf{x} \in \mathbb{R}^L$ proceeds as follows: first, \mathbf{x} is divided into overlapping segments via the multiplicative action of a (typically) smooth, symmetric window \mathbf{g} whose effective size l (typically $\ll L$) is chosen as a function of the sampling rate such that the analysis window length lies in the range of 15–40 ms, depending on the time-varying nature of the audio signal class under consideration. The discrete Fourier transform

(DFT) is then applied on each interval and the resultant spectral coefficients are modified according to the task at hand; the inverse DFT is then taken and a corresponding synthesis window applied to each segment. Finally, the overlapping segments are added together in an appropriately weighted manner in order to reconstitute the modified time series vector $\hat{\mathbf{x}}$.

2.1. Formalizing the Overlap-Add Method

As a prelude to the interpolation model presented below, it is helpful to understand the overlap-add procedure more formally as follows: using the pair (m, n) to denote modulation and translation indices respectively, and thus to index a (separable) lattice of points in the time-frequency plane, we may think of mapping each windowed segment of \mathbf{x} to a corresponding short-time spectral segment, or sampled “slice” of that signal’s time-frequency representation. In particular, this operation corresponds to a representation of \mathbf{x} in terms of a set of Gabor transform coefficients $\{c_{m,n}\}$ representing a sufficiently fine tiling of the time-frequency plane.

These so-called analysis coefficients are calculated as inner products of \mathbf{x} and translated, modulated versions of some chosen analysis window as $c_{m,n} = \langle \mathbf{x}, \mathbf{g}_{m,n} \rangle$, where $\mathbf{g}_{m,n}$ denotes a discretized, time-frequency shifted version of a window function $g(t)$:

$$g_{m,n}(t) = g\left(t - \frac{n}{N}L\right) e^{2\pi j \frac{m}{M}t}, \quad t \in \{0, 1, \dots, L-1\}.$$

Here M and N are positive, integer lattice constants chosen according to parameters a and b (representing time and frequency sampling intervals, respectively) such that $Na = Mb = L$, the length of the vector \mathbf{x} . The corresponding *Gabor expansion* in turn provides a means of reconstructing \mathbf{x} from its Gabor coefficients, which act as weights in the sum of translations and modulations of a dual (or *synthesis*) window function $\tilde{g}(t)$:

$$\mathbf{x}(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} c_{m,n} \tilde{g}_{m,n}(t) = \sum_{m,n} \langle \mathbf{x}, \mathbf{g}_{m,n} \rangle \tilde{g}_{m,n}(t).$$

In the discrete-time case, we may hence denote the Gabor transform of a vector \mathbf{x} as $\mathbf{c} = \mathbf{G}^* \mathbf{x}$, where \mathbf{G}^* denotes the Hermitian transpose of the $L \times MN$ Gabor analysis matrix \mathbf{G} having the time-frequency atom $\mathbf{g}_{m,n}$ as its $(m + nM)$ -th column, and the Gabor transform coefficients $\{c_{m,n}\}$ are written in the form of

a “stacked” column vector \mathbf{c} of length MN . Likewise, we may denote the Gabor expansion of \mathbf{x} by $\mathbf{x} = \tilde{\mathbf{G}}\tilde{\mathbf{c}}$, where $\tilde{\mathbf{G}}$ denotes the $L \times MN$ Gabor synthesis matrix having $\tilde{\mathbf{g}}_{m,n}$ as its $(m+nM)$ -th column, and the vector $\tilde{\mathbf{c}}$ represents the corresponding synthesis coefficients.

2.2. Overcompleteness and Gabor Frames

We distinguish between \mathbf{c} and $\tilde{\mathbf{c}}$ in the above discussion because an *overcomplete* representation admits an entire subspace of perfect-reconstruction synthesis coefficients. Indeed, if the column rank of $\tilde{\mathbf{G}}$ is equal to L , then the family (\mathbf{g}, a, b) will form a Gabor frame with redundancy MN/L (see, e.g., [10]). Owing to the overlap of the windowed time series segments in the scenario we consider here, $MN > L$ and this representation is in fact redundant, rather than being a simple change of basis (such as, e.g., the discrete cosine transform and its variants). In applications the redundancy rate is typically equal to two, corresponding to use of the DFT algorithm as described earlier and a “window overlap” in time of 50%.

In this case the analysis coefficients given by $\{\langle \mathbf{x}, \mathbf{g}_{m,n} \rangle\}$ are *not* the only choice of synthesis coefficients corresponding to a perfect reconstruction of the chosen signal, but in fact simply comprise the minimum-norm set in an ℓ^2 sense. Hence, it is possible to formulate statistical models in terms of a set of (unobserved) synthesis coefficients $\tilde{\mathbf{c}}$, rather than solely considering the analysis coefficients obtained via the standard methods short-time Fourier analysis.

3. INTERPOLATION OF MISSING DATA

3.1. A Gabor Regression Model

In light of the above exposition, consider first the standard additive observation model

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (1)$$

where $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_{L-1}]^T$ is the vector of the observed waveform, \mathbf{x} is that of the underlying (audio) signal we wish to estimate, and $\boldsymbol{\epsilon}$ represents samples of an independent, identically distributed, continuous Gaussian noise process with (potentially unknown) variance σ^2 .

Most audio signal processing approaches begin with the Gabor transform of the noisy data, given by $\tilde{\mathbf{G}}^* \mathbf{y}$ according to the chosen Gabor system. Here, however, we directly estimate the *synthesis* coefficient representation $\mathbf{x} = \tilde{\mathbf{G}}\tilde{\mathbf{c}}$. Indeed, by visualizing the overlap-add procedure on the operator level, we arrive at the following additive observation model formulated in terms of the Gabor synthesis coefficients $\tilde{\mathbf{c}}$:

$$\mathbf{y} = \tilde{\mathbf{G}}\tilde{\mathbf{c}} + \boldsymbol{\epsilon}. \quad (2)$$

It is important to recall once again that $\tilde{\mathbf{G}}$ need only exist in a *conceptual* sense, as only $l \ll L$ elements of each of its columns are non-zero. Hence it is never necessary in practice to construct such a matrix; the structure of a Gabor frame implies that columns of $\tilde{\mathbf{G}}$ may be obtained as translations and modulations of a synthesis window function $\tilde{\mathbf{g}}$ (with effective length l) according to the parameters of the chosen Gabor system.

3.2. Bayesian Estimation for the Gabor Regression Model

The noise variance term σ^2 appearing in (1), taken in conjunction with (2), immediately implies a Gaussian likelihood for \mathbf{y} centered on $\tilde{\mathbf{G}}\tilde{\mathbf{c}}$. If the synthesis coefficients $\tilde{\mathbf{c}}$ are in turn considered as (latent) random variables, we have by Bayes’ rule that their posterior distribution satisfies

$$p(\tilde{\mathbf{c}} | \mathbf{y}, \boldsymbol{\theta}) \propto p(\mathbf{y} | \tilde{\mathbf{c}}, \boldsymbol{\theta}) p(\tilde{\mathbf{c}}, \boldsymbol{\theta}), \quad (3)$$

where $\boldsymbol{\theta}$ denotes a collection of any additional probabilistic terms in the model. In particular, by considering a hierarchical prior of the form $p(\tilde{\mathbf{c}}, \boldsymbol{\theta}) = p(\tilde{\mathbf{c}} | \boldsymbol{\theta}) p(\boldsymbol{\theta})$, it is possible to formulate a variety of regularization schemes with regard to estimation of the synthesis coefficients. As detailed in [8], stochastic computation may then be carried out via Markov chain Monte Carlo (MCMC) methods in order to formulate point estimates of any model parameter. In particular, we shall be concerned here with the minimum mean-square error (MMSE) estimator $E[\tilde{\mathbf{c}} | \mathbf{y}]$, obtainable via Monte Carlo integration given a collection of random samples drawn from the joint posterior distribution according to (3).

As is well known, Bayesian models of this type constitute an implicit form of regularization; indeed, the very act of synthesis coefficient estimation in the overcomplete case [11] is an inherently ill-posed (many-to-one) inverse problem. Here we consider two approaches developed in [8, 11] for the specification of the prior structure of the synthesis coefficient vector $\tilde{\mathbf{c}}$. One case corresponds to an overcomplete estimation scheme; the other includes a latent binary indicator variable at each point (m, n) along the time-frequency lattice, resulting in the potential inclusion or exclusion of each time-frequency coefficient in the signal model. In both cases a (heavy-tailed) Student t prior distribution is assumed for $p(\tilde{\mathbf{c}})$; equivalently the coefficients are considered to be independent, zero-mean, and normally distributed, conditional on (unknown,) independent, inverse-gamma distributed variances a priori. The expected values of these variances (as a measure of coefficient power) may potentially be weighted in inverse proportion to frequency, in keeping with typical audio signal content; also, various Markov random field models may be formulated in conjunction with the set of indicator variables along the time-frequency lattice. The various forms of prior distribution considered will imply a valid joint distribution $p(\mathbf{y}, \tilde{\mathbf{c}}, \boldsymbol{\theta})$; see [8] for details.

3.3. Experimental Results

We now consider the application of the Gabor regression model to the interpolation of missing data values in audio time series. We present two preliminary results representing an exploration of the various model types described above, one (Fig. 1) using an overcomplete regression scheme as in [11], and the other (Fig. 2) using a model-averaged variable selection scheme in which a Markovian relationship over successive time blocks is assumed to govern Gabor coefficient activity at a given frequency (full details of which are given in [8]). With regard to the data, we consider a scenario typical of that encountered in audio restoration applications, in which short gaps resulting from severe impulsive noise degradations occur frequently and at random intervals [6].

To this end, simulations were performed in which audio time series were artificially degraded as follows: 16-bit signals sampled at a rate of 44.1 kHz were first downsampled to 11.025 kHz and then corrupted by a series of gaps of random length in the range 2–4 ms, spaced randomly with a minimum separation of 5 ms. These

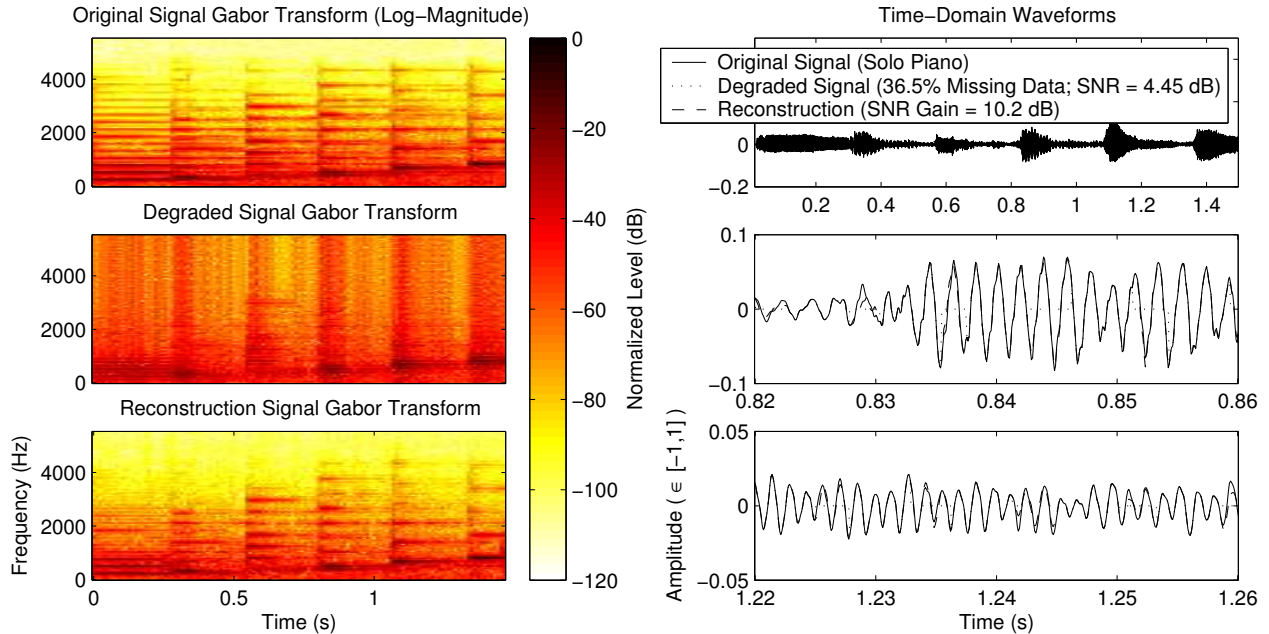


Fig. 1. Interpolation of gaps in a solo piano signal, showing the log-magnitude of the Gabor transform coefficients (left column) and the corresponding time series (right column). Time series comparisons are shown for both a note onset (middle) and a steady-state signal portion (bottom).

signals were in turn processed using the Gabor regression model described above, via a redundancy-two tight Gabor system derived from a 256-sample Hanning window [8]. Missing data values were imputed according to the signal reconstruction $\hat{x} = \tilde{G}\hat{c}$, where $\hat{c} = E[\tilde{c} | \mathbf{y}]$, the MMSE estimate of the synthesis coefficient set.

The first example comprises a solo piano signal, degraded to a SNR of 4.45 dB such that 36.5% of its data values were missing. An overcomplete Gabor regression model was applied to the corrupted signal, with a frequency-weighted prior distribution of coefficient variances used to penalize energy at high frequencies (see [11] for details). The results of this restoration are shown in Fig. 1. The estimated reconstruction appears qualitatively reasonable; this observation is supported by a corresponding increase in signal-to-noise ratio (SNR) of over 10 dB—although the effect of the prior weighting may be seen to result in some losses at higher frequencies.

The second example, shown in Fig. 2, consists of a jazz trumpet recording degraded to a SNR of 4.23 dB such that 37.2% of its data values were missing. This signal may be seen to have a less regular time-frequency structure; in particular the starting and stopping of notes necessitates a model capable of adapting to the local degree of signal non-stationarity. To this end a variable selection scheme was used as described above; the MMSE reconstruction shown here was obtained by averaging over the posterior distribution of models visited by the MCMC sampler. In this case the gain in SNR is 5.94 dB. While not as great as before, the qualitative aspects of the restoration are once again promising, both sonically and visually. Indeed, we encourage readers to audition these and other results for themselves; to this end, data and code for the reproduction of the experiments described herein will be posted at the first author’s home page:

<http://www.eecs.harvard.edu/~patrick>

We emphasize that the results presented here are obtained under an essentially “blind” and automatic estimation scheme, in the sense that all model elements are assigned hyperparameters corresponding to vague prior distributions. In the experiments described here, the parameters chosen a priori were those of the Gabor system, and (if used) the exponent for frequency-scaling of component variances or the transition parameters of the Markov chains along the time-frequency lattice (although we note that it is possible in principle to estimate even these elements from the data). Moreover, while the noise term ϵ is implicitly considered to be a measure of fitting error here, such a model may also be extended to cases where genuine global noise degradation exists. An additional advantage of our approach is that the missing values are imputed via consideration of the *joint* distribution of the audio waveform; both through the structure of the Gabor synthesis matrix and the prior dependencies formulated amongst the Gabor coefficients, information from surrounding areas of the time-frequency plane contributes to the interpolation of missing data values.

4. SUMMARY

Here we have presented two distinct examples of missing data interpolation for audio time series via a Gabor regression model. These examples were constructed in a manner which emulates audio restoration applications, where clicks and other impulsive degradations are so frequent and severe as to preclude the extraction of useful signal information in the areas of degradation. Of course, other important and related applications exist, including for instance the problem of packet loss in voice-over-Internet-protocol (VOIP) transmission.

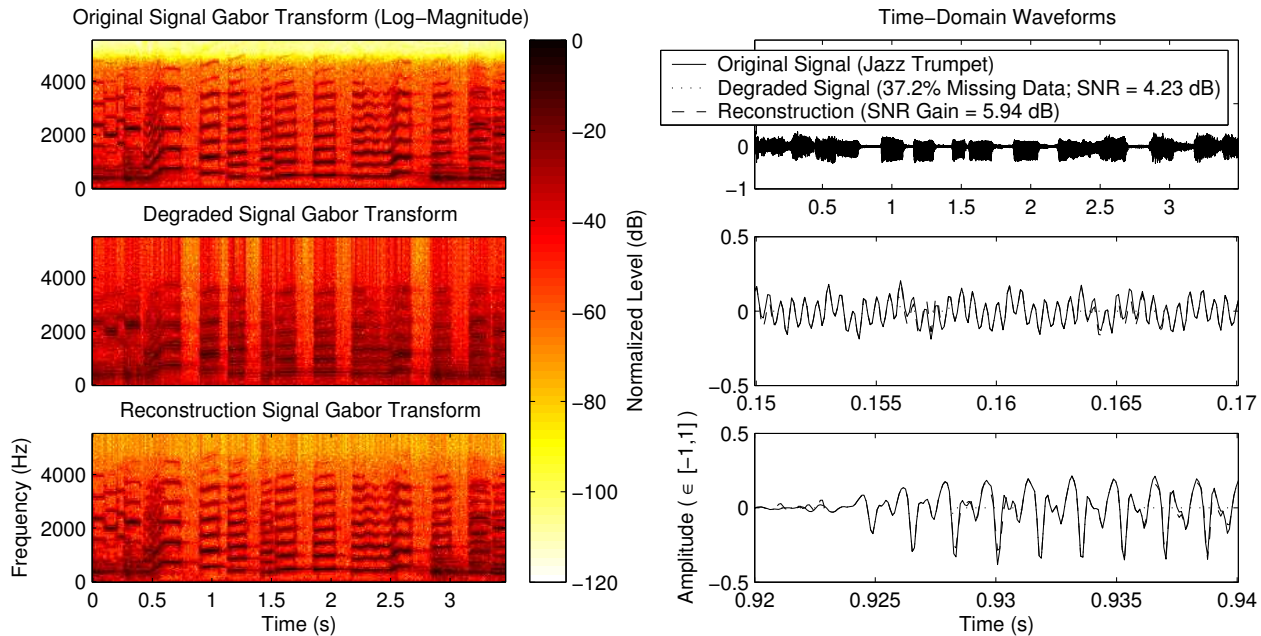


Fig. 2. Interpolation of gaps in a jazz trumpet signal, showing the log-magnitude of the Gabor transform coefficients (left column) and the corresponding time series (right column). Time series comparisons are shown for both a steady-state signal portion (middle) and a note onset (bottom)

The suitability of a Gabor regression model overall is evidenced by the fact that promising restorations (from both the objective and subjective points of view) may be obtained even in cases where over 35% of the data values are missing. Lastly, we note that while here the locations of missing data are assumed to be known a priori, the methodology may be extended to a fully Bayesian scheme for joint detection, interpolation, and noise reduction for signal enhancement, as in the earlier Bayesian approaches of [6, Chapter 12].

5. REFERENCES

- [1] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1124–1135, 1996.
- [2] I. Kauppinen and J. Kauppinen, "Reconstruction method for missing or damaged long portions in audio signal," *Journal of the Audio Engineering Society*, vol. 50, no. 7, pp. 594–602, 2002.
- [3] G. Cocchi and A. Uncini, "Subband neural networks prediction for on-line audio signal recovery," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 867–876, 2002.
- [4] P. A. A. Esquef, "Interpolation of long gaps in audio signals using line spectrum pair polynomials," Tech. Rep. 72, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Helsinki, 2004.
- [5] L. Lu, L. Wenyin, and H.-J. Zhang, "Audio textures: Theory and applications," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 2, pp. 156–167, 2004.
- [6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag, Berlin, 1998.
- [7] P. J. Wolfe and S. J. Godsill, "A Gabor regression scheme for audio signal analysis," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, pp. 103–106.
- [8] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, "Bayesian variable selection and regularisation for time-frequency surface estimation (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 66, no. 3, pp. 575–589, 2004.
- [9] M. Dörfler, "Time-frequency analysis for music signals: A mathematical approach," *Journal of New Music Research*, vol. 30, no. 1, pp. 3–12, 2001, Special Issue: Music and Mathematics.
- [10] T. Strohmer, "Numerical algorithms for discrete Gabor expansions," in *Gabor Analysis and Algorithms: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds., Applied and Numerical Harmonic Analysis, chapter 8, pp. 267–294. Birkhäuser, Boston, 1998.
- [11] P. J. Wolfe and S. J. Godsill, "Bayesian estimation of time-frequency coefficients for audio signal enhancement," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds., pp. 1197–1204. MIT Press, Cambridge, MA, 2003.