SUBBANDS AUDIO SIGNAL RECOVERING USING NEURAL NONLINEAR PREDICTION

Gianandrea Cocchi and Aurelio Uncini (*)

(*) Dipartimento INFOCOM[1] - University of Rome “La Sapienza”
Via Eudossiana 18, 00184 Rome – Italy, aurel@ieee.org

ABSTRACT

Audio signal recovery is a common problem in digital audio restoration field, because of corrupted samples that must be replaced.
In this paper a subbands architecture is presented for audio signal recovery, using neural nonlinear prediction based on adaptive spline neural networks. The experimental results show the mean square reconstruction error, and maximum error obtained with increasing gap length, from 200 to 5000 samples. The method gives good results allowing the reconstruction of over 100ms signal with low audible effects in overall quality.

1. INTRODUCTION

The introduction of digital systems in audio signal field let grow the problem of digital audio restoration. This is necessary whenever original signal is corrupted by noise, background or impulsive noise, or whenever a sequence of consecutive samples is missing. Situations of the first type are common with old gramophone recordings: background noise is due to transducing equipments, while impulsive noise appears as a result of natural surface irregularities, scratches and dust particles. Situations of the second type are common with magnetic tapes, when, by demagnetisation effects, a fragment of a musical track is completely missing. The case of impulsive noise and the case of a missing sequence are similar, because only few data are involved., while background noise involves all samples. So two different approaches must be adopted. We concerned on degradations of the first type, that is only a localized sequence must be elaborated to reconstruct the original signal. In fact this operation is also necessary when reduction of impulsive noise is required: a preprocessing stage localises noise positions and then a reconstruction stage of corrupted samples is needed.

The classical approach for removal of impulsive noise from audio data is median filtering. A window of predetermined length slides sequentially over the data, and the central sample within the window is replaced by the median of all the samples that are inside it. The main drawback derives from the fact that all samples are replaced causing a worsening of global quality, as a high pass filter is applied. Besides only short time duration impulses can be corrected. Improvements can be reached using an adaptive threshold [1], which avoids all samples to be replaced.

A different methodology for missing data reconstruction is based on the assumption that signal can be modelled as P order autoregressive process [2-3]. The model parameters, which are the taps of a FIR linear predictor filter, are calculated using a signal block including missing data. Then the missing samples are predicted in forward and backward mode and the attained results are combined in suitable manner to give missing sequence. The main drawback of autoregressive model based approach is an increasing model order for long sequences reconstructions. Model order must be two or three times the length of missing data sequence. This method is not very appropriate for reconstructing long sequence (over about a hundred of samples at 44.1KHz sampling rate [4]). An proposed improvement involves the addition of long term prediction to short term prediction [2]. The pitch period is calculated and prediction is carried out using P samples in proximity of missing data and Q ($Q<P$) a pitch period away. The method becomes more efficient specially with a clear signal periodicity, even if the number of samples which is possible to reconstruct remains low.

Linear prediction method has been extended to non linear case using neural networks [5]. A P inputs multilayer neural networks has been used, with hidden layer neurons with bipolar sigmoidal activation function, and a linear neuron in the output layer. The net has been trained using several examples extracted randomly from different uncorrupted signals. A forward and a backward prediction are performed in a similar manner as linear case. Some experiments have shown performances of the same level of the linear predictor. Improvement has been obtained adding a training stage using samples before and after the missing data, allowing the net to learn the local characteristics of the signal, although results remain comparable with those obtained using linear prediction.

All methods seen by now are based on a time domain signal elaboration. A dual approach is a missing data reconstruction on frequency domain [6]. Spectral components of signal are analysed using a Short Time Fourier Analysis (STFT) with overlapping windows. Modules and phases of spectral components concerning window containing missing data sequence are calculated using polynomial interpolators taking care for signal continuity. This method is reported to be successful for gap lengths of up to 40ms, or well over 2000 samples at 44.1KHz sampling rate.

The purpose of this work is to evaluate performances of a subbands signal reconstruction method, which realizes a trade off among techniques based on time domain and on frequency domain. Two different filter bank architectures have been

[1] This work is supported by the “Ministero della Ricerca Scientifica e Tecnologica” of Italy

implemented: 32 channels uniform filter bank and octave filter bank. For signal reconstruction a linear predictor and a non linear predictor have been used. Non linear predictor is realized using a recently proposed neural network architecture, very suitable for signal processing applications, based on an adaptive spline activation function called adaptive spline neural networks (ASNN) [9-10]. Experiments have been carried out using combinations of filter banks and prediction methods. The results show mean square reconstruction error, and maximum error obtained with increasing gap length, from 200 to 5000 samples. The method gives good results allowing the reconstruction of over 100ms signal with low audible effects in overall quality.

2. SUBBAND SIGNAL RECOVERING

The reconstruction of L consecutive missing samples in an audio signal may be considered an extrapolation problem, and is represented in Fig. 1.

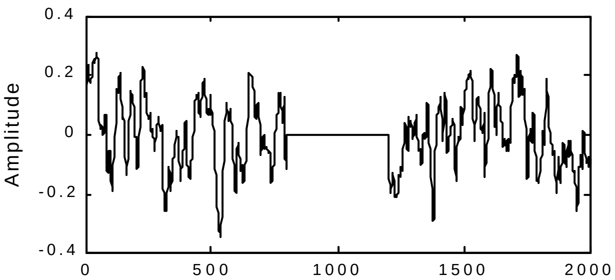

Figure 1 - Example of 400 samples (8ms) of missing samples of music audio signal.

Given the signal trend on the left and on the right of the missing sequence boundaries our task is to fill the gap. Using time domain based approaches known in literature it is not possible to solve the problem if the gap is more then one hundred of samples [2], [3], [5]. Better results can be obtained using frequency domain based approaches [6]. This is due to the fact that time-variant spectral amplitude varies slowly with respect time signal, so interpolation is easier. The drawback of this method is related to interpolation of complex numbers and phase discontinuities.

A multirate approach gives the advantage of a reduced sampling rate, and so a decrease of samples number, allowing a longer gap reconstruction. A direct implementation of multirate technique is a filter bank which offers two different advantages: former is related to the fact that the signal is split in several components. This is similar to a frequency analysis and so prediction is made easier by slower evolution of subband signal. The latter is due to the decimation process which reduces signal length, avoiding numeric problems.

Input signal $x[n]$ is properly selected so that missing sequence is centred. At the output of analysis filter bank we have M signals, $x_1[n]$, ...,$x_M[n]$, where M is the number of channels. To every $x_i[n]$ signal a forward and backward prediction algorithm is applied and the results are combined using weighing windows as in a cross fade operation. Signals with missing data gap filled by prediction results are then passed to synthesis filter bank to obtain full band reconstructed signal $y[n]$. In Fig.2 and Fig.3 are presented block schemes of subband signal reconstruction algorithm.

The method has been implemented using two different filter bank typologies, uniform filter bank and octave (or Q constant) filter bank. Two different prediction methods has been also used: the first based on linear prediction, the second based on neural network, realizing a non linear predictor. A brief presentation of filter bank architecture and prediction methods is given in the following sub sections.

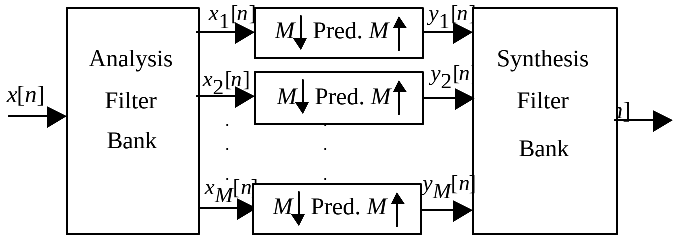

Figure 2 - Architecture of the proposed subband audio signal recovering.

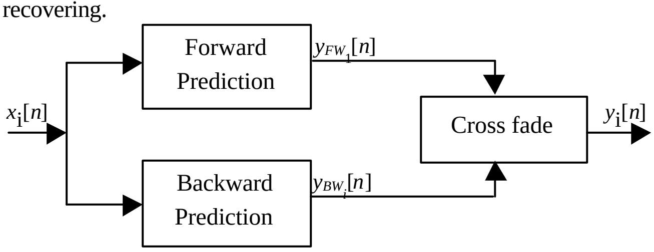

Figure 3 - The i-th channel forward and backward prediction blocks.

3. FILTER BANK ARCHITECTURE

3.1 Uniform filter bank

Uniform filter bank consists on low-pass, band-pass and high-pass filters, partitioning signal spectra into directly adjacent bands of equal width. All filters have the same bandwidth, and the *central* frequency are uniformly spaced in frequency. A M-channels pseudo-quadrature mirror filter (QMF) bank was implemented [7-8]. Input signal is split into M different signals, each representing signal progress in i-th band determined by i-th band-pass filter. An M factor decimator is applied to all signals, so output signals are M factor shorter than input signal.

3.2 Octave filter bank

Octave filter bank belongs to Q constant filter bank family; that is ratio between nominal bandwidth amplitude and its central frequency is constant. In a tree structure, two-channels filter banks consisting on a low-pass filter $H_0(z)$ and a complementary high pass filter $H_1(z)$ are used as band separating filters. By successively feeding just the low frequency output signal to a further band separating filter, the octave analysis filter bank is obtained [7]. Unlike uniform filter bank the sampling rate is reduced by a factor of two at each stage and taking the transfer functions in all stages of the filter to be the same, then the relative transition bandwidth (relative to the respective sampling rate) is the same in all stages. Reducing sampling rate by a factor of two at each stage output signal are different lengths. An eight stages filter bank was implemented, so signal length is reduced by a factor 2 for higher signal components to a 2^8 factor for lower components.

4. SUBBAND SIGNAL PREDICTION

Output signals from analysis filter bank are passed to predictor block to reconstruct missing data gap. We have a sequence with inside a gap of missing data that must be filled. Two different approaches are used: the first is based on a linear prediction algorithm [2-3], while the second is based on a non-linear prediction algorithm, realized using neural networks [5]. In this case a new neuron model is used characterized by an adaptive spline activation function [9]. The main difference with respect to related methods lies in input signal to the block: related methods use a full band signal while we use a sub band signal. In both, linear and non linear prediction, a forward and backward prediction is performed and results are combined by a cross-fade operation. Prediction quality decreases increasing the length of predicted samples [4], so weighing windows are used. Their expression is [5]:

$$\begin{cases} w_{fw}[n] = 0.5\left(1+\cos\left(\frac{n}{L_k}p\right)\right) \\ w_{bw}[n] = 0.5\left(1-\cos\left(\frac{n}{L_k}p\right)\right) \end{cases} \qquad (1)$$

where L_k is the number of missing data in j-th subband sequence and n=0,..., L_{k-1}. This operation is performed on every output signal from analysis filter bank and then subband reconstructed signals are passed to synthesis filter bank to obtain the full band output signal.

4.1 Linear prediction

Linear prediction is performed in two distinct stages: at first predictor parameters are estimated, then filter output is computed, repeating this operation L_k times. Filter parameters are computed using Levinson-Durbin algorithm. Forward prediction output is given by:

$$x_f[n] = \sum_{k=1}^{P} a_k x[n-k] \qquad (2)$$

where n=0,..., L_{k-1}, while backward prediction output is given by:

$$x_b[n] = \sum_{k=1}^{P} b_k x[n+k] \qquad (3)$$

where n= L_{k-1},...,0. Predictor order P=2 L_k was chosen.

4.2. Non linear prediction

An extension of linear prediction to non-linear case can be realized using neural networks, as proposed in [5]. It is known that to solve non linear problem with neural net a multi layer perceptron (MLP) is required. This net architecture causes an increase in computational cost, so it is used a net with, a new neuron model containing ad adaptive parametric spline activation function [9]. Networks designed build with such neurons are still universal approximators, maintaining good generalization capabilities [10]. Network architecture use is composed by only one such neuron, so forward and backward prediction outputs are given by:

$$x_f[n] = f\left(\sum_{k=1}^{P} w_k x[n-k] + bias \cdot w_0\right) \qquad (4)$$

$$x_b[n] = f\left(\sum_{k=1}^{P} w_k x[n+k] + bias \cdot w_0\right) \qquad (5)$$

where f(.) is neuron non linear adaptive activation function, which is initialised as bipolar sigmoidal. It is clear that if f(.) is linear and $bias$=0 equations (4), (5) fall into (2) and (3). As for linear predictor net inputs number is chosen equal to missing data length number L_k. Because missing data length L_k varies from application to another net inputs number is not constant. For this reason the nets used are not pre-trained, but they are run time created and all synaptic connections are initialised with linear predictor parameters. Then, for each j-th subband sequence all data, on the left and on the right of missing samples gap, are collected in a training set. Such initialisation is a good starting point for minimum searching on error surface and after few iterations (maximum 10) better results respect with linear predictor can usually be obtained.

5. EXPERIMENTAL RESULTS

Several tests have been carried out using four possible architectures obtained combining two filter banks and two prediction methods. Results obtained on different musical genres are reported, each for several different length of missing data, starting from 200 points , which correspond to 4.5ms at digital audio frequency sampling rate, to 5000 samples, which correspond to 113ms. For each test two numerical evaluating criterions are give and one subjective. Numerical criterions are give by maximum reconstruction error and mean square reconstruction error (MSE) given by:

$$ERR_{MSE} = 10\log_{10}\left[\frac{1}{L}\sum_{n=0}^{L-1}(x[n]-y[n])^2\right] \qquad (6)$$

where L is missing data length, $x[n]$ is original sequence and $y[n]$ is reconstructed one. Because human ear is considered a good judge in musical field, reconstructed sequences are submit to judgment of three musical experts, who gave a value from one (very bad) to five (optimum), and the mean is reported.

In all cases results show satisfactory reconstructions within 1500 missing data samples (about 34 ms). For longer gaps quality of reconstructed signal remains acceptable until over 110 ms gap, while audible effect have been detected in symphony orchestra signal. Fig. 4, 5 and 6 show the numerical results for different number of missing samples recovering.

The original and reconstructed signal waveform showed seem to be very similar, but not identical. An example of samples reconstructed sequence of missing data is presented in Fig. 7.

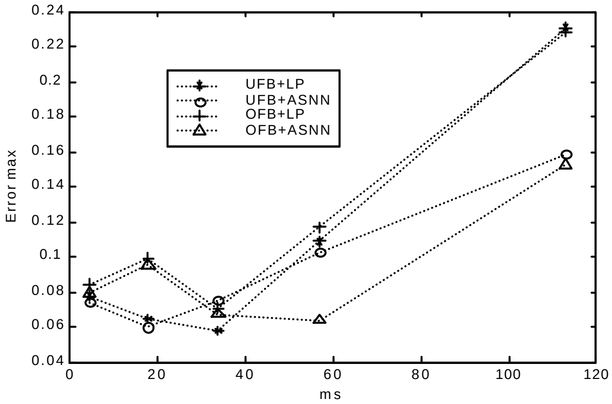

Figure 4 - The maximum error vs signal reconstructed length.

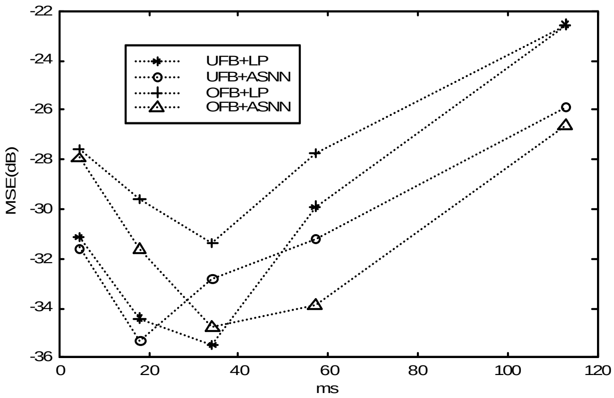

Figure 5 - The MSE error vs signal reconstructed length.

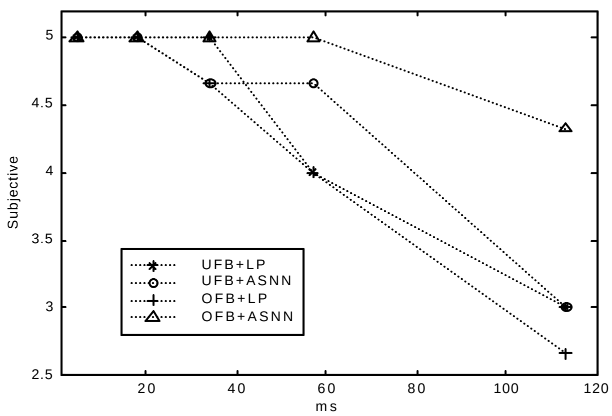

Figure 6 - The subjective opinion vs signal reconstructed length.

6. CONCLUSIONS

In this paper a subbands architecture for audio signal recovering, using neural nonlinear prediction based on adaptive spline neural networks has been presented.

Several experiments have been carried out demonstrating the effectiveness of the proposed method.

Although the numerical results are similar for linear and nonlinear approach, subjective analysis performed by expert musicians (see Fig. 6), shows better results for the neural networks method combined with octave filter bank.

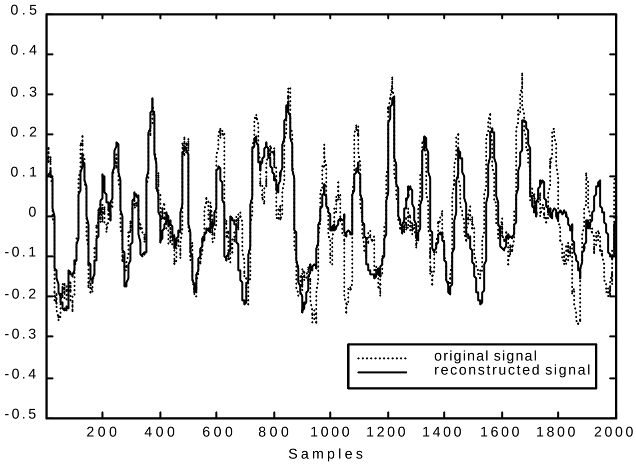

Figure 7 -Example of reconstruction of 2000 samples (45ms) by octave filter bank and spline neural networks.

7. REFERENCES

[1] Nieminem, P. Heinonen, Y. Neuvo, "Suppression and detection of impulsive type interference using adaptive median hybrid filter", *IEEE Proc. ICASSP 1987.*

[2] S.V. Vaseghi, P.J.W. Rayner, "Detection and suppression of impulsive noise in speech communication system", *IEE Proceedings*, Vol. 137, Pt. 1, No. 1, February 1990.

[3] S.V. Vaseghi, R. Franyling-Cork, "Restoration of old gramophone recordings", *J. Audio Eng. Soc.,* Vol. 40, No. 10, October 1992.

[4] S.V. Vaseghi, *Advanced signal processing and digital noise reduction*, Wiley-Teubner, 1996.

[5] A. Czyzewsky, "Artificial intelligence-based processing of old audio recordings", A.E.S. Preprint 3885 (F-6), San Francisco 1994

[6] R.C. Maher, "A method for extrapolation of missing digital audio data", *J. Audio Eng. Soc.*, Vol. 42, No. 5, May 1994.

[7] N.J. Fleige, *Mulirate Digital Signal Processing (Multirate systems, Filter Banks, Wavelet),* John Wiley 'n Sons, 1994.

[8] T.Q. Nguyen, "Near-Perfect Reconstruction Pseudo-QMF Banks", *IEEE Transactions on Signal Processing*, Vol. 42, No.1, January 1994.

[9] S. Guarnieri, F. Piazza and A. Uncini, "Multilayer Feedforward Networks with Adaptive Spline Activation Function", *IEEE Trans. on Neural Network*, Vol. 10, No. 3, pp.672-683, May 1999

[10] L.Vecci, F. Piazza and A. Uncini, "Learning and approximation capabilities of Adaptive Spline Activation Functions Neural Networks", *Neural Networks*, Vol.11 N.2, March 1998.