Alias minimization of 1-D signals using learning based method

Prashant Garg, Mohit Maheshwari, Sameer Dubey, Manjunath V. Joshi,
Vijaykumar Chakka, Asim Banerjee

Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, Gujarat, India.
e-mail: {prashant_garg, mohit_maheshwari, sameer_dubey, , mv_joshi, vijaykumar_chakka, asim_banerjee}@daiict.ac.in

ABSTRACT

In this paper, we propose a learning based approach for alias minimization of 1-D signals. Given an under-sampled (aliased) test speech signal and a training set consisting of several speech signals each of which are under-sampled (aliased) as well as sampled greater than or equal to Nyquist rate (non aliased), we learn the aliased frequencies in the test signal. Both the test signal samples and each of the under-sampled training set (database) signals are first interpolated to the size of the unaliased database signals. The test signal and each of the training set signals are then divided into a number of segments and the Discrete Cosine Transform (DCT) is computed for each segment. Assuming that the lower frequencies are unaliased and minimally distorted (due to interpolation), we replace the aliased DCT coefficients with those learnt from the training set. The unaliased test signal is then reconstructed by taking the inverse DCT. The comparison with the standard interpolation techniques in terms of both subjective and quantitative analysis indicates better performance while using the proposed approach.

1. INTRODUCTION

Sampling of a bandlimited signal is the first step towards digital world. A signal which is limited in time cannot be bandlimited and vice-versa. This makes aliasing inevitable whenever a finite duration signal is sampled. An infinite duration signal bandlimited to f_m Hz sampled at a rate greater than or equal to twice of f_m can be recovered using an ideal low pass filter. In practice this is not possible since the signal itself is time limited (hence band unlimited) and an ideal filter is non causal and hence cannot be realised. The sampling of a time limited signal with acceptable aliasing can be done by using an anti-aliasing filter. In our work, we show that even though we sample the output of the anti-aliasing filter at a rate less then the Nyquist rate, we can still learn the aliased region in frequency domain using the proposed learning based approach. The advantage of our approach is that it is possible to transmit a signal by undersampling it and recover the same at the receiver. It is a definite advantage as one need not sample the signal at a rate greater than or equal to Nyquist rate and hence the required bandwidth is reduced. This method can also be used for reducing the data storage on portable devices such as CD's, flash memory etc. However the approach needs a larger memory for learning the aliasing as we need a database of correctly sampled signals.

Our work is motivated by the work on image super-resolution by the image processing community. Super-resolution refers to algorithmic approach to attain a high spatial resolution image given a set of low resolution images. Here different approaches have been proposed for obtaining high spatial resolution image, given the image captured at low spatial resolution. Recently learning based approaches have been proposed for obtaining non-aliased and non-blur images using a database of training images at high resolution as well as at low resolution.

It is important to note that our work differs from the work on multirate signal processing where the researchers attempt to design efficient decimators and interpolators for sampling rate conversion in digital domain. Here the objective is to change the sampling rate without introducing aliasing and without going into the analog domain. However we allow aliasing by undersampling the given signal itself and still recover the aliased portion of the signal. To the best of our knowledge, we do not see such a work carried out in signal processing research community.

2. BLOCK DIAGRAM DESCRIPTION

The block diagram shown in Fig. 1 gives a brief overview of our approach. A 1-D test signal of finite duration 'T' is under-sampled after it is passed through an anti-alias filter using decimation thus giving an aliased output. Aliasing occurs because of overlapping of spectrum in the high frequency region. This aliased signal is then upsampled using interpolation approach so that it has the same number of samples as that of the properly sampled signals present in the database. The signal set in the database is also under-sampled and interpolated individually as the test signal. Thus we have an interpolated test signal for which aliasing has to be learned and a set consisting of interpolated versions of under-sampled signals and their corresponding properly sampled signals (non aliased). Once this is done, Discrete Cosine transform (DCT) of the interpolated test signal and the interpolated signals of the database is computed by dividing them into a number of segments. Assuming that the lower frequency values remain unaliased, the aliased DCT coefficients are decided by a threshold value (threshold is deduced experimentally) of the test signal and each of the database signals are compared using sum of the squared differences of their amplitudes to find a closest approximation for the aliased part of a particular segment. The corresponding DCT coefficients are picked up from the properly sampled database and are used to replace the aliased DCT coefficients. This is repeated for every segment in the test signal. In the end, the Inverse DCT is computed after learning all segments, which is then passed through a Low Pass Filter (LPF) to get the reconstructed analog signal.

The Fig. 2 illustrates the time and frequency domain representations when the signal is sampled at different rates. Fig, 2(d) shows the aliasing when sampled at a rate less than the Nyquist rate. In our approach, we learn the DCT coefficients in the aliased region by dividing the test signal as well as database signals into a number of segments and by picking up unaliased DCT coefficients in the properly sampled signals to reconstruct an undistorted test signal.

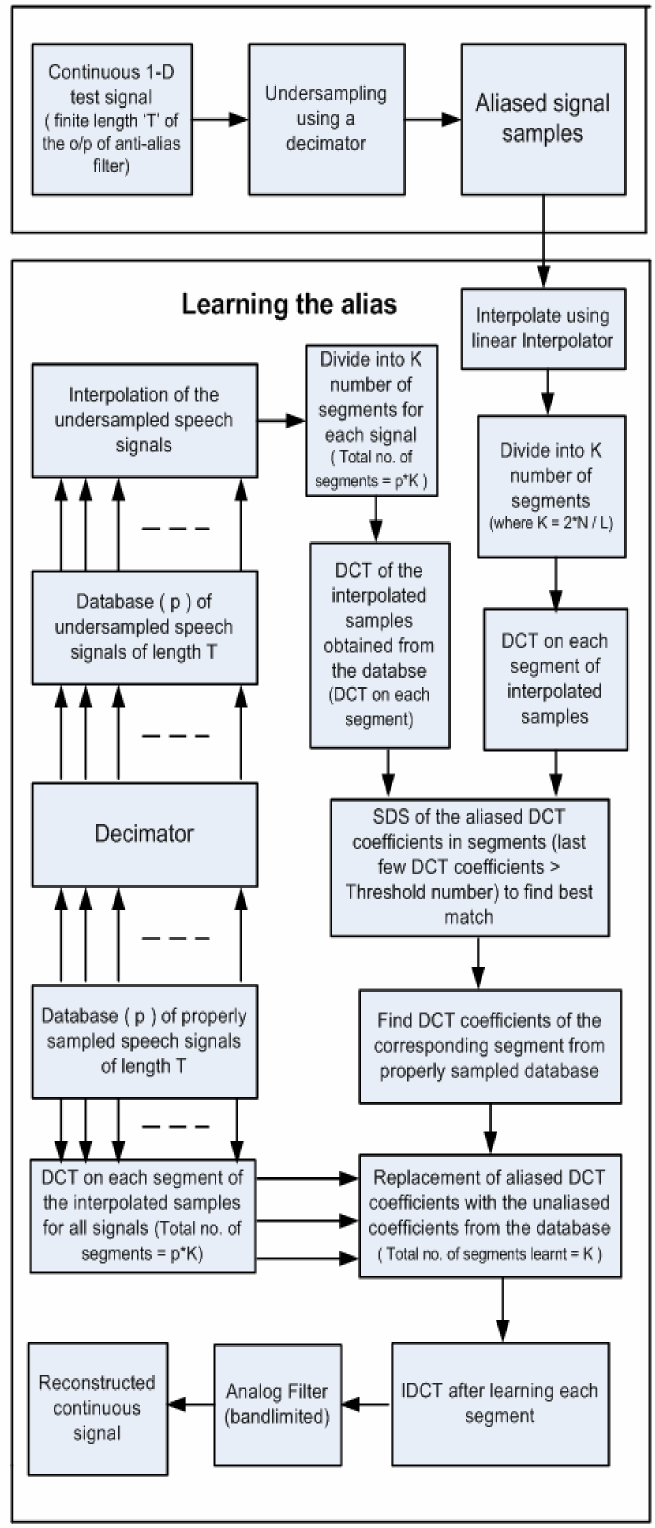

Fig1: Block Diagram illustrating the proposed approach

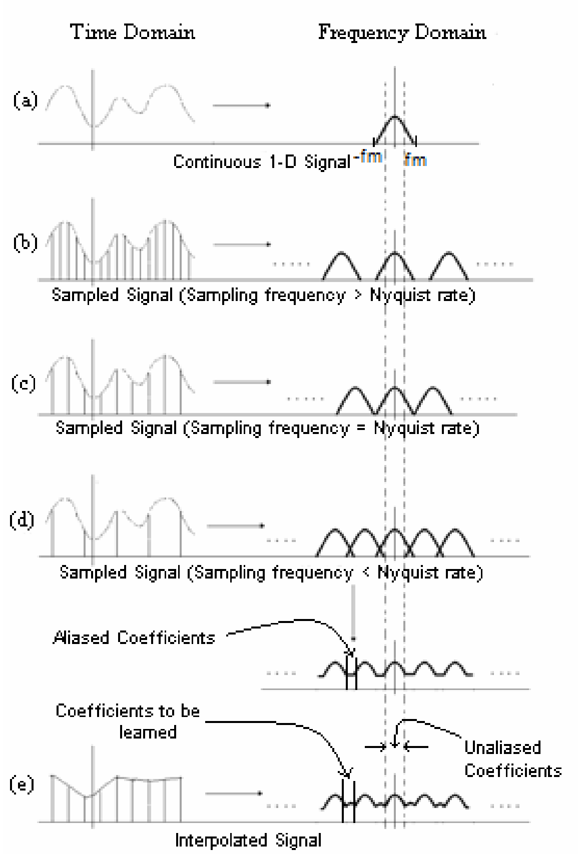

Fig 2: Diagram explaining variations in spectrum with different sampling rates and interpolation

3. LEARNING THE ALIASING USING DISCRETE COSINE TRANSFORM (DCT) BASED LEARNING

Here each set in the database consists of a pair of undersampled and properly sampled signals. After undersampling the signals of duration T sec., the test signal and the undersampled training signals in the database have a length of N samples. The corresponding properly sampled training signals have a size of 2*N. We first upsample the test signal and all the undersampled training images by a factor of α by using a standard interpolation technique. For example, If the test signal and the undersampled training signals have N samples for a duration of T seconds and the unaliased signals in the training set have 2*N samples then an interpolation factor of 2 is used to match all the signals to have same number of samples. We now divide each of the signal into segments of length L. The motivation for dividing into number of segments is due to the theory of JPEG compression where an image is divided into 8*8 blocks and the comparison is achieved by quantizing the DCT coefficients such that high frequency components tend to 0. In our case, we are interested in learning the aliasing for test signal (i.e. estimating the non-aliased frequency components corresponding to the aliased ones) using a training data set of aliased and non-aliased signals. This is done by taking DCT on each segment for all the sampled signals.

We learn DCT coefficients for each segment in the test signal by replacing the corresponding DCT coefficients from segments in properly sampled signals in the database Fig(3). It is reasonable to assume that the distortion caused due to interpolation is minimum for the low frequency. Hence, we learn only the DCT coefficients representing the high frequency.

Let $S_T(i)$, $1 \le i \le L$ represent the DCT coefficients in the interpolated segment of test signal. There are K (where K= 2*N/L) segments for each of the signals. Let

$$S_{Tu}^{(m)}(i), 1 \le i \le L$$

$$S_{TNyq}^{(m)}(i), 1 \le i \le L$$

where, m=1,2,……,p be the DCT coefficients undersampled and properly sampled signals respectively. If ‘p’ represents the total number of training pairs, then we have DCT segments for the training data set as 2*K*p.

Now the best matching unaliased segment for the considered test signal segment is obtained as

$$\hat{m} = \min_{in\_all\_K*P} \left[\sum_{i>threshold}^{L} (S_T(i) - S_{Tu}^{(m)}(i))^2 \right]$$

Here $\hat{m}$ is the index of the training signal which gives the minimum for the segment. Now these best matched DCT coefficients from the corresponding properly sampled segment are copied into corresponding locations in the considered segment of undersampled test signal (threshold $\le$ i $\le$ L). The learned test segment can be written as,

$$\hat{S}_T(i) = \begin{cases} S_{TNyq}^{(\hat{m})}(i); i \ge threshold \\ S_T(i); else \end{cases}$$

This is repeated for every test signal segment. We conducted the experiments by using the different values for segment length with different thresholds. We finally take the inverse DCT to get the unaliased signal samples. This is passed through a Low Pass Filter (LPF) to obtain the analog signal.

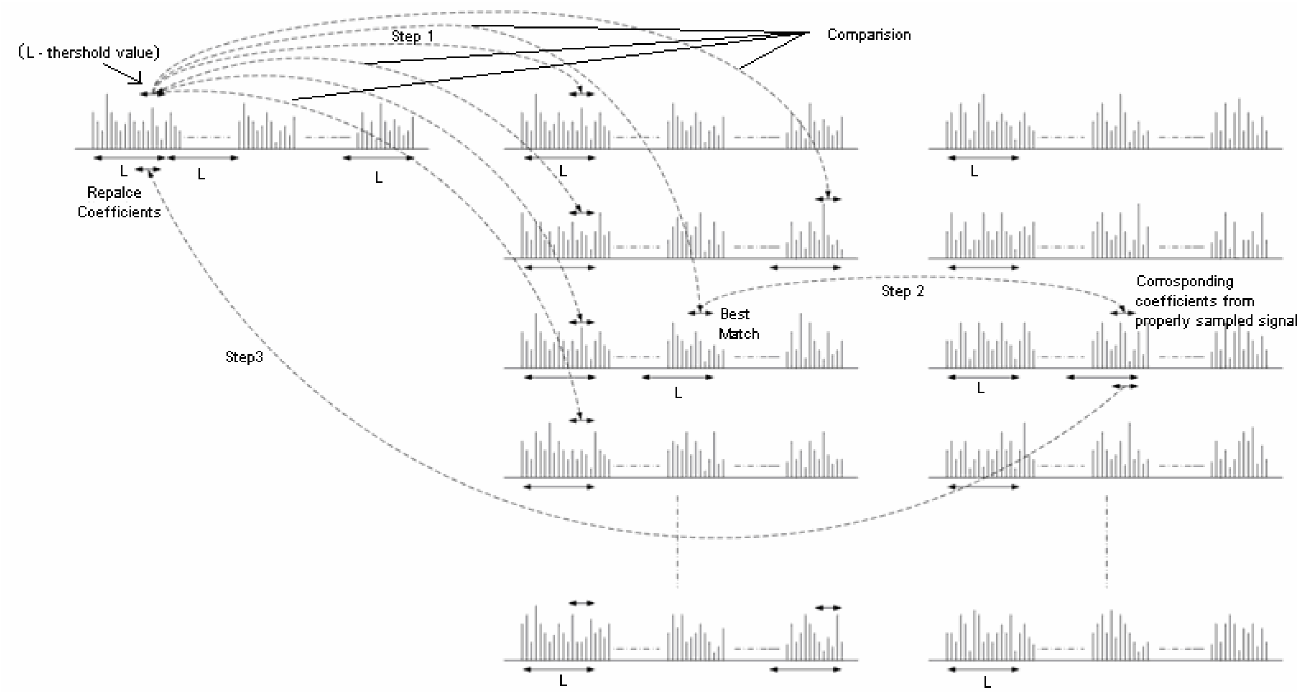

Fig 3: Pictorial representation of proposed learning approach

For quantitative analysis, Mean Squared Error (MSE) is then computed between interpolated and reconstructed by taking original as a reference, where MSE between two signals S_T and $\hat{S}_T$ is

$$MSE = \frac{\left(\sum_{j=1}^{L}\left(S_T(i) - \hat{S}_T(i)\right)^2\right)}{\left(\sum_{j=1}^{L}\left(\hat{S}_T(i)\right)^2\right)}$$

where, L= length of the signal

If MSE of reconstructed signal is greater than interpolated signal, then reconstructed signal is a closer match to the original signal as compared to the interpolated one and vice versa. Another Quantitative analysis measure taken is Signal-to-Noise Ratio (SNR) calculation. The SNR is computed using the following formula:

$$SNR = 20\log_{10}\left(\frac{\left(\sum_{j=1}^{L}\left(S_T(i)\right)^2\right)}{\left(\sum_{j=1}^{L}\left(S_T(i) - \hat{S}_T(i)\right)^2\right)}\right)$$

i.e. SNR = Total energy within original signal / Total Energy Error

S_T = original signal;

$\hat{S}_T$ = reconstructed signal or interpolated signal.

4. Experimental Results

The experiments have been conducted on a test signal from one of Gandhiji's speech. The test signal taken has a duration of 5 sec. and a maximum frequency of 4 KHz. The training data in the database consist of 10 properly sampled speech signals (Sampled at 12 KHz) and 10 under-sampled speech signals (Sampled at 6 KHz) of the same duration.

We first conducted the experiment to find an interpolation technique which gave us the best result in our case. We found that the Mean Square Error of Interpolated Signal and Test Signal turned out to be least in case of Linear Interpolation. So, we have used

Linear Interpolation technique for interpolating test signal and database signals.

Table 1 Comparison of Interpolation Techniques

Interpolation method	MSE Values
Spline	0.4361
Cubic	0.4187
Linear	0.4163

By repeatedly performing the experiment for different DCT points with the database consisting of speech signals in the voice of the same person, we obtain Normalized Mean Squared Error (MSE) values and SNR values for interpolated signal and reconstructed signal (Results obtained using Spline and Cubic Interpolation gave larger MSE values as compared to Linear Interpolation.).The final results for MSE and SNR comparisons are shown in Table 2 and 3 respectively. Graphs plotted for different stages of the experiments have been shown in Fig.4. The final results obtained are in accordance with the conceived approach.

Table 2 Quantitative Analysis using Mean Squared Error

DCT Point	Samples compared	Samples learned	MSE Values	
			Interpolated	Constructed
16	7	7	0.4163	0.4117
16	8	8	0.4163	0.4018
16	9	9	0.4163	0.4030
64	34	34	0.4163	0.3956
64	37	37	0.4163	0.3939
64	40	40	0.4163	0.3971
128	58	58	0.4163	0.3978
128	64	64	0.4163	0.3957
128	70	70	0.4163	0.3938
256	131	131	0.4163	0.3950
256	136	136	0.4163	0.3945
256	141	141	0.4163	0.3935

Table 3 Quantitative analysis using SNR

DCT Point	Samples compared	Samples learned	SNR values (in dB)	
			Interpolated	Constructed
16	7	7	10.8926	22.6686
16	8	8	10.8926	22.7808
16	9	9	10.8926	20.5825
64	34	34	9.3539	18.5161
64	37	37	9.3539	31.5404
64	40	40	9.3539	30.6090
128	58	58	9.3539	33.3615
128	64	64	9.3539	33.5280
128	70	70	9.3539	35.5787
256	131	131	9.3539	34.8428
256	136	136	9.3539	34.8895
256	141	141	9.3539	38.3260

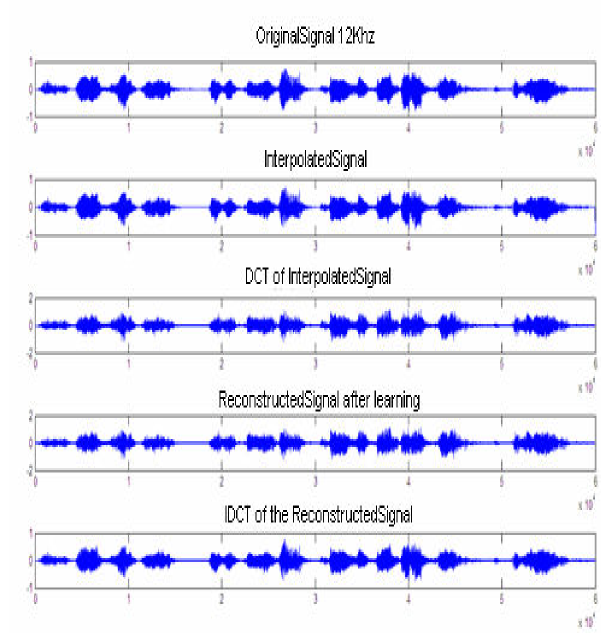

Fig 4: Graphs plotted for various stages of the experiment

5. CONCLUSION

It is evident that learning based techniques are not only useful for image restoration techniques but are equally effectively for speech signals as well. The results obtained can be significantly improved by building sufficiently large databases. Large databases can be used in learning aliased samples irrespective of the person's voice in test signal. However, increasing the database size beyond a limit may lead to an increase in computation time as well.

6. REFERENCES

[1]