DESIGN AND IMPLEMENTATION OF DIFFERENT AUDIO RESTORATION TECHNIQUES FOR AUDIO DENOISING APPLICATIONS

Merin K Mathai[1], Deepa. J[2]

[1]PG Scholar, Department of Electronics and Communication, College of Engineering Chengannur, Kerala, India
[2]Associate Professor, Department of Electronics and Communication, College of Engineering Chengannur, Kerala, India

Abstract

Audio signals are corrupted with many types of distortions. Major audio distortions are categorized into Globalized and Localized distortions. Localized distortion includes clipping and clicks where only certain samples are affected and globalized distortions include broadband noise where complete bandwidth is consumed by noise. Audio restoration is a technique for giving back the audio signals from these distortions. In this paper, audio restoration techniques for removing clipping, clicks and broadband noise are put forwarded. Recent approaches to solving audio restoration problem is with respect to sparse representation algorithms. Clipping distortion is addressed with a Sparse representation framework, it is treated as a reverse problem, where the distorted samples is estimated from the surrounding undistorted samples, they are embedded in frame based scheme, and reconstructed by using an overlap add method in conjunction with OMP algorithm and Gabor/DCT dictionary for modelling audio signals. Broadband denoising is done by using spectral subtraction and Click removal is done by using an adaptive filter method as the first step. Performance measures are done based on perception, average SNR calculation and defined parameter variations. This paper also targeting towards the software and hardware implementation of the restoration methods using TMS320C6713 DSK kit with help of tools mainly MATLAB and Code Composer studio.

Key Words: *Audio Distortions, OMP algorithm, Gabor/DCT dictionary, TMS320C6713DSK*

---***---

1. INTRODUCTION

Audio signals are subjected to a number of degradation which is common in audio sources. Generally, these degradations are classified into Localized degradations and Global degradations. Localized degradations are affect only certain group of samples; it includes clicks, crackles, scratches, breakages and clipping. Global degradations are which affects all samples of the waveform and include broadband noise and certain types of non-linear distortion [2].Audio restoration systems are aimed to reduce or eliminate these types of audible defects present in audio signals, introduced by the recording or reproduction mechanisms, or resulting from deterioration of the recording medium. Most common types of distortions are Clicks, Clipping and Broadband noise, among which localized distortions clipping and clicks are very difficult to resolve. This paper proposes the methods to dig out the effective ways for eliminating these degradation from audio signals.

1.1 Clipping

Clipping is a waveform distortion that limits a signal to reach its peak amplitude. It usually occurs in phone sound transmissions. It may occur in two forms,

[1]. Hard clipping- In cases where the signal is strictly limited at the threshold, producing a flat cutoff, Fig-1 shows Hard clipping and normal signal waveforms.

[2]. Soft clipping- In cases where the clipped signal continues to follow the original at a reduced gain. Hard clipping normally results in high frequency harmonics; soft clipping results in fewer higher order harmonics and intermodulation distortion components. A number of methods are available to restore the clipped signals [7], [8], [9]. Among which major are frequency domain methods. Here we are employing a time domain approach which provides more efficient restoration of the clipped samples.
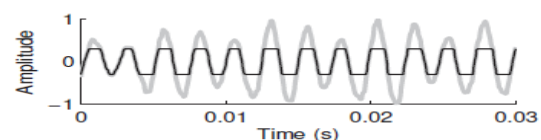

Fig -1: Clipping and normal signal waveforms

1.2 Clicks

Clicks in audio signals, shown in Fig-2; also known as Impulsive noise, consists of relatively short duration noise pulses, caused by different types of sources, such as switching noise, various channel environments in a communication system, dropouts or surface degradation of audio recordings like CD scratches, clicks from system keyboards, etc.
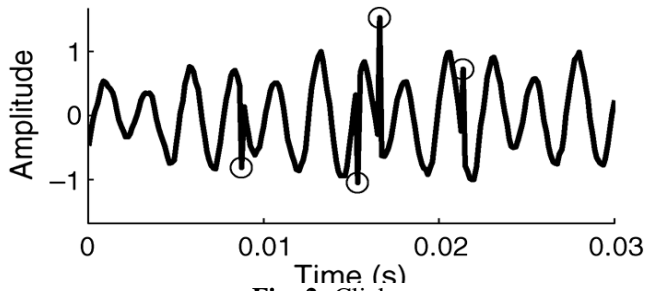
Fig -2: Clicks

1.3 Broadband Noise

Broadband noise is a noisy sound signal which has its energy distributed over a large section of the audible frequency range, also known as wideband noise. This noise spectrum consisting of a large number of frequency components, none of which is individually dominant.
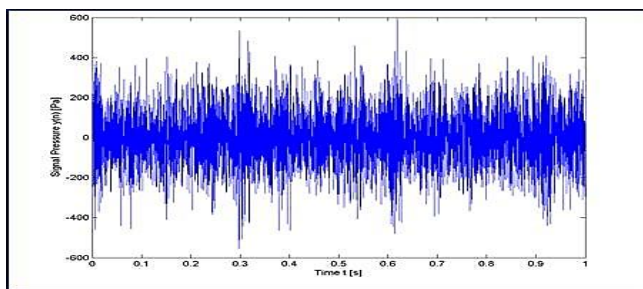

Fig -3: Broadband noise

Fig-3 shows the broadband noisy signal. This type of noise is often appeared as surface noise in old historical sound recordings.

2. PROPOSED METHOD

This project proposes the audio denoising methods for audio signals degraded by clipping, clicks and broadband noise by addressing these distortions individually. Clipping is addressed by SR framework leveraging from the concept of image inpainting, Clicks are addressed by using the basic principles of adaptive signal processing and Broadband noise is alleviated by spectral subtraction. All these methods are briefly described in this section, the Results and Discussion, Conclusion are presented in Section 3 and 4.

2.1 Declipping framework

Proposed framework is employed in time domain where the distorted data frames are estimated from the undistorted data surrounding by it. Sparse representations (SR) model frame work is used here to model the reliable and corrupted audio data co-efficients [1]. Currently, signal processing methods referred as the Sparse Representations are increasingly popular for solving underdetermined linear equation systems. Solving the problem of missing samples in signal processing using SR is termed as Inpainting.
Audio declipping problem can be analysed by using the SR modelling of the distorted audio signals. The proposed algorithm for addressing the declipping is defined below in steps. Here the audio data frames can be viewed as of vector of co-efficients like c ∈ R^L.

Proposed Algorithm for Declipping

Step 1. Segment the distorted audio samples into fixed number of frames of frame length N using analysis window w_a.
Step 2. Each frame is checked to find clipped and unclipped data.
Step 3. If the frame is corrupted, OMP algorithm used in conjunction with DCT/Gabor dictionary is used to reconstruct the corrupted frame using the reliable data in the respective frame.
Step 4. Restoration of complete audio signal using Synthesis window w_s by OLA method.

The audio declipping problem is defined as the recovery of the missing co-efficients based on the details of
1) Unclipped data in each frame
2) Clipping level
3) Observed signal properties such as sampling frequency, no: of bits, length etc.
In linear model, step 2 can be interpreted as $c_j(t) \triangleq c(t + t_j)w_a(t)$ is the windowed frame defined for $0 \leq t \leq N-1$. Step 3 followed by step 2 makes the estimate ($\hat{c}_j$) of clipped frame co-efficients using reliable data co-efficients with the help of an inpainting algorithm, Here Orthogonal Matching Pursuit (OMP) is used. The final phase, Step 4 the reconstruction of the full signal which is mathematically modelled as

$$\hat{c}(t) \triangleq \sum_j w_s (t - t_j)\hat{c}_j(t - t_j) \qquad (1)$$

where w_s is the synthesis window such that

$$\sum_j w_s (t - t_j)w_a (t - t_j) = 1, \forall t \ ,$$

a rectangular window for w_a and a sine window for w_s is used in this approach?
In the SR modelling, each frame can approximated as

$c_j \approx Dx_j$ where is $D \in R^{N \times K_D}$ the dictionary, $N \leq K_D$ and $x_j \in R^{K_D \times 1}$ is the representation vector of the jth frame. x_j is assumed to be sparse, i.e., to have few nonzero coefficients compared to in order to recover unknown samples.

Dictionary D used here is DCT/DGT dictionary. It is used by OMP algorithm [4] to perform the inpainting of an audio frame. General steps for OMP algorithm is described below. This approach emerges from the following optimization problem in Equation (2):

$$\hat{x}_i = \arg min\|x\|_0 \text{ s.t. } \|y_i^r - M_iDx\|_2^2 \leq \epsilon_i \qquad (2)$$

Inorder to solve Equation (2) , OMP algorithm is used here.

OMP algorithm

Input: $\phi \in R^{k \times N}$: The sampling matrix

$y \in R^k$: The measurements vector

m : The sparsity level of the signal x

ε_0 : OMP threshold

Output: $\tilde{x} \in R^N$: The estimate of the orginal signal

Procedure :

1. Intialize the residual $r_0 = y$ and the index set $\alpha_0 = \{\emptyset\}$. Set iteration counter , i=1
2. Find the index $\lambda_i = \text{argmax}_{j=1,...,N} |<r_{i-1} , \phi_j>|$
3. Augment the index set $\alpha_i = \alpha_{i-1} \cup \{\lambda_i\}$ and the matrix of chosen columns $\tilde{\phi}_i = [\tilde{\phi}_{i-1} \ \tilde{\phi}_{\lambda i}]$
4. Solve a least square problem for new estimate
   $\tilde{x} = \arg_x \min||y - \tilde{\phi}_i x||$
5. Compute the new residual $r_i = y - \tilde{\phi}_i x_i$
6. Increment counter i and return to step2 if i<m or $||r_i^2|| > \varepsilon_0$
   Algorithm stops when if i>m. or $||r_i^2|| > \varepsilon_0$
7. Retrieve the final estimate $\tilde{x}$.

In general audio data can either be waveform samples or coefficients in transforms like spectrograms. Here considering single dimensional signals only, this can also be extended to address formulation above can be used for multi-dimensional signals by simply considering the equivalent vectors.

2.2 Click Removal

In order to remove click from a degraded signal, an adaptive signal processing scheme is used. Therefore, adaptive filters can be used for noise cancellation in digital systems, in the same way, this can be applied to a clicked audio signal. Adaptive algorithms used for this purpose is LMS algorithm and NLMS algorithm [5]. Fig-4 shows an adaptive noise canceller designed for removing clicks
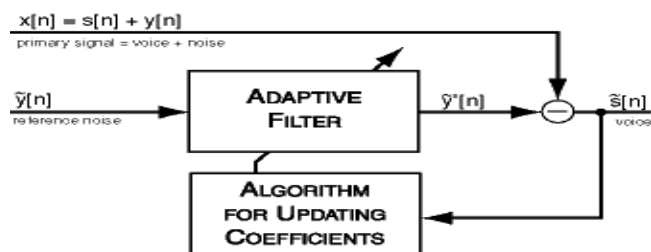

Fig-4: Click removal using ANC

where $s(n)$ is the clicked noisy signal and $y(n)$ is the clicked noise variant, the error signal $\hat{s}(n)$ is the desired output signal. Proper values for step size and filter order are carefully chosen for algorithm to obtain the desired output.

2.3 Broadband Denoising

Broad band denoising using spectral subtraction exploits the short time amplitude characteristics of speech and essentially performs a magnitude or power subtraction of the noise from the noisy speech. In its basic form it is a simple method that operates in the frequency domain to obtain a magnitude spectrum estimate of the noise and then use that estimate to filter the noisy signal[4]. Due to its popularity, many different variations of it have been developed. The degraded signal x(n) is modeled as a pure audio signal s(n) and a superimposed broadband noise d(n). The signal degradation is processed in the frequency domain using the short-time Fourier transform. For each frequency bin k and each frame m, a specific amount of the noise spectrum $|D(k)|^b$ is subtracted from the short-time spectrum $|X(m,k)|^b$ (b=1 for magnitude subtraction, and b=2 for power subtraction). The noise spectrum $|D(k)|$ has to be estimated from a noise only signal segment. For re-synthesis the denoised signal spectrum $|\ddot{S}(m,k)|$ is combined with the original phase spectrum $[X(m,k)]$. The spectral subtraction method can be modeled as time variant non-linear filter. Its transfer function depends on the signal-to-noise ratio $SNR(m,k)$ which is estimated from the short-time spectrum $|X(m,k)|$ and the noise spectrum $|D(k)|$. The noise variance within a single frame causes the SNR to be overestimated for some frequency bins. This results in a residual noise consisting of short sinusoidal impulses whose frequencies vary from frame to frame. This phenomenon is known as musical noise. The noise variance can be reduced by using a time-averaged signal spectrum instead of $|X(m,k)|$. This reduces the musical noise but without completely eliminating it. Block diagram of the designed broad denoiser is shown in Fig-5
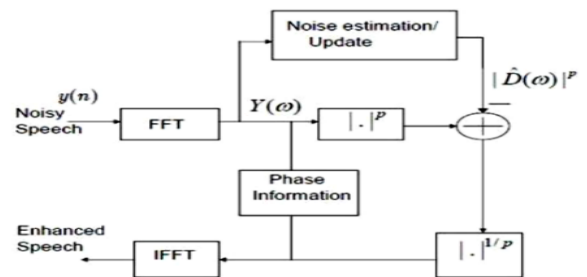

Fig-5: Broadband denoiser

Here, speech signals are only considered these are free from musical noise. Signal broken down into frequency bins, then processed and reconstructed using proper window. Here either Hamming/Hanning window can be used for obtaining the frequency bins, because they provide more frequency resolution than other windows.

3. RESULTS AND DISCUSSIONS

All the three distortions are addressed and the proposed restoration methods are implemented using MATLAB. Click removal is implemented as a part of hardware using TMS320C6713 DSK. Audio databases are downloaded and

also created using audio softwares. Database consisting of Speech and music signals   with a sampling frequencies of 16 KHz, 8 KHz, and 44.1 KHz are tested and verified the results by perception tests. Performance evaluation of each methods has been done and obtained results are shown below.

3.1 Declipping Results

Table-1 shows the average SNR value of the declipped estimate obtained as results. SNR value is calculated from the Equation (3). Datasets used are Speech signal at 8 KHz and Music signals at 16 KHz of 32ms frames are used.

$$SNR'_{full}(s, \hat{s}) \triangleq 10log\frac{\|s\|_2^2}{\|s-\hat{s}\|_2^2} \qquad (3)$$

Table -1: Average SNR of the restored signal as a function of clipping level

Clipping level	0.2	0.4	0.6	0.8	0.9
Music SNR(db.)	3.05	6.85	10.84	12.94	18.61
Speech SNR(db.)	2.66	4.17	16.07	24.39	20.57

It is clear that SNR of the declipped estimate is improved so better from and above the clipping level '0.4 in the case of both speech and music signals. Gabor dictionary is used for estimating the co-efficients. It is a fixed dictionary, hence much improved results can be obtained by using a learning dictionary. For the much better results are got with Speech signals , only slight variation is observed in the case of music signals.

Comparison between DCT and DGT dictionaries are also performed and its results for a speech signal is shown in Table-2

Table -2: Comparison of DGT and DCT

Clipping Level	DGT	DCT
0.1	1.0571	0.58740
0.2	4.5882	2.43983
0.3	8.63138	3.97389
0.4	7.85308	4.44919
0.5	15.9986	13.091
0.6	14.9121	13.7017
0.7	17.0139	12.8429
0.8	21.5961	12.3272
0.9	17.5716	12.7401

From the results, DGT is more apt to modelling the audio signals than DCT since it holds amplitude as well as free phase information. But at lower clipping levels (<0.5) DCT equalize and comparable with the performance of DGT.

Performance analysis of declipping is also done by varying the frame lengths depicted as f1, f2 and f3. Its values are f1=64, f2=128, f3=256. Chart-1 shows the comparative

change in average SNR value of the restored signal on the basis of clipping level and frame length. Music and speech signals is tested. Music signal tested results are shown here.
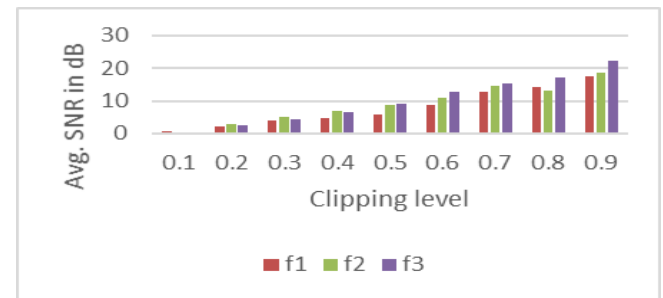

Chart-1: Comparison of frame lengths

It is clear from graph that at lower clipping levels (<=0.5), frames of small duration are more apt, here the adapted frame length from the graph is 128. But for higher clipping levels (>0.5), larger frames are suitable (adapted frame length=256), but frame length cannot be increased beyond a value, it increases the complexity level of the method, thus in turn increases execution time.

3.2 Broadband Denoising Results

From the plotted results of Fig-6 shows that by broadband denoising using spectral subtraction gives better results. This can be more improved by proper window placement spectral calculations. As the noise level increases(Gaussian noise), spectral subtraction performance also degrades, even though SNR level of the restored signal approaches about a constant value .It also maintains 5dB improvement from the noisy signal.
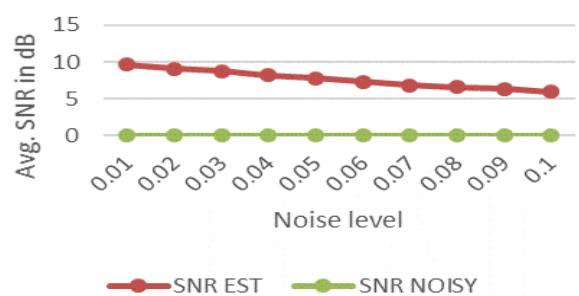

Fig-6: SNR plot of broadband denoised signal with variance to noise level

3.3 Adaptive Declicking Results

- Mean Square Error - MSE

The aim of an adaptive filter is to reduce the Mean Square Error. It is calculated using the following equation,

$$MSE = E (|d(n)-y_d(n)|^2) \qquad (4)$$

The values and graph of this quantity will be essential to evaluate the performance of the adaptive filter. If adaptive algorithm works well, after convergence time, the value of MSE should be reduced gradually to zero. Figures below shows the MSE plot of designed adaptive filter using the LMS and NLMS algorithm.
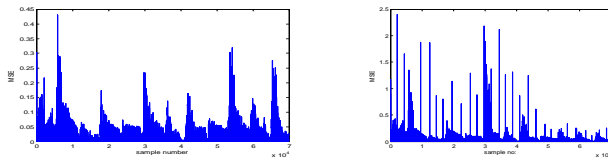
Fig-7: MSE plots for LMS and NLMS algorithms

From the results of Fig-7, the NLMS adaptive algorithm has less mean square error than that of LMS algorithm, since it has the feature of adapted step size than fixed step size in LMS Algorithm. More improved algorithms such as RLS can also be applied for the better adaptation of the signal.

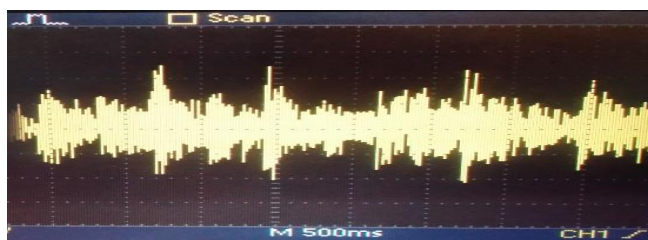## 3.4 Hardware Implementation Results

Adaptive filter method opted for removing the clicks is implemented using DSK [6]. Its implementation results are shown below in Fig-8 (a), (b), (c). (a) shows the click variant sent to the right channel of AIC23 codec in DSK as the input to adaptive filter , (b) shows the clicked noisy audio signal sent to the left channel and (c) obtained adaptive output resembles the desired signal which is observed using CRO. Fig.9 shows the hardware setup used for implementation purpose.
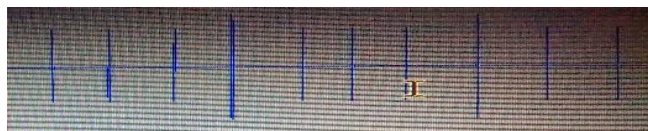
(a)   Click variant

(b)   Noisy signal

(c)   Output Signal

Fig-8: Hardware implemented waveform results

Fig-9: Hardware Setup

4. CONCLUSION

This paper is focused on audio inpainting algorithm using sparse representation of audio signals, removal of broadband noise and click using the proposed methods. The purpose was to study SR of audio signals and related audio inpainting problems. Proposed approach is worth studied and its software implementation is completed using Matlab. The project also targeting towards the hardware implementation of restoration methods using TMS320C6713 DSK .Overview of currently done work realized that audio inpainting problem is still current problem worth to study and resolve. Reconstruction of audio signals of commercial programs is not perfect and Sparse representation based audio inpainting algorithms can be nice tool for improving current situation. Spectral subtraction and adaptive algorithm approaches are also suitable for reducing the clicks and broadband distortion from the audio signals. It is still a current research problem. Adaptive click removal system is implemented in DSK and verified the results. In future, targeting towards the complete hardware implementation of complex OMP declipping algorithm using FPGAs and more implementations based on sparse signal representations for applications like Image Inpainting, Image denoising and Image Compression and addressing other major audio distortions.

REFERENCES

[1]. Adler A. "A Constrained Matching Pursuit Approach to Audio Declipping".Acoustics, Speech and Signal Processing, IEEE International Conference on (ICASSP 2011), IEEE, Prague, Czech Republic, 2011

[2]. S.J.Godsill. "Digital audio Restoration" Springer-Verlag, 1998.

[3]. J. Tropp, "Greed is good: Algorithmic results for sparse approximation," IEEE Trans. Inf. Theory, Oct. 2004.

[4]. S. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. Acoust. Speech, Signal Processing , vol ASSP-27, pp 113-120, 1979

[5]. Bernard Widrow and Samuel D. Stearns "Adaptive Signal Processing" [Book].

[6]. Texas Instruments Inc., TMS320C6713, TMS320C6713B Floating-Point Digital Signal Processors, SPRS186I, Dallas, TX, May 2004.

[7]. A.Janssen, R. Veldhuis, and "L. Vries, Adaptive interpolation of discrete- time signals that can be modeled as autoregressive processes", IEEE Trans. Acoustics, Speech and Signal Process., Apr.1986.

[8]. M. Lagrange and S. Marchand, "Long interpolation of audio signals using linear prediction in sinusoidal modeling", J. Audio Eng. Soc.2005

[9]. R. C. Maher, "A method for extrapolation of missing digital audio data," in Proc.95th AES Conv., 1993

BIOGRAPHIES

Merin K Mathai , M-Tech student in VLSI and Embedded Systems, College of Engineering Chengannur, Kerala, India. Current areas of interest is in ASIC design, Signal Processing , FPGA based design.
Email id: merinkmathai@gmail.com

Dr. Deepa J, Associate Professor, College Of Engineering, Chengannur, Kerala, India. Her Current areas of interests in Image processing, Signal processing and Bioinformatics
Email id: deepaj@ceconline.edu