# Projection Filter Based Framework Towards Improvement of Similarity Index Between Notes of Synthesizer and Indian Classical Instrument (Flute)

Dr. Abhijit Chitre[1], Dr. Kirti Wanjale[2], Dr. Araddhana Manisha Arvind Deshmukh[3], Dr. Shyamsunder P Kosbatwar[4], Dr. Sheela Hundekari[5], Avinash Somatkar[6]

[1]Associate Professor, Vishwakarma Institute of Information Technology, Pune.

[2]Associate Professor, Vishwakarma Institute of Information Technology, Pune.

[3]Associate Professor, Marathwada Mitra Mandal's College of Engineering, Pune.

[4]Associate Professor, Dnyanshree Institute of Engineering and Technology, Satara.

[5]Associate Professor, MIT ADT University, Loni kalbhor, Pune.

[6]Asstistant Professor, Vishwakarma Institute of Information Technology, Pune.

**Article Info**
**Page Number: 8325-8339**
**Publication Issue:**
**Vol. 71 No. 4 (2022)**

**Article History**
**Article Received: 25** March 2022
**Revised: 30** April 2022
**Accepted: 15** June 2022

**ABSTRACT:**
In this paper we propose an approach of signal synthesis which modifies the input signal to attain the quality of reference signal. The paper proposes the hybrid approach of adaptively modifying the input signal to attain the similarity of reference signal. Different approaches of signal modification using cascading configurations of adaptive filters prominently incorporating Affine Projection filters are presented with formulation the most optimum approach is presented. The adaptive filter approaches are demonstrated with inclusion of Empirical Mode decomposition technique to arrive at the most effective framework. The quality of the synthesized signal is evaluated from different experts to confirm the effectiveness of the proposed approach. The proposed framework is tested with the musical domain where the reference signal is recorded from flute and the input signal is recorded with electronic synthesizer. The effectiveness of the framework is also examined by calculating different mathematical distances and regression plots. The framework proposed to be a generic framework for signal modification as it is also tested with the signals recorded from sitar. To the best of authors' knowledge no work has been reported towards incorporating qualitative naturalness in the synthesized signal. This paper proposes solution for the same.

**Keywords:** Empirical mode decomposition (EMD), Least Men Sqaure (LMS), Recursive least square (RLS), Affine projection filter (AP), Interrater statistics, Kullback Leibler Divergence (KLD).

## 1. INTRODUCTION

Music is a key and vital aspect of human life. Different musical instruments gained popularity over the years. The role of different instruments started to be incorporated in a single instrument with the advent of synthesizers. A synthesizer is an electronic instrument capable of producing different frequencies. These synthesizers can imitate different instruments. Due to rapid development in audio processing techniques the accuracy of producing the tones of different instruments has been substantially increased. The reproduction of the western instruments like piano, saxophones, and drums is very effective in comparison with the Indian classical instruments. Recent advanced synthesizers are unable to mimic the music quality produced by the original Indian instruments which can be figured out by trained ear prominently when the instruments like Flute, Tabla are played on synthesizer. Indian classical music is highly influenced with the use of different instrument like Tabla, Sitar, Harmonium, Sarod, Flute etc. High fidelity reproduction of these instruments with synthesizer is still a need and challenging task.

Different approaches have been observed in the synthesis of music where most of the methods incorporate additive or subtractive synthesis [1]. Efforts have been observed in number of different approaches which include generating objective function for balancing the distance between synthesis of the target and smoothness in spectral domain [3], implementation of FM synthesis approach for instrument sound generation [4]. Literature also reveals different applications of adaptive filter configurations [2] where very few filter structures have been observed to be utilized in the direction of synthesis of instrument.

The concept of Empirical mode decomposition has been observed in regards with the analysis of the musical sound [6] but the use of the same technique for the purpose of synthesis seems to be possible. In this paper we have proposed a combination of different filter configurations along with the technique of empirical mode decomposition to achieve the synthesized signal which is qualitatively very close to the signal produced by the original instrument.

The significant highlights of this paper are as follows:

The technique of empirical mode decomposition along with different adaptive filter configurations to improve the quality of synthesis.
The unique proposed best suited configuration of combination of RLS and affine projection filters that will bring the synthesized signal qualitatively very close to the original signal. The effectiveness of the proposed approach is well supported when analysed using performance metrics such as structural similarity index (SSIM) and correlation coefficient along with the interrater statistics. The calculated mathematical distances between the density functions of the signals further support the utility of the proposed approach.

The rest of the paper is organized as follows: The section 2, throws light upon the different efforts observed in the literature to accomplish the synthesis. Various approaches taken with western and Indian classical instruments is been discussed. The following section focuses upon the different adaptive filter configurations which include Recursive Least square (RLS) and affine projection filter algorithm (AP). The section also discusses the empirical mode decomposition (EMD) technique to obtain Intrinsic Mode Functions (IMFs). Interrater statistics and four methods to evaluate the degree of agreement between different raters is also presented in the same section. The experimentation carried out with different proposed approaches with observations are presented in detailed in section 3. The discussion on experimentations, results and conclusions are placed in final section.

## 2. RELATED WORK

In last few years music technology has shown a rapid growth .The synthesizers are coming up with remarkable features and are able to mimic the instruments with great effect. Still it has been observed that the reproduction of the Indian instruments on synthesizer is substantially less effective in comparison with its western counterpart. The tonal quality can be differentiated with minimum efforts when the sound of synthesizer is compared with the sound of real Indian classical Instrument.

Different efforts have been observed in the literature which undertook the task of synthesis in variety of ways. Substantial efforts have also been observed in the domain of Raga and instrument identification. The approach of finite state models for the generation of Hindustani classical music was proposed in [7] where computer generated compositions were obtained using finite state model. Efforts were also seen in physically constructing the behaviour of instrument. The approach of building loaded tabla and mridanga drumheads are modelled as circularly symmetric composite membranes was discussed in [8]. Constructing an instrument electronically with the help of processor is also an approach which can be observed in some papers. Electronic sitar controller was built and discussed in [9] which uses sensor technology to extract gestural information from a performer, deducing music information such as pitch, pluck timing. Mathematical model based on fundamental frequency component along with extracted envelop is discussed in [10]. This work also presents mean square error as a performance parameter between the actual waveform and the waveform generated by the mathematical model.

The work proposed in this paper clearly contrasts from the different methodologies proposed and implemented in different literatures in the following way.

The effective use of different configurations of adaptive filter algorithms is made to generate the sound. Most of the approaches are focussed on producing the sound wherein the methodology proposed in the paper modifies the input sound of an instrument produced on synthesizer and brings is qualitatively much close to the sound recorded on the real instrument. The paper also evaluates the performance of different approaches and proposes the best combination for signal modification. The empirical mode decomposition (EMD) technique for generation of intrinsic mode functions is also taken into account to better the matching performance of produced sound. The interrater analysis is also presented in the paper in support to the proposed structure. Additional effort is made in terms of producing the data set necessary for the experimentation. Five Ragas were played on the synthesizer and the same rags were played on Flute. This is the most critical point for the entire philosophy of the paper. The demand of the experimentation was to record the same note on synthesize and on real instrument. So the dataset generated needed to be very specific. The framework proposed in this paper proves to be a generic one and can be applied to any domain which incorporates adaptive filter configurations. In this paper it is presented for musical domain. The applications which prominently need the tantalizingly close similarity with minimum mean square error can effectively use the proposed framework.

### 2.1 Empirical Mode Decomposition

This data analysis method is proposed by Huang et al. [16]. This EMD technique is thought of in regards with the experimentation described in this paper as it has a significant advantage over wavelet transform which is very effective technique for time frequency analysis. The advantage offered by EMD is the modes generated are completely empirical. Unlike wavelet transform the predefined basis functions are not available for EMD. The signal from musical instruments exhibits non linearity and non-stationary behaviour. Under these circumstances EMD proves to be better than wavelet as selecting appropriate wavelet is critical [16] from signal point of view which is tackled easily by EMD as it is generated from dataset so it is specific to dataset.

The heart of this technique is generating Intrinsic Mode Functions from the data which exhibits the energy extraction with different intrinsic time scales. The key feature of these IMFs is they have well defined Hilbert Transform which enables the calculation of instantaneous frequencies. These IMFs extraction is data dependent and is not governed by any predefined basis function. The extraction of local energy and the instantaneous frequency from the IMFs through the Hilbert transform can represent complete energy-frequency-time distribution of the data. The philosophy of Empirical Mode Decomposition (EMD) [18] begins by considering the local level oscillations in the data. If the evolutionary behaviour of a signal x(t) between two consecutive extrema is observed, then a (local) high-frequency part, or local detail, can be heuristically defined.

Figure 1 illustrates IMFs extracted from the signal in successive stages

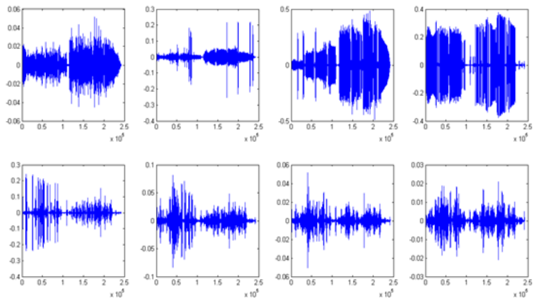


Fig.1 Empirical modes of signal

### 2.2 RLS Algorithm

It is based on the principle of least squares method. RLS algorithm with a large forgetting factor demonstrates a reduced steady-state misalignment at the cost of poor readaptation capability and with a small forgetting factor they exhibit an improved readaptation capability at the expense of an increased steady-state misalignment [14]. The key concept of RLS algorithm is the weight vector update formula which is the outcome of solution of minimization problem [14,15]

$$\text{minimize } \sum_{i=1}^{k} \lambda^{k-i} \ (d_i - x_i^T w_k)^2 \qquad (1)$$

In this equation $0 << \lambda << 1$ is the forgetting factor, $d_i$ and $x_i \in R^{M\times 1}$ are the desired signal and input signal vector at iteration i respectively and $w_i \in R^{M\times 1}$ is required weight vector at iteration k. The solution of the maximization problem in [1]is obtained by

$$w_{k=}R_k^{-1}P_k \qquad (2)$$

where $R_{k=\sum_{i=1}^{k}\lambda^{k-i} x_i x_i^T}$ and $P_{k=\sum_{i=1}^{k}\lambda^{k-i} x_i\ C}$ are approximations of autocorrelation. Matrix R and cross correlation vector of Weiner filter respectively. The autocorrelation and cross correlation vector can be expressed as

$$R_k = \lambda R_{k-1} + x_k x_k^T \qquad (3)$$

and

$$P_k = \lambda P_{k-1} + x_k d_k \qquad (4)$$

respectively

**2.3 Affine Projection Filter**

This algorithm originally emerged for convergence speed improvement when input signal does not exhibit flat spectrum [22,23]. The common thread of all AP algorithms if the filter update equation which considers N vectors of input signal (N is referred to as Projection order). Let Equation (5) represents the filter coefficients update equation.

$$\Delta w_{L1}[n] = w_{L1}[n] - w_{1L}[n-1] \qquad (5)$$

The Equation (6) needs to be modified under N constraints as given by Equation (7).

$$||\Delta w_{L1}[n]||^2 = \Delta w_{L_1}^T[n]\Delta w_{L1}[n] \qquad (6)$$

$$w_{L1}^T[n]X_{L1}[n-k] = d1[N-1] \quad k = 0,1 \ldots\ldots N-1 \qquad (7)$$

The solution of the problem leads to AP update equation given by

$$w_{L1}[n] = w_{L1}[n-1] + A^T[n](A[n]A^T[n])^{-1}e_{N1}[n] \qquad (8)$$

where

$$A[n] = (x_{L1}[n], x_{L1}[n-1], \ldots\ldots\ldots x_{L1}[n-N+1])^T \qquad (9)$$

and $e_{N1}[n]$ is a vector $N \times 1$ given by

$$e_{N1}[n] = d_{N1}[n] - A[n]w_{L1}[n-1] \qquad (10)$$

and $d_N[n]$ represents desired signal vector $N \times 1$

$$d_{N\tau}^T[n] = (d[n], d[n-\tau] \ldots\ldots\ldots d[n-(N-1)\tau])^T \qquad (11)$$

Weight update equation is given by

$$w_{L1}[n] = w_{L1}[n-1-\alpha(N-1)] + \mu A_\tau^T[n](A_\tau[n]A_\tau^T[n] + \delta I)^{-1}e_{N\tau}[n] \qquad (12)$$

where μ is step size and

$$e_{N\tau}[n] = d_{N\tau}[n] - A_\tau w_L[n - 1 - \alpha(N - 1)]$$

**2.4 Interrater Statistics**

The quantification of degree of agreement between different raters is indicated by interrater statistics. For the quality evaluation of the reconstructed signals the reproduced sounds were evaluated from 10 experts (7 experts and 3 non experts). Three correlation coefficients Spearman, Pearson, and Kendall were calculated.

**Pearson coefficient**

It is measure of association between two observations. The coefficient is given by

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{\bar{y}})}{\sqrt{\sum_{i=1}^{n}(xi - \bar{x})2\ (y_i - \bar{y})2}} \qquad (13)$$

where

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

**Spearman coefficient**

It is measure of degree of agreement between two observations. The coefficient is given by

$$\frac{\sum_{i=1}^{n}(\mathrm{rank}(x_i) - \mathrm{rank}(\overline{x}))(\mathrm{rank}(y_i) - \overline{\mathrm{rank}(\overline{y})})}{\sqrt{\sum_{i=1}^{n}(\mathrm{rank}(xi) - \mathrm{rank}(\overline{x}))2\ (\mathrm{rank}(y_i) - \mathrm{rank}(\overline{y}))2}} \qquad (14)$$

**Kendall coefficient**

The discrepancy between concordant and discordant pairs is given by Kendall coefficient. The coefficient is given by

$$\frac{\sum_{i=1}^{n} \mathrm{sgn}(x_i - x_j) \sum_{j=1}^{n} \mathrm{sgn}(y_i - y_j)}{n(n-1)} \qquad (15)$$

## 3 EXPERIMENTATION

### 3.1 Recording of Musical Note Samples

The experimentation first aims at generating data from original instrument. The field decided for experimentation is musical field where the reference signals are recorded from actual instruments which served the role of destination and the signals which are recorded from synthesizer play the role of input .Sitar and flute were considered for experimentation. The aim is to convert the input signal from synthesizer to the target signal recorded from actual Sitar and flute. Five Ragas were first recorded from sitar and flute which are Bhairav, Bhairavi, Todi, and Bhoop and a plain note Sa-Re-Ga-Ma-Pa-Dha-Ni-Sa. The selection of above Ragas serve the reference to generate the same notes on synthesizer. Instead of generating random notes these Ragas serve the reference datum for recording. The same Ragas were generated on synthesizer by playing the respective notes in flute mode of synthesizer. The tonal difference reflects the quality of synthesized musical note with original instrument note as a reference. The methodology of experimentation is showcased in figure 2.

### 3.2. Development of Synthesis Frameworks

To qualitatively modify the synthesizer signal, five different approaches were developed progressively and synthesis results were compared. Following diagram reveals the overall methodology implemented for the development of the framework approaches. For all these proposed and evaluated configurations the input is the signal from synthesizer and the target is the flute signal recorded from actual flute. The output synthesized signal is compared with the destination signal. For further analysis nomenclature of Approach 1 to Approach 5 is mentioned. Figure 3 shows the different approaches adopted for synthesis using adaptive filter configurations. The numerical values in the figure show Structural Similarity Index (SSIM) coefficient between original and reconstructed signal. Mathematical model for the most optimum approach is provided in Appendix.

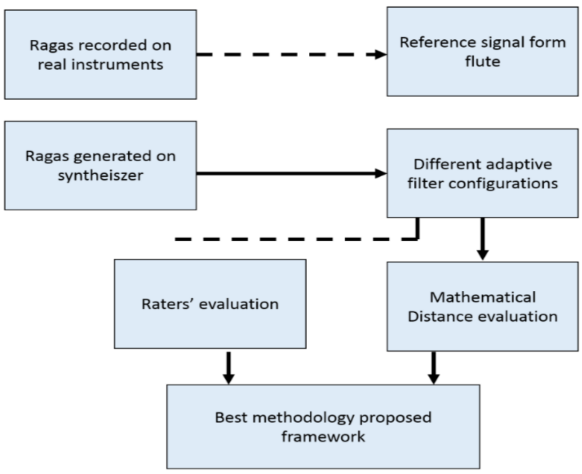
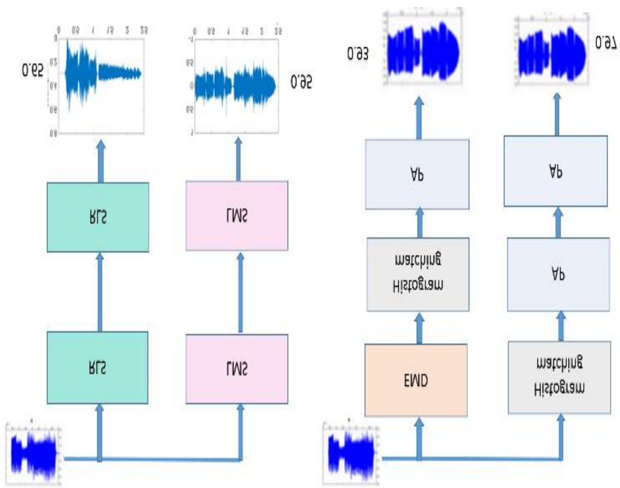
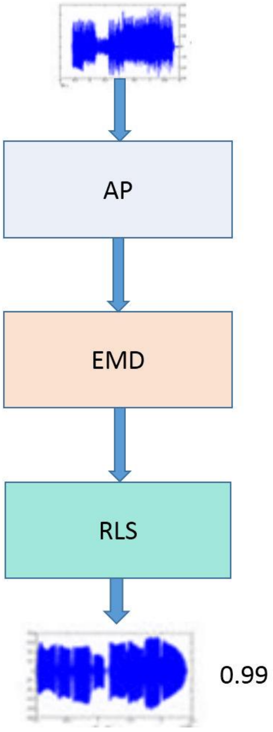



Fig.2 Methodology adopted for synthesis



(a) (b) (c) (d)




(e)

Fig. 3. Different adaptive configurations for synthesis with coefficient of similarity

Any signal is prominently characterized by spatial, spectral, and temporal features. These features mark the distinction between two signals[21]. Attack slope, Zero crossing rate, Roll off, Brightness, Irregularity, centroid, spread and skew are the prominent features considered for extractionm.Following table represents the parameters along with their significance, which are calculated and mentioned in Table 1. These parameters are evaluated for all the reconstructed signals from above mentioned approaches in Table 1.

Table 1. Signal parameters (Statistical, spatial, spectral).

| **Parameter** | **Significance** |
|---|---|
| Attack slope | Slope of attack |
| Zero crossing rate | Rate of sign changes along signal |
| Roll off | Indication of frequency bin F below which 85% of magnitude distribution is concentrated. |
| Brightness | Represents harmonic composition( high frequencies) |
| Irregularity | Variations in spectral behavior |
| Centroid | Represents spectral brightness of signal (spectral shape) |
| spread | Spread of frequencies in a signal |
| skew | Signifies Asymmetry of distribution |

Different parameters calculated for experimented approaches are demonstrated in Table 2. First row in Table 2 presents parameters of original signal and subsequent rows indicate parameters for Approach 1 to Approach 5 respectively. Last row of table indicates parameters evaluated for musical note of synthesizer.

Table 2. Parameters calculated for econstructed signals.

| **Attack_sl** | **ZCR** | **Rolloff** | **Regular** | **Centro** | **Sprea** | **Skew** | **Kurtos** | |
|---|---|---|---|---|---|---|---|---|
| 1060476 | 508.89 | 3794.5 | 0.34045 | 1904.7 | 3.3657 | 1.8E+0 | 7E+12 | 0.1136 |
| 1221763 | 512.81 | 4114.0 | 0.33567 | 2016.2 | 3.5462 | 1.74E+ | 6.54E+ | 0.1271 |
| 1074707 | 496.58 | 3792.8 | 0.34231 | 1907.7 | 3.3629 | 1.8E+0 | 7.01E+ | 0.1135 |
| 1228614 | 491.48 | 3489.5 | 0.34093 | 1754.0 | 2.9808 | 1.67E+ | 6.43E+ | 0.0775 |
| 1693558 | 523.02 | 3021.0 | 0.25502 | 1636.9 | 2.7875 | 2.79E+ | 1.31E+ | 0.0987 |
| 1091961 | 500.78 | 3888.0 | 0.34207 | 1941.9 | 3.4208 | 1.73E+ | 6.63E+ | 0.1145 |
| 593108.8 | 1067.7 | 7277.5 | 0.35204 | 2858.8 | 3.3513 | 3.45E+ | 1.41E+ | 0.1921 |

Mathematical distances, Bhattacharyya and Kullback Leibler are computed between PDF of original signal and PDFs of reconstructed signals.

Table 3. Mathematical distances between density functions of original and reconstructed signal by different approaches [24]

| | **Appr1** | **Appr2** | **Appr3** | **Appr4** | **Appr5** | **Synth** |
|---|---|---|---|---|---|---|
| **BHD** | 6.8exp(- 4) | 1.8exp(- 4) | 1.1exp(- 4) | 8.3exp(- 5) | 2.5exp(- 5) | 0.1173 |
| **KLD** | 0.113 | 0.005 | 0.0081 | 0.006 | 0.0021 | 0.275 |

The regression plots of features of original signal to the features of reconstructed signals are shown in figure 5

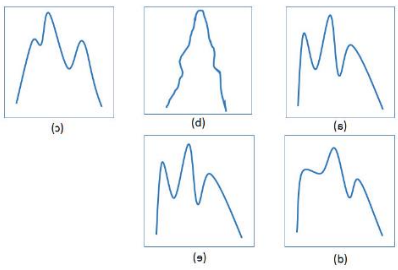


Fig. 4. Density functions of original signal and reconstructed signals with synthesis approach mentioned

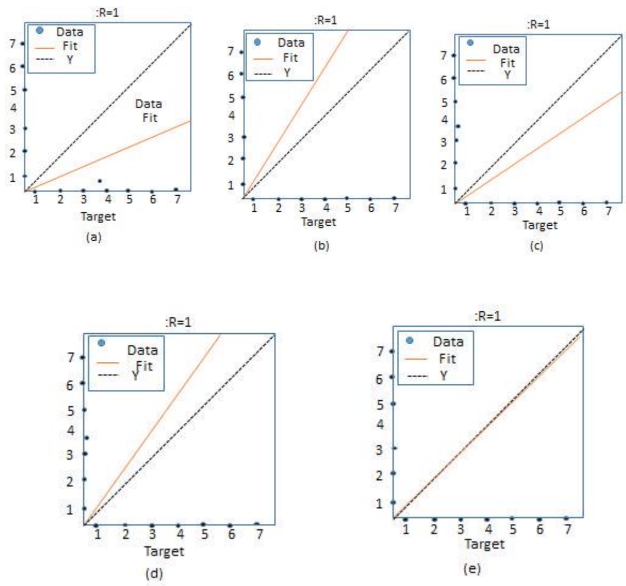



Fig.5. Regression Plots of the proposed synthesis approaches

### 3.3 Interrater Statistics

Along with the mathematical distances the reconstructed signals were also evaluated with respect to relative tonal quality of original signal from different raters. For the quality evaluation of the reconstructed signals the reproduced sounds were evaluated from 10 raters. (7 experts and 3 non experts). Initially only 3 evaluators were considered. 3 correlation coefficients were evaluated viz. Spearman, Pearson, and Kendall. Following table shows the values of these coefficients for Flute.

Table 4: Correlation Coefficients for rater

| **Rater** | **Spearman** | **Pearson** | **Kendall** |
|---|---|---|---|
| 1-2 | 0.9377 | 0.9654 | 0.8522 |
| 1-3 | 0.8866 | 0.9452 | 0.785 |
| 2-3 | 0.906 | 0.9462 | 0.804 |

Above observations were evaluated by four methods of calculating interclass correlation. The above experimentation is extended to 10 raters where 7 raters were experts and 3 raters were non experts in musical field. Table 5 shows the evaluation on the scale of 1:10 for reconstructed signals from various approaches with 10 as a benchmark for original note. Each rater is made to listen to the generated signal from the synthesis frameworks number of times along with the original signal. The ratings are the score of these signals with original signal as reference. So after number of iterations the score obtained. These score are validated by interater correlation and also by mathematical distances and regression plots.

Table 5: Raters evaluation for synthesis approaches

| | | | | | |
|---|---|---|---|---|---|
| Rater 1 | 6 | 7 | 7 | 8 | 10 |
| Rater 2 | 5 | 8 | 7 | 9 | 9 |
| Rater 3 | 6 | 8 | 8 | 8 | 10 |
| Rater 4 | 7 | 8 | 8 | 9 | 9 |
| Rater 5 | 7 | 7 | 8 | 9 | 9 |
| Rater 6 | 6 | 8 | 7 | 8 | 9 |
| Rater 7 | 6 | 8 | 7 | 8 | 9 |
| Rater 8 | 7 | 7 | 8 | 8 | 9 |
| Rater 9 | 7 | 8 | 7 | 8 | 9 |
| Rater 10 | 6 | 7 | 7 | 8 | 9 |

### 3.4 Evaluation of Similarity Measures between Original and Re- constructed Signals

MSSIM is the standard measure which indicates these structural similarity between two signals (The similarity between original signal and each reconstructed signal is calculated) [17]. Also Correlation coefficient is considered as the measure of similarity. Plot of similarity index between original signal and reconstructed signals with multiple approaches is shown in Figure 5.29 whereas plot of correlation coefficient between original signal and reconstructed signals with multiple approaches is shown in Figure 5

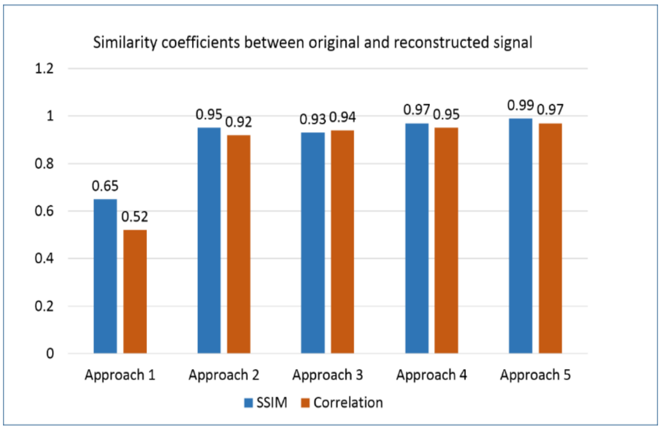



Figure 6: Similarity Index Evaluation between Signals generated by different Synthesis Approaches and Original Signal

### 3.5 Results and Discussion

Figure 3 shows the different frameworks adopted for modifying the signal from synthesizer to the desired signal generated form the real instrument. The figures also show the values of the similarity coefficients between reconstructed and original signal. From these values we can see that the reconstruction with RLS and cascaded RLS combination does not produce effective reconstruction. Even if the cascading stages are increased the similarity coefficient will saturate at a certain value and shows very less improvement. The similarity index of 0.65 is not at all acceptable when it come the perceived tonal quality and the discrepancy can easily be noticed. LMS algorithm gives better reconstruction than RLS and with cascading the similarity goes up to 95% with original signal. But still the difference of 5% is clearly noticeable especially when it is heard be expert ears. The difference of 5% is also significant in perceived tonal quality. The reconstruction quality shows saturation and does not reflect and signs of improvement with increase in the number of stages. Affine projection algorithm shows significance improvement but it also shows the saturation around similarity index of 0.965, after which increase in the number of stages does not reflect significant improvement in the quality of reconstructed signal. Even if this similarity is quite high and reconstruction seems to be very efficient, the hearing perception is not satisfactory. The difference of 0.04 is vital from tonal perception particularly for trained ears. These results reflect the fact that any one adaptive filter configuration is not able to produce reconstructed signal tantalizingly close to original signal. So we need a configuration which will exploit the positives from different basic filter structures .This fact can be observed by the cascaded combination of affine projection and RLS filter. The framework AP-EMD –RLS generates a reconstructed signal which is 0.993 similar to original and is the best reconstruction. Another approach of combining two domains for modification is also worked out and it proves effective. The combination of probabilistic and temporal domain combining Histogram matching with Affine Projection yields a signal reconstruction which is 97% similar to original. The result is less superior compared to AP-EMD_RLS combination but still very effective and generates signal which is also reasonably good from listeners point of view. One more configuration is worked out which also involve empirical mode decomposition the first stage followed by Histogram matching and EMD. Compared to only LMS and RLS it produces very good reconstruction with 92% similarity. So this experimentation shows that the combination of affine projection along with EM and RLS algorithm is the best combination for synthesis. The results of similarity index can be seen in figure 7. This structure is also tested with sitar and proved to be excellent as the reconstructed signal is almost 100% qualitative with original recorded signal. The results were verified with number of ways before confirming the fact. Figure shows the PDFs of the original signal and the reconstructed signal. The PDF corresponding to the reconstructed signal from approach (AP-EMD-RLS) shows complete match with the PDF of the original signal and it can be clearly seen that it is combination of multiple Gaussians which resembles the characteristics of the original instruments as the produce signal which shows Gaussian distribution. The same observations can be made about the PDFs of signals generated form other above mentioned approaches. The PDFs of approach 5 also resembles similarity with the PDF of original whereas the approaches which were not effective show PDFs of reconstructed signals which shows no similarity with PDF of original signal. The figure which show the regression plots for the signals also confirms the fact where the regression plot for the AP-EMD-RLS approach shows almost overlap. Also Table 4 demonstrates the mathematical distances for these reconstructed signals with the original signal. Approach showcasing combination of AP-EMD-RLS shows the minimum value for Bhattacharyya and Kullback Leibler distance. The reconstructed signals were also opined by the raters where 70% were experts and 30% were non experts. Table 6 shows the raters evaluation which also confirms the fact that the proposed framework of AP-EMD-RLS is the most effective framework which got maximum ratings. Figure 6 showcases the agreement of raters and it reflects almost 98% agreement among the raters. The experimentation produced three effective frameworks which produce the signal which resembles the original with similarity index greater than 0.96. One of the framework was based upon the novel approach of combination of probabilistic and temporal domains. It also reveals a generic framework to modify the musical note produced on synthesizer to qualitatively match with the note produced on the real instrument is proposed and the network is combination of Affine projection, EMD and RLS structures (AP-EMD-RLS).The best approach has been qualitatively treated with parameters similarity index, correlation coefficient and also with the mathematical distances Bhattacharyya and Kullback Libeller. All these parameters indicate the effectiveness of the proposed approach. The proposed approach outperforms other methods for combination of reasons. It includes combination of affine projection algorithm with EMD and RLS and it is evident from the fact that affine projection algorithm improves the performance of other adaptive filter configurations. The characteristic of affine projection is it combines the advantages of LMS and RLS algorithms. LMS gives weightage to current samples while RLS prominently focuses on the current samples and with the forgetting factor reduces the importance of the previous samples. The characteristics of Indian classical instruments especially flute does not show disconnects of present sample with the previous. The air column created inside the flute generates a memory based structures which also has a relation with the previous note. So any single algorithm shows a saturation in the improvement of similarity index after a certain value. The proposed combination first uses Affine projection filter which has an inherent property of giving weightage to present as well as past samples. Not only that it selects the past N best samples and not only the near past samples. This shows significant performance improvement. Also the stage following the AP is EMD which exploits oscillatory modes of signal and followed by RLS which again has a different property which in a way creates a disturbance and it does not allow the output to saturate and boosts quality significantly. Also initially the signals are completely uncorrelated and that is an undesirable scenario for RLS. Once Affine projection improves the correlation between the signals RLS works effectively.

So the combination proved to be the best as a generic combination which is also verified with sitar.

## 4. CONCLUSION

The experimentation produced three effective frameworks which produce the signal which qualitatively resembles the original with similarity index greater than 0.94. One of the framework was based upon the novel approach of combination of probabilistic and temporal domains. It also revealed a generic framework to modify the musical note produced on synthesizer to qualitatively match with the note produced on the real instrument is proposed and the network is combination of Affine projection, EMD and RLS structures (AP-EMD-RLS).The best approach has been qualitatively treated with similarity index, correlation coefficient and also with the mathematical distances Bhattacharyya and Kullback Libeler. All these parameters indicate the effectiveness of the proposed approach. The proposed approach outperforms other methods for combination of reasons. It includes combination of affine projection algorithm with RLS and it is evident from the fact that affine projection algorithm improves the performance of other adaptive filter configurations. The characteristic of affine projection is it combines the advantages of LMS and RLS algorithms. LMS gives weightage to current samples while RLS prominently focuses on the current samples and with the forgetting factor reduces the importance of the previous samples.

The characteristics of many practical signals does not show disconnects of present sample with the previous. This generates a memory based structures which also has a relation with the previous samples. So any single algorithm shows a saturation in the improvement of similarity index after a certain value. The proposed combination first uses Affine projection filter which has an inherent property of giving weightage to present as wel as past samples. Not only that it selects the past N best samples and not only the near past samples. This shows significant performance improvement. Also the stage following the AP is EMD which decomposes the signal into its different modes which are specifically data dependent and does not work with predefined basis functions like conventional transforms. So this stage work at the different components of the signal and give more accuracy in temporal matching of the signals. The final stage of the proposed model is RLS which again has a different property which boosts quality significantly. Also initially the signals are completely uncorrelated and that is an undesirable scenario for RLS. So once affine projection improve the correlation between the signals RLS works effectively.

The proposed framework is also experimented and tested with sitar and it resulted into output signal with similarity of 99.6% with actual signal recorded from sitar. So the proposed framework proves to be the generic framework of signal matching with least mean square between original and reconstructed signal.

## 5. REFERENCES

1. Matthias, R., Robert S., & Sylvain M. 2006. Fast additive sound synthesis using polynomials. In Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06) (pp. 18-20). Montreal, Canada.
2. Thakor, N. V., & Zhou, Y. S. 1991. Applications of adaptive filters to ECG analysis: noise cancellation and arythmia detection. IEEE Transaction on Biomedical Engineering, 38(8), 785-794.
3. Wager, S., Chen, L., Kim, M., & Raphael, C. 2017. Towards expressive instrument synthesis through smooth frame by reconstruction from string to woodwind. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017) (pp. 391-395). New Orleans, LA, USA.
4. Chowning, H. M. 1973. The synthesis of complex audio spectra by means of frequency modulation. Journal of Audio Engineering Society, 21(7).
5. Pakarinen, J., Karjalainen, M., V¨alim¨aki, V., & Bilbao, S. 2005. Energy behaviour in time varying fractional delay filters for physical modelling and synthesis of musical instruments. In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. (ICASSP 2005) (pp. iii/1-iii/4), 3, Philadelphia, PA, USA.

ISSN: 2094-0343
2326-9865


6. Fulton, M., & Soraghan, J. J. 2008. Modelling the I.M.Fs produced by the E.M.D of musical audio for synthesis and manipulation. In proceedings of the 15th International Conference on Systems, Signals and Image Processing (IWSSIP 2008) (pp. 229-232). Bratislava, Slovakia.
7. Das, D., & Choudhury, M. 2005. Finite state models for generation of Hindustani classical music. In Proceedings of Symposium on Frontiers of Research in Speech and Music.
8. Gaudet, S., Gauthier, C., & Léger, S. 2016. The evolution of harmonic Indian musical drums: A mathematical perspective. Journal of Sound and Vibration, 292, 388-394.
9. Kapur, A., Lazier, A. J., Davidson, P., Wilson, R. S., & Cook, P. R. 2004. The electronic sitar controller. In the Proceedings of 2004 Conference on New Interfaces for Musical Expression (NIME04) (pp.147-163). Hamamtsu, Japan.
10. Gunasekera, S. N., Wijayakulasooriya, J. V., & Perera, P. C. 2009. Natural synthesis of north Indian musical instruments. In proceedings of Fourth International Conference on Industrial and Information Systems, (ICIIS 2009) (pp. 369-373). Sri Lanka.
11. Ramirez, R., Hazan, A., Maestre, E., & Serra, X. 2008. A genetic rule based model of expressive performance for jazz saxophone. Computer Music Journal, 32(1), 38-50.
12. Gunasekera S. N., Wijayakulasooriya J. Y., & Perera S. P. C. 2015. Physically inspired synthesis of santoor and sarod. In proceedings of 10th IEEE International Conference on Industrial and Information Systems. (ICIIS 2015) (pp. 164-169). Sri Lanka.
13. Herrera-Boyer, P., Peeters G., & Dubnov S. 2003. Automatic classification of musical instrument sounds. Journal of New Music Research, 32(1), 3–21.
14. Bhotto, M. Z. A., & Antoniou, A. 2013. New improved recursive least-squares adaptive-filtering algorithms. IEEE Transactions on Circuits and Systems-I: Regular Papers, 60(6), 1548-1558.
15. Gonzalez, A., Ferrer, M., Albu, F., & Diego, M. D., 2012. Affine projection algorithms: evolution to smart and fast algorithms and applications. In proceedings of 20th European Signal Processing Conference (EUSIPCO 2012) (pp. 1965-1969). Bucharest, Romania.
16. Huang, N. E., Shen, Z., Steven R., Long, Manli C., Wu, Shih, H. H., Zheng, Q., Yen, N-C, Tung, C. C., & Liu, H. H. (998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. The Royal Society, 454, 903-995.
17. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612.
18. Peng, Z. K., Tse, P. W., & Chu, F. L., 2005. A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. Mechanical Systems and Signal Processing, 19(5), 974–988.