

# Mutual Information Based Framework of Quality Parameter Formulation for Similarity Index between Two Musical Notes



Abhijit V.Chitre, Aditya Abhyankar

**Abstract**—The paper presents an approach of formulating a parameter to access the relative quality of a musical note with reference to the standard musical note. The tonal quality perceived by an individual is prominently subjective. The work in the paper focusses on the objectivity of this comparison. The objectivity introduced essentially will be helpful to come up with at least basic conclusive step regarding the relative tonal quality. The proposed parameter is also evaluated against the rating given by the raters to the tonal quality of a musical note with reference to reference note to prove its effectiveness and agreement with raters. Harmonium and flute are taken as the instruments for experimentation. The approach proposed can be applied to any two musical notes produced with different instruments with one instrument reference. The quality parameter is proposed by evaluating different parameters of musical note and then using the technique of minimum redundancy maximum relevance (mRmR) technique of feature ranking. The Empirical way is formulated to develop the quality parameter which is also put to test against rankings of raters. Thus the paper provide a systematic way of deducing the quality parameter which will be in agreement with the tonal quality perceived by human ears.

**Keywords**— Bland Altman, Structural similarity index, Interrater Statistics, Kullback Leibler Divergence(KLD), Minimum redundancy maximum relevance(mRmR)

## I. INTRODUCTION

In the era of digital music the use of synthesizers has gained a remarkable level. Synthesizers can mimic any instrument. Still it is noteworthy to mention the observable gap in producing the tonal quality of Indian classical instruments on synthesizers when compare to real instrument. The quality perception of music is a subjective aspect. The tonal quality produced with reference to the reference note needs some objectivity. This is also required when one instrument is tuned with reference to another. There are some existing parameters like SNR, SSIM, correlation coefficients which compare the quality of two signals [1], but they produce contradictory results specifically when it comes to tonal perception perceived by human ear.

This fact is presented and justified in the paper. So to address this need an approach to generate a quality parameter which will be based upon the characteristics of signal from instruments is proposed. The proposed parameter is also evaluated against the rating of raters to prove its effectiveness. The parameter is tested for harmonium and flute.

To formulate the quality measure a reference signal is recorded from sitar and the same musical note is recorded from synthesizer. Additionally 3 different signals are generated which exhibit different tonal quality than the original signal, where two signals are of better quality than the signal recorded from synthesizer and two signals are of lower quality than synthesizer signal. All the signals are evaluated with the similarity index with correlation coefficient and with Kullback leibler divergence. [16] These signals are also tested with raters to obtain rankings.

The features of these signals are calculated and based upon minimum redundancy maximum relevance technique the ranking of the features is evaluated. [14]. From these features heuristic preferential weightage approach is taken to formulate the quality measure. The derived parameter is also verified against raters evaluation and the agreement between raters is established with Bland Altman plot.

The overall methodology of experimentation is shown in fig 1

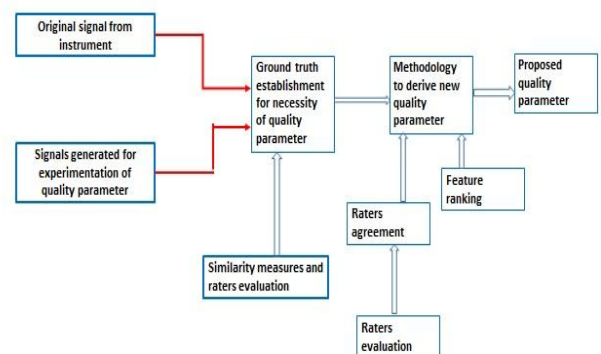


Fig.1. Overall Methodology

## II. LITERATURE REVIEW

Number of different approaches can be found in literature regarding quality assessment techniques. Different means and methods for subjective quality assessment of audio programme material are mentioned in [10]. The paper also described evaluation scale at 6 levels which describes the level of impairment in the given musical scale.

Manuscript published on January 30, 2020.

\* Correspondence Author

Dr. Abhijit V.Chitre\*, Associate Professor in Vishwakarma Institute of Information Technology, Pune, India

Dr. Aditya Abhyankar, Dean, Faculty of Technology and Professor in Department of Technology of SP Pune University, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The existence of a correlation between perceived audio quality and dynamics, distortions, tempo, spectral features and emotional predictions is demonstrated by [11]. They have proposed a new objective signal parameter to unify crest factor. A profile of music perception skills was proposed to help the researchers with an instrument to enable them to assess the level of listener's perceptual musicality objectively. [12].

Psychoacoustic analysis method is proposed in [2] which discusses different results of category of music from psychological perspective. SSIM (Structural Similarity Index) as an objective parameter for comparison of quality of an image with reference to original image is proposed and supported by the comparison with the subjective ratings. [1] New methods of extending existing linear embedded feature selection methods were proposed by [13]. The extension of the SVM feature selection method and achieved improvement in regularization and constructed feature selection methods with nonlinear classifiers. [13]. Minimal redundancy maximal relevance method is proposed to be combined with other feature selector methods such as wrappers. The paper mentioned the above method very efficient to find subset of features at low expense. [14]

Tan, Chin Tuan et al devised a new method to predict the perceived quality of nonlinearly distorted signal based on outputs of array of gamma tone filters in response of original signal and distorted signal. They have proposed this method based on cross correlation of frames. [8]

The literature survey reveals different approaches of feature classification and ranking. It also reveals the quality measures to establish structural similarity between two signals. The literature does show the different efforts on quality based on structural similarity, different feature ranking techniques, development of profiles of listener.

These approaches are found to be derived at different levels but a systematic way of arriving at a quality parameter with embedding all above approaches is not reported.

**Minimum redundancy and Maximum Relevance (mRmR) based ranking:**

This method executes filter based approach. It aims at minimizing the redundancy among the features simultaneously.

The key concept in the technique is Mutual Information which is a measure of similarity between random variables. [13,14]

The mutual Information between two random variable  $x_1$  and  $y_1$  is given by the equation below.

$$I(x_1, y_1) = \sum_{x_1 \in X_1} \sum_{y_1 \in Y_1} p(x_1, y_1) \log \left( \frac{p(x_1, y_1)}{p(x_1)p(y_1)} \right) \dots \dots \dots (1)$$

$p(x_1, y_1)$  represents the joint density function of  $x_1, y_1$

and  $p(x_1)p(y_1)$  represents marginal density functions of  $x_1, y_1$  respectively.

For feature set: Let  $f(p)$  denotes the  $p^{th}$  feature and let there are N observations of  $f(p)$

sof  $f(p)$  can be written as N dimensional vector given by

$$fp = [fp^1, fp^2, \dots, fp^N]$$

$fp$  represents the instance of discrete random variable  $Fp$ .  $I(Fp, Fq)$  is the mutual information between features  $p$  and  $q$  where  $p, q = 1, 2, \dots, d$  and  $d$  is the input dimensionality which is same as number of features in data set.

Let  $h = [h^1, h^2, \dots, h^N]$  be the target class label and  $I(Fp, H)$  be the mutual information between target class and feature  $p$ .

$S_1$  denotes the feature set with cardinality  $|S_1|$

In mRmR essentially two conditions need to be satisfied. Minimum Redundancy condition which is given by equation (2)

$$W1 = \frac{1}{|S_1|^2} \sum_{Fp, Fq \in S_1} I(Fp, Fq) \text{ which the total mutual information is between features } \dots \dots \dots (2)$$

Maximum Relevancy condition given by equation (3)

$$V1 = \frac{1}{|S_1|} \sum_{Fp \in S_1} I(Fp, H) \dots \dots \dots (3) \text{ is the mutual information between features and target}$$

The simplest combinations for above two conditions are

$$\begin{aligned} \max(V1 - W1) \dots \dots \dots (4) \\ \max(V1/W1) \dots \dots \dots (5) \end{aligned}$$

According to eq (3) first feature selection is done. From second feature onwards the feature  $i$  satisfying equations (6) and (7) is selected at every step and it is kept as a part of feature set  $S_1$ . So  $S_1$  forms the set of selected features.

$$\min_{Fp \in \Omega S_1} \frac{1}{|S_1|} \sum_{Fq \in S_1} I(Fp, Fq) \dots \dots \dots (6)$$

$\Omega S_1$  is feature subset which consists of features except features which are already selected.

$$\max_{Fp \in \Omega S_1} I(Fp, H) \dots \dots \dots (7)$$

where  $\Omega S_1$  is the feature subset of all features except those features which are already selected.

**Bland Altman Plot:** It is a graphical analysis to check agreement between two observations. The plot is the graph of differences in measurements against average of the measurements. Horizontal lines are drawn at the mean difference, and at the limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation of the differences.

**III. EXPERIMENTATION**

**The experimentation first aims at generating data from original instrument. Sitar is considered for experimentation.**

An octave is recorded from original Sitar and the same note is recorded from synthesizer in Sitar mode



To build upon the quality parameter concept additional signals were generated from synthesizer signal using the framework of Adaptive filter with different algorithms .Some of these signals are close to original signal form actual instrument and some of the signals differ from original signal.

In all five signals were generated out of which one is from original sitar ,second from synthesizer played in sitar mode and remaining three are modified versions of synthesizer signals with adaptive filter algorithms.

All of them were evaluated from Raters and also on the basis of Correlation coefficient and PSNR calculation. Kullback Leibler distance [16] is also evaluated for the generated signals with reference to the original signal recorded form sitar. Regression plots are obtained from the features to observe degree of similarity between features. Density functions are also obtained for these signals to observe similarity between the signals under experimentation.

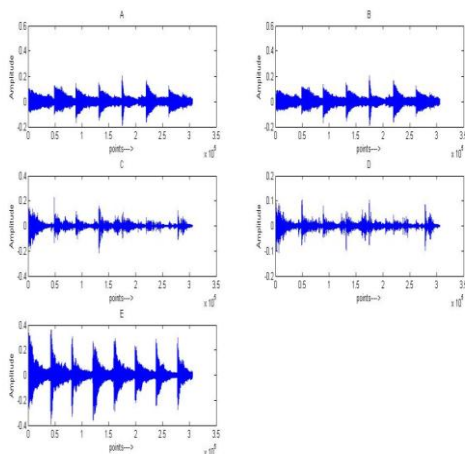
**A:** In the experimentation in addition to the signals tonal quality with reference to original signal are also generated for relative tonal perception comparison.

Different features were calculated. These features include statistical, temporal and spatial parameters. Table shows values of these features.[2][3][4][5][6]

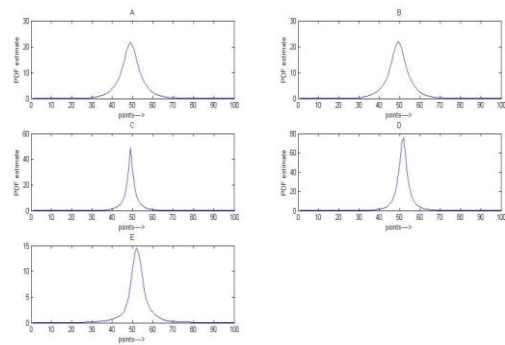
The regression plots shown below demonstrates the quotient of similarity between original signal parameters and reconstructed signals parameters.

**IV. RESULTS AND DISCUSSION**

The results of the experimentation is discussed in the section. Fig. 2 shows original sitar signal recorded and signals obtained for the experimentation of arriving at the quality parameter by beginning with the ground truth establishment of the necessity of quality measure. In the second stage of experimentation the actual procedure of formulating the parameter is mentioned and implemented.



**Fig. 2. Original and reconstructed Sitar signals**



**Fig.3: Density Functions of signals**

**Table I: Raters ratings of signals**

	Sig 2	Sig3	Sig4	Sig5
Rater1	9	5	1	1
Rater2	9	5	2	1
Rater3	8	6	2	1
Rater4	9	5	2	1
Rater5	9	6	2	1
Rater6	8	6	2	2
Rater 7	9	6	2	1
Rater 8	9	5	1	2
Rater 9	8	6	1	2

**Table. II: Feature ranking with mRmR**

Signal	PSNR	Correlation coefficient	KLD	Rating
Signal2	51.9	0.98	0.00242	9
Signal3	29.58	0.16	0.6	1
Signal4	30.9	0.33	0.332	2
Signal5	24.6	0.08	0.017	6

**Table.III: PSNR and Correlation and Quality Parameter Comparison for Sitar**

7	10	11	14	8	12
0	0.1	0.3	0.26	0.16	0.1
Centroid	Kurtosis	Flatness	Std.dev	spread	entropy

Table 5 demonstrates the different parameters calculated for different signals. Table 2 shows the values of different similarity measures of signals with reference to original signal. Signal 2 is very close to original signal. This is demonstrated by Kullback Leibler distance which reflects very small values. Also Correlation coefficient is supporting this fact with value 0.98 indicating high degree of similarity. This is also supported by Raters with strong correlation of 9. In contradiction to this, signal5 is a signal from synthesizer which has a perceptual quality much better than signals 3 and 4 but correlation coefficient is showing disagreement to the perceptual basis as well as to Kullback Leibler distances.

The KLD is indicating value 0.19 which actually states that the signal is better than signals 3 and 4 which is confirms to the raters. Signal5 is almost 50-60% good in perceived quality. It may have a structural dissimilarity with the original signal but perceived quality scale is satisfactory. The above observations can be confirmed by observing regression plots in figure 5, where plots of signal2 and exhibits maximum similarity as it is very close in perceptual quality to original signals whereas regression plots for signal 3 and 4 are deviated form original benchmark substantially compared to signal 5 which is a signal from synthesizer. The same observation can be confirmed from figure 3 fig A indicates the density distribution of original signal. This experiment demonstrates that when a signal has high degree of correlation with the original signal correlation coefficient works effectively but when this degree of similarity is not substantial then it is producing results which do not demonstrate agreement with the perceived tonal quality of the signal. This demands the need of some quality parameter which will be in agreement with the perceived quality and also with the mathematical distances mentioned in the experimentation.

$$s1(t) = f(s(t))$$

$$r1(t) = f\{r(t)\}$$

$$Rank: \{si(t) | i = 0,1 \dots N - 1 \dots \dots \dots (1)$$

$$\{ri(t) | i = 0,1, \dots N - 1$$

Rank  $ri(t)$  is generated based on MRmR based

$$IRi(t) = \frac{s1(t) + r1(t)}{2}, s1(t) - r1(t) \dots \dots \dots (2)$$

$$IRi(x, y) = \frac{\log_2(s1(x, y)) + \log_2(r1(x, y))}{2}, \log_2(s1(x, y))$$

$$- \log_2(r1(x, y)) \dots \dots (3)$$

$$Q = f\{IRi(x, y)\} \dots \dots \dots (4)$$

Let  $p1$  is the parameter (feature) of original signal and  $p2$  is the parameter of the signal to be evaluated for quality with reference to original signal,

The comparison parameter of the corresponding feature is given by

$$p = f(p1, p2) \text{ where } p1 \forall R \text{ and } p2 \forall R \text{ then}$$

$$p = p1 * \frac{p2}{q}; p \text{ is correlation coefficient}$$

where  $q = p1^2 + p2^2$  which is normalizing factor

when  $p1 = p2$

$$p = 0.5 \text{ which shows max correlation}$$

To make max correlation we write

$$q = 2 * p$$

Applying similar rationale for above calculated parameters we can write

Signal	PSNR	Correlation coefficient	KLD	Derived Q	Raters rating
Sig2	51.9	0.98	0.00242	0.99	9
Sig3	29.58	0.16	0.28	0.049	1
Sig4	30.9	0.33	0.332	0.28	3
Sig5	24.6	0.08	0.017	0.60	6

$$z = 2 * \frac{(z1z2)}{(z1^2 + z2^2)}$$

$$y = 2 * \frac{(y1y2)}{(y1^2 + y2^2)}$$

$$x = 2 * \frac{(x1x2)}{(x1^2 + xz2^2)}$$

$$w = 2 * \frac{(w1w2)}{(w1^2 + w2^2)}$$

$$v = 2 * \frac{(v1v2)}{(v1^2 + v2^2)}$$

wherez,y,x,w,v are prominent features ranked

by mRmR

These features are independent

$$Q = f(z,y,x,w) = z * y * x * w * u \text{ for Sitar (max } Q=1)$$

Table. IV .Comparison of Quality parameter with existing parameters



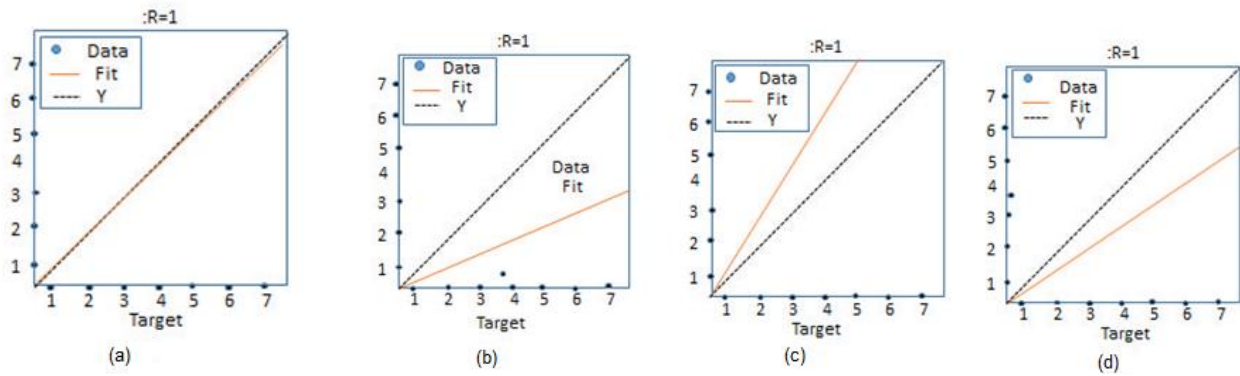


Fig.4. Regression plots of features of signals

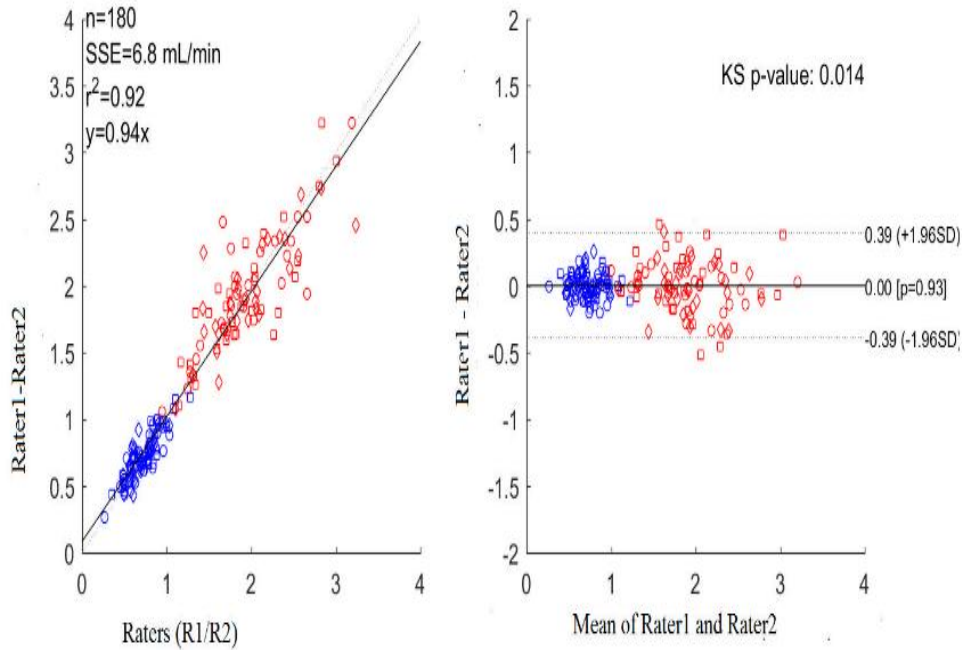


Fig.5. Bland Altman Plot of Raters Agreement

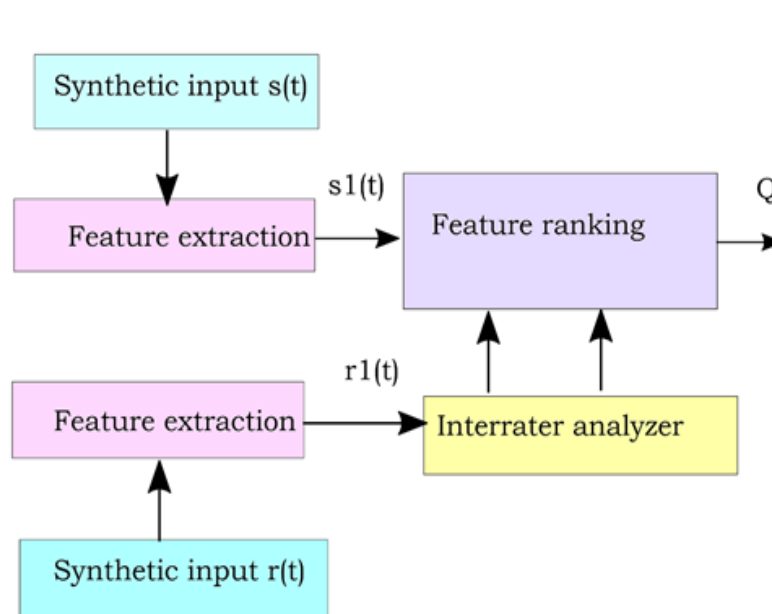


Fig.6. Methodology of mathematical procedure for quality measure formulation

Table.V: Features calculated for the signals

Attack time	Zero Crossings	Roll off	Brightness	Centroid	Spread	Skew	Kurtosis	Flatness	Entropy	Std.devn
4.87E-02	14700	5490	0.5.01	2.78E+03	4.70	1.04E+09	3.37E+12	1.76E-01	8.53E-01	-3.92E-06
4.06E-02	14800	5280	0.5.00	2.59E+03	4.05	9.73E+08	3.45E+12	1.31E-01	8.48E-01	-6.48E-06
3.27E-02	76800	1440	0.1.40	9.30E+02	1.96	1.26E+09	8.50E+12	1.74E-02	7.92E-01	7.62E-05
3.29E-02	78200	1670	0.1.74	1.05E+03	2.94	5.12E+08	2.37E+12	2.41E-02	8.09E-01	5.31E-05
4.96E-02	12900	5330	0.4.45	2.62E+03	2.09	8.73E+09	5.94E+13	1.46E-01	8.47E-01	-6.92E-06

Table.VI: Quality parameter in terms of derived feature relation parameters

No.	Z	Y	X	W	V	
Signal2	1	1	1	1	0.990	0.990
Signal3	0.59	0.497	0.908	0.797	0.997	0.214
Signal4	0.65	0.603	0.867	0.936	0.998	0.318
Signal5	0.76	0.9701	0.909	0.898	0.991	0.601

IV CONCLUSION:

The overall conclusion of the work carried out in the paper can be stated in two stages. The paper presented a ground truth of necessity of a quality parameter specifically for the signals in the musical domain. The paper demonstrated an experimental work to establish the requirement of a quality measure by comparing the correlation coefficient and the Kullback Leibler distance. The discrepancy in the perceived tonal quality is also supported by raters’ evaluation. In the second stage a systematic approach was demonstrated to formulate the quality parameter by evaluating the different features of the signals and calculating the prominent features with mRmR technique. The work demonstrated and formulated the quality parameter which is in agreement with the raters’ evaluation of the perceived quality of the musical note.

Table 4 exhibits the comparison of the proposed parameter with the existing parameter correlation coefficient Kullback Leibler distances. The proposed quality parameter shows the quality quotient of 0.6 on the normalized scale of 0-1 which is in agreement with the raters and also this value is close to the value given for the synthesizer and this signal is very similar to synthesizer. So the proposed parameter confirms this fact. The behavior of existing quality parameters shows degree of agreement only when the similarity of the reconstructed signal with the original signal is very high or extremely low which can be figured out for signal 4 and signal 3 respectively but for the signals with intermediate level of similarity these parameters don not exhibit the agreement with raters.

So from this we conclude that the proposed parameter is exhibiting the behavior which is in agreement with the raters for all the signals with different perceived tonal quality with respect to the original tone.

The formulation of the parameter needed the signals features to be captured in different domains and the proposed parameter is formulated with the features in temporal, spectral and probabilistic domain for sitar. The methodology adopted to different instrument as well. The parameter extracts the best of three domains and formulates a value which is in agreement with the raters as it eventually contributes to the perceived quality of the

musical tome. The agreement among the raters is also confirmed with Bland Altman plot as shown in fig 4. So the paper effectively devised a method of formulating a quality parameter essential for perceptual tonal quality of signals from instruments. This quality parameter is also tested with Flute and proved to be in absolute agreement with raters.

The formulation of the parameter needed the signals features to be captured in different domains and the proposed parameter is formulated with the features in temporal, spectral and probabilistic domain for both sitar and flute.

The methodology adopted to different instrument as well. The parameter extracts the best of three domains and formulates a value which is in agreement with the raters as it eventually contributes to the perceived quality of the musical tome.

FUTURE SCOPE

The formulation of parameter in the paper is based upon the features calculated in different domains o extract the best combination of different domains. This approach can be modified in future with Deep learning where more effective parameters can be evaluated automatically form the recorded signals. Also in the paper minimum redundancy technique is implemented which can also be combine with SVM or PCA based ranking to come up with more similar ranked features contributing to perceived tonal quality.

REFERENCES

1. Zhou Wang, Alan ConardBovik, Hamid Rahim Sheikh, Eero P Simoncelli ,”Image Quality Assessment: From Error Visibility to Structural Similarity,” IEEE Trans. Image Processing, vol. 13, No. 4, April 2004.
2. Ni Putu Amanda NITIDARA,JokoSarwono,I.G.Nyoman,” Psychoacoustic Analysis of Preference Reverberation Time for Gamelan Bali Concert Hall,” InternationalConferenceInternoise2014, MelbourneAustralia,Vol 249(1).
3. Juan José Burred, Axel Röbel, and Thomas Sikora, “Dynamic spectral envelope modelling for timber analysis” speech & language processing.vol 18 ,no 3 March 2010.



4. Sang Hyun Park," Musical Instrument Extraction through Timbre Classification", NVIDIA Corporation Santa Clara, CA 95050
5. Bartsch, M.A., and Wakefield, G.H. "Singing Voice Identification Using Spectral Envelope Estimation,"IEEE Transactions, Speech and Audio Processing, VOL. 12, pp.100-109, March 2004.
6. A.Eronen, A.Klapuri, "Musical Instrument Recognition using Cepstral Coefficients and Temporal Features," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2000.
7. Zhang, Q. D., "Classification of Recorded Musical Instruments Sounds Based on Neural Networks," IEEE Symposium on Computational Intelligence In Image and Signal Processing,(CIISP), 2007.
8. Chin Tuan, Brain C.G.Moore, Nick Zacharov , Ville VeikkoMattila , "Predicting the Perceived Quality of Non Linearly distorted Music and Speech signals," Journal of Audio engineering Society , Vol.52, Issue 7/8 ,Pages 699-711.
9. J.Vargo ,J.C.Nesbit ,K.Belfer , A. Archambault , "Learning Objective Evaluation : Computer Mediated Collaboration and Inter Rater Reliability ," International Journal of Computer Applications , Vol .25,Issue 3,2003.
10. W.Hoeg, L.Christensen, Walker,"Subjective Assessment of Audio Quality -The Means and Methods within the EBU," EBU Technical Review,winter 1997.
11. Alex Wilson, Bruno Fazenda,"Perception and Evaluation of Audio Quality in Music Production," Proceeding .on 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, September 2-5, 2013.
12. Lily N.C.Law, Marcel Zentner,"Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills,"
13. Julia Neumann,ChristophSchnorr, GabrieleSteidl,"Combined SVM Based Feature Selection and Classification," Springer Journal of Machine Learning, November 2005,Vol.61, Issue 1-3, pp 129-150.
14. Isabelle Guyon, AndreElisseeff,"An Introduction to Variable Feature Selection," Journal of Machine Learning Research 3 (2003), 1157-1182.
15. Alain Hore ,DjemelZiou , "Image Quality Metrics :PSNR vs. SSIM," IEEE Conference on Pattern Recognition (ICPR) 23-26 Aug. 2010,DOI: 10.1109/ICPR.2010.579
16. Mary McHugh,"Interrater Reliability: The Kappa Statistic,"Biochem Med (Zagreb), 2012 Oct, 22(3) : 276-282

## AUTHORS PROFILE



**Dr. Abhijit V. Chitre**, is presently working as Associate Professor in Vishwakarma Institute of Information Technology, Pune. He has more than 17 years of teaching and industry experience. He possesses B.E (Electronics), M.Tech (Digital Systems), and Ph.D. degrees from University of Pune. He has supervised several M.Tech Dissertation Projects. He has attended several seminars, workshops and conferences at various levels. He has more than 25 research papers published in various

national and international journal and conferences. His area of interest in teaching and research includes Signal and Image Processing, VLSI Design and Machine Learning.



**Dr. Aditya Abhyankar**, is currently working as Dean, Faculty of Technology and Professor in Department of Technology of SP Pune University. Aditya Abhyankar received the BE degree in Electronics and Telecommunication Engineering from Pune University, India in 2001. He received the MS and Ph.D. degrees from Clarkson University, NY, USA in 2003 and 2006 respectively. He worked as a post-doctoral fellow at Clarkson University, NY, USA in the

academic year 2006-07. He worked as a consultant for Biometrics LLC, WV, and USA in 2007. His research interests include pattern recognition, signal and image processing, wavelet analysis and biometric systems.