Abstract

Musical instruments and systems are some of the earliest human inventions that intuitively made use of the relationships between harmonics before the mathematics of such relations were understood and formalized. Study of sounds produced by musical instruments has a wide range of applications including modeling and reproduction of the acoustics of music instruments, digital music synthesizers, digital audio auditors, digital audio mixers, spatial-temporal sound effects for home entertainment and cinemas, music content classification and indexing and music search engines for internet.

Sound information obtained from various musical instruments is analog in nature. In modern communication systems, signals are converted to digital form via ADC before transmission. However, the signal obtained at the receiver end is often corrupted with different types of noise. Hence received sound signal needs signal denoising, which involves the manipulation of the signal data to produce a very high quality hearing perception before it can be used for applications.

Appropriate signal representation is a fundamental issue in any signal processing problem. Representation of a musical instrument sound signal is either dictated by certain embedded features of the signal or is driven by application at hand. The time impulse functions and complex exponentials are two extreme bases. The wavelet basis for the signal representation offers infinite number of possibilities between these two extreme cases and uses the scaled and translated version of some basic function. The non uniqueness in the choice of basis is one of the important reasons for the wavelet transform (WT) to find widespread application in signal processing. Depending on the application, a wavelet has to be suitably chosen, but finding a suitable basis is a nontrivial task and many researchers in recent past have attempted to answer this. Since wavelet representation is particularly useful for representing non stationary signals, it is interesting to find a basis that represents the signal in optimal way.

Audio coding or audio compression algorithms are used to obtain compact digital representations of high fidelity (wideband) audio signals for the purpose of efficient transmission over larger distances and storage. This becomes a necessity
because of the bandwidth constraint and to reduce the transmission time between any two systems communicating with each other. The primary objective in audio coding is to represent the signal with a minimum number of bits while reproducing at the receiver end as output that cannot be distinguished from the original input signal.

In this work, an effort has been made to systematically explore the area of denoising, coding and identification of sounds produced by four major types of traditional Indian musical instruments viz. woodwind, string, percussion and keyboard by using latest state-of-the-art technology, software and algorithms.

In our work, we have carried out the performance comparison of adaptive filter algorithms (LMS, NLMS, RLS) and wavelet based (DWT, DWPT) robust adaptive block algorithms. Further, optimal threshold for denoising the sound signals produced by various Indian musical instruments of different categories has been investigated. The sound signal is corrupted with variable percentage of additive white Gaussian noise (AWGN), and then different methods are compared on the basis of peak signal to noise ratio (PSNR). In adaptive filter approach, the signal is applied to the algorithm as one piece and PSNR is calculated. While in wavelet approach the signal is first segmented, then thresholding methods are used for each segment/block. The optimal size of the block is decided on the basis of minimum mean square error criteria. The PSNR values, thus obtained, depend upon the level of decomposition and shape of the wavelet. All the blocks obtained after individual block denoising are concatenated to get the final denoised signal. The optimal wavelet, which is chosen by a systematic search and the level of decomposition for denoising percussion, wind, string and keyboard instrument sound, is different because maximum PSNR values are obtained with different wavelets and level of decomposition. The wavelet based algorithm is superior in terms of quality of the signal which is close to the original signal, along with the time required for denoising, hence this algorithm can also be used for real time application.

In this dissertation, a technique based on Discrete Wavelet Transform has been used to identify the optimal wavelet that best matches the sounds produced by Indian musical instruments. It is observed that when the sound signal is decomposed for different levels and the energy in the wavelet coefficients is
calculated for each level, the energy retained in wavelet coefficients varies with the type of wavelet and its decomposition level. The maximum level of decomposition for various types of sound signals is different and the signals are also reconstructed with the wavelet coefficients only up to the maximum level of decomposition with optimal wavelet.

Filter bank theory has been used to compute the approximation and detail coefficients of the wavelet filter for identifying the scaling and wavelet function of the wavelet. A knowledge of the filter bank coefficients of the sounds of musical instruments is helpful in identifying the signature wavelet of the sound signal, which can be used to reconstruct the original signal with negligible error. Among the three algorithms used for computing filterbank coefficients viz. LMS, NLMS and RLS, RLS algorithm performs better in most respects; also, it converges very fast i.e. number of iterations are less. Hence an algorithm based on RLS algorithm is developed to find out scaling and wavelet functions of the sounds of musical instruments with better accuracy and speed of convergence.

This work is further extended to identify biorthogonal compactly supported wavelet from the sound signal. This work is based on maximizing projection of the sound signal onto successive scaling subspaces, which result in minimization of energy of sound signal in the wavelet subspace. The analysis wavelet filter is identified similar to a sharpening filter in image processing. First, 2-band FIR biorthogonal perfect reconstruction filter bank is identified from the given sound signal, which leads to the design of biorthogonal compactly supported wavelet. Second, the wavelet with desired support as well as desired number of vanishing moments is also identified.

The structured arrangement of sounds in musical pieces, results in the unique creation of complex acoustic mixtures. The analysis of these mixtures, with the objective of estimating the individual sounds which constitute them, is known as musical instrument sound signal separation, and has applications in audio coding, audio restoration, music production, music information retrieval and music education. In this work, the problem of sound signal separation from the mixture in the presence of additive white Gaussian noise (AWGN) is investigated. Because of the additive noise independent component analysis (ICA) algorithm does not give a consistent estimation of separated signals. The different sound
signals are first down sampled and then mixed with the help of mixing matrix found iteratively. The mixed signal is added with variable percentage of AWGN. Resulting noisy mixed signal is further denoised with block denoising algorithm developed previously, then decomposed into wavelet coefficients with DWT. Fast ICA algorithm is further applied on these wavelet coefficients for identification of individual sound signal, and then IDWT is used to reconstruct the separated sound signals. The hearing perception of the separated sound signals is close to the original signals.

In this work, a new hybrid multi-stage musical instrument sound signal coding method, based on discrete wavelet packet transform (DWPT) and modified discrete cosine transform (MDCT) using efficient psychoacoustic models, is investigated. The primary objective is to perform lossy and perceptually transparent compression on the sounds of various Indian musical instruments. The original sound signal is first decomposed into wavelet packets using an optimal wavelet basis. Then, the wavelet packets are run through psychoacoustic model in wavelet domain to determine auditory masking level for thresholding, which, in turn, is used to perform thresholding of wavelet coefficients in respective sub-bands. Further, audio signal coefficients are partitioned into frames which overlap in such a way that each block gets 512 samples and windowed using a Hanning window with 1/16 frame overlap. Additionally, MDCT is applied to each block of data to de-correlate the spectral information. Removal of spectral redundancy is achieved by compressing the weak components more than the dominant components. The resulting signal is quantized with variable number of bits, which are determined based on the results of the psychoacoustic model in FFT domain. This technique provides an efficient way to exploit key strengths of both DWPT and MDCT. In addition, using wavelets, the compression ratio can easily be varied in accordance with audio quality, while most other compression techniques have fixed compression ratios.