MUSIC SYNTHESIS USING PITCH SHIFTING AND ADSR MATCHING

Mallikarjun Chandrasha Bulla[1], S.M.Mukane[2]

[1] Student M.E., Electronics & Telecommunication Engineering, SVERI's College of Engineering, Pandharpur, Maharashtra, India
[2] Head of Department, Electronics & Telecommunication Engineering, SVERI's College of Engineering, Pandharpur, Maharashtra, India

ABSTRACT

The sound received at the ears is processed by humans using signal processing that separates the signal along intensity, pitch and timbre dimensions Sound can be represented in time domain as a waveform or in frequency domain as a set of spectra. Along with the pitch analysis, timbre, the vertical dimension of sound drives the research in the field of classification and identification of sounds. The pitch and timbre analysis, hand in hand, also run the manipulation- synthesis processes of musical instruments. And here it becomes the matter of crux whether to highlight pitch analysis or timbre analysis of musical sounds. Timbre, called the color of sound, distinguishes the two different instruments playing the same note. Pitch will be same for these instruments. Timbre analysis involves modeling of spectral characteristics of sound when we think of the transformations, may be time or frequency domain. The spectral parameters of musical sound help the researchers to capture the expressiveness of sound. Considering these aspects, this work helps to select the timbre features across a wide variety and categories and is focused on monophonic sounds. It contributes to study of tumbrel features which are helpful for manipulation of musical notes. Pitch analysis and manipulation talks about the tuning of instruments as well as modifying the human voices. The applications lie with giving sounds to animated characters. Conventional Fourier-based signal processing, while endowed with fast algorithms, is unable to easily represent signal along these attributes Also it's difficult to process the pitch and timbre separately using these methods. In this paper we use a recently proposed cortical representation (simulation to the processing method of human brain) to represent and manipulate sound. We briefly overview the summary of the process for obtaining, manipulating and inverting cortical representation of a sound and describe algorithms for manipulating signal pitch and timbre separately. The algorithms are first used to create sound of an instrument between a "guitar" and a "trumpet". Applications to creating maximally separable sounds in auditory user interfaces are discussed.

Keyword: - *Pitch, Timbre, ADSR, Zero-Crossing Rate, Pitch and timbre manipulation, Pitch Detection Parameters, Timbre detection Parameters*

1. INTRODUCTION

1.1 What is Pitch?

Any sound that we hear as a tone is made of regular, evenly spaced waves of air molecules. Some sounds are higher, some sounds are lower. These are the differences in the pitch caused by different spacing in the waves. Closer the waves, shorter are the lambda, higher is the frequency and hence higher will be the tone. (Spacing of the waves is the wavelength, lambda). Range of the audible sound is 20 Hz to 20 kHz. Longer waves have lower frequency and higher lambda, they sound low.

Music pitch: Musicians name the pitches that they use most often e.g. they might call a note "middle C" or "2 line G". Shorter instruments make shorter waves and higher sound whereas larger or longer instruments make longer waves and lower sounds. E.g. a bird makes high pitch and lion makes low pitch.

Fig -1 High-Frequency Sound Wave

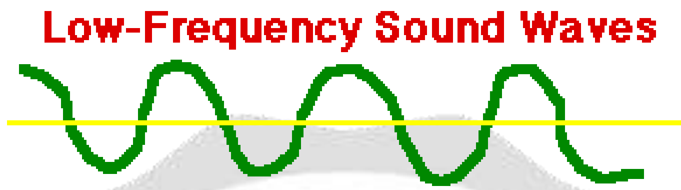


Fig – 2 Low-Frequency Sound Waves

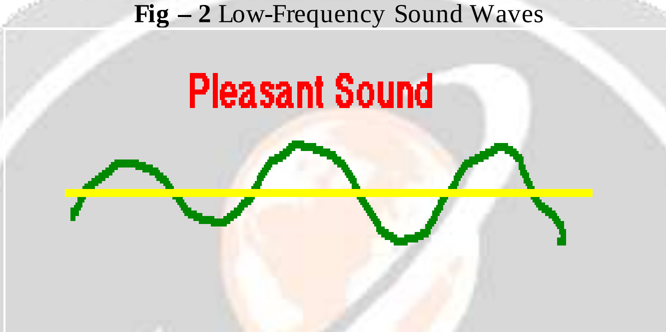


Fig – 3 Pleasant Sound Waves

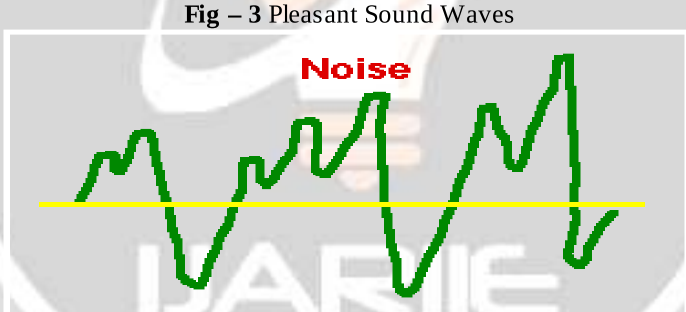


Fig – 4 Noise Waves

Table – 1 Octave Table Explaining the relationship between the note frequency

Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz	Note	Hz
C1	32.7	C2	65.4	C3	130.8	C4	261.6	C5	523.3	C6	1046.5	C7	2093.0
C#1	34.6	C#2	69.3	C#3	138.6	C#4	277.2	C#5	554.4	C#6	1108.7	C#7	2217.5
D1	36.7	D2	73.4	D3	146.8	D4	293.7	D5	587.3	D6	1174.7	D7	2349.3
D#1	38.9	D#2	77.8	D#3	155.6	D#4	311.1	D#5	622.3	D#6	1244.5	D#7	2489.0
E1	41.2	E2	82.4	E3	164.8	E4	329.6	E5	659.3	E6	1318.5	E7	2637.0
F1	43.7	F2	87.3	F3	174.6	F4	349.2	F5	698.5	F6	1396.9	F7	2793.8
F#1	46.2	F#2	92.5	F#3	185.0	F#4	370.0	F#5	740.0	F#6	1480.0	F#7	2960.0
G1	49.0	G2	98.0	G3	196.0	G4	392.0	G5	784.0	G6	1568.0	G7	3136.0
G#1	51.9	G#2	103.8	G#3	207.7	G#4	415.3	G#5	830.6	G#6	1661.2	G#7	3322.4
A1	55.0	A2	110.0	A3	220.0	A4	440.0	A5	880.0	A6	1760.0	A7	3520.0
A#1	58.3	A#2	116.5	A#3	233.1	A#4	466.2	A#5	932.3	A#6	1864.7	A#7	3729.3
B1	61.7	B2	123.5	B3	246.9	B4	493.9	B5	987.8	B6	1975.5	B7	3951.1

Continuing with the semi-automatic approach for classification based on pitch detection. The following figure shows the classification tree indicating separation of instrumental from human singing voice.

1.2 What is Timbre?

Timbre is the quality of sound. It allows the ear to distinguish sounds that have same pitch and loudness. Timbre is determined by the harmonic content of a sound and the dynamic characteristics of the sound such as Vibrato and Attack 'n' decay of the sound. Some investigators report that it takes duration of about 60 ms to recognize the timbre of a tone and any tone shorter than about 4 ms is perceived as atonal click. It's suggested that it takes about 4 dB changes in mid or high-harmonics to be perceived as a change in timbre whereas about 10 dB changes in one of the lower harmonics is required.

Harmonic contents:

1) Attack and decay:

E.g. plucked guitar string and striking a cymbal with stick.

2) Vibrato and tremolo:

Vibrato is the periodic change in the pitch of the tone. It could be considered as FM. Tremolo is used to indicate the periodic changes in the amplitude or loudness of the tone. Vibrato is considered to be a desirable characteristic of the human voice (if it is not excessive) and can be used for expression; it also adds richness to the voice. If the harmonic content of a sustained sound from a voice or wind instrument is reproduced precisely, the ear can readily detect the difference in timbre because of the absence of vibrato. More realistic synthesised tones will add some type of vibrato and/or tremolo to produce a realistic tone.

Timbre expressed by guitar and cymbal:

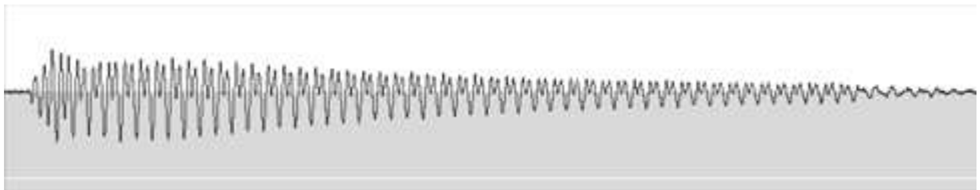


Fig - 5 Plucked guitar string

Fig - 6 Striking a cymbal with a stick

1.3 Need for Pitch and Timbre Manipulation

Why to manipulate Pitch and timbre?

Pitch is related to harmonic structure and timbre is related to the quality of sound. Take an example of two musical instruments: they might have the same pitch if they play the same note but it's their different timbre that allows us to distinguish between them.

Reasons to manipulate Pitch:

1. for assessment of speech intelligibility

2. for prediction of the cortical response to an arbitrary stimulus

3. Pitch information of speech signal has received importance especially in

    a. Speaker identification (e.g. males have low pitch and females have high pitch)

    b. Speech Recognition

    c. Speech Synthesis

    d. Speech articulation training aids for the deaf

Well-known method for pitch extraction is cepstral method but it yields unsatisfactory results for noisy speech Latest method is sinusoidal auto correlation model for pitch extraction. It gives better results at low SNR but it is suitable for only short duration sentences uttered by limited number of speakers. Pitch detection –useful in new areas like wave to MIDI converters such as "Digital Ear" A singer can make better progress on pitch accuracy if he is trained by a hybrid of traditional teaching methods and real time visual feedback. Now-a-days computer music research is dedicated and extended to making music (synthesis), or manipulating music to make new music.

A PERFECT ANALOGY FOR REASONING PITCH DETECTION

Having an objective 'listener', who can show you important parameters about the sound you produce, is like a navigation system to a pilot; for example, even though a pilot may be able to see the ground, an altimeter is still useful. Moreover, a musician has numerous things to concentrate on at the same time, such as pitch, volume and timing, making it possible for unsatisfactory aspects of their sound to slip by without their noticing.

Pitch detection serves the purpose of that objective 'listener'.

Another reasoning is even though the musical intervals at any moment throughout the song may be correct, if the absolute pitch of the notes is changed then this error may only become obvious when playing with others, or with a teacher; however, a pitch tool can enable students to detect pitch drift while practicing at home. It is concerned with finding the parameters of a voice with efficiency, precision and certainty, so it can be used in real world environments such as teaching.

Reasons to manipulate Timbre

For speech and musical instruments timbre is conveyed by envelop of the spectrum. It allows us to distinguish between speakers or instruments.

2. DISCUSSION OF THE PITCH AND TIMBRE DETECTION

2.1 Block Diagram

Methodology for pitch and timbre manipulation:

- **Need**
  - Speech Processing: Male, female voice recognition.
  - Analysis of Musical Signals: Finding the pitch and calculating the ADSR i.e. Attack, decay, sustain and release timing parameters.
  - Identification of musical instruments: Identifying the type of instrument.
- **Methodology**
  - Time Domain Approaches: analyses signal in time domain using autocorrelation
  - Frequency Domain Approaches: analyses signal in frequency domain using autocorrelation calculation with FFT or LPC can also be used for calculation of fundamental frequency.

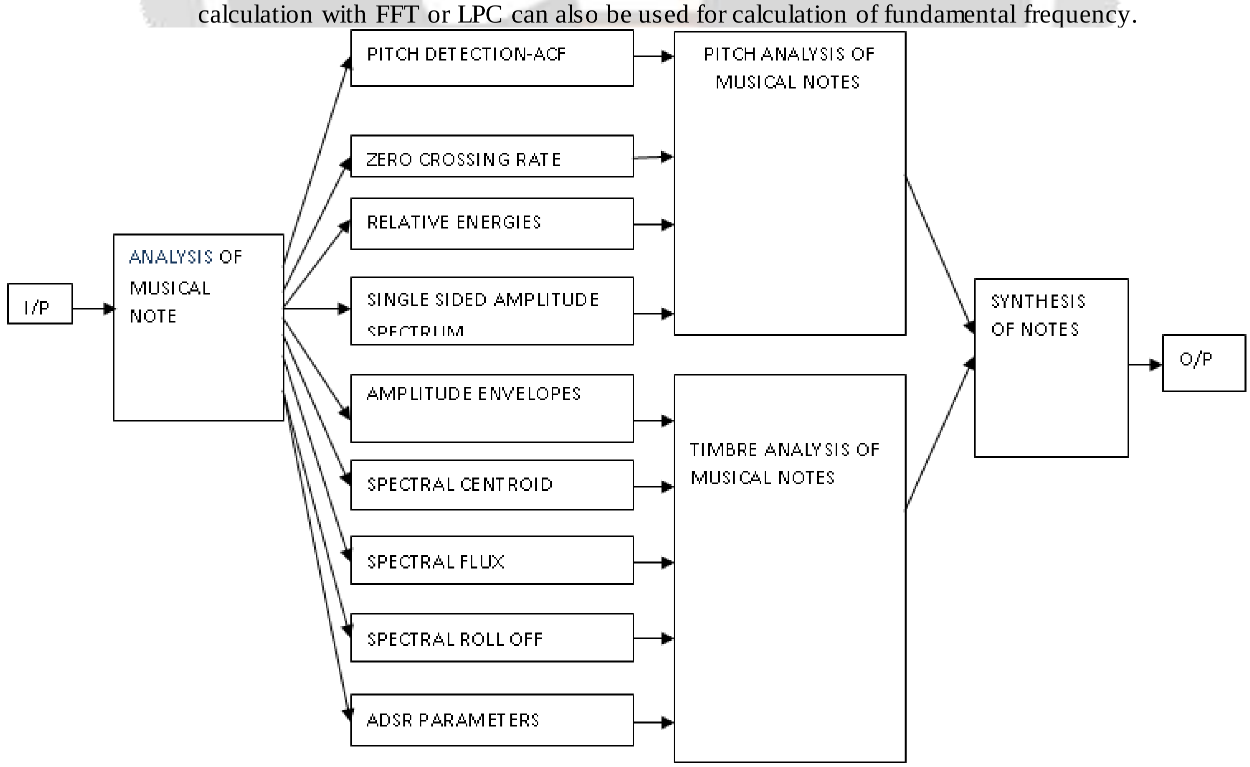


Fig – 7 Block Diagram of Pitch and Timbre Detection

2.2 Pitch detection

For musical instruments and other objects which vibrate in regular and periodic fashion, the harmonic frequencies are related to each other by simple whole number ratios. The lowest frequency produced by any particular instrument is known as the fundamental frequency or first harmonic of the instrument. Pitch detection is done in time as well as frequency domain. The purpose of pitch detection is to identify the musical notes. Various pitch detection methods are available for pitch detection and many of them can be used to validate the results. This work has progressed through the pitch detection using autocorrelation in time domain, fft i.e. frequency domain and lpc coefficients. All algorithms worked well for pitch calculation of notes played by trumpet and clarinet.

1) Time domain approach:

It consists of autocorrelation, Zero-crossing rate (ZCR), Peak rate and Slope event rate methods for pitch detection. The mostly used method is autocorrelation in time domain.

Autocorrelation: Fundamental frequency is generally seen as the frequency of the first strong partial (the fundamental) or as the frequency difference between two adjoining harmonic overtones. The correlation between two waveforms is a measure of their similarity. The waveforms are compared at different time intervals, and their "sameness" is calculated at each interval. The autocorrelation function itself is periodic. Problems with this method arise when the autocorrelation of a harmonically complex, pseudo periodic waveform is taken.

$$R_x(\nu) = \sum_{n=-\infty}^{\infty} x[n]x[n+\nu]$$

$$R_{x'}(\nu) = \sum_{n=0}^{N-1-\nu} x'[n]x'[n+\nu]$$

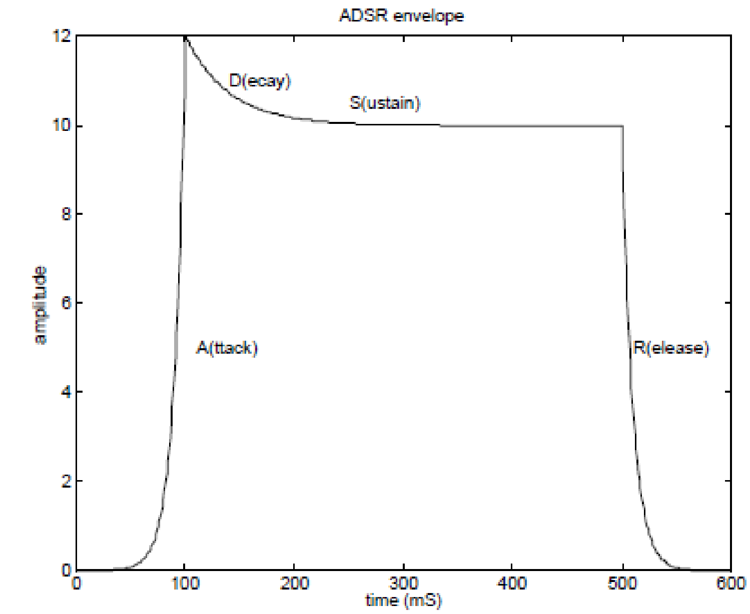
2) Frequency domain approach:

In the frequency domain, polyphonic detection is possible, usually utilizing the Fast Fourier Transform (FFT) to convert the signal to a frequency spectrum. This requires more processing power as the desired accuracy increases, although the well-known efficiency of the FFT algorithm makes it suitably efficient for many purposes. Popular frequency domain algorithms include the harmonic product spectrum, Cepstral analysis and maximum likelihood which attempts to match the frequency domain characteristics to pre-defined frequency maps (useful for detecting pitch of fixed tuning instruments); and the detection of peaks due to harmonic series.

2.3 Pitch manipulation

It is another important task in music signal processing which helps to reduce the noise and detect the note played. The work focuses on manipulation of pitch of musical note for tuning instruments. This also helps to learn the music at home for a naïve or beginner person. The relativity intervals of notes don't change, ear can't perceive it but if the absolute pitch of the note is changed then it can be prominent. Guitar tuning is done using the pitch manipulation concept. This shift is always in the power of two.

2.4 Timbre Detection Parameters



2.4.1 ADSR PARAMETERS

ATTACK TIME: Time taken by the note energy to reach to maximum value is called the attack time.

DECAY TIME: Time taken by the note energy to drop to 50 % of the maximum energy is called the decay time.

SUSTAIN TIME: Time taken by the note energy to drop to 25 % of the maximum energy is called the sustain time.

RELEASE TIME: And finally time taken by the note energy to drop to minimum value is called the release energy.

All parameters are calculated on the basis of relative energy of musical notes. Spectral features are also equally important which give the frequency domain behaviour of the musical notes.

2.4.2 SPECTRAL PARAMETERS

SPECTRAL ROLL OFF:

Spectral Roll off is the frequency boundary where 85 % of the total power spectrum energy resides. It's also known as skew of the spectral shape and helps to distinguish the percussive instruments from wind instruments.

$$\sum_{k=0}^{R} X[k] = 0.85 \sum_{k=0}^{N-1} X[k]$$

SPECTRAL CENTROID:

Spectral Centroid is assumed to be the centre of gravity for the frequency components of a signal.

$$SC_{Hz} = \frac{\sum_{k=1}^{N-1} f[k] X[k]}{\sum_{k=1}^{N-1} X[k]}$$

where $X[k]$ is the magnitude corresponding to frequency bin $k, f(k)$ is the center frequency of that bin, N is the length of the DFT and SC is the Spectral Centroid in Hertz. Brightness is amplitude form of the Spectral Centroid equation. Darker the sound more the dominance of lower frequencies and brighter the sounds more are the higher frequencies.

SPECTRAL FLUX:

The amount of frame-to-frame fluctuation in time is given by the spectral flux and it is computed by the 2-norm difference between consecutive STFT frames.

$$SF = [\![X_t[f] - X_{t-1}[f]]\!]$$

By observing the two consecutive values of the spectral flux, we notice whether it's changing faster or slower. Speech signals change faster than music signals and hence it is useful for separation of instruments from voices. All these parameters help to distinguish the musical instruments.

2.5 Timbre manipulation

It is analysed on the basis of time domain, spectral and frequency domain parameters. Each domain has it peculiarity. Time domain simply gives the envelope of the note. By observing the amplitude envelope we can identify the instrument by practice. Same way we can also take the help of ADSR parameter calculation, attack, decay, sustain and release timings for that instrument note. Frequency domain parameters give the odd to even harmonics energy or power ratio, number of harmonics and also the relative power. These parameters help us to distinguish the instruments played even if the note played is same by those instruments. So timbre manipulation is basically targeted to either generate an altogether a different note or smooth the distorted note of an instrument. In short, identify whether the note played is distorted because of the instruments' deformation or not.

3. IMPLEMENTION OF PITCH AND TIMBRE FEATURES

3.1 Time Domain Methods

Autocorrelation

Definition: The correlation between two waveforms is a measure of their similarity. The waveforms are compared at different time intervals, and their "sameness" is calculated at each interval. The autocorrelation function itself is periodic. It is less susceptible to noise and sensitive to sampling rate. Though expensive through computational point of view it is the most preferred method for calculation of the pitch.

ZCR (Zero Cross Rate Detection)

Definition: The thought was that the ZCR should be directly related to the number of times the waveform repeated per unit time. If the waveform contains higher-frequency spectral components, then it might cross the zero line more than twice per cycle. Detect zero-crossing pattern and hypothesize a value for f0 based on these patterns. Calculate mean and variance of the zero crossing rate to increase the robustness and of a feature extractor. It is simple and inexpensive but inaccurate while dealing with highly noisy signals.It is weak to deal with harmonic partials which are stronger than fundamental frequency. It presents poor results in case of oscillating signals around zero. In such cases we can always think of modified ZCR – considers threshold value, signal rectification and then mean calculation. The time domain analysis methods are applicable to monophonic pitch detection.

3.2 Frequency Domain Methods

HPS (Harmonic Product Spectrum)

Definition: Series of peaks corresponding to F0 with harmonic components at integer multiples of it . It is computationally inexpensive and reasonably resistant to additive and multiplicative noise. It's resolution as good as length of FFT

Linear Predictive Coding

It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters. Here we apply this method to sound analysis as well.

Maximum Likelihood

It matches the frequency domain characteristics to predefined frequency maps. And useful for detecting pitch of fixed tuning instruments.

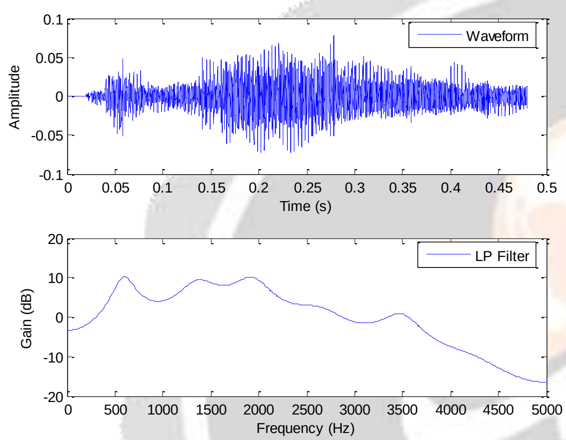
Autocorrelation using FFT

It is also one of the popular methods to give good results for calculation of fundamental frequency. The frequency domain methods are applicable to polyphonic pitch detection

4. RESULT

4.1 Pitch Detection Result

Analysis Method: LPC

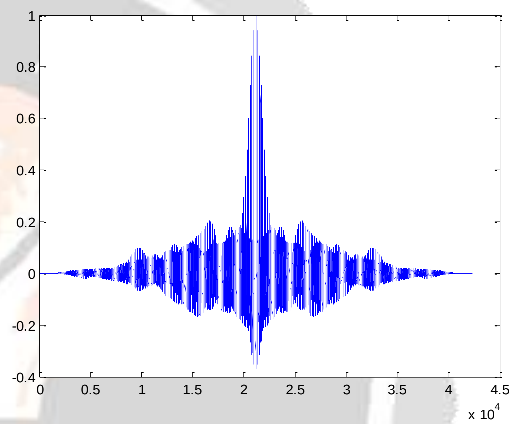


Filename: adsr1 (trumpet sa + clarinet sa)

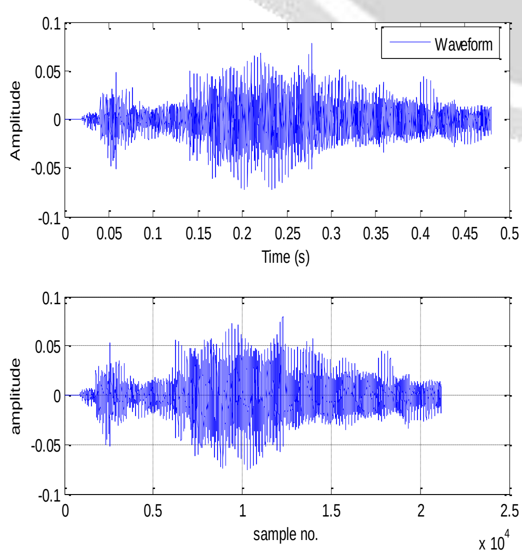
Formant 1 Frequency 588.4

4.2 Pitch Detection For ADSR1

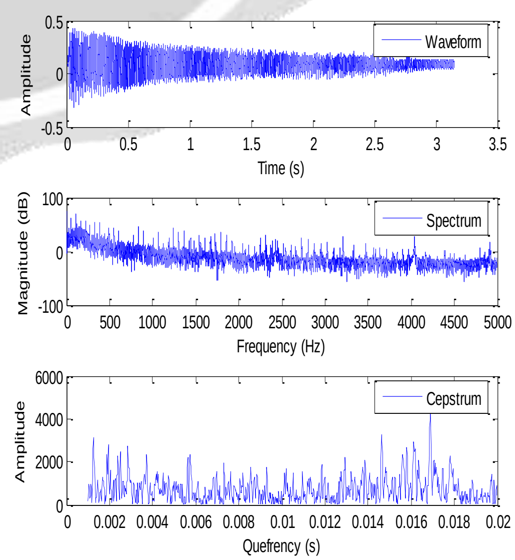
Analysis Method: Autocorrelation in Time Domain



4.3 Time Domain Representation of ADSR1

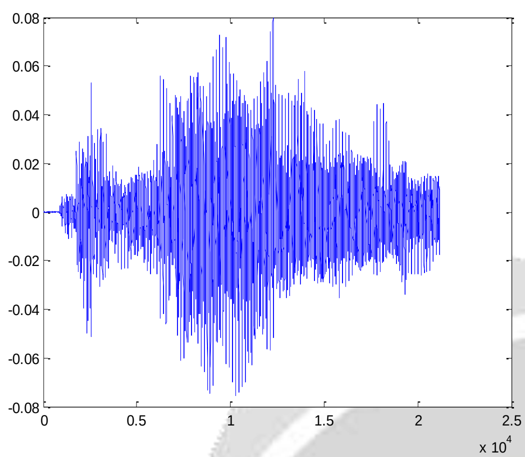


4.4 Pitch Detection for Frequency Domain Method



4.5 Timbre Manipulation Results: For Clatrumpet

Amplitude Envelop



Energy Plot

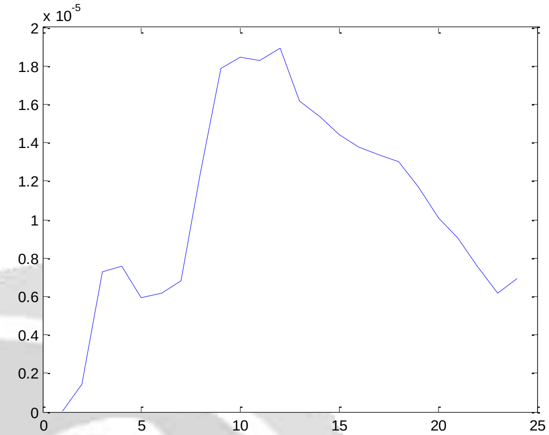


Table – 2 ADSR table For The Note 'SA'

Instrument	ADSR	Pitch Frequency
Trumpet	**19 2 0 0**	**265.662651 Hertz**
Clarinet	**12 9 0 0**	**265.662651 Hertz**
Clatrumpet	**9 3 10 15**	**272.222222 Hertz**

Table – 3 Analysis of ADSR parameters, Number of Harmonics & Relative Power

NAME OF INSTRUMENT NOTE	ADSR Parameters	Number of Harmonics	Relative Power
TRUMPET SA	19 2 0 0	13	0.0032 1.2733 1.3698 0.7265 1.1633 1.0500 1.0110 0.9642 0.8169 0.8909 1.1094 0.6875 1.2449
CLARINET SA	12 9 0 0	6	0.0053 1.9962 1.6720 1.9910 1.4855 1.5564
TRUMPET RE	19 3 0 0	12	0.0079 0.3850 0.4585 0.5510 0.4798 0.9916 0.9134 0.8671 0.8743 1.1234 1.1432 1.1729
CLARINET RE	18 7 0 0	4	0.0062 1.0251 0.5092 0.5359

5. COCLUSION

The features extracted prove useful for distinguishing clarinet and trumpet. The time domain ADSR envelope is useful for identifying the notes as well as instruments. ADSR parameters are used for Musical Signatures also. This study has already been extended to distinguish the instrument families like wind, string and percussion types [2][3]. Further the work is progressed through the development of new musical instrument note by manipulating timbre. The ADSR parameters are calculated by percent method on the basis of relative energy ratio. The work has tried to convert trumpet ADSR into clarinet ADSR parameters by manipulating the energy ratios. And above mentioned set of, time as well as frequency domain parameters, is used to compare the values of all, to check whether the resultant instrument is falling closer to trumpet or clarinet. We give it a name "Clatrumpet" for this instrument.

Apart from manipulating the timbre for generation of different musical notes, we have also manipulated the pitch. The applications are wide from manipulating the voice of male sounding like female to giving sounds to animated characters. As this work has focused on musical instruments, it can be used for tuning of musical instruments. We have tried to synthesize the shifted musical notes. In short, a single musical note is taken and it is periodically shifted to generate a complete octave for an instruments.

6. REFERENCES

[1] D. N. Zotkin, S. A. Shamma, P. Ru, R. Duraiswami, L. S.Davis. "Pitch and timbre manipulations using cortical representation of sound:- Perceptual Interfaces and Reality Laboratory, UMIACS, University of Maryland, College Park 20742 D. N. Zotkin, S. A. Shamma, P. Ru, R. Duraiswami, L. S.Davis

[2] Chang-Hsing Lee Dept. of Comput. Sci. & Inf. Eng., Chung Hua Univ., Hsinchu,Taiwan Hwai-San Lin Chih-Hsun Chou, Jau-Ling Shih " Modulation Spectral Analysis of Static and Transitional Information of Cepstral and Spectral Features for Music Genre Classification" Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP '09. Fifth International Conference on Date of conference: 12-14Sept.2009 .

[3] Takehiro Abe, Katsutoshi Itoyama, Kazuyoshi Yoshii,Kazunori Komatani, "Analysis and manipulation approach to pitch duration of musical instrument sound without changing tumbrel characteristics". Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, September 1-4, 2008

[4] Jianfeng Xu, Koichi Takagi, Shigeyuki Sakazawa Media Solutions Laboratory, KDDI R&D Laboratories Inc. {ji-xu, ko-takagi, kazawa}@kddilabs.jp "Motion synthesis for synchronizing with streaming music by segment-based search on metadata motion graphs" :- 978-1-61284-350-6/11/$26.00 ©2011 IEEE.

[5] S.N. Gunasekera+, J.V. Wijayakulasooriya*, P.C. Perera* +Open University of Sri Lanka, *Faculty ofEngineering, University ofPeradeniya email:jan@ee.pdn.ac.lk "Natural synthesis of North Indian musical instruments". Fourth International Conference on Industrial and Information Systems, ICIIS 2009,28 - 31 December 2009, Sri Lanka.

[6] Yasuda, K. Graduate Sch. of Eng., Osaka City Univ. Hama, H. **Volume:** 1 "Timbre Analysis and Synthesis of Stringed Musical Instruments". Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on Date of Conference: Aug. 30 2006-Sept. 1 2006.

[7] Juan José Burred, Member, IEEE, Axel Röbel, and Thomas Sikora, Senior Member, IEEE. "Dynamic Spectral Envelope Modeling for Timbre Analysis of Musical Instrument Sounds" Ieee transactions on audio, speech and language processing vol.18, no.3, march 2010.

[8] Rafael George Amado Jozue Vieira Filho. "Pitch Detection Algorithms Based on Zero-Cross Rate and Autocorrelation Function for Musical Notes"Department of Electrical Engineering (UNESP) Sao Paulo State University Campus of Ilha Solteirargamado@ieee.org jozue@dee.feis.unesp.br

[9] ZHU Bin 1, WANG Xin 2.1. College of information science of technology, Zhejiang Shuren University, Hangzhou 310015, P.R.China E-mail: Zhubinxx@sohu.com

[10] David Gunawan #1, D. Sen #2# School of Electrical Engineering and Telecommunications.The University of New South Wales, Sydney Australia1 d.gunawan@unsw.edu.au. 2 dsen@ee.unsw.edu.au "Music source separation synthesis using Multiple Input Spectrogram Inversion":- MMSP'09, October 5-7, 2009, Rio de Janeiro, Brazil. 978-1-4244-4652-0/09/$25.00 c! 2009 IEEE.

[11] Marek Bohác, Jan Nouza M. Inst. of Inf. Technol. & Electron., Tech. Univ. of Liberec, Liberec, Czech Republic Nouza, J. Volume: 1. "Direct magnitude spectrum analysis algorithm for tone identification in polyphonic music transcription 2011". Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on Date of Conference:15-17Sept.2011

[12] Jari Kleimola, Jussi Pekonen, Henri Penttinen, Vesa Välimäki* Dept. of Signal Processing and Acoustics Helsinki University of Technology (TKK), Espoo, Finland jari.kleimola,jussi.pekonen,henri.penttinen,vesa.valimaki}@tkk.fi Jonathan S. Abel CCRMA, Dept. of Music Stanford University, Stanford, CA, USA abel@ccrma.stanford.edu."Sound synthesis using an all pass filter chain with auto rate coefficient modulation" Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09), Como, Italy, September 1-4, 2009

[13] Chung Hua University, Hsinchu, Taiwan "Music genre classification using modulation spectral features and multiple prototype vectors representation 2011" - 2011 4th International Congress on Image and Signal Processing. Chang-Hsing Lee, Chih-Hsun Chou, Cheng-Chang Lien, and Jen-Cheng Fang. Department of Computer Science and Information Engineering

[14] Nelson Lee, Julius O. Smith, III, and Vesa Välimäki, Senior Member, IEEE. "Analysis and Synthesis of Coupled Vibrating Strings Using a Hybrid Modal-Waveguide Synthesis Model" IEEE TRANSACTIONS ON AUDIO, SPEECH, AND ANGUAGE PROCESSING, VOL. 18, NO. 4, MAY 2010.