TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PULCHOWK CAMPUS

Final Year Project Proposal On

Music Organizer and REcommender (MORE)

Submitted By:

Name	***Roll No***	***Email***	***Mobile No***
Basanta Raj Onta	061BCT507	*basanta_onta@hotmail.com*	9841472370
Binit Amatya	061BCT510	*binitamatya@hotmail.com*	9741001666
Deepa Singh Dongol	061BCT516	*medeepa@hotmail.com*	9841409096
Samita Lalchan	061BCT540	*esa_lale@hotmail.com*	9841369496

Submitted To:

Department of Electronics and Computer Engineering
Pulchowk Campus,
Institute of Engineering,
Pulchowk, Lalitpur

August 8, 2008

Abstract

People listen to songs all the time and usually have a fondness for certain specific types of songs, but they can never find out what the song sounds like without listening to it first. Also the genres of music are very loosely defined so they cannot really depend on the genre to decide if they are going to like the song or not. Also a person may want to listen to a specific type of song at a time depending on their mood.

So what the project proposes to do is analyze the songs using their characteristics like frequency, beats etc to build a Self Organizing Map (SOM) which places similar types of songs together.

These SOMs can then be used to build playlist of songs. These SOMs can also be uploaded to a central server. This central server will then accumulate SOMs of all the users and combine them to into a detailed SOM. This SOM can be accessed by users to find out about new songs that are similar to the songs they already like. Thus the program allows users to experience new songs that they most probably will like because they like similar songs.

Acknowledgement

We would like to thank **Mr. R. L. Rajbhandari**, B.E project coordinator for his coordination and relevant information regarding the project selection and proposal preparation.

The course “Innovation for future team leaders of BE project” organized by **Dr. Jyoti Tandukar** gave us chance to present and refine our project ideas. We thank him for organizing this course and acknowledge his effort that encouraged us to take this challenging project.

We would also like to offer our gratitude to all our **teachers** whose lectures and ideas were the basis for our project research and finally we’d like to thank all our **friends** who gave us their valuable input on our initial project ideas and helped us build on them to create the final proposal as shown here.

Basanta Raj Onta (061BCT507)

Binit Amatya (061BCT510)

Deepa Singh Dongol (061BCT516)

Samita Lalchan (061BCT540)

Letter of Transmittal

Date: 8th August, 2008

The Project Coordinator,

Department of Electronics and Computer Engineering,

Institute of Engineering (IOE),

Pulchowk, Lalitpur

Subject: Submission of proposal for Final Year Project

Dear Sir,

This is to inform you that we have selected "Music Organizer and REcommender (MORE)" as Final Year Project. The project is an automated music organizer which sorts a user's local music collection according to its similarity with each other and also suggests similar new songs via an online service. We hope that this system will be an effective method to organize songs according to its type and an easy way to find new songs of the user's taste.

We look forward to your accepting our proposal so we can move forward with it. We also expect full cooperation as well as necessary resources and guidelines from your side. We hope that with your help and support we can complete this project effectively and in the given time constraint.

Thank you.

Sincerely,

Basanta Raj Onta (061BCT507)

Binit Amatya (061BCT510)

Deepa Singh Dongol (061BCT516)

Samita Lalchan (061BCT540)

Table of Contents

List of Figures

1. Introduction

1.1 Background

MP3 music has become the most popular music format in use today. There are millions of songs available in this format for users to listen to. However the music is never organized according to its features. They are usually organized by artists who in turn are sorted by alphabet or genre. Genre in music is very loosely defined term and thus, the classification of music into genre is also loosely defined and sometimes the classification is not correct or enough to find the music you like. Sometimes, even the artists cover a lot of genres during their career and thus finding a song which has been classified according to artists who in turn are loosely grouped into different genres is a time consuming job. In this loosely sorted world, finding a song that you might like is like finding a needle in a haystack.

It would be much easier to search for new songs if the songs were grouped according to the way they sound. That would put similar songs together and the users will be more likely to find songs that they might like as they can choose songs that are near the songs they already like. This classification method will also allow users to browse their own collection according to the similarities in the songs and thus find a group of songs that match their mood at the time without any hassle.

1.2 Problem Statement

People listen to songs all the time and usually have a fondness for certain specific types of songs, but they can never find out what a new song sounds like without listening to it first. Also the genres of music are very loosely defined so they cannot really depend on the genre to decide if they are going to like the song or not. So if a person wants to listen to new songs he has to try them based on their past experience with the artist or has to just try it before deciding if s/he likes it.

Even the songs in the user's own music collection are usually organized according to the artists and thus if a person wants to create a playlist, they cannot create a playlist of similar songs automatically. They have to do it manually or add all songs of the library. This is a tedious process that if automated can save a lot of hassle for the user.

1.3 Objectives

Some of the major objectives are listed below:

- Develop a Self Organizing Map for a user's music archive that organizes the user's songs according to its similarity with other songs in the user's archive.
- Build an online Self Organizing Map by combining individual user's Self Organizing Maps.
- Allow people to search for new or unheard songs that they most probably will like due to the fact that it is similar to the songs they already like

2. Literature Review

2.1 MP3 Music format

MPEG-1 Audio Layer 3, more commonly referred to as MP3, is a digital audio encoding format using a form of lossy data compression.

It is a common audio format for consumer audio storage, as well as a de facto standard encoding for the transfer and playback of music on digital audio players.

MP3 is an audio-specific format that was co-designed by several teams of engineers at Fraunhofer IIS in Erlangen, Germany, AT&T-Bell Labs in Murray Hill, NJ, USA, Thomson-Brandt, and CCETT. It was approved as an ISO/IEC standard in 1991.

MP3's use of a lossy compression algorithm is designed to greatly reduce the amount of data required to represent the audio recording and still sound like a faithful reproduction of the original uncompressed audio for most listeners, but is not considered high fidelity audio by audiophiles. An MP3 file that is created using the mid-range bit rate setting of 128 kbit/s will result in a file that is typically about 1/10th the size of the CD file created from the original audio source. An MP3 file can also be constructed at higher or lower bit rates, with higher or lower resulting quality. The compression works by reducing accuracy of certain parts of sound that are deemed beyond the auditory resolution ability of most people. This method is commonly referred to as perceptual coding. It internally provides a representation of sound within a short term time/frequency analysis window, by using psychoacoustic models to discard or reduce precision of components less audible to human hearing, and recording the remaining information in an efficient manner. This is relatively similar to the principles used by JPEG, an image compression format.

2.2 Psychoacoustics

Psychoacoustics is the study of subjective human perception of sounds. Alternatively it can be described as the study of the psychological correlates of the physical parameters of acoustics. The psychoacoustic model provides for high quality lossy signal compression by describing which parts of a given digital audio signal can be removed (or aggressively compressed) safely - that is, without significant losses in the (consciously) perceived quality of the sound.

It can explain how a sharp clap of the hands might seem painfully loud in a quiet library, but is hardly noticeable after a car backfires on a busy, urban street. This provides great benefit to the overall compression ratio, and psychoacoustic analysis routinely leads to compressed music files that are 1/10 to 1/12 the size of high quality original masters with very little discernible loss in quality. Such compression is a feature of nearly all modern audio compression formats. Some of these formats include MP3, Ogg Vorbis, WMA, MPEG-1 Layer II (used for digital audio broadcasting in several countries) and ATRAC, the compression used in Minidisc and walkman.

Psychoacoustics is based heavily on human anatomy, especially the ear's limitations in perceiving sound as outlined previously. To summarize, these limitations are:

- High frequency limit
- Absolute threshold of hearing
- Temporal masking
- Simultaneous masking

2.3 Audio Analysis

Audio Analysis requires the following steps

2.3.1 Preprocessing

Since the majority mp3 is a compressed data format such, the audio must be converted to a raw data format such as Pulse Code Modulation (PCM) [4]. In addition, since the music must be simply recognizable for feature extraction rather than high quality, it has been determined that mixing stereo signals to mono and down-sampling to 11 kHz does not significantly affect accuracy, yet greatly increases performance.

2.3.2 Feature Extraction

Some of the major Features of audio that need to be extracted for analysis are:

2.3.2.1 Timbre

Timbre is the quality of a musical note or sound that distinguishes different types of sound production, such as voices or musical instruments. The physical characteristics of sound that mediate the perception of timbre include spectrum and envelope. Timbre is also known in psychoacoustics as *sound quality* or *sound color*.

2.3.2.2 Rhythm Patterns

Rhythm is the variation of the length and accentuation of a series of sounds and is a major difference between different types of songs.

2.3.3 Similarity Calculations

The feature vectors produced for each song represent a vast amount of data. In order for these data to be compared on-the-y for dynamic playlist generation, efficient comparison algorithms must be employed. Various techniques are adopted, most commonly Earth Mover's Distance (EMD), asymptotic likelihood approximation, or Monte Carlo sampling.

2.3.4 Optimization

Each step in the process of determining musical similarity requires a significant amount of computational resources. A variety of parameters can be varied in order to achieve increased performance without significantly affecting accuracy.

Speed is a key component in order for a system to be appealing to the user; it must be both accurate and efficient. However, the speed of one component may be more important than others. In most cases, speed of comparison may be most important, as feature extraction is only performed once.

2.4 Self Organizing Map

The Self-Organizing Map is a powerful tool for explorative data analysis and in particular to visualize clusters in high-dimensional data. Method with similar abilities includes Principle Component Analysis, Multi-Dimensional Scaling, Sammon's mapping, or the Generative Topographic Mapping. One of the main advantages of the SOM is that new information can easily be placed on the map according to the existing organization. Furthermore, the SOM is a very efficient algorithm which has proven to be capable of handling huge amounts of data. It has a strong tradition in the organization of large text archives, which makes it an interesting choice for large music archives. The SOM usually consists of units which are ordered on a rectangular 2-dimensional grid. A model vector in the high-dimensional data space is assigned to each of the units. During the training process the model vectors are fitted to the data in such a way that the distances between the data items and the corresponding closest model vectors are minimized under the constraint that model vectors which belong to units close to each other on the 2-dimensional grid, are also close to each other in the data space. For our experiments we use the batch-SOM algorithm. The algorithm consists of two steps that are iteratively repeated until no more significant changes occur. First the distances between all data items and the model vectors are computed and each data item is assigned to the unit that represents it best.

It is an unsupervised neural network with applications in various domains including audio analysis. The objectives of the SOM are to map high-dimensional data to a 2-dimensional map in such a way that similar items are located close to each other. The SOM consists of an ordered set of units which are arranged in a 2-dimensional visualization space referred to as the man. Common choices to arrange the map units are rectangular or hexagonal grids. Each unit is unit closest to a data item gets n points, the second n-1, the third n-2 and so forth, for the n closest map units.

3. Feasibility Analysis

3.1 Technical Feasibility

This project will require a good knowledge of Audio Analysis and Self-Organizing Maps. There are many Audio analysis methods which may be useful to a varying degree for our specific type of project. These methods will have to be studied and analyzed to find the best fit or we might have to use a hybrid of these methods. Self-Organizing Maps is one of the most promising fields in Artificial Intelligence and thus we expect to get good support and supervision from the department on this subject. The Internet also has many resources on these subjects and many research papers have been published on these subjects. Thus with a bit of in depth research on the subjects, we believe we will become able to build the necessary modules for the project without much hassle. The online modules are all straightforward and technology to create them has been in operation for a number of years already. Thus we believe that with the help of technical supervision from the department, access to research papers and other information in the internet as well as different books available on the subject, the project is technically feasible.

3.2 Operational Feasibility

The main concern about this project's operational feasibility is the possible inaccuracy of the audio features extraction process and the time it may take to extract and analyze the features. However the inaccuracy may be minimized by trying different feature extraction methods and choosing the best of the best combination of the methods. The extraction and analysis time may be decreased by extracting and analyzing sample features from the audio file instead of extracting the features from the whole audio piece. And since the initial version of this project is going to be academic in nature, the analysis time is not a big restriction from doing the project. The process can be refined as the project is developed and we get an exact idea of the limitations that might occur. So this project does not have any major operational infeasibility and even its apparent problems at this hour can be minimized as we do more research and start to apply the research on to the project.

3.3 Economic Feasibility

This project is being developed as an academic project, thus there is no cost of manpower. We are not going to require any extra hardware, thus there is no extra cost for hardware. The only hardware that we require is a personal computer which all project members already own. The possible source of cost may be any proprietary codes or tools we may have to use during project development but we will try to find free alternatives to the tools as much as possible. Another source of cost may come in implementation of the project, if it is deployed for public use. The project will then require web hosting and processing services. However since this project is going to be deployed for academic purpose in its initial phase, we expect the web costs to be virtually nil. The income from this project will have to come from online ad revenues, as this will be a music searching application, we believe recording companies and artists will love to have advertisements in this project. But the income will directly depend on the wide deployment and massive success of the project. However since, the cost of developing the project is very low, the possible income, while welcome, is not really necessary and the lack of it will not hamper the feasibility of this project economically. Thus, we believe that the project is economically feasible.

3.4 Risk Analysis

The success or failure of this project will primarily depend on the robustness and accuracy of audio feature extraction algorithms. As this is a field that is still being researched in, the algorithms are not perfect. However according to the research we have carried out so far, the currently available algorithms give acceptable results on audio feature extraction. Also, Self Organizing Map is a major part of our project and as it is a very complex field, extensive research will be required on the issue. Thus these are the major risks involved in this project. But since signal analysis and artificial intelligence are some of the most advanced and cutting edge technologies in the field of computing today. Thus the knowledge gained by doing this project outweighs the risks associated with this project.

4. Proposed System

4.1 Overview

With the availability of high quality compression system we can find a large collection of music archives electronically stored in our computer or in web. Such music archives are organized either according to the artist, album, title or genre. Such organization offer user to locate the title he or she is looking for or to find out which types of music are available in general. A straightforward approach is to use text-based search for the title, artist, album or some phrases in the lyrics. The next most used approach is to explore the music archive, searching for musical styles such as *classical, rock, jazz, pop* etc. but such styles are broadly classified and may or may not verifies the liking of the user. However, such organizations rely on manual categorizations and usually consist of several hundred categories which involve high maintenance costs, in particular for dynamic music collections.

Alternative approach of organization which we proposed is the organization of the musical archives without relying on meta-data on the various pieces of the music, or musical stores, rather relying on the sound information, present in the form of any acoustical wave format, as it is available. Based on the sound signal we extract low level features based on frequency spectra dynamics, and process them using psychoacoustic models [1]of our auditory system. The resulting representation allows us to calculate to a certain degree the perceived similarity between different pieces of music. We use this form of data representation as input to the *Self-Organizing Map*. This neural network provides cluster analysis by mapping similar data items close to each other on a map display, representing various styles of music into which the pieces of music are actually organized.

4.2 System Description and Architecture

4.2.1 Local System Architecture

The proposed local system can be divided into 4 stages: *preprocessing, feature extraction, analysis, visualization on Self-Organizing Map.*

In a preprocessing stage, the audio signal is transformed, down-sampled and split in individual pieces to be sampled. The system then extracts features which are robust towards non-perceptive variations and on the other hand resemble characteristics which are critical to our hearing sensation i.e. rhythmic patterns in various frequency bands. Finally, the data may be optionally converted, organized into clusters using the SOM and uploaded to the central catalog.

The pieces of music are given in audio file format such as MP3 files. We first decode these to the raw Pulse Code Modulation (PCM) audio format. The raw audio is down-sampled to lower frequency and subsequently segmented to be able to optimize for computations.

To sense the loudness in audio, the power spectrum of the audio signal is calculated. The audio signal is then passed through the series of band-pass filters whose center frequencies are closely related to critical-band rates. A total of 24 critical-bands have been defined according to Bark scale.

Using the spreading matrix the power spectrum is spread across the critical bands. The pressure level of the audio signal is calculated in decibel value. Leveling of the signal as per our hearing sensation is done and the specific loudness sensation is calculated.

The loudness of a critical band usually rises and falls several times resulting in a more or less periodical pattern, also known as the rhythm. We find the rhythmic patterns per frequency band by calculating fluctuation strength of each sampled signal and comparison with the other sampled signal.

Using the rhythm pattern we apply the Self-Organizing Map algorithm to organize the pieces of music on a 2-dimensional map display in such a way that similar pieces are grouped close together.

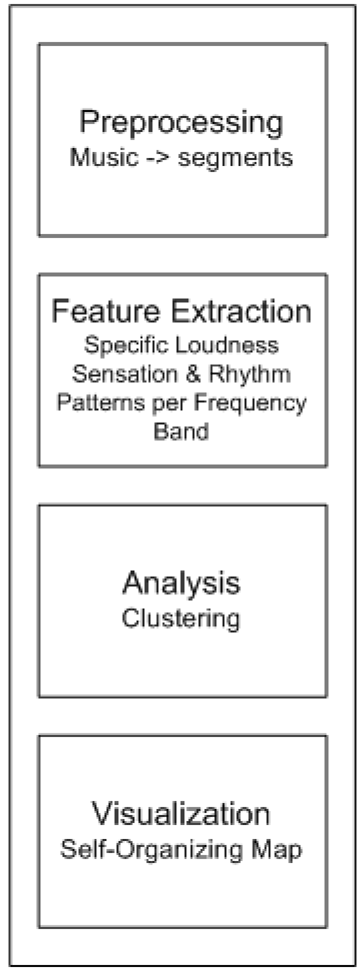

Figure 1: Local System Overview

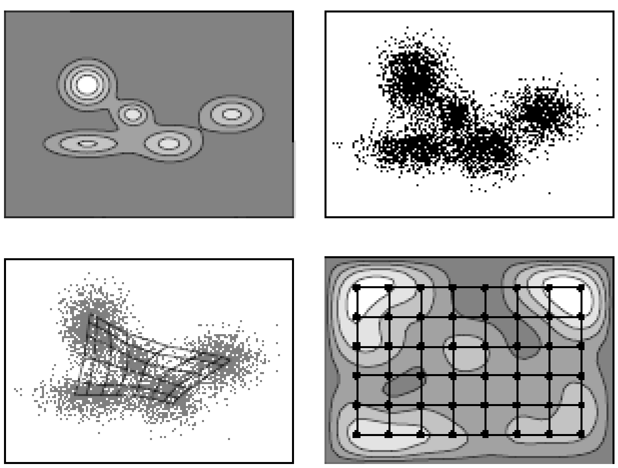

Figure 2: Simple demonstration of SOM

4.2.2 Online System Architecture

The Self-Organizing Map created at the local system is then uploaded to a central server which contains an online Self-Organizing Map which contains data from all the Local Self-Organizing Maps uploaded by the users. Thus this map is bound to be very detailed and contain songs that have never been heard by all the users.

This online Self-Organizing Map can be queried by all users to find songs similar to the song that they have listened to and like. Thus this online system acts as a song recommender to users who want to try new songs that sound similar to the songs that they like. A simple system architecture diagram is given below:

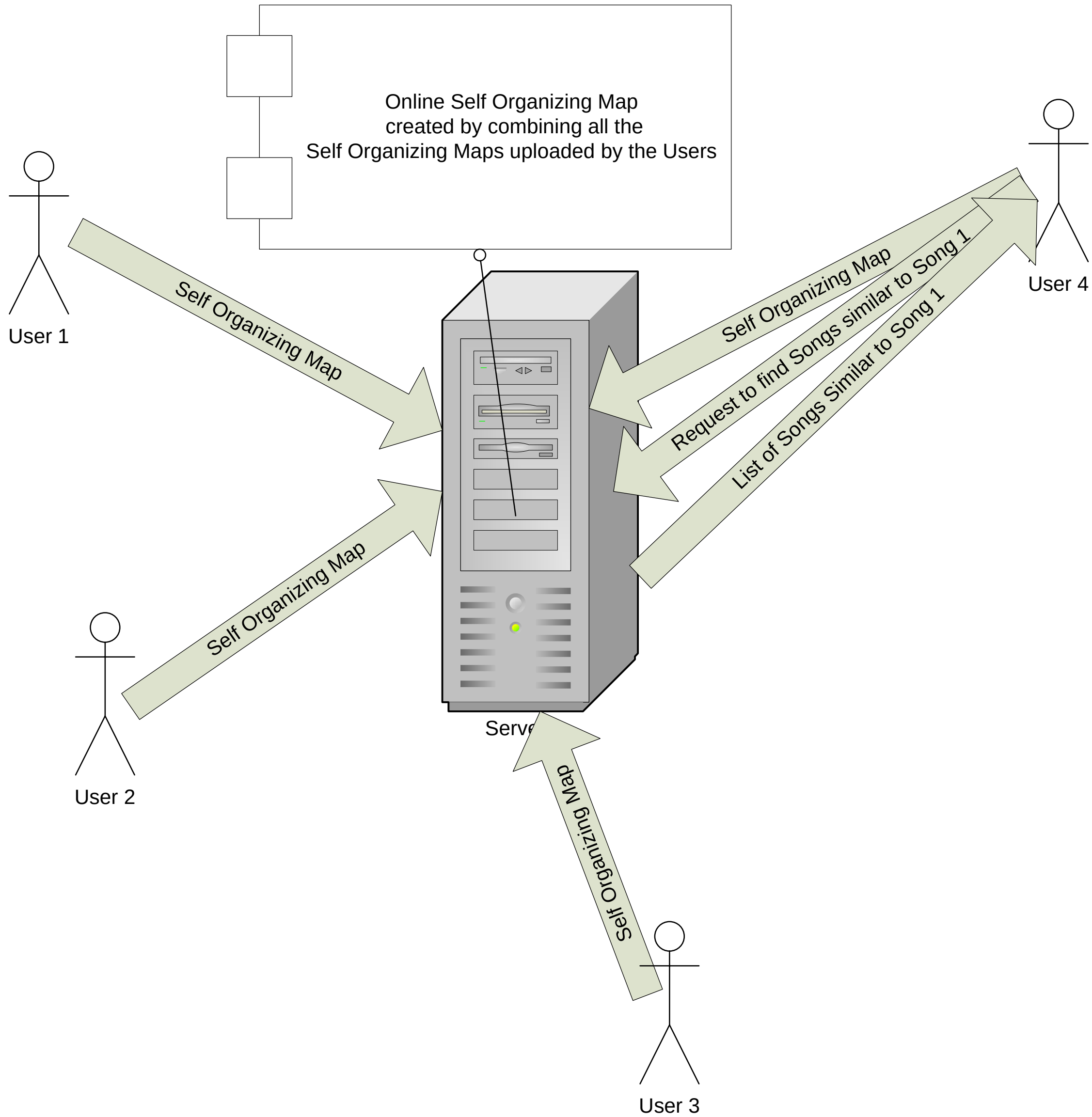

Figure 2: Online System Overview

5. Project Scheduling

Our tentative project schedule is as follows:

ID	Task Name	Start	Finish	Duration
1	Initial Research and Study	9/16/2008	10/6/2008	3w
2	System Design	10/6/2008	10/10/2008	1w
3	Music Analysis Module	10/13/2008	10/31/2008	3w
4	Self Organizing Map Module	10/24/2008	11/6/2008	2w
5	Online Services Module	11/7/2008	11/20/2008	2w
6	Debugging	11/3/2008	12/8/2008	5.2w
7	Testing	12/5/2008	1/5/2009	4.4w
8	Implementation	1/5/2009	1/13/2009	1.3w
9	Documentation	9/16/2008	1/12/2009	17w

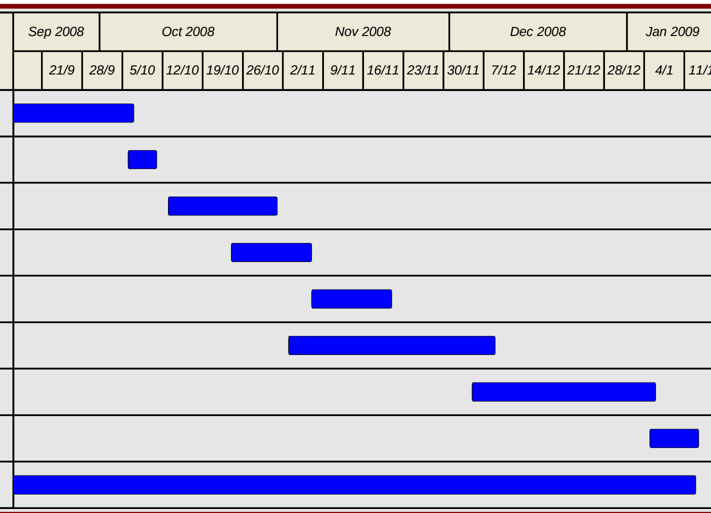

Figure 4: Tentative Project Schedule

6. Expected Results

Basically by the end of this project we expect to have a software that

1) Allows users to automatically build playlists consisting of similar songs, one of each type or totally dissimilar songs as they wish.

2) Create a complex song information archive using inputs from many users giving it the detail that would have been impossible had it been done manually. More popular the software gets, more detailed its content becomes.

3) Build an online repository of similar songs allowing unknown songs to become noticed.

4) Allow people to try new or unheard songs that they most probably will like due to the fact that it is similar to the songs they already like

5) After the SOM becomes really detailed, every user will not need to analyze each song as the information can be accessed via web from the server, the user will have to analyze the song if it is not already present in the server, and when he analyzes the song, its info will also be added to the server, increasing its information content.

7. Conclusion

Finally, we believe that this project is suitable for our final year major project as the project has research value as well as social value. The research part of the project requires research in audio analysis as well as self-organizing maps both of which require a lot of research and the project has social value as it requires social participation and can give benefit to the users by giving them different songs to listen to that they would never have listened to otherwise.

8. References

1. Andreas Rauber, *Automatically analyzing and organizing music archives*, In Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries (ECDL), 2001

2. Elias Pampalk, Andreas Rauber, Dieter Merkl , *Content based Organization and Visualization of Music Archive,* 2002

3. Martin, K. D., Scheirer, E. D., and Vercoe, B. L., *Music content analysis through models of audition*. Presented at the 1998 ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications, Bristol, England, September, 1998.

4. Chih-chin Liu, *Content-based retrieval of mp3 music objects*, In Proc of the Int'l Conf on Information and Knowledge Management (CIKM 2001)

5. Stefan Leitich, Andreas Rauber, *Information Retrieval in Digital Libraries of Music,* Department of Software Technology and Interactive Systems, Vienna University of Technology

6. http://citeseerx.ist.psu.edu

7. http://www.wikipedia.org

8. http://sound.media.mit.edu