# Project: Summarizing and Analyzing Research Papers

## Submission Template

**Learner Name**: P.JAYA LAKSHMI VALLI DEVI

**Learner Email**: vallidevi076@gmail.com

**Topic**: A Machine Learning based Music Retrieval and Recommendation System

**Research Paper**: Shih, H.-H., Narayanan, S. S., and Kuo, C. J. (2002).

### Initial Prompt

**Description (50 words max)**:

we present a music retrieval and recommendation system using machine learning techniques. We propose a query by humming system for music retrieval that uses deep neural networks for note transcription and a note-based retrieval system for retrieving the correct song from the database. We evaluate our query by humming system using the standard MIREX QBSH dataset. We also propose a similar artist recommendation system which recommends similar artists based on acoustic features of the artists’ music, online text descriptions of the artists and social media data. We use supervised machine learning techniques over all our features and compare our recommendation results to those produced by a popular similar artist recommendation website.

**Generated Summary (100 words max)**:

Faster computational speed and increasing number of online users have resulted in a dramatic increase in music consumption. It is getting more and more difficult for the general public, especially non-experts, to find and retrieve music from the millions of songs available online. A lot of research is being done these days to find efficient music retrieval and recommendation methods. One music retrieval method that is gaining a lot of popularity these days due to its convenient usage is query by humming, which is a content-based music retrieval method that can retrieve melodies using users’ hummings as queries. This allows users to find old songs that they only remember the tune of or retrieve obscure songs heard in public places. The Music Information Retrieval (MIR) community has also been doing a lot of work on automatic recommendation systems ranging from the content-based methods to social tagging and similarity networks (Cohen and Fan, 2000; Hong et al., 2008). One of the key research topics in this area that has gained a lot of traction is automatic similar artist recommendation. Currently, there are several musical retrieval and similar artist recommendation apps. There are apps such as SoundHound, MusixMatch etc, that can retrieve songs using humming as a query, and websites such as All Music Guide (AMG)1 and last.fm2 that give similar artist recommendation.

### Iteration 1

**Description (50 words max)**:

The overall system takes a hummed tune as an input, which is then fed to the Query by Humming (QBH) system. The QBH system uses the input to output a ranked list of songs with highest similarities to the query. The user can then either manually choose the correct song from the ranked list or use the default setting of choosing the most highly ranked song as the song to be retrieved. The retrieved song along with its metadata is then used as an input to the similar artist recommendation system, which then outputs a list of most similar artists.

**Generated Summary (100 words max)**:

The Query by Humming system takes a few notes from a melody hummed or sung by the user as the query. The notes of the query is transcribed using our note transcription method and is then passed onto the retrieval system, which uses the transcribed query and the melody database, which refers to the entire list of pre-transcribed melodies or songs that can be recognized by our system, to give a ranked list of melodies that match the input query.

### Iteration 2

**Description (50 words max)**:

We can then measure the similarity between two artists using spectral distance measurements. We extracted musical (Tzanetakis and Cook, 2000), psychoacoustic (Cabrera, 1999) and speech features (Eyben et al., 2010) from Bollywood songs for each artist in a spectral vector representation. Su et al. (2013) used these features successfully in categorizing musical genres and moods. The musical features include timbre, chroma, spectral flatness; psychological features include loudness, sharpness; sound features include frequency and speech characteristics. For each artist si , the feature dimension is 865. Each entry vi(k) is the mean value of the corresponding feature for artist si . The distance of two artists over the acoustic feature space is calculated as follows: d(i, j) = kvi − vjk (1) where i, j stand for two artists, d(i, j) is the distance of artist i and j over acoustic feature space, k · k is the L2- norm, vi is the audio feature vector of artist i, dimension is 865. Note that each feature in the acoustic space has been normalized by its mean and variance.

**Generated Summary (100 words max)**:

This research was partially supported by Grant Number 16214415 of the Hong Kong Research Grant Council and partially supported by Bai Xian Asian Institute. We would like to thank our colleagues from Human Language Technology Center of The Hong Kong University of Science and Technology, who provided expertise that greatly assisted the research. We thank Anik Dey for assistance with his Bollywood knowledge which improved the precision of our work and Ricky Chan for assistance with the acoustic modeling of note transcription system. We would also like to thank the three annotators who helped us label our dataset

### Final Prompt

**Description (50 words max)**:

We show that QBH system overall shows encouraging results and can be improved with additional data. We also propose a similar artist recommendation system and experiment the system on an exhaustive list of 116 Bollywood artists and show that the recommendations based on spectral distance, cooccurrence and degree measures give better results on average for all artists compared to popular similar artist recommendation website. We plan to collect more data in the future and test the system on a larger dataset.

**Generated Summary (100 words max)**:

For training note models, we have used humming data from the IOACAS corpus 8 and TCS corpus 9 with additional humming data collected by us. We annotated the humming data manually using the Tony software 10 . For the evaluation of the overall query by humming system, we have used the standard set used by MIREX for this purpose. It uses the Roger Jang corpus 11, consisting of 4431 queries and 48 ground-truth MIDI files. So, for our experiments we first transcribe the notes in all the ground truth MIDI files. The queries are then each transcribed and passed onto our retrieval system, which generates a list of most likely candidate melodies. The system is evaluated using mean Reciprocal Ranking (MRR): MRR = 1 |Q| X |Q| (i=1) (1/ranki)

### Insights and Applications

**Key Insights (150 words max)**:

The musicological model controls transitions among the note models and the rest in a manner similar to the language model used in speech recognition. The transition probabilities among note HMMs are defined by note bi-grams, which were estimated from a large database of MIDI files containing melodies similar to Ryynanen and Klapuri (2004). Since key is important in de- ¨ termining note transitions as some note sequences are more common than others in a certain musical key

**Potential Applications (150 words max)**:

query by humming, similar artist recommendation, music information retrieval, machine learning

The Query by Humming system takes a few notes from a melody hummed or sung by the user as the query. The notes of the query is transcribed using our note transcription method and is then passed onto the retrieval system, which uses the transcribed query and the melody database, which refers to the entire list of pre-transcribed melodies or songs that can be recognized by our system, to give a ranked list of melodies that match the input query.

### Evaluation

**Clarity (50 words max)**:

Once we have done the logistic regression, we can apply the trained weights to the test set. We evaluate each artist si and calculate the related artists set Cˆ i . Then we compare it to the standard candidate set Ci and calculate the precision, recall and F-score. Table 5 shows parts of our results. There are five columns. The first column is the baseline of Last.fm’s results. The other four are the results from combination of features. Section 5.4. shows that artists with higher degree contain more correlation links to other artists, which partly reflect their influences. We show artists with the highest degree and lowest degree. From Table 5, we can see that Last.fm does not work very well for artists with low degree. Actually, we can not find similar artists information for artists with low degree on the last.fm’s website. On the contrary, our method compensates this shortage.

**Accuracy (50 words max)**:

For training note models, we have used humming data from the IOACAS corpus 8 and TCS corpus 9 with additional humming data collected by us. We annotated the humming data manually using the Tony software 10 . For the evaluation of the overall query by humming system, we have used the standard set used by MIREX for this purpose. It uses the Roger Jang corpus 11, consisting of 4431 queries and 48 ground-truth MIDI files. So, for our experiments we first transcribe the notes in all the ground truth MIDI files. The queries are then each transcribed and passed onto our retrieval system, which generates a list of most likely candidate melodies. The system is evaluated using mean Reciprocal Ranking (MRR): MRR = 1 |Q| X |Q| (i=1) (1/ranki)

**Relevance (50 words max)**:

We can see that our method performs smoothly when dealing with both high degree and low degree artists and performs 40% better on average in F-measure. Actually, the sole spectral mean distance feature has already reached a pretty good precision and recall for some artists. Combined with co-occurrence features, we can see that precision and recall increased for the high degree artists whereas they did not decrease on the low degree artists. However, the co-occurrence feature alone does not perform so well. This may be due to the fact that our corpus is not large enough, so we will continue to collect data in order to improve our results in the future.

### Reflection

**(250 words max)**:

AI has exploded on the scene and everyone is trying to figure it out at the same time which is exciting. A lot of educators are afraid that kids will use technologies like ChatGPT to cheat and have even banned its use. The reality is that kids have been cheating since the beginning of school and we have an exciting opportunity as educators to shape the narrative around this new technology. Biscotti plans to continue to learn as much as she can about how to use AI to better support her students. She says as an educator, “I feel that I am obligated to prepare my students for their future, not my past. These tools will only improve and they are here to stay. It’s imperative that kids are familiar with them and know how to use them or we risk sending graduates out into the workforce at a competitive disadvantage.”