A Statistical Approach to Retrieval under User-dependent Uncertainty in Query-by-Humming Systems

Erdem Unal* Shrikanth S. Narayanan* Elaine Chew

Speech Analysis and Interpretation Laboratory, [sail.usc.edu]*
Integrated Media Systems Center, Viterbi School of Engineering, University of Southern California, CA

unal@usc.edu shri@sipi.usc.edu echew@usc.edu

ABSTRACT

Robustly addressing uncertainty in query formulation and search is one of the most challenging problems in multimedia information retrieval (MIR) systems. In this paper, a statistical approach to the problem of retrieval under the effect of uncertainty in Query by Humming (QBH) systems is presented. Direct transcription of audio to pitch and duration symbols is performed. From the transcribed data vector, finger prints that carry a fixed length of information from characteristic local points of the hummed melody are extracted. Instead of employing the humming input as a whole, extracted characteristic information packages are used for search through the database. The distance for each finger print to the original melodies in the database is calculated and converted to probabilistic similarity measures. Melodies with the highest similarity measures are returned to the user as the most likely query result. This algorithm is tested with manually annotated data comprising 250 humming samples in conjunction with a database of 200 pre-processed midi files. Retrieval accuracy of 94 percent is demonstrated for the samples of subjects that have some musical training/background compared to 72 percent accuracy achieved for the samples of non-trained subjects. Results also show that extracting finger prints with respect to characteristic local points of the hummed tune is an effective and robust way for search and retrieval under the effect of uncertainty.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - *Query formulation, Retrieval model, Search process, Selection process.*

General Terms

Algorithms, Design, Human Factors.

Keywords

Query-by-Humming, Retrieval, Uncertainty

1. INTRODUCTION

Content based multimedia information retrieval (MIR) is a rapidly developing field of research. With the availability of large storage spaces in web servers, and exploding information generation worldwide, this field has gained tremendous momentum. Digital music and other music related information is a key example of such information and the underlying complexity and variety makes interacting with them increasingly challenging. Enabling natural human interactions with these large databases hence becomes a critical component of MIR systems. Humming/Singing is one of the most interesting cases of these human interactions where, a system can take a user's audio input and search using it through the databases. This requires audio information retrieval techniques to be developed for mapping the human humming waveform to pitch and rhythm symbols for the underlying melody. A query engine needs to be developed in order to search with the converted symbols. The search process is required to be robust to user variability in query formulation, since the hummed input may not be as accurate as its original (expected) form in the database. This paper focuses on our statistical approach to the uncertainty problem in the retrieval process of a Query by Humming System.

Ghias et al. [6] was one of the first to propose Query by humming in 1995, and coarse melodic contours were used to represent melodic information. The coarse melodic contour was widely used and discussed in several query by humming systems that followed. Original pieces are also stored in a contour based representation rather than audio format, that enables symbol to symbol comparison. Autocorrelation was used to track pitch and convert humming into coarse melodic contours. String matching algorithms were used to search converted symbols in the database and best ranked songs were returned to the user depending on how well they matched the query. McNab et al. [7, 8] improved this framework by introducing duration contours for rhythm representation. McNab also used a retrieval system similar to Ghias' that ranked the songs in the database according to the string similarities. Blackburn et al. [9], Roland et al. [10] and Shih et al. [11] improved McNab's system by using tree based database searching. Jang et al. [12] used the semitone (half step) as a distance-measure and removed repeating notes in their melodic contour, which was one of the first efforts to implement an algorithm against uncertainty in rhythmic structures of the input. Lu et al. [13] proposed a new melody string which contained pitch contour, pitch interval and duration as a triplet. Lu's retrieval system combined string matching algorithms with dynamic programming and represented a hierarchical search to improve results. Haus et al. [14] implemented rules for correcting errors of contour transcription that are caused by the uncertainty in the humming. As a counter approach to note segmentation, Zhu et al. [15] used dynamic time warping indexes to compare audio directly to the database. Unal et al. [3] created a large database of humming samples collected from people with various backgrounds and analyzed the performance of both musically trained and untrained human subjects with statistical techniques. Building a system that performs pitch and time information based retrieval from a humming piece using statistical data-driven methods has been shown to be feasible [1]. However, since the input is totally user dependent, and includes high rates of variability and uncertainty, the challenge that remains is achieving robust performance under such conditions. The goal of this paper is to present a statistical approach to address this problem by considering the humming performance as a combination of different aspects such as critical pitch and duration changes, rather than a whole.

The rest of the paper is organized as follows. In Section 2, the relevant theory of musical perception of humans is discussed. Following this, in section 3, the proposed Query by Humming system is explained, including discussions about transcription techniques and the proposed search algorithms. Section 4 gives information about the retrieval experiment results and their discussion, and is followed by future research plans and conclusion in section 5.

2. PERCEPTUAL CUES: RELATIVE PITCH AND DURATION CHANGES

Human perception of music is one of the key elements that should be considered while developing a user-centric QBH system. According to previous published research in music perception, music is deemed to be stored in the human brain as relative information of pitch (frequency) and duration of consecutive notes. For example, some researchers have pointed out that while only less than 0.01% of the population possesses the ability to identify absolute pitch [17], almost any musical person has the ability to identify relative pitch – a skill necessary to accurately recognize intervals, that is to say, the distance between two pitches [18]. Similar skills exist for identifying relative durations of different notes. This internal skill for recognizing intervals and relative durations may be internalized without any formal education during the early ages [4].

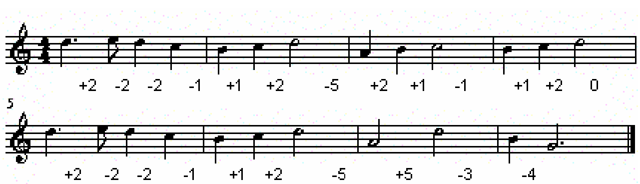


Figure 1. Illustration of the semitone transitions of consecutive notes in the popular nursery song "London Bridge is Falling Down."

Figure 1 shows the musical notation of a well known song called "London Bridge is Falling Down". A person without any prior knowledge of musical notation will not be able to recognize this melody by looking at this figure. However, if this person has a chance to listen to the melody, s/he will be able to recognize it fairly easily. This is because it is generally extremely difficult for most people to memorize absolute information about notes; rather, most of us recognize and remember relative note (pitch and duration) information automatically. In the illustration of Fig. 1, the relative distance between the pitches of two consecutive notes are shown as positive and negative integers that correspond to the semitone level transitions of consecutive notes. The memory of relative durations is stored in a slightly different way in that relative durations are more likely to be mapped into fixed rhythmic structures rather than single duration changes [5].

Finally, when a person listens to a melody, it is believed that the brain automatically performs preprocessing of the pitch contours and rhythmic structures, and compares the information that is processed to those that are previously stored. Our goal in this study is to build a retrieval system that attempts to functionally mimic some features of the mental processing that occurs in human brain when listening to music.

3. QUERY BY HUMMING SYSTEM

To design a robust QBH system, knowledge based methods of signal processing, probability and music theory have to be incorporated. As seen from the Figure 2, we have divided the overall system into small subsystems corresponding Recognition, Transcription and Retrieval.

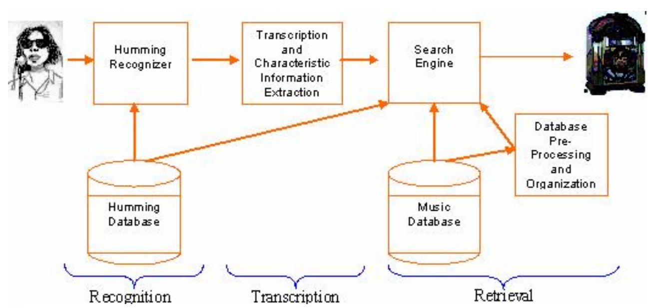


Figure 2. The Proposed Query by Humming System

3.1 Recognition

In our previous studies, we proposed statistical approaches to humming recognition to yield note level decoding. Conceptually, the approach tries to mimic a human's perceptual processing of humming as opposed to attempting to model the production of humming. Such statistical approaches have had great success in automatic speech recognition and can be adopted and extended to recognize human humming and singing [1, 2]. The basic goal in this part of the system is to train Hidden Markov Models with representative human humming data, and perform note level decoding of an unseen humming waveform.

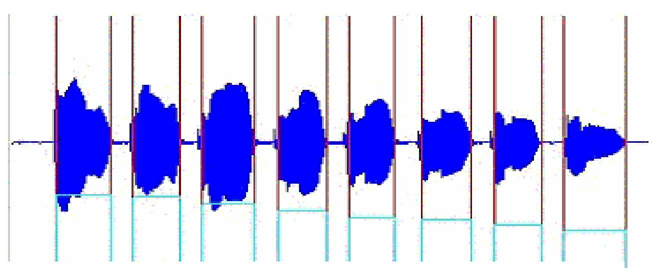


Figure 3. Automatic segmentation of notes in the Humming Recognizer

As seen from Figure 3, recognizer's main task is to separate humming notes from each other and automatically extract duration and pitch information for each note segment as accurately as possible. Vertical lines represent the start and end points of a humming note, which is detected by energy, spectral and pitch change features. Following the segmentation process pitch extraction is handled by standard Fourier transform techniques [1].

3.2 Humming Transcription and The Proposed Notation

Music theory provides a comprehensive way to represent notes and with specific notations. The goal here is to utilize a scheme that is both comprehensive and provides a "inter lingua" between the query formulation front end and the representations in the database. Hence, for query related calculations, the duration and pitch information extracted from the front-end recognizer is transcribed into a specific notation that the proposed QBH system will commonly use in both front-end and the back-end. Furthermore, this representation will allow us incorporate our statistical findings about the humming data that we previously analyzed [3]. This includes statistical analysis of humming samples of 100 various people, performing a pre-defined melody list of songs that cover almost every single pitch interval in one full octave.

A music note has two main attributes, namely pitch and duration. In traditional music notation, absolute pitches are given names such as "A B C D…" or are more commonly known (in absolute solfege) as "la, ti, do, re…" However, this notation is not helpful in a user centric domain, since the humming can be performed in any different key that in turn results in shifting of notes up or down according to the starting note of the humming. While a humming piece can be performed starting with any note, the fact that makes it sound like the original melody is the accurately hummed note transitions, which is the way music perception theory supposes that human brain stores information about a music piece. In the case of simple example of "London Bridge is Falling Down" in Fig. 2.1, the original piece starts with a "re" and continues with "mi re do ti do re". The original note sequence follows a semitone transition path (+2 -2 -2 -1 +1 +2) as shown with integer values on the figure. The exact same melody can be still be recognizably reproduced (sung or hummed) starting with any pitch, as long as the semitone differentials are performed correctly.

More errors at the note transitions decrease the perceived quality of humming, and also the humming piece may start to resemble some other melody rather than the target one, the subject meant to perform. That is one of the reasons that motivated our statistical approach to this problem with probabilistic measures that will be discussed in the retrieval part in Sec. 3.4.

After extracting duration and pitch values for each humming note, we use the following calculations in order to label the note transitions. The Relative Pitch Difference (RPD) information is calculated with the formula:

$$\text{RPD} = \frac{\log(f_k) - \log(f_{k-1})}{TDC}$$

where f is the calculated frequency value of the humming note, k is the index of the note and TDC is the known musical theoretical distribution constant of semitones, $\log \sqrt[12]{2}$ .Similarly, Relative Duration Difference (RDD) is calculated as the ratio of durations of two consecutive notes where t is the duration of one humming note and can be represented as:

$$\text{RDD} = \frac{t_k}{t_{k-1}}$$

In contrast to other commonly used transcription techniques, in our transcription, we allow float number ratios because, incorporating the details of user-centric variability is the main task of building a robust system. A human can not be expected to hum a note transition perfectly, meaning that the logarithmic pitch difference will not necessarily correspond to integer ratios. That's why we want to record the variability, as a measure of error that is encountered during the humming.

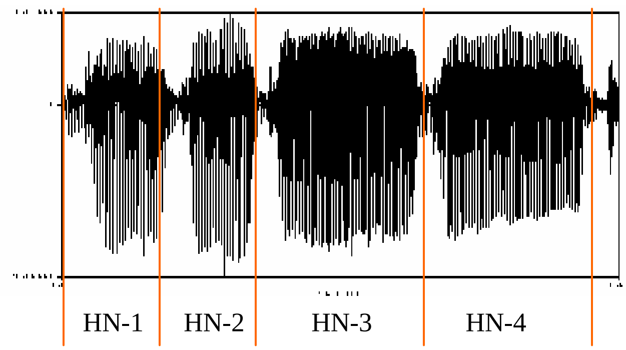


Figure 4. Segmented Notes and Labeling

Figure 4 shows a 4 note humming waveform and the vertical lines represent the temporal boundaries that are calculated by the recognizer. The recognizer also extracts the numeric information for pitch and duration that will be used in transcription. Table 1 shows the corresponding transcription of the extracted information for each humming note (HN) using the notation we used in this study.

Table 1. Pitch and Duration Transcription

	HN#1	**HN#2**	**HN#3**	**HN#4**
Pitch (Hz)	111.7	123.25	141.98	128.63
Duration (sec)	0.326	0.255	0.456	0.520
Pitch Transcription(***PT***)	0	1.703	2.449	-1.709
Duration Transcription(***DT***)	1	0.78	1.14	0.94

A positive pitch transcription corresponds to increase in the previous hummed note's pitch and a value that is larger than 1 in the duration transcription refers to an increase in the duration.

3.3 Extracting Finger Prints at Local Characteristic Points

Our proposed solution to retrieval under the effects of uncertainty is to collect finger prints, i.e., the salient melodic information, from the transcribed information vector and search using them through the database. For identifying the characteristic points in the humming piece, aspects of musical composition are considered. Fluctuating and monotonic note and rhythmic transitions, and repeated transition structures are the most common factors that make a music tune unique and easy to remember. However, high levels of pitch and rhythmic transitions are vulnerable to uncertainty which can be explained by Fitt's Law [16]. It means that the above referred characteristic points may potentially include high levels of uncertainty.

Finger prints should spot these local points in the humming and extracting these information combinations will be helpful in an accurate and robust retrieval engine. In our study, we only selected extracting information at monotonic and fluctuating pitch and rhythm patterns and created information packages at these local characteristic points. Each information package is a $2 \times 2k+1$ matrix, in which the first column carries the pitch transcription information and the second row has the duration transcription, centering on the desired cell.

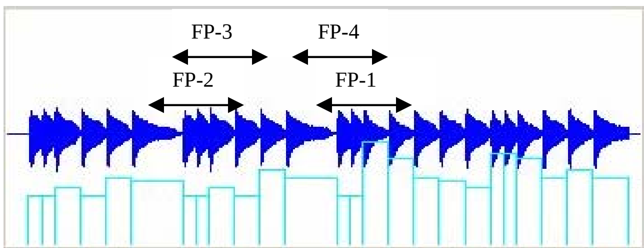


Figure 5. Graphical representation of a humming sample with labeled finger prints (FP's)

Figure 5 shows a full humming piece with segmented notes. The rectangular bars under each humming note is the graphical representation, where the width of each rectangular is the duration and the height is the extracted pitch information of the corresponding note. Fallowing that, in table 3, finger prints (FP1-4) that are extracted form the above humming piece are shown. Third columns of each finger print spots the desired characteristic local point. FP1, shown in table 2.1, centers the highest absolute note transition in the whole humming sample and creates a $2 \times 2k+1$ matrix for the most closest 2 cells. Variable k is the length of the information package which, in this case, selected to be 2. Similarly, FP2 is created for the smallest pitch transition point, which means it centers the point where two consecutive notes are the closest with respect to their pitch levels. FP3 and FP4 shows the extracted information packages for the largest and the smallest absolute duration change, respectively.

Table 2.1. FP1: largest pitch transition

PT	-4.62	-0.33	10.99	-2.21	-3.67
DT	0.48	0.86	2.42	0.94	1.06

Table 2.2. FP2: smallest pitch transition

PT	-0.81	-4.79	0.07	2.44	-1.92
DT	1.05	0.56	0.77	2.86	1.18

Table 2.3. FP3: largest duration change

PT	-4.79	0.07	2.44	-1.92	6.54
DT	0.56	0.77	2.86	1.18	0.91

Table 2.4. FP4: smallest duration change

PT	-1.92	6.54	-1.82	-4.62	-0.35
DT	1.18	0.91	1.01	0.48	0.86

3.4 Search and Retrieval

Figure 2 shows a music database of original melodies and a database organizer cooperating with a query engine. The initial task of the query engine is to perform a robust search of the extracted finger prints through the appropriately organized database and to pull out the matched melodies from the database. Hence, for this search to be accomplished, the database needs to pre-process the original music melodies and store the transcribed information.

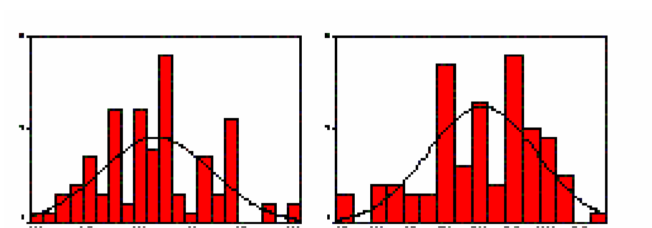
At this stage of our research, this task has not been automated yet, so we had to manually extract main melodies from the original midi files we selected for small scale testing. Since our humming database will be primarily used for testing the accuracy of the retrieval, we have prepared a melody database of 200 melodies including the pieces we used in our data collection experiments discussed in [3] and pieces from some famous bands such as Beatles.

Each characteristic information package (FP) is searched through the preprocessed database where all pitch and duration transcription of the 200 original midi files are present.

3.4.1 Prediction Intervals for Search

Since data is from a number of human subjects with different musical backgrounds, high levels of uncertainty and variability is expected in their humming. However, while uncertainty is increased, especially in the case of untrained subjects, there is a trade off between accepting this variability against the accuracy of the retrieval engine. High levels of uncertainty may lead to radical changes of the structure of the melody that eventually result in mismatches. Our approach to this problem is data driven.

For each level of pitch and duration transitions, we build prediction confidence intervals (PCI) that are used in guiding of the search engine as shown in figure 5.



Transition: 4 semitones
Mean: 3.80
Std: 1.08
PCI[4] (p=0.05):1.603< x <6.042

Transition: 12 semitones
Mean: 10.08
Std: 1.80
PCI[12] (p=0.05): 8.316< x < 12.766

Figure 6. Histogram of training data set, normal distribution curve and Prediction Confidence Intervals(PCI) for 4 and 12 semitone pitch transitions.

The note transitions performed by the subjects are tested ($p<0.05$ K-S test) to be normally distributed for each level ranging from 1 to 12. In general, 100 samples which are randomly chosen from our humming database are used for training of these prediction intervals. In some cases, we had smaller number of samples due to data problems. Figure 6, shows the prediction intervals for the case of a 4 and 12 semitone note transitions where, subjects hummed in the melody London Bridge, and Happy birthday, respectively. The prediction interval of a 4 semitone (PCI[4]) implies that, a performed transition which falls into the given limits may belong to a semitone level difference of 4.

In case of FP1, shown in Table 2.1, since the center value (10.99) does not belong to the given limits of PCI[4], the query engine will not search for a similar pattern around a 4 semitone level pitch transcription value in the database.

Table 3. Calculated Prediction Intervals

Semitones	#of Samples	Lower Confidence Limit	Upper Confidence Limit
1	100	-.91920	3.52570
2	100	-0.12576	4.23163
3	100	.76287	5.20306
4	100	1.60343	6.04220
5	100	2.44369	6.88166
6	18	3.01297	7.69543
7	100	4.12326	8.56150
8	100	4.96258	9.40189
9	100	6.12077	10.45102
10	24	7.67491	11.23122
11	-	-	-
12	100	8.31676	12.76655

Table 3 shows the calculated prediction intervals for each of the semitone level transitions. The prediction interval for 11 semitones transition that corresponds to a major 7^{th} is not calculated due to lack of data. According to the table, the center value for the given example FP1, 10.99, may belong to the 10, 11(expected) or the 12 semitone transition.

3.4.2 Error Calculation and Probability Assignment

For this particular case, the query engine will search through the values (10, 11 and 12) and will create candidate (C) matrices of length $2 \times 2k+1$. The numerical difference of the two matrices is then calculated which represents the total error (TE) for that particular local characteristic point.

$$TE_1 = \sum_{n=1}^{2} \sum_{m=1}^{2\times 2k+1} \{[FP_1]_{nm} - [C]_{nm}\}^2$$

A smaller TE value means, the candidate information package C, is more similar to FP1. The search Engine calculates TE values for every single candidate matrix and these values are stored in a list sorted from the lowest TE value to the highest. The melody "*q*", which carries the smallest TE value, is assigned the highest probability for the correct match, that is $P(CM_q \mid FP_1)$. Other candidate songs receive smaller probability values according to their increasing TE values. These steps are repeated for all mentioned finger prints and finally a similarity measure (S_q) that is the product of probabilities of each finger print is calculated for the candidate song *q*.

$$S_q = \prod_{j=1}^{4} P(CM_q \mid FP_j)$$

In our system, the 5 songs that received the 5 highest similarity measures are returned to the user as the query result (this can be generalized to the top N best matches).

4. EXPERIMENTS and RESULTS

The aforementioned algorithm was tested with a database of 200 preprocessed midi files. For testing, 250 humming pieces from our humming database [3] were also selected, equally distributed between both musically trained and non-trained subjects. Some samples from non-trained subjects carried so much uncertainty that, we had difficulty in recognizing what song they were trying to hum. This in turn affected the retrieval results for non-trained samples. On the other hand, results are promising in the case of musically trained subjects.

Table 4.1 Results for Non-Trained Subjects

Size of Database	50		100		200	
	Top of the list	Within first 5	Top of the list	Within first 5	Top of the list	Within first 5
K=1	%54	%92	%42	%86	%38	%80
K=2	%84	%100	%78	%90	**%72**	%88
K=3	%82	%100	%80	%86	%72	%86

Table 4.2 Results for Trained Subjects

Size of Database	50		100		200	
	Top of the list	Within first 5	Top of the list	Within first 5	Top of the list	Within first 5
K=1	%82	%96	%80	%94	%76	%88
K=2	%100	%100	%98	%100	**%94**	%100
K=3	%100	%100	%98	%100	%98	%100

The result cells are divided into two columns, the first column represents the success ratio when the correct match is at the top of the result table, and the second column gives the success ratio when success is considered as giving the correct match within the top five results. Table 4.1 shows the experiment results for non-trained subjects' samples. When *k*, the length of the finger prints, is selected to be 2, 72% accuracy is achieved. On the other hand, Table 4.2 shows the results for trained-subjects for the same selection of *k* and 94% accuracy is presented.

The effect of musical training is obvious as expected. As mentioned earlier, the increase in the uncertainty that is caused by the subject in their humming, results in mismatches with respect to what is expected in the canonical melody structure, and this can be easily observed from the Table 4.1.

Also the effect of the length of the information matrix on the retrieval performance is considerable. Increasing *k* directly improved the results in the case of humming pieces of musically trained subjects; on the other hand, this effect is non-monotonic for the case of non-trained subject's samples. This is because increasing *k* not only increased the amount of information stored in the finger prints but also the effect of the uncertainty in the query calculations.

Increasing the size of the database has a negative effect on the retrieval accuracy and this effect is more significant in the case of non-trained subjects. For trained subject's samples, this effect is considerably small.

5. FUTURE WORK and CONCLUSION

This paper presented a statistical approach to the problem of retrieval under uncertainty for query by humming. A primary source of motivation for our research came from observations regarding the way humans perceive music. We incorporated the knowledge based methods regarding human understanding of music with data based models. Our statistical approach required extracting finger prints from hummed melodies and searching these finger prints under the guidelines regarding inter-subject characteristics that we calculated from previously collected data. Results showed that the size of the database, the size of the finger prints and subject's musical training are the main factors that affected the success ratio of our approach.

Future work includes automating the preprocessing of the original melodies, so that we can test the validity of our ideas in the case of larger databases. We would also like to include the repeated pattern extraction feature as a new finger print and see its effects on retrieval accuracy. Since scaling is a problem in terms of retrieval accuracy, we would like to handle this with adding new features to the query formulation and feeding our statistical approach with appropriate database organization algorithms.

6. ACKNOWLEDGEMENTS

This work was funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152, in part by the National Science Foundation Information Technology Research Grant NSF ITR 53-4533-2720, and in part by ALi Microelectronics Corp. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation and ALi Microelectronics Corp.

7. REFERENCES

[1] Shih H.-H., Narayanan, S. S. and Kuo, C.-C. J. An HMM-based approach to humming transcription. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME2002),* August 2002.

[2] Shih H.-H., Narayanan, S. S. and Kuo, C.-C. J. Multidimensional Humming Transcription Using Hidden Markov Models for Query by Humming Systems. *In Proceedings of IEEE International conference on Acoustics Speech and Signal Processing,* 2003

[3] Unal, E., Narayanan, S. S., Shih, H.-H., Chew, E., Kuo, C.-C. J. Creating data resources for designing user-centric front-ends for Query by Humming systems. In *Proceedings of 5th ACM Multimedia Information Retrieval Conference'03,* (Berkeley CA, November, 2003)

[4] Bamberger, J. Turning Music Theory on its Ear. *International Journal of Computers for Mathematical Learning* Vol.1, No.1, 1996

[5] Desain, P, Honing, H. The formation of rhythmic categories and metric priming. *Music Perception*, 2003, Vol 32, pp 341-365

[6] Ghias, A., Logan, J., Chamberlin, D. and Smith B.C. Query by humming: musical information retrieval in an aoudio database. In *Proceedings of ACM Multimedia Conferenece'95* (San Francisco, California, November 1995)

[7] McNab, R. J., Smith, L. A., Witten, I.H., C.L. Henderson, C.L., and Cunningham, S.J Towards the digital music library: Tune retrieval from acoustic input. In *Proceedings of Digital Libraries Conference* 1996.

[8] McNab, R. J., Smith, L. A., Witten, I.H., Henderson, C.L. Tune Retrieval in multimedia library. Multimedia Tools and Applications, vol.10, 2000.

[9] Blackburn, S. and DeRoure, D. A tool for content based navigation of music. In Proceedings of ACM Multimedia 98, 1998, pp. 361-368

[10] Rolland, P.Y., Raskins, G., and Ganascia, J.G. Music content-based retrieval: an overview of melodiscov approach and systems. In P*roceedings of ACM Multimedia 99* (November 1999)

[11] Shih, H.-H. , Zhang, T. and Kuo, C.-C. J. Real-time retrieval of song from music database with query-by-humming. In *Proceedings of ISMIP* (1999), 251-57.

[12] Chen B. and Roger Jang, J.-S. Query by Singing. In *Proceedings of 11th IPPR Conference on Computer Vision, Graphics and Image Processing* (Taiwan, 1998).

[13] Lu, L., You, H., and Zhang, H.-J. A new approach to query by humming in music retrieval. In Proceedings of *IEEE International Conference on Multimedia and Expo* (2001)

[14] Haus, G. and Pollstri, E. An Audio Front End for Query-by-Humming Systems. *In Proceedings of ISMIR 2001*(Bloomington, Indiana, October 2001)

[15] Zhu, Y. and Shasha, D. Warping Indexes with Envelope Transforms for Query-by–Humming. In *Proceedings of ACM SIGMOD 2003* (San Diego, CA, June 2003)

[16] Huron, D. Tone and Voice: A Derivation of the Rules of Voice-leading from Perceptual Principles. *Music Perception*, Vol. 19, No. 1 (2001) pp. 1-64.

[17] Rossing, T. D., *Science of Sound*, 3rd ed. (with F. Richard Moore, Paul A. Wheeler), Addison-Wesley, San Francisco, 2002

[18] Capleton., B. *Perfect Pitch* http://www.amarilli.co.uk/piano/perfectp.asp