Creating Data Resources for Designing User-centric Front-ends for Query by Humming Systems

Erdem Unal* S. S. Narayanan* H.-H. Shih* Elaine Chew C.-C. Jay Kuo

Speech Analysis and Interpretation Laboratory, [http://sail.usc.edu]*

Integrated Media Systems Center

University of Southern California Viterbi School of Engineering, California, USA

unal@usc.edu shri@sipi.usc.edu maverick@aspirex.com echew@usc.edu cckuo@sipi.usc.edu

ABSTRACT

Advances in music retrieval research greatly depend on appropriate database resources and their meaningful organization. In this paper we describe the data collection efforts related to the design of query by humming (QBH) systems. We also provide a statistical analysis for categorizing the collected data, especially focusing on inter-subject variability issues. In total, 100 people participated in our experiment resulting in around 2000 humming samples drawn from a predefined melody list consisting of 22 different well known music pieces, and over 500 samples of melodies that were chosen spontaneously by our subjects. These data will be made available for the research community. The data from each subject were compared to the expected melody features, and an objective measure was derived to quantify the statistical deviation from the baseline. The results showed that the uncertainty in the humming varies with respect to the melodies' musical structure and subjects' musical background. Such details are important for designing robust QBH systems.

Categories and Subject Descriptors

H.3.2 **[Information Storage and Retrieval]:** Information Storage – *file organization.* H.5.5 **[Information Interfaces and Presentation]:** Sound and Music Computing – *methodologies and techniques.*

General Terms

Design, Human Factors.

Keywords

Humming database, uncertainty quantification, query by humming, statistical methods.

1. INTRODUCTION

Content based multimedia data retrieval is a developing research area. Integrating natural interactions with multimedia databases is a critical component of these kinds of efforts. Using humming, a natural activity of humans, for querying data is one of the options.

This requires audio information retrieval techniques to be developed for mapping the human humming waveforms to pitch number strings representing the pitch and rhythm contours of the underlying melody. A query engine needs to be developed in order to search the converted symbols in the database and it should be precise and robust to inter-user variability and uncertainty in query formulation.

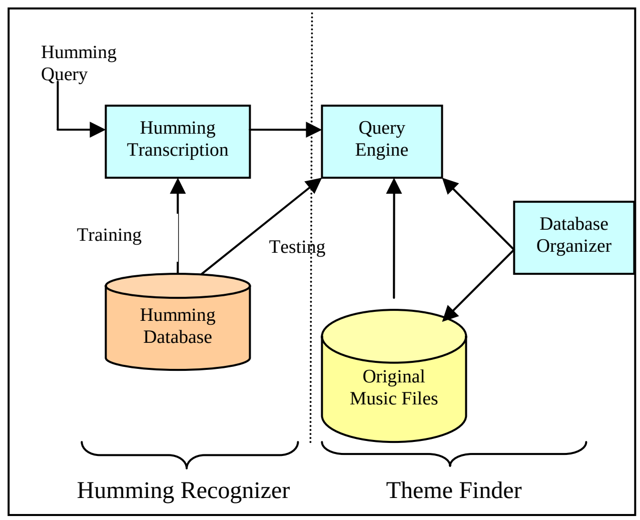

Figure 1.1: Flowchart of our Query by Humming System.

Ghias et al. [6] was to 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. Autocorrelation was used to track pitch and convert humming into coarse melodic contours. McNab et al. [7, 8] improved this framework by introducing duration contour for rhythm representation. 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. Lu et al. [13] proposed a new melody string which contained pitch contour, pitch interval and duration as a triplet. All these efforts had significant contribution to the topic.

1.1 The Role of the Study in QBH Systems

Our proposed statistical approach to humming recognition aims at providing note level decoding. Since it is data-driven, it provides more robust processing in terms of handling variability in humming. 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]. In order to achieve this, a humming database needs to be developed that captures and represents the variable degrees of uncertainty that can be expected by the front-end of the Query by Humming System.

Our goal in this study is to create a humming database that includes samples of people with various musical backgrounds in order to make statistical categorization of inter-subject variability and uncertainty in the collected data. Our research contributes to the community, by providing a publicly available database of human humming, one of the first efforts of its kind.

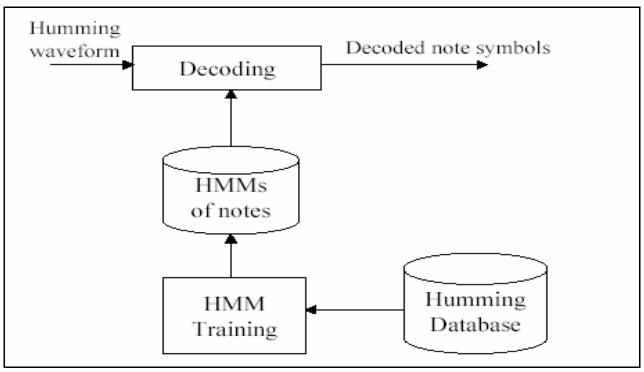

Figure 1.1.2 The role of Humming Database in statistical humming recognition approach.

As seen from the figure 1.1.2, the collected data will be used to train the Hidden Markov Models that we used to decode the humming waveform. From the uncertainty analysis we performed, we will be able to select which data is going to be used in the training set so that; inaccurate data will not effect the decoding accuracy. On the other hand, the whole data can also be used to test the accuracy of the retrieval algorithms.

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. In section 2, we will discuss our hypothesis about the sources of uncertainty in humming performance. Since our proposed approach is based on statistical pattern recognition, it is critical that the test and training data adequately represent the kinds of variability expected.

In section 3, we describe the experimental methodology detailing the data collection procedure. The information about the data and its organization is explained in section 4. In section 5, we present statistical analysis aimed at quantifying the sources and nature of user variability. Results are presented in section 6 in the context of our hypothesis.

2. HYPOTHESIS

The data collection design was based on certain hypotheses regarding the dimensions of user variability. We hypothesized that the main factors contributing to variability include the musical structure of the melodies that are being hummed, the subject's familiarity to the song and the subject's musical background, and that these effects can be modeled in an objective fashion using the audio signal features.

2.1 Musical Structure

The succession of notes and the rhythm of a melody are the features that greatly influence how well a human can faithfully reproduce it through humming. Some melodies have a very complex musical structure in that they have difficult note transitions and complex rhythmic structures that make them difficult to hum. When we create a database, we wish to have samples reflecting a range of musical structure complexity. The note succession as notated in the score of the melodies was the main feature that we used to categorize the musical structure.

Pitch range is an important factor affecting the difficulty of a melody. We measured the pitch range of the songs according to two statistics: the difference between the highest and the lowest note of the melody and, more importantly, the largest semitone differential (interval) between any two consecutive notes. For example, two of the well known melodies we asked our subjects to hum; "Happy Birthday" and "Itsy Bitsy Spider" have different musical structures. The range where the all notes in "Happy Birthday" is one full octave (12 semitones), while the range in "Itsy Bitsy Spider" is only 5 notes (7 semitones). Moreover, the highest absolute pitch change between two consecutive notes in "Happy Birthday" is again 12 semitones while this same quantity is only 4 semitones in "Itsy Bitsy Spider". On the other hand, one of the melodies in our melody list was the "United States National Anthem." It has notes covering 19 semitones, and the highest differential between two consecutive notes is 16 semitones, not an easy interval to be sung by untrained people. If we want to compare these three songs, we can speculate that the average performance of the humming of "Itsy Bitsy Spider" will be better than the performance of the humming of "Happy Birthday" or of the "United States National Anthem".

Apart from pitch range, difficulty can also be a function of "perceived closeness" of intervals in terms of fractions between pitch frequencies. For example, the interval of 7 semitones (corresponding to a perfect fifth and approximately a frequency ratio of 2:3) is a simple relationship to make and thus sing, whereas the interval of 6 semitones (corresponding to an augmented fourth or diminished fifth and approximately a frequency ratio of 5:7), although closer in terms of frequency, is usually more difficult to sing. Familiarity

The quality of reproducing a melody (singing or humming) also depends on the subject's familiarity with that specific melody. The less familiar the subject with the melody, the higher the uncertainty is expected. On the other hand, even though a melody may be very well known, it does not mean that it would be hummed perfectly, as evidenced by many performances at karaoke bars. Therefore, we prepared a list of well-known pieces

(Happy Birthday, Take Me to the Ball Game…) and nursery rhymes (Itsy Bitsy Spider, Twinkle Twinkle Little Star…) and asked our subjects to rate their familiarity to each melody. We hypothesize that the humming performance will be better when our subjects hum the melodies with which they are more familiar.

2.2 Musical Background

We can expect musically trained people to hum the melodies we ask with a high accuracy rate, while musically non-trained people are less likely to hum the melodies with the same accuracy. By musically trained, we mean that the subject has taken some professional music classes of any kind such as, diction, instrumental instruction and singing lessons. Whether or not the instruction is related to singing, even a brief period of instrumental training affects one's musical intuition. On the other hand, we also know that music intuition is a basic cognitive ability that some non-trained subjects may already possess [4, 5]. We in fact experienced very accurate humming from some non-trained subjects. Hence another goal of the data design was to sample subjects of varied skills.

3. EXPERIMENT METHODOLOGY

Given the aforementioned goals, the actual corpus creation was done according the following procedure.

3.1 Subject Information

Since our project does not target a specific kind of user population, we encouraged everyone to participate in our humming database collection experiment. However, in order to enable the performance of informed statistical analysis, we asked our subjects to fill out a form that asks information about their age, gender, and their linguistic and musical background. The personal identity of the subjects was not documented in the database. Most of the participants were university students who were paid a fee for their participation.

3.2 Melody List and Subjective Familiarity Rating

We prepared a list of 22 melodies that included folk songs, nursery rhymes and classical pieces. These melodies were categorized with respect to their musical structure, in total covering most of the possible note intervals in their original score (perfects, majors, minors). Table 3.2.1 shows the number of intervals we covered for each interval type in both ascending and descending format. The melody list only misses a Major 7th interval that corresponds to an 11 semitone transition.

The melodies with large intervals were assumed to be the more complex and difficult melodies (United States of America National Anthem, Take Me to the Ball Game, Happy Birthday) and the ones that cover small intervals, were assumed to be the less complex melodies (Twinkle Twinkle Little Star, Itsy Bitsy Spider, London Bridge…) The full melody list used for this corpus collection is available online at the project webpage [14].

These melodies were randomly listed on the same form where we asked our subjects to give their personal background information. The form template is also available online [14].

At this stage, we asked our subjects to rate their familiarity using a scale of 1 to 5 after hearing the songs played from the computer as MIDI files, with 5 being the highest level of familiarity. Subjects used "1" for rating melodies that they were unable to recognize from the MIDI files.

During the rating process, we asked our participants to disregard the lyrics and the name of the melody, as we believe that the tune itself is the most important feature.

Table 3.2.1 Intervals covered in the full melody list

Semi-tones	Interval Type	Frequency		
		Ascending	Descending	Total
0	Perfect Unison	199		199
1	minor 2nd	43	39	82
2	Major 2nd	185	48	233
3	minor 3rd	27	43	70
4	Major 3rd	15	33	48
5	Perfect 4th	22	14	36
6	Aug 4th/dim 5th	2	-	2
7	Perfect 5th	9	10	19
8	minor 6th	4	4	8
9	Major 6th	7	4	11
10	minor 7th	2	-	2
11	Major 7th	-	-	-
12	Perfect Octave	4	-	4

3.3 Humming Query

After the familiarity rating process, we picked ten melodies which are rated highest by the subject. We asked them to sing each of these melodies twice using "…da, da, da…" a stop consonant-vowel syllable that will be used in training note-levels in the front-end recognizer [1, 2].

3.4 Equipment and Recording Environment

A digital recorder is a convenient way of recording audio data. We used a Marantz PMD690, a digital recorder, which provides a convenient way to store the data to flash memory cards. The ready-to-process humming samples were transferred to a computer hard disk and the data was backed up on CDR's.

Martel, a tie-clip electret [17] condenser microphone is preferred here for its own built-in filters which lower the ambient noise level. The whole experiment was performed in a quiet office room environment to keep the data clean.

4. DATA

In total, we have acquired this far, a humming database from 100 participants, whose musical training varies from none to 25+ years of professional piano performing. These people were mostly college students whose ages are over 18 and hail from different countries. Each subject performed 20 humming pieces from the predefined melody list and, 6 humming piece of their own choice, giving us a total of up to over 2500 samples. This humming database will be made available online at our website in the near future and will be completely open source. The instructions for accessing the database will be posted in the website [14].

For convenient access and ease of use, the database needs to be well organized. We gave unique file names to each humming sample. These file names include a unique numerical ID for each subject, the id of the melody that was hummed and the personal information of the subject (gender, age, and whether s/he is musically trained or not). We also included an objective measure of uncertainty at the end (See Sections 5 and 6). Here is the file format:

txx(a/b)(+/-)pyyy(m/f)zz_uw

xx is an integer value that gives the track number of the song that is hummed in the melody list, (a/b) defines the first and second performances, (+/-) indicates if the subject is musically trained or not, yyy stands for the personal id number, (m/f) defines the gender of the subject and zz tells us the age of the subject. "w" is a float number that shows the average error per note transitions in semitones, which does not necessarily correspond to the quality of humming.

5. DATA ANALYSIS

One of the main goals of this study is to implement a way to quantify the variability and uncertainty that appears in the humming data. We needed to distinguish between good and bad humming, not only subjectively but also objectively from the viewpoint of automatic processing. If a person is musically trained and listens to the humming samples that we collected, s/he can easily make a subjective decision about the quality of the piece with respect to the (expected) original. However, this is not the case in which we are primarily interested.

For objective testing, we analyzed the data with a signal processing free software named PRAAT [15], and retrieved information about the pitch and the timing of the sound waves for each of the notes that the subject produced by humming. Each humming note is segmented manually and for each segmented part, we extracted the frequency values with the help of Praat's signal processing tools. Rather than the notes themselves, we analyzed the relative pitch difference between two consecutive notes [1, 6]. The pitch information we obtained, allowed us to quantify the pitch difference at the semitone level by using the theoretical distribution of semitones in an octave.

Relative Pitch Difference (RPD) is defined as Two Consecutive Notes in semitones

$$RPD = \frac{\log(f(k+1)) - \log(f(k))}{TDC} \qquad [6]$$

where

f : frequency of the hummed note

K : index of the hummed note

TDC : Theoretical Distribution Constant ($\log \sqrt[12]{2}$)

(The logarithmic distribution constant of semitones in an octave)

5.1 Performance Comparison in Key Points

During data collection, we observed various performance levels at particular parts of each melody. The most common parts that subjects make significant errors are the wide range note transitions, the first couple of notes of each melody where subjects make key calibration, and some specific intervals defined as inharmonic such as augmented/diminished intervals.

5.1.1 Wide range note transitions

Humming sample as a whole is most highly affected by large interval leaps in the original melody. While large interval transitions are difficult for non-trained subjects to sing accurately, the case is not so for musically trained people. A musically trained subject will not necessarily hum the melody perfectly. However, their performance at these transitions is expected to be more precise.

Figure 5.1.1.1 shows the distribution of the actual intervals sung by 20 randomly selected sample subjects at the point of the melody when the largest interval occurs in "Itsy Bitsy Spider." Each subject hummed the melody twice. This particular melody, shown in Figure 5.1.1.2, is one of the easiest melodies we have in our database, having a maximum note-to-note transition interval of "4" semitones (labeled with <*> in the score).

Ten of the subjects for this particular test group are musically trained so we analyzed a total of 20 (each participant hummed a melody twice) samples from musically trained subjects and 20 samples from untrained subjects.

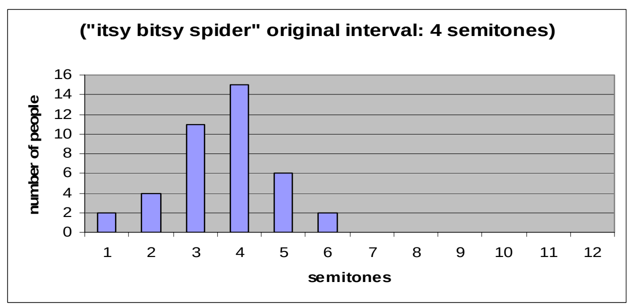

Figure 5.1.1.1: humming performance of the selected control group for song "itsy bitsy spider" (first two phrases) at the highest semitone level difference

As seen from the figure, the mode (highest frequency) of the performance for this interval is 4, the actual value. 15 out of 40 samples showed accurate singing of this interval and 10 of these accurate samples were performed by musically trained people. The average absolute error made by musically trained subjects in humming that interval transition was calculated to be 0.63 semitones while this value was 1.29 semitones for non-trained subjects. As expected, the largest interval singing by musically trained subjects was 104.8% better than the performance of non-trained subjects.

Figure 5.1.1.2 "Its Bitsy Spider"

To further investigate, this time we analyzed the humming samples performed by the same control group for the melody "Happy Birthday" which is shown in figure 5.1.1.3. The largest interval skip in "happy birthday" is 12 semitones (one octave labeled with "<*>"), which is a relatively difficult melodic leap for untrained subjects. "Happy Birthday" was one of the examples containing a large interval in our predefined melody list. Figure 5.1.1.4 shows the performance distribution of the previous control group for the humming of "Happy Birthday".

Figure 5.1.1.3 "Happy Birthday"

The mode for the singing of the largest interval is 12 the size of this largest interval in "happy birthday". 15 out of 40 samples were accurate in reproducing this particular interval and 11 of these were musically trained subjects. The average absolute error calculated for musically trained subjects is 0.845 semitones and, the average absolute error in non trained subject's performance is 1.963 semitones. These values show that, musically trained subjects performed 132.3% better than the non trained subjects in singing the largest interval in happy birthday.

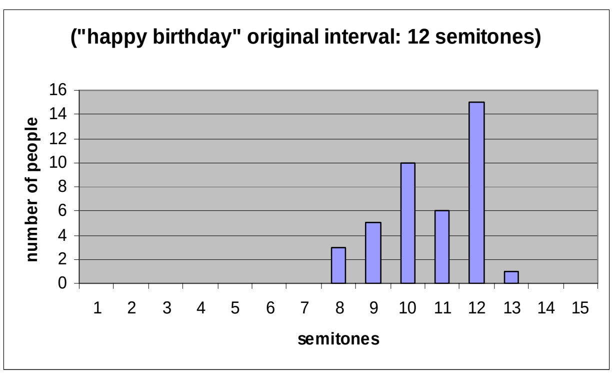

Figure 5.1.1.4: humming performance of the selected control group for "happy birthday" at the highest semitone level difference

A simple factor analysis of variance (ANOVA) for the songs, "itsy bitsy spider" and "happy birthday" indicates that the effect of musical training on the accurate singing of the largest intervals is significant. ["itsy bitsy spider"→ $F(1,39)=8.747$ $p=0.005$; "happy birthday"→ $F(1,39)=10.630$ $p=0.002$]

5.1.2 Key calibration

Subjects experienced key calibration problems at the start of each humming and they performed high levels of errors. This is because, for a particular time at the beginning, subjects try to decide in what key they want to continue their humming performance and this decision period results in unexpected levels of error in frequency contour. This orientation period is mostly obvious in non-trained subjects.

To investigate this, we analyzed the first interval of each humming sample, and compared the performance of subjects at the same interval this time occurred in the other parts of the same melody.

Consider the melody "London Bridge" shown in Figure 5.1.2.1. As seen from table 5.1.2.3, the analysis showed that, for "London Bridge", the error value calculated for the performance of the first interval of the melody (a major 2nd interval or 2 semitones labeled with "<*>" in the score figure) is 0.542 semitones and the error value for the performance of the same interval that occurred elsewhere (randomly selected from major 2nd intervals with labeled with "<%>") in the same melody is calculated to be 0.138 semitones. The performance improvement is 74.5%.

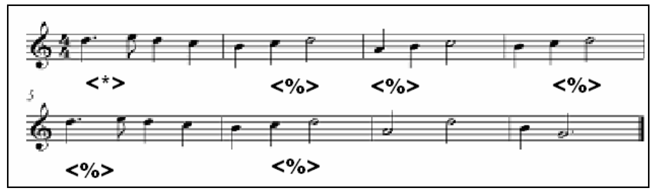

Figure 5.1.2.1 "London Bridge

We present another example, "Did you ever see a Lassie," shown in Figure 5.1.2.2. Because of the key calibration problem, subjects performed 52.5% better at the major 3rd intervals (labeled with "<%>") that are within the melody when compared to the one that is at the beginning which is labeled with "<*>".

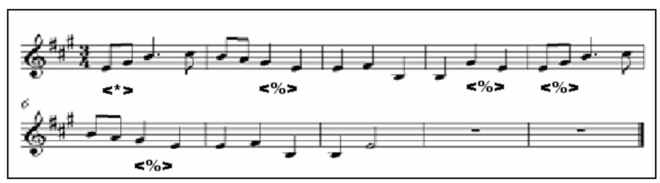

Figure 5.1.2.2 "Did You Ever See a Lassie"

A simple factor analysis of variance (ANOVA) for the songs, "London Bridge" and "Did you ever see a Lassie," indicates that the effect of key calibration at the beginning of the humming is significant. ["London Bridge"→ $F(1,47)=12.800$ $p=0.001$; "Did you ever see a Lassie"→ $F(1,39)=10.473$ $p=0.002$] The results are summarized in Table 5.1.2.3.

Table 5.1.2.3 Calculated errors at various locations vs interval kinds

	Interval, beginning of the song	Interval, elsewhere	Performance Improvement
"2 semitones: Major 2nd" London Bridge	0.542	0.138	74.5 %
"4 semitones: Major 3rd" Did you Ever See a Lassie	0.773	0.367	52.5 %

5.1.3 Special Intervals

We also had a chance to observe the affect of dissonance that is referred as the quality of sounds which seem “unstable” and have a need to resolve to “stable” sounds [16]. As discussed in section 2.1, it is hard to sing an augmented fourth interval (6 semitones) versus the wider perfect fifth interval (7 semitones). To investigate this, the performance of a control group of 20 people’s humming of a perfect fourth (5 semitones, frequency ratio approximately 3:4), an augmented fourth (6 semitones), and a perfect fifth interval (7 semitones) are analyzed and average error values are calculated for each interval. For statistics on the singing of the perfect fourth (labeled with “<%>”) and perfect fifth intervals (labeled with “<*>”) we analyzed the song “Twinkle Twinkle Little Star” shown in Figure 5.1.3.1, and for the augmented fourth interval (labeled with “<@>”) we analyzed the song “Maria,” shown in Figure 5.1.3.2. The average error values in semitones are summarized in figure 5.1.3.3.

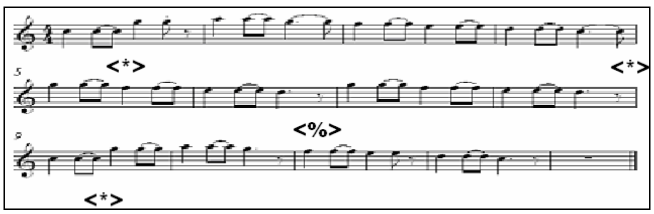

Figure 5.1.3.1 “Twinkle Twinkle Little Star”

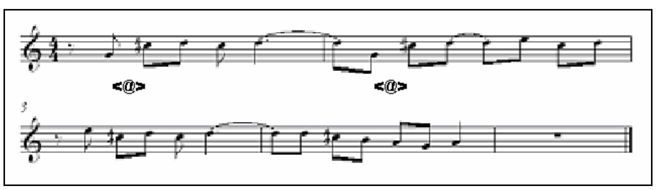

Figure 5.1.3.2 “Maria”

A simple factor analysis of variance (ANOVA) for the singing of the perfect fourth, augmented fourth and perfect fifth intervals indicate that the effect of dissonance over calculated error per interval is significant. [“Perfect and Augmented Intervals”→ F(1,47)=13.700 p=0.001]

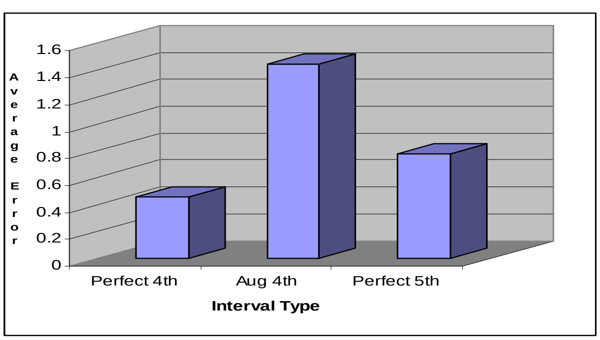

Figure 5.1.3.3: comparison of the average error calculated with the type of intervals

5.2 Performance Comparison across the Whole Piece

In the melody “Itsy Bitsy Spider” (see Figure 5.1.1.1), there are 24 notes and 23 transitions. For each interval, Figure 5.2.1 compares the interval sung an untrained subject to that in the original piece.

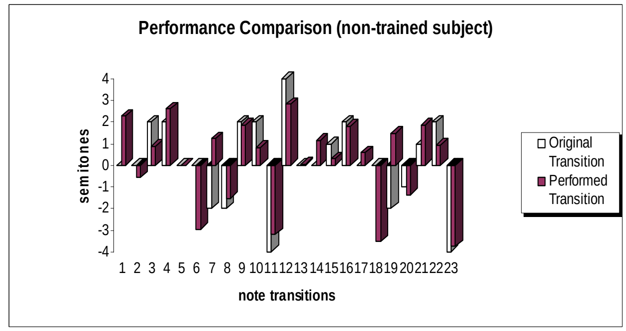

Figure 5.2.1: comparison of humming data to the base melody at each note transitions for non-trained subject for “Itsy Bitsy Spider.”

For each interval transition, we calculated the error between the data and the original expected values in semitones. The sum of all these values will give us a quantity that serves as an indicator for the quality of this particular humming sample. In this case, this subject performed an error average of 1.16 semitones per each note transition interval.

Figure 5.2.2 shows the comparison of a musically trained subject’s humming in comparison to the original melody. The analysis showed that, the average error in this musically trained subject’s humming is 0.28 semitones per transition, expectedly lower than the error that we calculated in the non-trained subject’s humming.

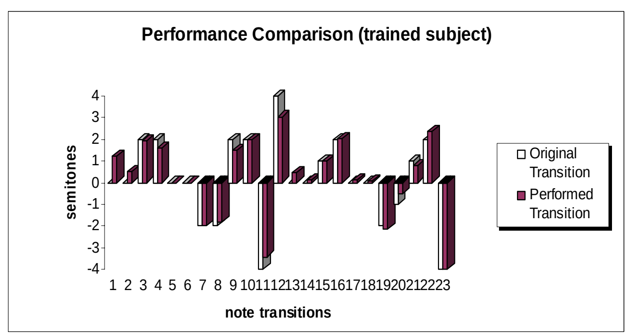

Figure 5.2.2: comparison of humming data to the base melody at each note transitions for non-trained subject for “Itsy Bitsy Spider.”

6. RESULTS AND DISCUSSION

Assuming that the final average error value per transition gives information about the accuracy of the humming, we analyzed and compared the error values of the humming performances of the same control group that we discussed before. For the melodies “Itsy Bitsy Spider” and “Happy Birthday”, the results are as follows.

From Table 6.1, one can easily see that, the uncertainty in the musically trained subject's humming is smaller than the uncertainty in the non-trained subject's humming of a particular song.

Table 6.1 Average Error values in Semitones in trained and non-trained subject's humming data for the melodies "Itsy Bitsy Spider" and "Happy Airthday"

	Itsy Bitsy Spider	Happy Birthday
trained	0.43	0.47
non-trained	0.63	0.70
All subjects	0.53	0.58

The average error value in the humming of the musically trained subjects in our control group is 0.43 semitones per transition for the melody "itsy bitsy spider". The average error value for the non trained subjects is 0.63 semitones per transition.

"Happy Birthday", previously claimed to be a more difficult melody to hum because of its musical structure, has the expected results as well. The average error value for trained subjects is calculated to be 0.47 semitones per note transition, larger than the value of the same subjects performed while humming "itsy bitsy spider" and the average error that is calculated for the non trained subjects is 0.70, which was also larger than the error value that same non-trained subjects performed during the humming of "itsy bitsy spider".

We conclude that one can expect larger error values in the humming performance of musically untrained subjects, when compared to musically trained subjects, as explained in section 2.3. The ANOVA analysis shows that the effect of musical background is also significant for humming performance. ["itsy bitsy spider" → $F(1,39)=12.062$, $p=0.001$; "happy birthday" → $F(1,39)=8.646$, $p=0.006$]. In addition, we also need to expect more uncertainty when the hummed melody contains intervals that are hard to sing as previously discussed and explained in section 2.1. The ANOVA analysis of humming performance of "itsy bitsy spider" and "happy birthday" showed that the effect of musical structure is also significant. [$F(1,79)=5.91$, $p=0.017$]

Moreover, all these average error values are calculated to be lower than the error values that are calculated at the largest interval transitions that we discussed in section 5.1. It also signifies that, most of the error values in the whole piece are dominated by the large interval transitions where subjects make the most pitch transition errors. This implies that, non-linear weight functions for high level versus low level note transitions should be implemented by the Query by Humming System at the back-end part where search engine performs the query.

7. FUTURE WORK AND CONCLUSION

In this paper, we discussed our corpus creation for designing user-centric front-ends for Query by Humming Systems. We first created a list that included the melodies to be hummed by the subjects. This list was created based on specific underlying goals. We included some melodies that are deemed difficult to hum as well as some familiar and less-complex nursery rhymes. The experimenter decided what songs a subject was going to hum with the help of the musical background of the subject and the familiarity ratings that the subject assigned at the beginning of the experiment. After collecting data for this specific melody list, the subjects were asked to hum some self-selected melodies not necessarily in the original list. The data was organized by subject details and quality measures and will be made available to the research community. We performed preliminary analysis of the data and tried to implement a way to quantify the uncertainty in the humming performance of our subjects, with the help of signal processing tools and knowledge of the physical challenges in humming large intervals. We believe that this procedure increases the validity of the data in our database.

Ongoing and future work includes integrating this organized and analyzed data into our Query by Humming music retrieval System. The front end recognizer will use this data for its training [1]; we can decide what data to include in the training with respect to quantified uncertainty. Moreover, we can also test our query engine using this data, so that we can test the performance of our whole system against data that have variable degrees of uncertainty.

8. ACKOWLEDGEMENTS

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.

9. REFERENCES

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

[2] H.-H. Shih, S. S. Narayanan, and C.-C. J. Kuo, "Multidimensional Humming Transcription Using Hidden Markov Models for Query by Humming Systems" IEEE Transactions on Speech and Audio Processing, Submitted, 2003

[3] Desain, Honing, van Thienen and Windsor, "Computational Modeling of Music Cognition: Problem or Solution," Music Perception vol. 16, 1998

[4] Jeanne Bamberger, "Turning Music Theory on its Ear," International Journal of Computers for Mathematical Learning vol. 1 No.1 1996

[5] L. Taelte and R. Cutietta, In R. Colwell and C. Richardson (eds), "Learning Theories Unique to Music" Chap17: Learning theories as roots of current musical practice and research. NY: Oxford University Press, pp.286-298, 2002.

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

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

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

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

[10] P.Y Rolland, G Raskins, and J.G Ganascia, "Music content-based retrieval: an overview of melodiscoc approach and systems," in Proceedings of ACM Multimedia 99, November 1999, pp. 81-84

[11] H.-H. Shih, T.Zhang, and C.-C. Kuo, "Real-time retrieval of song from music database with query-by-humming," in ISMIP, 1999, pp. 251-57.

[12] B. Chen and J.-S. Roger Jang, "Query by Singing" in 11th IPPR Conference on Computer Vision, Graphics and Image Processing, Taiwan, 1998, pp.529 536.

[13] Lie Lu, Hong You, and Hong-Jiang Zhang, "A new approach to query by humming in music retrieval," in 2001 IEEE International Conference on Multimedia and Expo, 2001.

[14] "USC Query by Humming project homepage," URL://sail.usc.edu/music/

[15] "Praat: Doing Phonetics by Computer" URL://www.praat.org/

[16] "Wikipedia"

URL://en.wikipedia.org/wiki/Music

[17] "Martel Electronics" URL://www.martelelectronics.com