

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*,

Integrated Media Systems Center

USC Viterbi School of Engineering, University of Southern California, CA, 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 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 are being 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 human humming varies depending on the musical structure of the melodies and the musical background of the subjects. 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,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MIR'03, November 7, 2003, Berkeley, California, USA.

Copyright 2003 ACM 1-58113-778-8/03/00011...\$5.00.

a natural human activity, for querying data is one of ways for facilitating such interactions.

Mode of interaction in music databases 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 then needs to be developed in order to search for the converted symbols in the database. The query engine should be precise and robust to inter-user variability and uncertainty in query formulation.

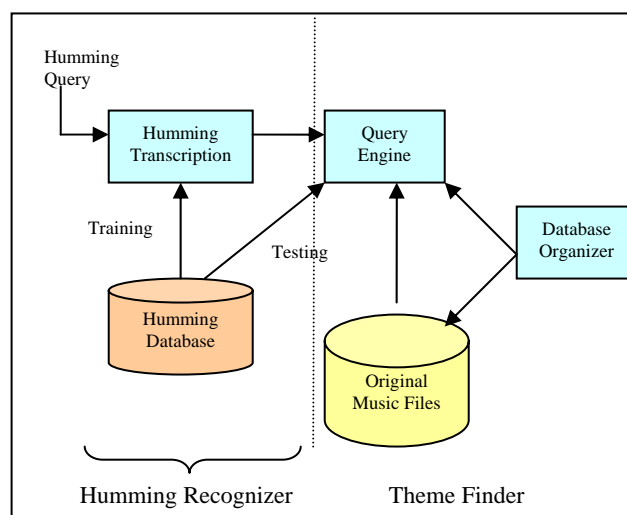


Figure 1.1: Flowchart of our Query by Humming System.

Ghias et al. [6] have been credited for being the first to propose the idea of Query by humming in 1995. They used coarse melodic contours to represent melodic information. Autocorrelation was used to track pitch and convert humming into coarse melodic contours. Coarse melodic contour has been widely used and discussed in several query by humming systems that followed. McNab et al. [7, 8] improved this framework by introducing the concept of duration contour for rhythm representation. Blackburn et al. [9], Roland et al. [10] and Shih et al. [11] extended McNab's system by using tree based database searching. Jang et al. [12] used the semitone (half step) as a distance-measure and removed repeated notes in their melodic contour. Lu et al. [13] proposed a

new melodic string representation which contained pitch contour, pitch interval and duration as a triplet. Haus et al. [15] implemented rules for correcting contour transcription errors caused by uncertainty in the humming. Counter to the previous note segmentation algorithms, Zhu et al. [16] used dynamic time warping indices to compare audio directly with the database. Unal et al. [17] used a statistical approach to the problem of retrieval under the effect of uncertainty. In her fault-tolerance studies, Doraisamy et al. [18] used McNab’s findings to classify different types of humming errors that a person can make. She compared extracted n-gram windows from the original melody to the ones that are performed in the humming input and studied their correlation. All these efforts have made significant contributions to the topic of Query by Humming.

1.1 The Role of this Study in QBH Systems

Our proposed statistical approach to humming recognition aims at providing note level decoding using statistical models (we favor hidden Markov models or HMMs) of audio features representing melodies. Since the approach is data-driven, it promises robustness in terms of handling human 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 comprehensive 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 by a cross-section of people with various musical backgrounds in order to make statistical assessments 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.

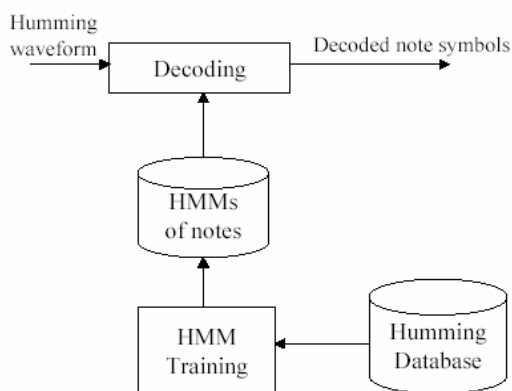


Figure 1.1.2 The role of humming database in statistical humming recognition approach (an HMM based approach is illustrated).

As seen from the Figure 1.1.2, the collected data will be used to train the Hidden Markov Models (HMMs) that we use to decode the humming waveform. From the uncertainty analysis we perform, we can determine the appropriate data to be used in the

training set so that inaccurate data will not adversely affect the decoding accuracy. On the other hand, the entire data set can also be used to test and optimize the accuracy of the retrieval algorithms.

Building a statistical system that performs pitch and time information based retrieval from a humming sample has been shown to be feasible [1]. However, since the quality of the input is greatly dependent on the user, and includes high rates of variability and uncertainty, a key challenge is achieving robust performance under such conditions. In Section 2, we will discuss our hypothesis on 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. 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 hypothesize that the main factors contributing to humming 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 them through humming. Some melodies possess a very complex musical structure such as difficult note transitions and complex rhythmic structures that make them difficult to hum. When we create a database, a criterion is to populate it with samples reflecting a range of musical structure complexity. In this regard, the note succession as notated in the score of the melodies was the information that we used to determine the musical complexity.

Pitch range is an important factor affecting the difficulty with respect to humming 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 characteristics according to those measures. The range of notes in “Happy Birthday” spans 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 the 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.” Its note collection spans 19 semitones, and the highest differential between two consecutive notes is 16 semitones, not an easy interval to be accurately sung by non-professionals. 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 an 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. Hence it is important to incorporate information about the type of intervals.

2.2 Familiarity

The quality of the reproduced melody (singing or humming) also depends on the subject’s familiarity with the specific melody. The less familiar the subject is with the melody, the higher the expected uncertainty. 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 with each melody. In her studies about relevance assessment Uitdenbogerd believed that it was a very difficult task for users to compare and process unknown pieces of music [14]. This result also supports our hypothesis that the humming performance will be better when our subjects hum the melodies with which they are more familiar.

2.3 Musical Background

We can expect musically trained subjects to hum the melodies we ask with a higher accuracy, while musically non-trained subjects are less likely to hum the melodies with the same degrees of accuracy. By musically trained, we mean that the subject has had some formal music training, for example through classes such as diction, instrumental instruction or 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 in our database. Hence another goal of the data acquisition was to sample subjects of varied skills.

3. EXPERIMENT METHODOLOGY

Given the aforementioned goals, the actual corpus creation was done according to 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 requested for 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 compensated for their participation per institutional review board approval for human subjects.

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 set only lacks a major 7th interval which corresponds to an 11 semitone transition.

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	Aug4th/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

The melodies containing large interval leaps 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 contains smaller 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 is available online at the project webpage [19]. 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 [19].

At this stage, we asked our subjects to rate their familiarity with each melody using a scale of 1 to 5 after hearing the melodies 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 details regarding the lyrics and the name of the melody, as we believe that the tune itself is the most important feature.

3.3 Humming Query

After the familiarity rating process, we picked ten melodies which were 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 CDRs.

Martel, a tie-clip electret [22] condenser microphone is preferred for its built-in filters which lower the ambient noise level. The whole entire experiment was performed in a quiet office room environment to keep the data as clean as possible.

4. DATA

In total, we have acquired thus far, a humming database from 100 participants, whose musical training varied from none to 25+ years of professional piano performance. 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 pieces of their own choice, giving us a total of over 2500 samples. This humming database is being made available online at our website and will be completely open source. The instructions for accessing the database will be posted in the website [19].

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). The file format is as shown:

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

xx is an integer value that gives the track number of the song that is hummed in the melody list, (a/b) specifies whether the sample is the first or second performances, (+/-) indicates if the subject is musically trained, yyy stands for the personal id number, (m/f) gives the gender of the subject and zz tells us the person's age of. "ww" 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 need 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 with which we are primarily interested.

For objective testing, we analyzed the data with a signal processing freeware software named PRAAT [20], 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 absolute values of the notes themselves, we analyzed the relative pitch difference (RPD) 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.

In this study, we defined humming error as numerical semitone level difference between the hummed note transition and the target note transition. For this we use the following formula.

$$RPD = \frac{\log(f(k+1)) - \log(f(k))}{\log \sqrt[12]{2}} \quad [6]$$

The logarithmic difference of the pitch values of two humming notes, divided by the theoretical distribution constant gives the RPD. This calculated value can be compared to the baseline transition to see how well the performance for that specific interval is. The absolute distance between the RPD and the target semitone transition is the measure of the humming error that will be used in our analysis.

5.1 Performance Comparison in Key Points

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

5.1.1 Wide range note transitions

The 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 same is not true for musically trained people. A musically trained subject will not necessarily hum the melody perfectly. However, their performance at these challenging transitions can be expected to be more precise.

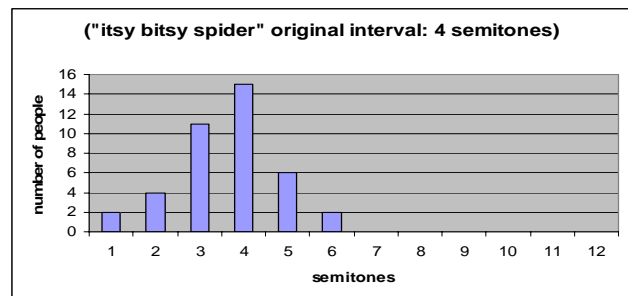


Figure 5.1.1.1: Humming performance of the selected control group for the song "itsy bitsy spider" (first two phrases) at the highest semitone level difference.

Figure 5.1.1.1 shows the distribution of the actual intervals sung by 20 randomly selected subjects at the point of the largest interval leap 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 in our database, having a maximum note-to-note transition interval of “4” semitones (marked by <*> in the score).



Figure 5.1.1.2 “Itsy Bitsy Spider” melody.

Ten of the subjects in 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.

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 sung by musically trained subjects was 104.8% better than the performance of non-trained subjects.



Figure 5.1.1.3 “Happy Birthday” melody.

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 is 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”.

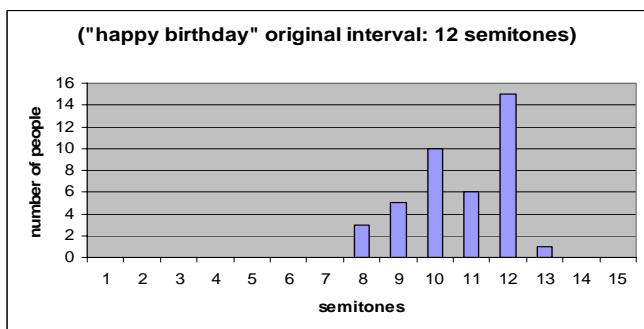


Figure 5.1.1.4: humming performance of the selected control group for “happy birthday” at the highest semitone level difference

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 by musically trained subjects. The average absolute error calculated for musically trained subjects is 0.845 semitones and, the average absolute error in non trained subjects’ 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.”

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 with higher levels of errors at the beginning of the melody. This may be because, for a particular time at the beginning, subjects try to adjust their humming to the key they have in their mind, and this transition period results in unexpected levels of error in the fundamental frequency contour. This orientation period is mostly obvious in non-trained subjects.

To investigate this hypothesis, we analyzed the first interval of each humming sample, and compared the performance of subjects at the same interval in later 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) is 0.542 semitones and the error value for the performance of the same interval that occurred later (randomly selected from major 2nd intervals labeled with “<%>”) in the same melody is calculated to be 0.138 semitones. The performance improvement is a remarkable 74.5%.



Figure 5.1.2.1 “London Bridge” melody.

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 minor 3rd intervals (labeled with “<%>”) that are within the melody as compared to the one at the beginning (labeled with “<*>”).



Figure 5.1.2.2 “Did You Ever See a Lassie” melody.

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" <i>London Bridge</i>	0.542	0.138	74.5 %
"4 semitones: Major 3rd" <i>Did you Ever See a Lassie</i>	0.773	0.367	52.5 %

5.1.3 Special Intervals

We also had a chance to observe the effect of dissonance which refers to the perceptual quality of sounds which seem "unstable" and have a need to resolve to "stable" sounds [21]. As discussed in section 2.1, it is hypothetically more difficult to sing an augmented fourth interval (6 semitones) versus the wider perfect fifth interval (7 semitones).



Figure 5.1.3.1 "Twinkle Twinkle Little Star"

To investigate this, the performance of a perfect fourth (5 semitones, frequency ratio approximately 3:4), an augmented fourth (6 semitones) and a perfect fifth interval (7 semitones) using humming samples from a control group of 20 subjects, were analyzed and average error values were 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," from "West Side Story" shown in Figure 5.1.3.2.



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 indicated that the effect of dissonance on the calculated error per interval is significant. ["Perfect 4th&5th Intervals and Augmented 4th intervals" → $F(1,47)=13.700$ $p=0.001$]

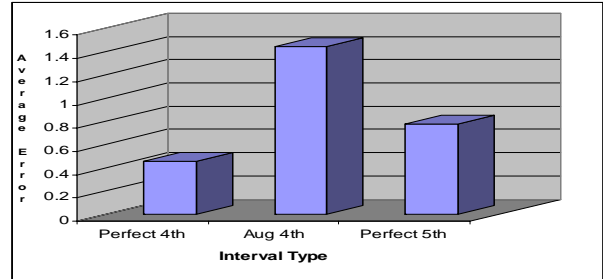


Figure 5.1.3.3: comparison of the average error calculated with the type of intervals

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 by an untrained subject with that occurring in the original piece.

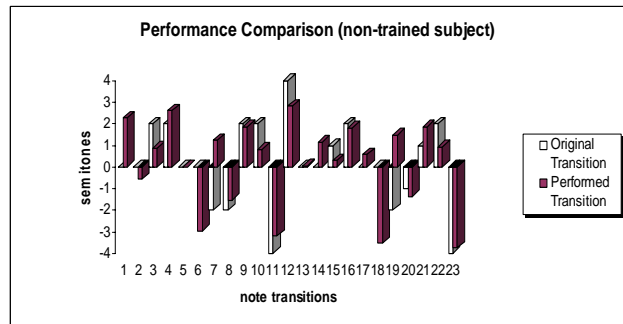


Figure 5.2.1: Comparison of humming data to the base melody at each note transition for non-trained subjects (shown for "Itsy Bitsy Spider.")

For each interval transition, we calculated the error between the observed data and the original expected values in semitones. The sum of all these values gives us a quantity that serves as an indicator for the quality of this particular humming sample. In the case shown in Figure 5.2.1, this subject performed with an average error of 1.16 semitones per interval.

Figure 5.2.2 compares a musically trained subject's humming with 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.

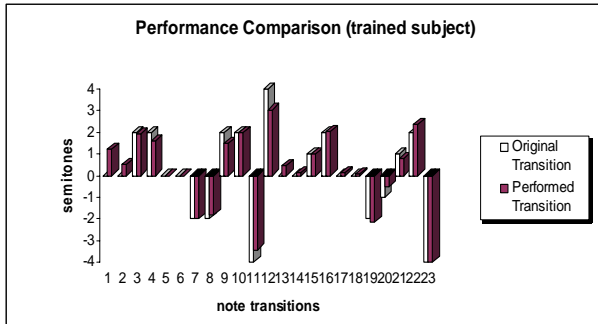


Figure 5.2.2: comparison of humming data to the base melody at each note transitions for non-trained subject for “Itsy Bitsy Spider.”

5.3 Retrieval Analysis

In our QBH experiments, the humming database serves two purposes: that of training the note models in the front-end recognizer and that of testing the QBH system. For the front end humming recognizer, statistical speech recognition techniques are used in order to automatically segment hummed notes one from another. To do this robustly and accurately, a large data set is necessary.

Since the data samples have great variability, it is also possible to test the performance of the retrieval engine against various levels of uncertainty in the query sample. In order to compensate for the negative effects of uncertainty in the input, we developed our retrieval engine algorithms according to the statistical findings we gathered from the data analysis.

The retrieval engine aims to define statistical prediction intervals for the performance of each possible note transition, so that an incoming sample can be checked to see if it belongs to expected limits for specific intervals [17].

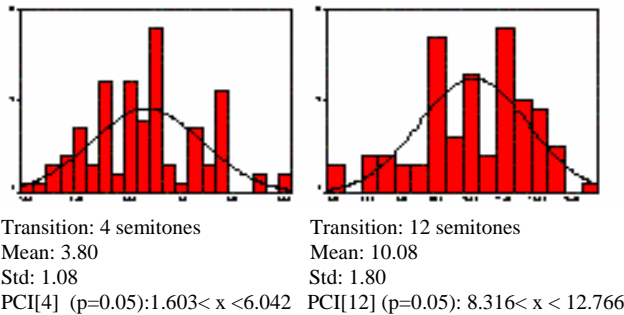


Figure 5.3.1 Histogram of training data set, normal distribution curve and Prediction Confidence Intervals (PCI) for 4 and 12 semitone pitch transitions.

In our studies, we calculated the required statistical prediction interval limits by using the collected samples as the training set and used these limits in our similarity measurement tests. Figure 5.3.1 shows the histogram of the performance of a randomly selected group of 24 subjects humming the 4 semitone transition 100 times. The graph is tested to be normally distributed (KS test $p < 0.05$) around a mean of 3.80 and with the calculated prediction interval limits of 1.603 and 6.042. The second graph shows the

histogram of the performance of a test sample of 38 subjects humming a 12 semitone transition 100 times. This time the statistical prediction interval limits are 8.316 and 12.766. All statistical prediction limits are calculated in the same manner to produce the following [17].

Table 5.3.1 Calculated Prediction Intervals

Semitones	#of Samples	Lower Confidence Limit	Upper Confidence Limit
1	100	-0.91920	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 5.3.1 shows the prediction intervals for each semitone level transition in our database. By using this table, one can statistically determine which semitone transition a sample may belong to and the certainty of the prediction. For example, a 6.155 pitch difference in semitones between two humming notes may belong to 5,6,7,8 or 9 semitone transitions with a statistical confidence level of $p < 0.05$.

5.3.1 Retrieval Experiment Results

Constructed limits are used as guidelines in finger print search algorithms that are explained in Unal et al. [17]. Finger prints are used to extract characteristic information from the input humming. Rather than considering the entire humming input, this characteristic information is used to search the database. The proposed search method is tested with 250 humming samples within an original music database of 200 pieces that includes our original melody list and pieces from the Beatles. 94% retrieval accuracy is observed within a test sample of trained subjects, while 72% retrieval accuracy is achieved by a test sample of non-musically trained subjects. The decrease in performance is an expected result as mentioned in Section 5.2; the increased uncertainty in non-trained subject’s humming is statistically significant.

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 previously discussed control group. For the melodies “Itsy Bitsy Spider” and “Happy Birthday”, the results are as follows:

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 Birthday.”

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

From Table 6.1, one can see that the uncertainty in the musically trained subject’s humming is less than that in the non-trained subject’s humming of the same song.

The average error value in the humming of the musically trained subjects in our control group is 0.43 semitones per transition in the melody “Itsy Bitsy Spider”. The average error value for the non trained subjects is 0.63 semitones per transition.

“Happy Birthday”, previously hypothesized to be a more difficult melody to hum because of its intervals and range produces the expected results as well. The average error for trained subjects is calculated to be 0.47 semitones per note transition, which is 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 for the same subjects humming “Itsy Bitsy Spider”.

We conclude that one can expect larger error values in the humming of musically untrained subjects, compared to that of musically trained subjects, as explained in Section 2.3. The ANOVA analysis shows that the effect of musical background is also significant for humming quality. [“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 expect more uncertainty when the hummed melody contains intervals that are difficult 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, these average error values are determined to be lower than the error values calculated at the largest interval transitions as discussed in Section 5.1. This result shows 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, a non-linear weight function for high level versus low level note transitions should be implemented by the Query by Humming System at the back-end where the search engine performs the query.

7. FUTURE WORK AND CONCLUSIONS

In this paper, we discussed our corpus for designing user-centric front-ends for Query by Humming Systems. We first created a list of melodies to be hummed by the subjects 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 which songs a subject should hum based on an initial assessment of the musical background of the subject and the familiarity ratings that the

subject assigned to each melody at the beginning of the experiment. After collecting data for the melody list, the subjects were asked to hum some self-selected melodies not necessarily in the original list. The data was organized by subject information and objective quality measures, and is being made available to the research community. We performed some preliminary analysis of the data and implemented 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 annotated data into our Query by Humming music retrieval System. The front end recognizer will use this data for its training [1]; we can decide which data to include in the training with respect to quantified uncertainty. Moreover, we can also test our query engine using this data; we can assess the performance robustness of our whole system against data that have variable degrees of uncertainty. Preliminary testing shows that, the designed retrieval algorithms that are trained by the statistical findings of this study achieved 83 percent accuracy when tested on a database of 200 melodies. We would plan to evaluate the performance of our system using a larger database and to build up a web-based system that will be publicly accessible.

8. 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, National Science Foundation Information Technology Research Grant NSF ITR 53-4533-2720, and 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 or 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 Proceedings of 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” in Proceedings of IEEE International conference on Acoustics Speech and Signal Processing, 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 audio

- database,” in Proceedings of ACM Multimedia Conference’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 Proceedings of Multimedia Tools and Applications, 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 melodic 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 Proceedings of ISMIP, 1999, pp. 251-57.
- [12] B. Chen and J.-S. Roger Jang, “Query by Singing” in Proceedings of 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 Proceedings of IEEE International Conference on Multimedia and Expo, 2001.
- [14] Alexandra L. Uitdenbogerd and YawWah Yap, “Was Parsons right? An experiment in usability of music representations for melody-based music retrieval,” In Proceedings of International Conference in Music Information Retrieval (ISMIR), October 2003.
- [15] G. Haus and E. Pollstri, “An Audio Front End for Query-by-Humming Systems,” in Proceedings of International Conference in Music Information Retrieval (ISMIR), 2001.
- [16] Y. Zhu and D. Sasha, “Warping Indexes with Envelope Transforms for Query-by-Humming Systems,” in Proceedings of ACM SIGMOD, June 2003.
- [17] E. Unal, S.S. Narayanan and E. Chew, “A statistical Approach to Retrieval under User-dependent Uncertainty in Query-by-Humming Systems,” in Proceedings of ACM MIR04, October 2004.
- [18] S. Doraisamy and S. Ruger, “A Comparative and Fault-tolerance Study of the Use of N-grams with Polyphonic Music,” in Proceedings of International Conference in Music information Retrieval (ISMIR), October 2002.
- [19] “USC Query by Humming project homepage,”
URL://sail.usc.edu/music/
- [20] “Praat: Doing Phonetics by Computer”
URL://www.praat.org/
- [21] “Wikipedia”
URL://en.wikipedia.org/wiki/Music
- [22] “Martel Electronics”
URL://www.martelelectronics.com