Multimedia Systems (2005) 00: 1–9
DOI 10.1007/s00530-005-0176-5

REGULAR PAPER

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

Creating data resources for designing usercentric frontends for query-by-humming systems

Published online: 10 May 2005

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 intersubject 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 to 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.

Keywords Humming database · Uncertainty quantification · Query by humming · Statistical methods

E. Unal (✉) · S. S. Narayanan · H.-H. Shih
Speech Analysis and Interpretation Laboratory, USC Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
E-mail: unal@usc.edu, shri@sipi.usc.edu, maverick@aspirex.com

E. Chew · C.-C. Jay Kuo
Integrated Media Systems Center, USC Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA
E-mail: echew@usc.edu, cckuo@sipi.usc.edu

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

Interaction with music databases requires that audio information retrieval techniques be developed for mapping the human humming waveforms to numeric 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 interuser variability and uncertainty in query formulation.

Ghias et al. [6] have been credited for being the first to propose the idea of QBH in 1995. They used course contours to represent melodic information. Autocorrelation was used to track pitches and convert humming into coarse melodic contours. Coarse melodic contour has been widely used and discussed in several QBH systems that followed. McNab et al. [7, 8] improved this framework by introducing the concept of a 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 that consisted of the 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 their fault-tolerance studies, Doraisamy et al. [18] used McNab's findings to classify different types of humming errors that a person can make. They compared extracted n-gram windows from the original melody to those performed in

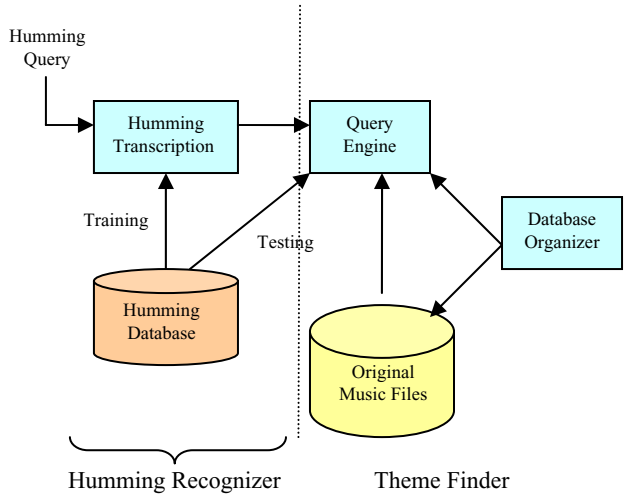


Fig. 1 Flowchart of our query-by-humming system

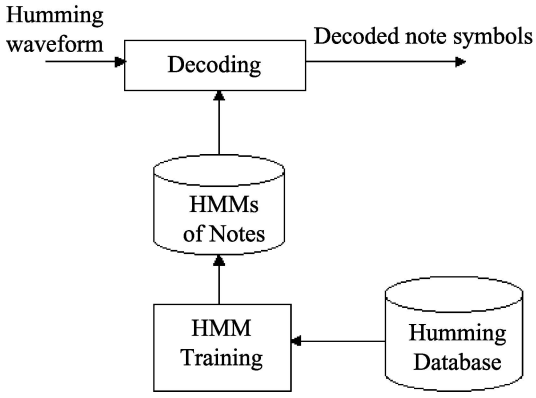


Fig. 2 Role of humming database in statistical humming recognition approach (an HMM-based approach is illustrated)

the humming input and studied their correlation. All these efforts have made significant contributions to the topic of QBH.

1.1 The role of this study in QBH systems

Our proposed statistical approach to humming recognition (Fig. 1) 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 frontend of the QBH 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 intersubject 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.

As seen from Fig. 2, the collected data will be used to train the 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 depends largely 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 features of the melodies being hummed, the subject's familiarity with 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, one 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 we used to determine the musical complexity.

Pitch range is an important factor affecting the difficulty of humming 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 these 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 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 for nonprofessionals to sing accurately. 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 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 Out to the Ball Game," etc.) and nursery rhymes ("Itsy Bitsy Spider," "Twinkle Twinkle Little Star," etc.) 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 nontrained 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 musical intuition is a basic cognitive ability that some nontrained subjects may already possess [4, 5]. We, in fact, experienced very accurate humming from some nontrained 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 information about their age, gender, and 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 by their musical structure, in total covering most of the possible note intervals in their original score (perfects, majors, minors). Table 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 seventh interval, which corresponds to an 11-semitone transition.

The melodies containing large interval leaps were assumed to be the more complex and difficult melodies ("United States of America National Anthem," "Take Me Out to the Ball Game," "Happy Birthday"), and those containing smaller intervals were assumed to be the less complex melodies ("Twinkle Twinkle Little Star," "Itsy Bitsy Spider," "London Bridge. . ."). The full melody list

Table 1 Intervals covered in the full melody list

		Frequency		
Semitones	Interval type	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

used for this corpus is available online at the project Web page (`http://sail.usc.edu/music`). 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 (`http://sail.usc.edu/music`).

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 that were rated highest by the subjects. 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 frontend 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 were backed up on CDRs.

Martel, a tie-clip electret condenser microphone, is preferred for its built-in filters that lower the ambient noise level.[1] The 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 over 18 years of age and 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 Web site and will be completely open source. The instructions for accessing the database will be posted at the Web site (`http://sail.usc.edu/music`).

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 level of musical training). We also included an objective measure of uncertainty at the end (Sections 5 and 6). The file format is as shown:

$$txx(a/b)(+/-)pyyy(m/f)zz\_ww\,,$$

where xx is an integer value that gives the track number of the song from the melody list being hummed, (a/b) specifies whether the sample is the first or second performance, $(+/-)$ 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. ww is a float number that shows the average error per note transition 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 appear 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 concerned.

For objective testing, we analyze the data with a signal processing freeware software named PRAAT[2] and retrieve information about the pitch and timing of the sound waves

[1] `http://www.martelelectronics.com`

[2] Praat: doing phonetics by computer, `http://www.praat.org`

for each of the notes that the subject produced by humming. Each humming note is segmented manually, and for each segmented part we extract the frequency values with the help of Praat's signal processing tools. Rather than the absolute values of the notes themselves, we analyze the relative pitch difference (RPD) between two consecutive notes [1, 6]. The pitch information we obtained allows us to quantify the pitch difference at the semitone level by using the theoretical distribution of semitones in an octave.

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

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

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 made 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 affected by large interval leaps in the original melody. While large interval transitions are difficult for nontrained 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 3 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 Fig. 4, 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).

Ten of the subjects in this particular test group are musically trained, so we analyze 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.

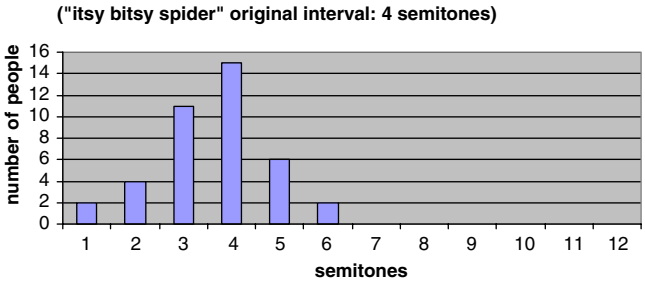


Fig. 3 Humming performance of the selected control group for the song "Itsy Bitsy Spider" (first two phrases) at the highest semitone level difference

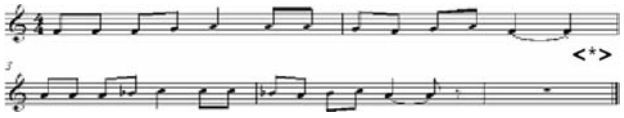


Fig. 4 "Its Bitsy Spider" melody

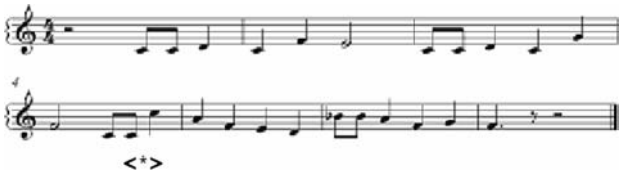


Fig. 5 "Happy Birthday" melody

Fifteen out of 40 samples show accurate singing of this interval, and 10 of these accurate samples are performed by musically trained people. The average absolute error made by musically trained subjects in humming that interval transition is calculated to be 0.63 semitones, while this value is 1.29 semitones for nontrained subjects. As expected, the largest interval sung by musically trained subjects is 104.8% better than the performance of nontrained subjects.

To further investigate, we then analyze the humming samples performed by the same control group for the melody "Happy Birthday," which is shown in Fig. 5. The largest interval skip in "Happy Birthday" is 12 semitones (one octave is labeled "⟨∗⟩"), which is a relatively difficult melodic leap for untrained subjects. "Happy Birthday" is one of the examples containing a large interval in our predefined melody list. Figure 6 shows the performance distribution of the previous control group for the humming of "Happy Birthday."

The mode for the singing of the largest interval is 12, the size of the largest interval in "Happy Birthday." Fifteen out of 40 samples are accurate in reproducing this particular interval, and 11 of these are by musically trained subjects. The average absolute error calculated for musically trained subjects is 0.845 semitones and, the average absolute error in nontrained subjects' performance is 1.963 semitones. These values show that musically trained subjects performed 132.3% better than the nontrained 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

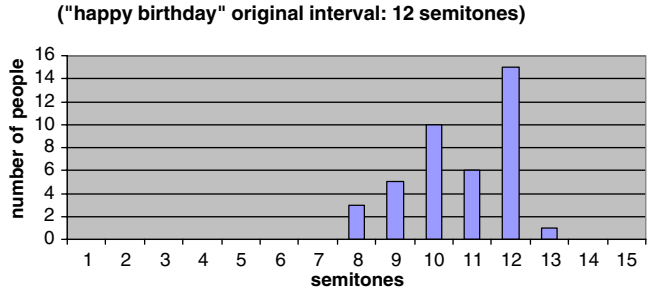


Fig. 6 Humming performance of the selected control group for "Happy Birthday" at the highest semitone level difference

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 error levels at the beginning of the melody. This may be because, for a certain 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 most obvious in nontrained subjects.

To investigate this hypothesis, we analyze the first interval of each humming sample and compare the performance of subjects at the same interval in later parts of the same melody.

Consider the melody "London Bridge" shown in Fig. 7. As seen from Table 2, the analysis shows that, for "London Bridge," the error value calculated for the performance of the first interval of the melody (a major second interval or 2 semitones labeled "⟨∗⟩" in the score) is 0.542 semitones, and the error value for the performance of the

Table 2 Calculated errors at various locations vs. interval types

	Interval, beginning of 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%

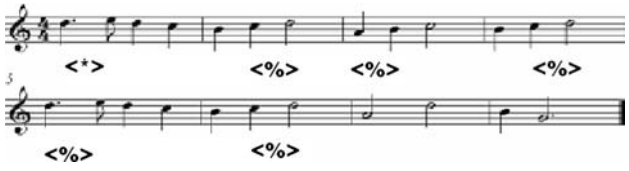


Fig. 7 "London Bridge" melody

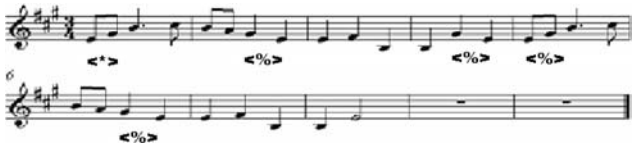


Fig. 8 "Did You Ever See a Lassie" melody

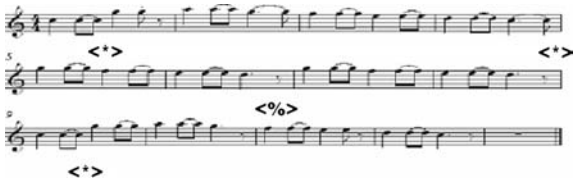


Fig. 9 "Twinkle Twinkle Little Star"

same interval that occurred later (randomly selected from major second intervals labeled "⟨%⟩") in the same melody is calculated to be 0.138 semitones. The performance improvement is a remarkable 74.5%.

We present another example, "Did You Ever See a Lassie," shown in Fig. 8. Because of the key calibration problem, subjects performed 52.5% better at the minor third intervals (labeled "⟨%⟩") that are in the melody as compared to the one at the beginning (labeled "⟨∗⟩").

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 2.

5.1.3 Special intervals

We also had a chance to observe the effect of dissonance, which refers to the perceptual quality of sounds that seem "unstable" and have a need to resolve to "stable" sounds.[3]. 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).

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 are analyzed, and average error values are calculated for each interval. For statistics on the singing of the perfect fourth (labeled "⟨%⟩") and perfect fifth intervals (labeled "⟨∗⟩"), we analyze the song "Twinkle Twinkle Little Star" (Fig. 9), and for the augmented fourth interval (labeled "⟨@⟩") we analyze the song "Maria" from "West Side Story" (Fig. 10).

A simple factor analysis of variance (ANOVA) for the singing of the perfect fourth, augmented fourth, and perfect fifth intervals indicates that the effect of dissonance on the calculated error per interval is significant. ["Perfect 4th and

[3] Wikipedia: `http://en.wikipedia.org/wiki/Music`

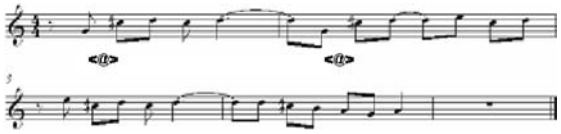


Fig. 10 "Maria"

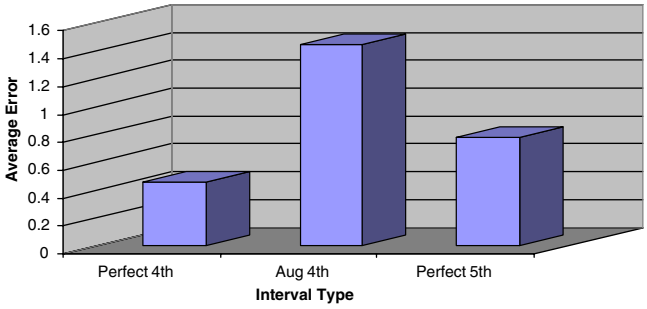


Fig. 11 Comparison of the average error calculated with the interval type

5th Intervals and Augmented 4th intervals"→ $F(1,47) = 13.700\ p = 0.001$.]

5.2 Performance comparison across the whole piece

In the melody "Itsy Bitsy Spider" (Fig. 3), there are 24 notes and 23 transitions. For each interval, Fig. 12 compares the interval sung by an untrained subject with that occurring in the original piece.

For each interval transition we calculate 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 of the quality of this particular humming sample. In the case shown in Fig. 12, this subject performs with an average error of 1.16 semitones per interval.

Figure 13 compares a musically trained subject's humming with the original melody. The analysis shows 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 nontrained subject's humming.

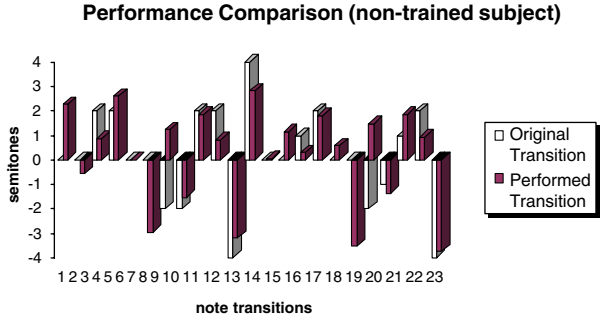


Fig. 12 Comparison of humming data with the base melody at each note transition for nontrained subjects (shown for "Itsy Bitsy Spider")

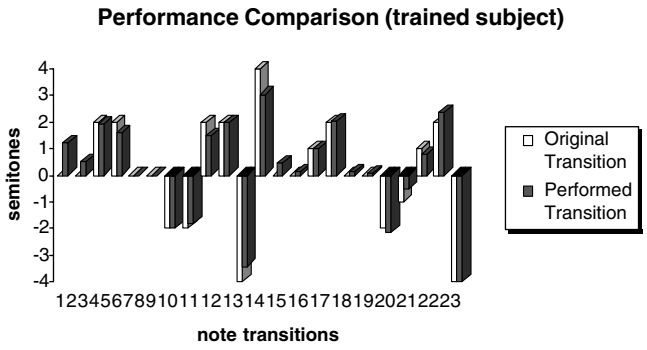


Fig. 13 Comparison of humming data with the base melody at each note transition for nontrained subject for "Itsy Bitsy Spider"

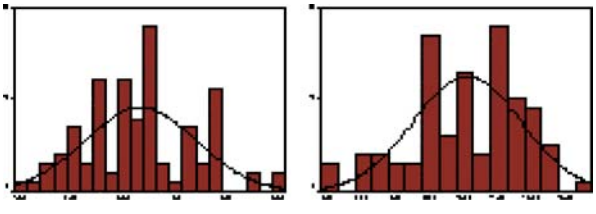
5.3 Retrieval analysis

In our QBH experiments, the humming database serves two purposes: that of training the note models in the frontend recognizer and that of testing the QBH system. For the frontend humming recognizer, statistical speech recognition techniques are used to automatically segment hummed notes from one 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 falls within expected limits for specific intervals [17].

In our studies, we calculate 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 14 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



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: 1.08
PCI[12] (p=0.05): 8.316< x < 12.766

Fig. 14 Histogram of training data set, normal distribution curve and prediction confidence intervals (PCIs) for 4- and 12-semitone pitch transitions

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

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 results documented as [17].

Table 3 shows the prediction intervals for each semitone level transition in our database. 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 fingerprint search algorithms explained in Unal et al. [17]. Fingerprints 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 in an original music database of 200 pieces that includes our original melody list and melodies from the Beatles' songs. 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 nontrained subjects' 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 analyze and compare 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 shown in Table 4.

Table 4 Average error values in semitones in trained and nontrained subjects' humming data for the melodies "Itsy Bitsy Spider" and "Happy Birthday"

	Itsy Bitsy Spider	Happy Birthday
Trained	0.43	0.47
Nontrained	0.63	0.70
All subjects	0.53	0.58

From Table 4 one can see that the uncertainty in the musically trained subjects' humming is less than that in the nontrained subjects' 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 nontrained 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 nontrained subjects is 0.70, which is 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 nontrained 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" $\rightarrow F(1, 39) = 12.062$, $p = 0.001$; "happy birthday" $\rightarrow F(1, 39) = 8.646$, $p = 0.006$.] 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" shows 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 nonlinear weight function for high-level versus low-level note transitions should be implemented by the QBH system at the backend where the search engine performs the query.

7 Future work and conclusions

In this paper, we discussed our corpus for designing user-centric frontends for QBH 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 were organized by subject information and objective quality measures and are 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 or unusual 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 QBH music retrieval system. The frontend recognizer will use these data for 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 these data and assess the performance robustness of our whole system against data that have varying degrees of uncertainty. Preliminary testing shows that the designed retrieval algorithms that are trained by the statistical findings of this study achieved 83% accuracy when tested on a database of 200 melodies. We 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.

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.

References

1. Shih, H.-H., Narayanan, S.S., Kuo, C.-C.J.: An HMM-based approach to humming transcription. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME2002) (2002)
2. Shih, H.-H., Narayanan, S.S., Kuo, C.-C.J.: Multidimensional humming transcription using hidden markov models for query by humming systems. In: Proceedings of the IEEE International conference on Acoustics Speech and Signal Processing (2003)
3. Desain, H., van Thienen, W.: Computational modeling of music cognition: problem or solution? Music Percept. **16**(1),151–166 (1998)
4. Bamberger, J.: Turning music theory on its ear. Int. J. Comput. Math. Learn. **1**(1) 33–55 (1996)
5. Taelte, L., Cutietta, R.: In: Colwell, R., Richardson, C. (eds.): Learning Theories Unique to Music, Chap 17: Learning theories as roots of current musical practice and research, pp. 286–298. New York: Oxford University Press (2002)
6. Ghias, A., Logan, J., Chamberlin, D., Smith, B.C.: Query by humming: musical information retrieval in an aoudio database. In: Proceedings of the ACM Multimedia Conferenece'95, San Francisco (1995)
7. McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L., Cunningham, S.J.: Towards the digital music library: tune retrieval from acoustic input. In: Digital Libraries Conference (1996)
8. McNab, R.J., Smith, L.A., Witten, I.H., Henderson, C.L.: Tune retrieval in multimedia library. In: Proceedings of Multimedia Tools and Apllications (2000)
9. Blackburn, S., DeRoure, D.: A tool for content based navigation of music. In: Proceedings of ACM Multimedia, vol. 98, pp. 361–368 (1998)
10. Rolland, P.Y., Raskins, G., Ganascia, J.G.: Music content-based retrieval: an overview of melodiscoc approach and systems. In: Proceedings of ACM Multimedia, vol. 99, pp. 81–84 (1999)
11. Shih, H.-H., Zhang, T., Kuo, C.-C.: Real-time retrieval of song from music database with query-by-humming. In: Proceedings of ISMIP, pp. 251–257 (1999)
12. Chen, B., Roger Jang, J.-S.: Query by singing. In: Proceedings of the 11th IPPR Conference on Computer Vision, Graphics and Image Processing, Taiwan, pp. 529–536 (1998)
13. Lu, L., You, H., Zhang, H.-J.: A new approach to query by humming in music retrieval. In: Proceedings of the IEEE International Conference on Multimedia and Expo (2001)
14. Uitdenbogerd, A.L., Yap, Y.: Was Parsons right? An experiment in usability of music representations for melody-based music retrieval. In: Proceedings of the International Conference in Music Information Retrieval (ISMIR) (2003)
15. Haus, G., Pollstri, E.: An audio front end for query-by-humming systems. In: Proceedings of International Conference in Music Information Retrieval (ISMIR) (2001)
16. Zhu, Y., Sasha, D.: Warping indexes with envelope transforms for query-by-humming systems. In: Proceedings of ACM SIGMOD (2003)
17. Unal, E., Narayanan, S.S., Chew, E.: A statistical aproach to retrieval under user-dependent uncertainty in query-by-humming systems. In: Proceedings of ACM MIR04 (2004)
18. Doraisamy, S., Ruger, S.: A comparative and fault-tolerance study of the use of n-grams with polyphonic music. In: Proceedings of the International Conference in Music Information Retrieval (ISMIR) (2002)