# Analysis of a Contour-based Representation for Melody

Youngmoo Kim, Wei Chai, Ricardo Garcia, and Dr. Barry Vercoe
Machine Listening Group • MIT Media Lab

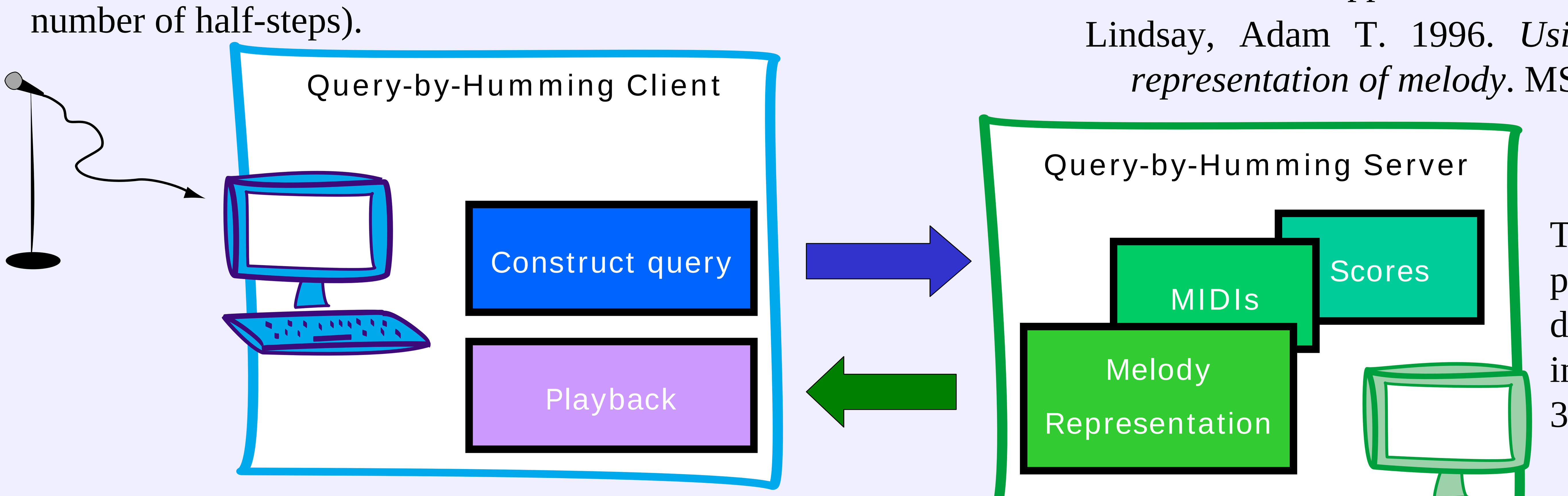
## Abstract

Identifying a musical work from a melodic fragment is a task that most people are able to accomplish with relative ease. For some time now researchers have worked to give computers this ability as well. To accomplish this, it is reasonable to study how humans are able to perform this task, and to assess what features we use to determine melodic similarity. Research has shown that melodic contour is an important feature in determining melodic similarity, but it is also clear that rhythmic information is important as well. The goal of this research is to explore what variation of contour and rhythmic information can result in the most efficient, robust, and scalable representation for melody. This is the basis of a query-by-humming system that will be used to test the validity of our proposed representation.

## Proposed melody representation

The literature suggests that a coarse melodic contour description is more important to listeners than strict intervals in determining melodic similarity. Experiments have shown that interval direction alone (i.e. the 3-level +/-/0 contour representation) is an important element of melody recognition (Dowling, 1978). There is, of course, anecdotal and experimental evidence that humans use more than just interval direction (a 3-level contour) in assessing melodic similarity (Lindsay, 1996). From a practical standpoint, a 3-level representation will generally require longer queries to arrive at a unique match. Given the perceptual and practical considerations, we chose to explore finer (5- and 7-level) contour divisions for our representation.

We used a triple <*T P B*> to represent each melody, where *T* is the time signature of the song, *P* is the pitch contour vector, and *B* is the beat number vector. The range of values of *P* vary depending on the number of levels of contour used, but follow the pattern of 0, +, -, ++, --, +++, etc. The first value of *B* is the location of the first note within its measure in beats (according to the time signature). Successive values of *B* are incremented according to the number of beats between successive notes. Values of *B* are quantized to the nearest whole beat. We used a vector *Q* to represent different contour resolutions and quantization boundaries. The length of *Q* reveals the number of levels of contour being used, and the individual values of *Q* indicate the absolute value of the quantization boundaries (in number of half-steps).




The Query-by-Humming System

This system uses the best-performing 5-level contour described above, and currently includes a database of about 3000 songs.

## Results

For this analysis, we assembled a data set of 50 multi-track MIDI files, containing a mixture of popular and classical music. All selected songs had a separate monophonic melody sound track. Random queries were chosen from within individual songs of the data set and matched against the entire set. Our results showed that the performance of 5-level contours are generally better than the 3-level contour, and 7-levels is better than that. For the 5- and 7-level contours, we also examined a variety of quantization boundaries. What is illuminating is that the best 5-level contour was able to equal the performance of the 7-level contour. This suggests that a 5-level contour may be an optimal tradeoff between efficiency and robustness to query variation (more levels will cause more variations in queries).

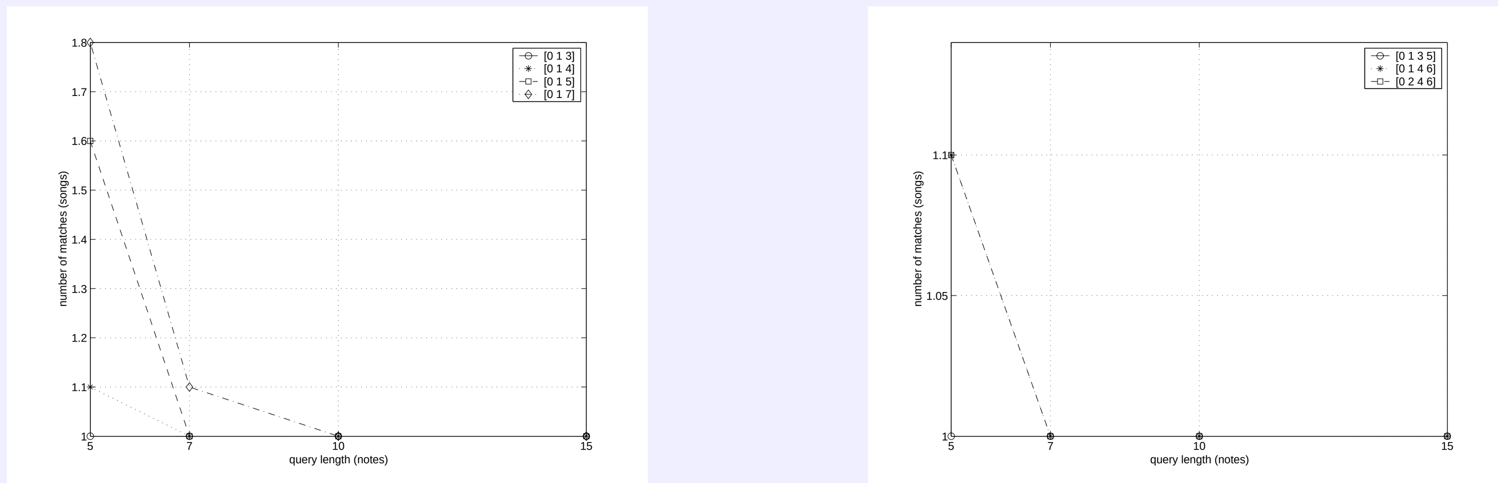


**Figure 1: Performance of 5-level and 7-level contours**

Given this result, it is revealing to examine the histogram of interval occurrences in our data set. An optimal quantizer would divide the histogram into sections of equal area. This was approximately true for the *Q* = [0 1 3] case, which has the best performance. Other choices for quantization boundaries clearly have less-optimal probability distributions, which is why they do not perform as well.

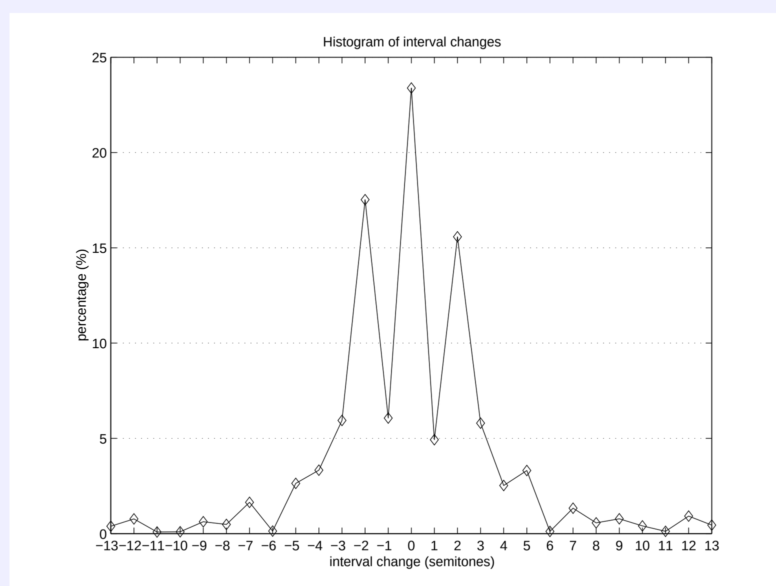



**Figure 2: Interval distribution of initial data set**

## References

Dowling, W. J. "Scale and contour: Two components of a theory of memory for melodies." *Psychological Review*, vol. 85, no. 4, pp. 341-354, 1978.

Lindsay, Adam T. 1996. *Using contour as a mid-level representation of melody*. MS thesis. MIT Media Lab.

For more information, contact <chaiwei@media.mit.edu>