Music fingerprinting is a related approach to finding a particular piece [2], where a portion of the sought-after recording is provided as the search example. Unlike QBH, music fingerprinting uses precise measurements of acoustic features of the original recording as the search key, or fingerprint. By design, music fingerprinting systems work only when the search query is exactly the same as the recording in the database. An alternate performance (such as a hummed melody) will stymie a fingerprint system, preventing it from making a match.

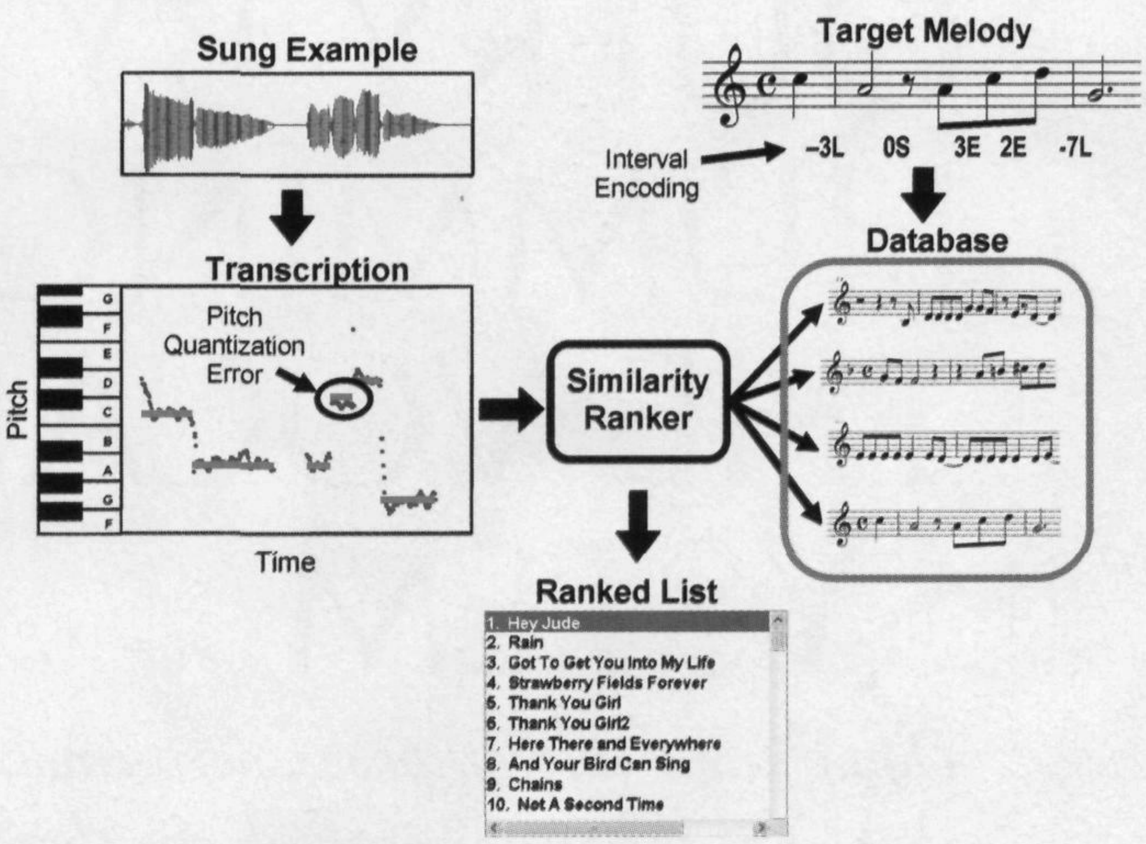
Music librarians attest to the usefulness of QBH, as they often take phone calls in which patrons hum a tune over the phone to try to identify its title. This musical query is natural for many people and requires no access to a copy of the desired recording at the time of the query. Moreover, QBH systems may provide more general results than fingerprinting systems, as a user may have a tune in mind but not a particular artist (such as "Retrieve both Phish's and ZZ Top's versions of 'Jesus Just Left Chicago'").

Here, we discuss QBH using the VocalSearch [9] system developed at the University of Michigan and Carnegie Mellon University (with support from a grant from the National Science Foundation) as a running example because it provides performance on par with the best current systems and involves a simple design and user interface (see the figure here). Queries are usually in the form of a user-supplied melody, theme, hook, instrumental riff, or some other memorable part of a piece [9]. Most QBH systems are therefore designed for music that is melodic. Music that is predominantly rhythmic or timbral is less amenable to QBH techniques, though some interesting recent work has addressed rhythmic recognition (www.songtapper.com) [4].

The typical QBH melody-comparison techniques—string alignment, n-grams, Markov models, dynamic time warping—compare monophonic (a single melody line) melodies in the database to a monophonic query. Since most Western music is harmonic or polyphonic (multiple concurrent voices), individual monophonic melodies must first be drawn from music to make it searchable. These melodies are not extracted directly from the audio, since audio with multiple concurrent instruments render standard pitch-tracking methods unreliable. This open problem may be avoided by using symbolically encoded music (such as musical instrument digital interface, or MIDI, protocol files).

A MIDI file contains instructions detailing which notes to play, the order of the notes, and the duration and volume of each of them. MIDI files are available for tens of thousands of pieces, thanks (in part) to the popularity of MIDI-based cell phone ringtones. Polyphonic MIDI files may be automatically converted to sets of monophonic files in various ways [8, 12]. The resulting files may then be used as search keys in a melodic database.

To increase precision and recall, QBH systems



How queries are processed in the VocalSearch QBH system.

typically search over a database of melodic themes (such as the first four notes of Beethoven's "Fifth Symphony" or the verse of a song) rather than over a full piece. Restricting the database this way—single voice, representative themes—reduces the chance of a serendipitous match between the query and an obscure component of some unrelated piece of music (such as a descending scale in a query that matches a walking bass line). Having a reduced amount of data to search can speed operation by a factor of 100 or more. For example, each song in VocalSearch's Beatles database is represented by a theme for the verse and a different theme for the chorus. The figure outlines an example of these themes in the form of standard music notation.

Developing a system to identify themes is a special challenge. Themes can occur anywhere in a piece; they can be played by any instrument or voice; for example, Dvorak's "American Quartet" introduces the theme with a viola, an instrument rarely used to introduce thematic material. Themes can be repeated with variation, sometimes subtle, sometimes not. Some themes are simple and easily confused with the nonmelodic components of a