piece. Historically, theme databases have been built by highly skilled musicians, a labor-intensive process that does not scale well. However, automatic thematic extraction methods have also been shown to be effective [7].

### What Is That Noise?

Query quality is wide ranging, and even trained musicians do not necessarily present "better" queries to QBH systems [9]. Individuals may not recall the theme correctly, may mix themes, or mix voices (such as lead guitar and vocalist) into a composite query. Individuals also have production problems; they may be unable to sing the desired pitch and may laugh and cough, all while presenting the query. The computer "hears" all the unintended artifacts and considers them part of the query.

Transcription, or converting audio to musical notation, is another source of noise. Transcribers often introduce artifacts, drop notes, and botch rhythms. Thus, even a "perfect" query may not appear perfect after transcription. The transcription portion of the figure is an example of what the transcription of several well-sung notes looks like. The blue dots at the top of the figure represent a sequence of fundamental frequency estimates of the sung query; note the pitch glides and wavering pitch. The red lines represent the transcriber's note estimates (pitch and duration) quantized to the nearest piano key; note the quantization error on the fourth note, raising what should be a C to a C sharp.

Such errors, whether due to the poor pitch of an untrained singer or to a poor transcription must be included when matching the transcribed query to targets in the database. One approach to handling quantization error is to avoid it altogether by matching the pitch contour directly to database themes without quantization [5]. VocalSearch handles transcriber and singer error through prior estimation of error likelihood. It does this by training the system on a set of sung melodies of known pieces. The transcribed melodies are compared to the original, and the differences are recorded. Training builds a probabilistic model of the singer and transcriber error.

### Best Guess

Many researchers have looked into melodic-comparison techniques [3, 6, 12], while others have looked into polyphonic comparison techniques [8, 11]. See [1] for a comparative analysis of the most popular melodic-similarity measures for QBH, including n-grams, dynamic time warping, Markov models, and string alignment. Many articles on the subject are also available in the International Conference on Music Information Retrieval research paper repository at www.ismir.net/all-papers.html.

> MANY RESEARCHERS HAVE reported results showing good performance in terms of processing time, precision, and recall.

VocalSearch [9] uses a probabilistic string-alignment algorithm to measure similarity between targets and queries. String-alignment systems determine how similar the query is to each of the targets based on edit distance; the "cost" (such as number of character changes) is what it takes to transform one string (the query transcription) into another string (the database theme). The theme with the lowest-cost transformation into the query is considered the best match.

In VocalSearch, these costs are based on the trained error model discussed earlier, which involves the type of errors singers make when singing. For example, it is very likely that when attempting to jump the interval of a minor 3rd, a singer will instead jump a major 3rd; it is very unlikely the singer will jump a major 7th. Thus, if we are trying to match a query to two themes, one with a minor 3rd and one with a major 7th, and the query presents a major 3rd, we would consider the first query to have the lower edit cost.

The melodic encoding used for queries and targets has a strong effect on the kinds of errors a QBH system can handle. For example, to get around the problem of key mismatch between query and target(s),