

# What is a Sung Query?

Colin Meek  
University of Michigan  
Ann Arbor MI 48109 USA  
meek@umich.edu

William P. Birmingham  
University of Michigan  
Ann Arbor MI 48109 USA  
wpb@umich.edu

Bryan Pardo  
University of Michigan  
Ann Arbor MI 48109 USA  
bryanp@umich.edu

## ABSTRACT

If we are to collect queries from disparate sources for a benchmark, it is critical that we state and control the procedures and assumptions underlying the query gathering process. Many MIR systems make explicit or implicit assumptions about the form a *sung* query takes. We would like to table these assumptions, warts and all, and suggest ways in which we can reconcile them with the *query-by-humming* portion of a MIR benchmark.

## 1. INTRODUCTION

In query-by-humming systems, a user sings or hums a memorable bit of a song he's trying to track down. The system then compares that query (reduced to some abstracted form) to a database of music (usually in some abstracted form). Even within the domain of monophonic queries – assuming that the user is not playing along on his guitar, or singing with her barbershop quartet – there can be great variety in the manifestations of a query. This variety lies both in constraints (explicit or otherwise) imposed by a system or developer, and in the query style.

## 2. SYSTEM CONSTRAINTS

In some cases, constraints are imposed by system developers, to simplify the retrieval process:

- The “Graffiti™” query: “Da Da Da”. Many MIR systems employ note-based abstractions of a query. In practice, it is difficult – using amplitude thresholding and frequency analysis data – to *segment* a query, identifying points where new notes begin. Just as we are sometimes required to learn a new way of writing to communicate with our PDA devices (e.g., the Graffiti™ system for Palm OS), some systems require [7] (or politely ask [2]) us to carefully delimit notes using hard constants (e.g., “Da”).
- The metronome query: Beat tracking is tough at the best of times [4][3], and on short monophonic queries

– with frequently clipped rests and lousy rhythm in our test query collection – it could prove extremely problematic. Giving the singer the beat – through a metronome click [2], a samba percussion track or some visual or aural conductor – allows for the recovery not only of *timing* information but *metrical* information as well.

## 3. INTERACTIVITY

In some cases, there may be an interactive element to the query gathering process:

- The spell checker: In our project, we are developing interfaces that allow users to iteratively correct the abstraction of their singing to minimize occurrences of error.
- Music browser: Users may be interested in “browsing” music databases, based on aural, visual and textual cues.

The “benchmark” in these examples would have to be some collection of actual people, making packaging and distribution rather problematic. For this reason, we believe that queries gathered by relatively non-interactive (or at least pre-defined) means will have to be the norm.

## 4. STYLISTIC ISSUES

We have observed some legitimate stylistic “quirks” in queries we have gathered. Notably, in the context of MIR systems, we have the “pseudo-polyphonic” query: “Start spreading the news, *ba-du-da-du-da*, I’m leaving today...” A rendition of Frank Sinatra’s “New York, New York” recreates not only the crooner’s smooth voice, but the trumpet line from Count Basie’s orchestra. Several systems work with databases composed of “themes” or single lines, with no mechanism for hypothesizing about jumps between voices.

## 5. MIR SYSTEM COMPONENTS

In the interest of isolating sources of error in a MIR system, some system developers pose idealized queries [6], in the “this - is - the - query - we - would - see - if - system - component -  $x$  - were - 100% - effective” sense (e.g., manually segmented queries for note-based retrieval).

Building on this idea, it would be useful to have, at each level of processing, standard performance evaluations. For example, a transcriber from the audio query to a representational system would be measured against a hand-transcription

that is presumed correct. Multiple evaluative measures will likely be useful, since various retrieval systems will likely require different performance specifications (for instance, one method may be tolerant of pitch error, while another is tolerant of rhythm error). That said, for a given task, a standard evaluation measure should be specified, so that systems performing the same task may be measured by the same yardstick. To avoid bias, evaluation should be performed automatically, without human intervention.

## 6. THE QUERY CORPUS

How do we deal with these query forms? Should we standardize on a particular context for the collection of queries, ignore the context altogether, or group queries according to the underlying style or interace constraint? While it may be tempting to make an appeal to the Graffiti<sup>TM</sup> handwriting recognition analogy, and argue that singers will quickly adapt in the interest of improving performance, allowing interface constraints at this point may be dangerous for two reasons:

- First, to our knowledge, no comprehensive study of user query behavior and preferences has been undertaken. As such, we are not currently in a position to intelligently decide which assumptions are reasonable, and;
- Second, any decision at this point may remove the motivation for the pursuit of *unrestricted* designs.

Admitting that “enlightened self-interest” may influence our decisions about queries we contribute to a benchmark, we suggest the following guidelines:

- Allow no constraint on the method used to gather the query. If we disallow (pseudo-)polyphony, we are making a statement about the validity of polyphonic MIR

systems [1]. If we use metronome or Graffiti<sup>TM</sup> queries, then we underplay the advantages of subtle edit models [6] and the robustness of frame-based approaches [5] in the face of difficult to quantize or segment input.

- Reach a consensus on “classes” of query. Each class would then have a *shared* set of tools and procedures for query gathering, to insure consistency.

## 7. CONCLUSION

We have likely only scratched the surface of this issue. Even in the absence of a standard for the query gathering procedure, we emphasize that assumptions about, and restrictions on that process must be made explicit and clear.

## 8. REFERENCES

- [1] D. Byrd and T. Crawford. Problems of music information retrieval in the real world. 2002.
- [2] W. Chai. Melody retrieval on the web. Master’s thesis, Massachusetts Institute of Technology, 2001.
- [3] R. Christopher. Automated rhythm transcription. In *Proceedings of ISMIR 2001*.
- [4] S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 2001.
- [5] D. Mazzone. Melody matching directly from audio. In *Proceedings of ISMIR2001*.
- [6] C. Meek and W. Birmingham. Johnny can’t sing: A comprehensive model for error in sung queries. In *Proceedings of ISMIR 2002*.
- [7] J. Shifrin, B. Pardo, C. Meek, and W. Birmingham. Hmm-based musical query retrieval. In *Proceedings of JCDL 2002*.