Perceptual and Cognitive Applications in Music Information Retrieval

David Huron
Cognitive and Systematic Musicology Laboratory
School of Music
Ohio State University

huron.1@osu.edu

Abstract

Music librarians and cataloguers have traditionally created indexes that allow users to access musical works using standard reference information, such as the name of the composer or the title of the work. While this basic information remains important, these standard reference tags have surprisingly limited applicability in most music-related queries.

Music is used for an extraordinary variety of purposes: the restaurateur seeks music that targets a certain clientele; the aerobics instructor seeks a certain tempo; the film director seeks music conveying a certain mood; an advertiser seeks a tune that is highly memorable; the physiotherapist seeks music that will motivate a patient; the truck driver seeks music that will keep him/her alert. Although there are many other uses for music, music's preeminent functions are social and psychological. The most useful retrieval indexes are those that facilitate searching according to such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information.

In attempting to build such musical indexes, two general questions arise: (1) What is the best taxonomic system by which to classify moods, styles, and other musical characteristics? (2) How can we create automated systems that will reliably characterize recordings or scores?

Internet-based music distribution has brought these two questions to the fore. In the case of proprietary musical databases, the second problem can be centrally managed, and perhaps addressed using manual methods. However, past experience with Internet access to text documents implies that non-proprietary decentralized indexing is likely to prove more popular and successful. That is, future music indexes will likely resemble web-wide search engines (like Infoseek or Google) rather than closed proprietary systems (like Beatscape or Encyclopedia Britannica).

The problem of building musical web crawlers that traverse the web and index music-related files (such as MP3) should prove challenging. The enabling technology for such musical web crawlers will need to draw extensively on research in music perception and cognition. Consider two sample problems: music summarization and mood characterization.

Music Summarization

Before downloading or streaming an entire work, there is great benefit to hearing a brief illustrative excerpt — a musical equivalent of the "thumbnails" commonly used in electronic picture galleries. Not all

portions of a musical work are equally representative of a given piece, and so the practice of extracting the initial few seconds (*incipit*) is not optimum for identifying or recognizing a work. Most well-known musical themes or hooks do not appear within the first ten seconds. Moreover, the optimum musical thumbnail may consist of two or more brief passages edited into a single five-second sound bite.

One way of viewing this problem is as a musical equivalent of the problem of text summarization (Mani & Maybury, 1999). In our lab, we have initiated a research project intended to develop useful algorithms for automatic music summarization. We are currently building a database of listener responses indicating the most important segments in various musical works. Specifically, we are marking those segments in a sample of sound recordings which are most prototypical (Rosch & Lloyd 1978). Salient passages tend to be (1) easily recalled by listeners from one day to the next, and (2) generate many false-positive recognition responses (i.e., listeners incorrectly claim to have heard the passage before, having been exposed to other excerpts from the same work). This database will provide a test suite that allows us to assess the effectiveness of different summarization algorithms.

Mood Characterization

The most well-known models of mood entail two factors: *valence* (happy/anxious) and *arousal* (calm/energetic) (e.g., Thayer, 1989). Musical moods can be be usefully characterized using the resulting two-dimensional space. Exemplars of the four quadrants might be Rossini's *William Tell Overture* (happy/energetic), Bach's, *Jesu, Joy of Man's Desiring* (happy/calm), Berg's *Lulu*, (anxious/energetic), and the opening of Stravinsky's *Firebird* (calm/anxious). Of the two factors, arousal is more computationally tractable and can be estimated using simple amplitude-based measures (e.g. Huron, 1992).

More recent research has focussed on specific emotional connotations, such as auditory correlates of cuteness, fear, disgust, aggressivity, submissiveness, and pleasure (e.g., Ohala, 1980; Huron, Kinney & Precoda, MS).

Suggested Readings

Huron, D. 1992. The ramp archetype and the maintenance of auditory attention. *Music Perception,*10*(1)*83-92.

Huron, D., Kinney, D. & Precoda, K. (2000). Relation of pitch height to perception of dominance/submissiveness in musical passages. In review.

Mani, I. & Maybury, M.T. (eds.) 1999. *Advances in Automatic Text Summarization.* Cambridge, Massachusetts: MIT Press.

Ohala, J. 1980. The acoustic origin of the smile. *Journal of the Acoustical Society of America,* 68, S33.

Rosch, E. & Lloyd, B. (eds.) 1978. *Cognition and categorization.* Hillsdale, N.J.: Erlbaum.

Thayer, R.E. 1989. *The Biopsychology of Mood and Arousal.* New York: Oxford University Press.