Audible Magic Corporation
985 University Ave Suite 35
Los Gatos, CA 95032
P: 408 399-6405

Bibliography of Audio Identification Technology

Mel Frequency Cepstral Coefficients for Music Modeling

Beth Logan; International Symposium on Music Information Retrieval (2000)
http://ciir.cs.umass.edu/music2000/papers/logan_abs.pdf

Abstract: We examine in some detail Mel Frequency Cepstral Coefficients (MFCCs) - the dominant features used for speech recognition - and investigate their applicability to modeling music. In particular, we examine two of the main assumptions of the process of forming MFCCs: the use of the Mel frequency scale to model the spectra; and the use of the Discrete Cosine Transform (DCT) to decorrelate the Mel-spectral vectors.

Content-Based Retrieval of Music and Audio

Jonathan T. Foote; In C.-C. J. Kuo et al., editor, *Multimedia Storage and Archiving Systems II, Proc. of SPIE*, Vol. 3229, pp. 138-147, 1997 http://www.fxpal.xerox.com/people/foote/papers/spie97.abs.html

Abstract: Though many systems exist for content-based retrieval of images, little work has been done on the audio portion of the multimedia stream. This paper presents a system to retrieve audio documents by acoustic similarity. The similarity measure is based on statistics derived from a supervised vector quantizer, rather than matching simple pitch or spectral characteristics. The system is thus able to learn distinguishing audio features while ignoring unimportant variation. Both theoretical and experimental results are presented, including quantitative measures of retrieval performance.

Overview of Audio Information Retrieval

Jonathan T. Foote, In *Multimedia Systems*, vol. 7 no. 1, pp. 2-11, ACM Press/Springer-Verlag, January 1999

Abstract: The problem of audio information retrieval is familiar to anyone who has returned from vacation to find an answering machine full of messages. While there is not yet an "AltaVista" for the audio data type, many workers are finding ways to automatically locate, index, and browse audio using recent advances in speech recognition and machine listening. This paper reviews the state-of-the-art in audio information retrieval, and presents recent advances in automatic speech recognition, word spotting, speaker and music identification, and audio similarity with a view towards making audio less "opaque."

Classification, Search, and Retrieval of Audio

Erling Wold, Thom Blum, Douglas Keislar, James Wheaton; *CRC Handbook of Multimedia Computing*; CRC Press LLC (1999)
http://www.musclefish.com/crc/

Abstract: Many audio and multimedia applications would benefit if they could interpret the content of audio rather than relying on descriptions or keywords. These applications include multimedia databases and file systems, digital libraries, automatic segmentation or indexing of video (e.g., news or sports footage), surveillance, as well as sound browsers for effects designers and musicians. This chapter describes an audio analysis, search, and classification engine that reduces sounds to acoustical and perceptual features. This analysis allows the sounds to be classified or queried by their audio content. Queries can based on any one or a combination of the acoustical features, by specifying previously learned classes based on these features, or by selecting or entering reference sounds and asking the engine to retrieve sounds that are similar or dissimilar to them.

Content-Based Classification, Search, and Retrieval of Audio

Erling Wold, Thom Blum, Douglas Keislar, James Wheaton; *IEEE Multimedia* 1070-986X/96/$5.00 Vol. 3, No. 3: FALL 1996, pp.27-36; http://www.musclefish.com/crc/

Abstract: Many audio and multimedia applications would benefit from the ability to classify and search for audio based on its characteristics. The audio analysis, search, and classification engine described here reduces sounds to perceptual and acoustical features. This lets users search or retrieve sounds by any one feature or a combination of them, by specifying previously learned classes based on these features, or by selecting or entering reference sounds and asking the engine to retrieve similar or dissimilar sounds.

Audible Magic Corporation
985 University Ave Suite 35
Los Gatos, CA 95032
P: 408 399-6405

Audio Databases with Content-Based Retrieval

Thom Blum, Douglas Keislar, James Wheaton; Erling Wold, *Intelligent Multimedia Information Retrieval*; Edited by Mark T. Maybury; AAAIPress/The MIT Press; ISBN: 0262631792 (1997)

Abstract: Despite vast research and development efforts in such diverse domains as digital signal processing, psychoacoustics, speech recognition, computer music, and multimedia databases, there is a paucity of literature that addresses the issue of automatic classification of sounds. Many audio and multimedia applications would benefit from the ability to classify and search for audio based on the characteristics of the audio rather than by resorting exclusively to keywords. This chapter describes an audio analysis, search, and classification engine which reduces sounds to perceptual and acoustic attributes. Sounds can then be searched or retrieved by any one or a combination of the attributes, by specifying previously learned classes based on these attributes, or by selecting or entering reference sounds and asking the engine to retrieve sounds that are similar (or dissimilar) to them. After surveying some related research, we examine this engine and describe a specific application for browsing sounds in a database.

A Content-Aware Sound Browser

Douglas Keislar, Thom Blum, James Wheaton, and Erling Wold; *Proceedings of the 1999 International Computer Music Conference* (1999)

Abstract: The SoundFisher TM browser is a cross-platform, single- or multi-user sound-effects database application. It incorporates an audio analysis engine that permits retrieval of sounds based on their acoustical similarity, as well as on traditional keywords and file attributes. Databases can be constructed from sounds on a local files system and/or the World Wide Web.