Audio Similarity Retrieval Engine

SISAP ’10, September 18–19 2010, Istanbul, Turkey

Pavel Jurkas  Milan Stefina  David Novak  Michal Batko
{xjurkas|xstefina|david.novak|batko}@fi.muni.cz
Masaryk University, Brno, Czech Republic

Overview
• general engine for similarity search in audio data
• primary indexing based on metric model
• modular architecture

Data
• short streams with various sound samples
• four low-level MPEG-7 Audio descriptors:
  – Audio Power (AP) – power of the signal
  – Audio Spectrum Centroid (ASC) – the central frequency
  – Audio Spectrum Spread (ASS) – frequency spread in the audio signal
  – Audio Waveform (AW) – min & max value

System Overview
• both data & query descriptors are split into fixed-size segments
• metric similarity ($L_2$) on level of segments
• overall similarity measured for “longest overlapping subsequence”
• segments organized in a metric index
• individual descriptors can be aggregated

Current Demo and Vision

Current Demo Evaluation
• 1,100 short sound samples
  – from Partners in Rhyme and Freesound
• human-judged evaluation of effectiveness
  – individual descriptors tested
  – not a clear winner for all test cases
• aggregation (1:1:1:1) gives best results
  – see http://mufin.fi.muni.cz/audio/

System Potential and Vision
• multiply size of the database
• implement better metric similarity functions
• test different descriptors (high-level)
• add semantics by keywords reranking
• create a universal sequence-matching module
• stream processing