Multimedia Editing in the Cloud: Treating Audio as ‘Big Data’

Adam T. Lindsay
Multi-Service Networks,
Cosener’s House, 2009
and now for something completely different...
Some context

- I’m interested in...
  - Metadata-assisted multimedia editing
  - Using music analysis service to get sample-accurate, hierarchical event pointers
  - Using these for music remixing
Remixing using metadata

Song
Bars
Beats
Tatums
Splitting things into tiny pieces

- Make the programming paradigm as declarative/functional as possible
- Rendering output can be done independently of handling metadata
- A content description can act as proxy for the underlying content
- Don't need to handle the data at all
- Rendering instructions form a small vocabulary
Smells a lot like SMIL 1.0

Rendering

<sequence duration="57.11676" source="847e7a3146fb790ccfa4a071f7395775">
  <trackinfo filename="../music/aha.mp3" id="847e7a3146fb790ccfa4a071f7395775"/>
  <trackinfo filename="../music/SLadies.mp3" id="1630307ae0eea4a380ab2213827eec6f"/>
  <trackinfo filename="../music/BJean.mp3" id="2d539b439ec027e73abd2390c5611d2f">
    <parallel duration="0.33472">
      <beat duration="0.33472" start="0.21285"/>
      <beat duration="0.31216" source="1630307ae0eea4a380ab2213827eec6f" start="0.38352"/>
    </parallel>
  </trackinfo>
</sequence>
Audio is easy

- Transparent
- Linear superposition of waveforms = mixing
- Simple information set:
  - Source, source-start-time, source-duration
  - Destination, destination-start-time
  - Return samples, destination-start-time
Source independence

- Feels like MapReduce
- Data-intensive (CD audio = 10 MB/min)
- Need a strategy to cope with data-heavy nature
  - One Map task per source
  - One Reduce per job
  - Collection/Reduce best at biggest contributing source
Implementation

- Proof of concept
- Using small-scale (flat) P2P network
- (Based on Thrift libraries for RPC implementation)
- Consistent hashing for matching content with node
- No experiments to test gains and costs yet
The Future

- Video can fit into this model too
- ...if we apply alpha channel adjustments first
- Editing in the local network
- Push content addressing down network stack
- Maybe push aggregation to content processor nodes
Mixing on a web scale

- Frame/sample-accurate fragment addressing
- Better content-centric addressing
- Some access control/billing models
Fin