

PHILIPS

Perceptual Evaluation of Music Similarity

Alberto Novello, M. McKinney

Philips Research Laboratories Eindhoven (The Netherlands)

Goals

- Develop a model of music similarity perception, based on musical audio characteristics, to support MIR applications.
 - Collect “ground truth” data on music similarity perception
 - Develop an algorithmic model operating on features extracted from the audio to predict the similarity data.

Music Similarity

- Can be defined along different criteria:
 - Musicological
 - Algorithmic
 - Perceptual**
- And each one of these is multidimensional !
- No absolute yardstick (ground truth)

Previous work:

Music theory:

- Deliege (segmentation-cue abstraction-categorization)

Perceptual:

- Lamont (dimensions): 2 classical songs
- Chupchik (dimensions): 3 genres (pop, classical, jazz)

Algorithmic:

- Pampalk 'Computational Models of Music Similarity and their Application in Music Information Retrieval'
- Logan, Berenzweig (evaluation of best acoustical features, MFCCs)

Some open issues:

- Incomplete verification of theoretical models
- Algorithmic work is not linked to human perception
- Experimental work is fragmented or too specific (e.g. melodic similarity). Need for a global description.
- Music Similarity has typically been analyzed on small sets of music (few genres or few songs)
- Absence of widely accepted ground truth

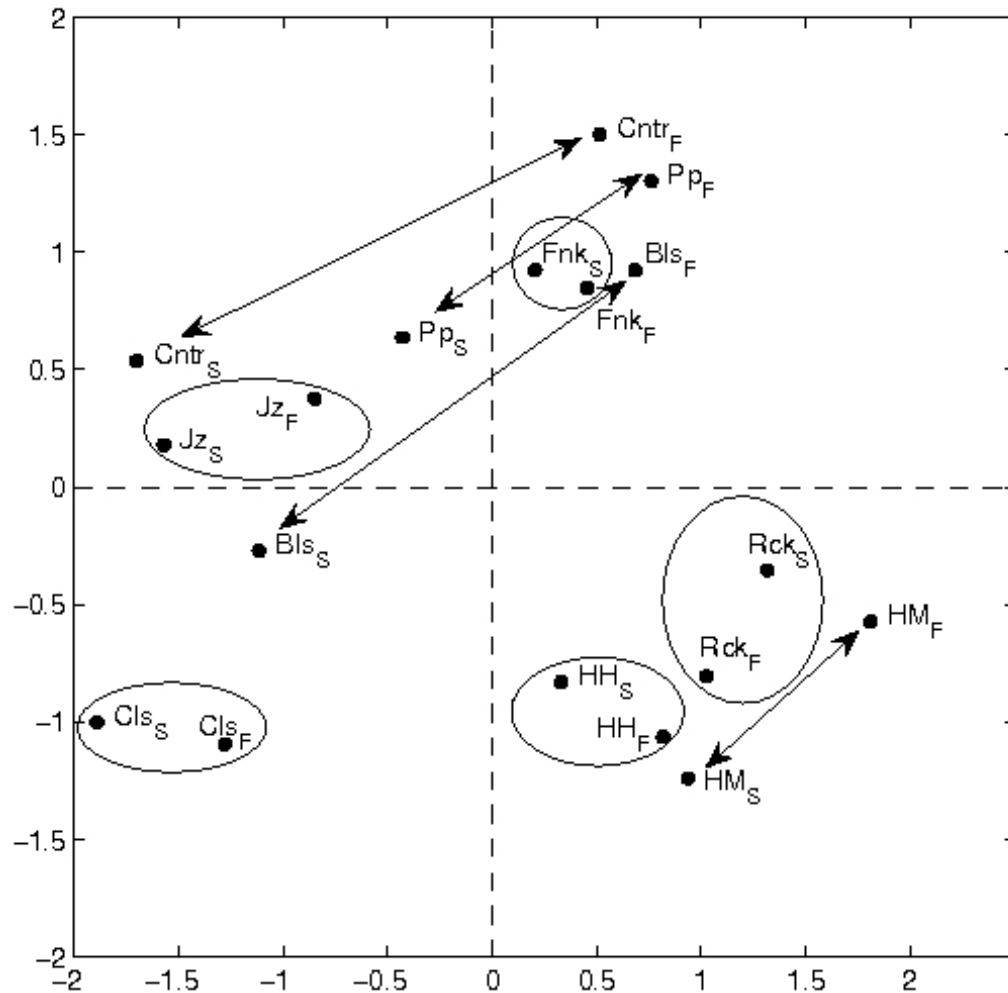
Specific Experimental Questions:

- Is there consistency of similarity ratings within and across subjects?
- Does music experience influence similarity perception (musician/non musician, known/unknown music)?
- What are the relevant musicological cues used by people in similarity ratings? How salient are they?

Results: Multi Dimensional Scaling

Similarity matrix (non metric MDS)

The Stress Factor = 0.244 (2D) / 0.157 (3D)




Interpretation of Dimensions (?)

Conclusion:

- Combination of triadic comparison method with MDS is a useful technique for evaluating music similarity perception.
- Rating music similarity can be difficult: large number of outliers (8/36 subjects)
- High consistency in similarity ratings across subjects (95% of the triads yield significant concordance)
- No influence of musical experience or familiarity with the music on similarity judgments
- Genre clustering is evident in MDS-generated topologies: Similarity is likely correlated with genre.

Current and future work

- Large scale web experiment (January, February /07)
 - 78 subjects (~ 2h each)
 - 78 songs (13 genres, 2 tempi, 3 timbres)
 - We need subjects!!!! (please )
http://simexp.hightechcampus.nl/index_en.html
- Algorithmic development
 - Predicting perceptual data from acoustic features.