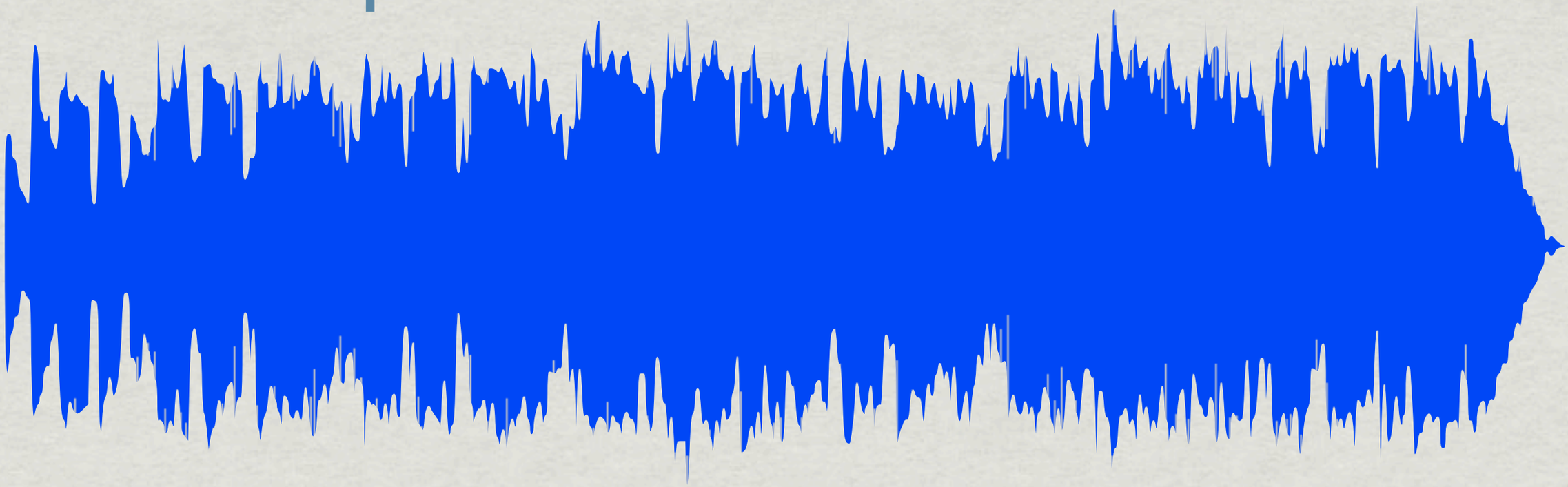
Methods for the automatic structural analysis of music

Jordan B. L. Smith

CIRMMT Workshop on Structural Analysis of Music

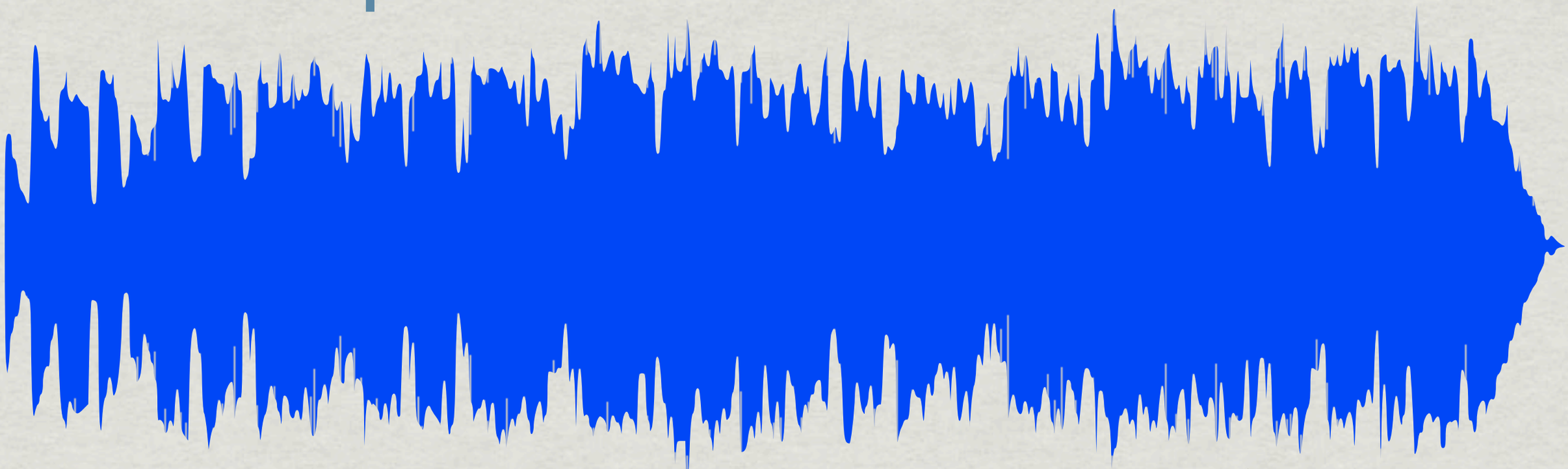
26 March 2010

The problem



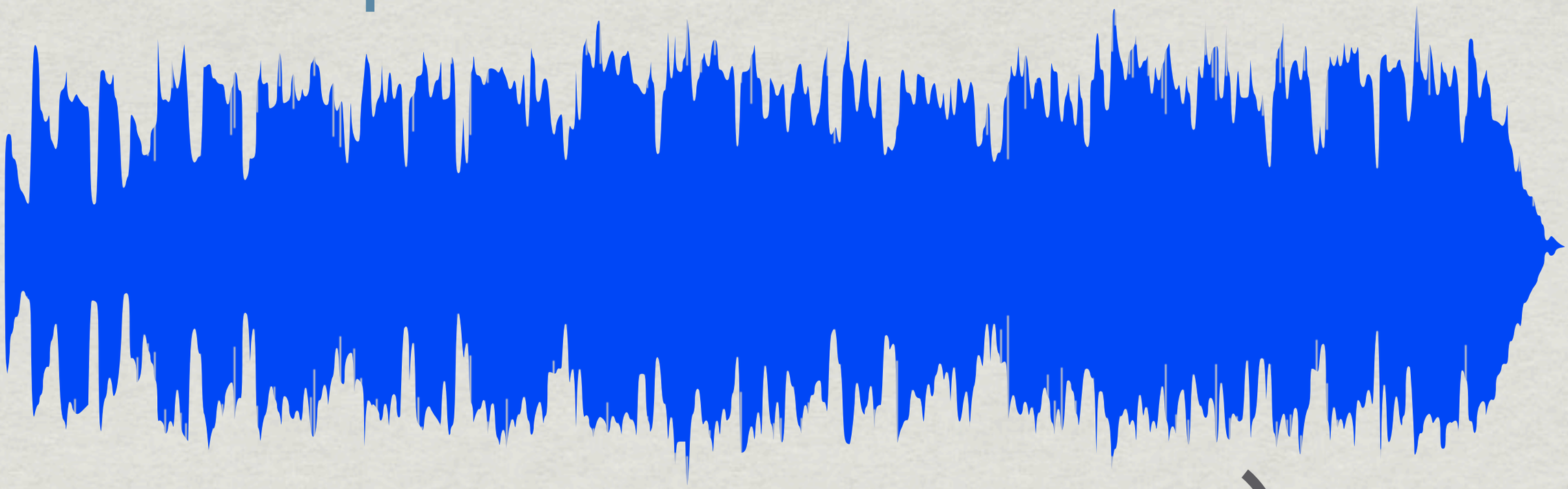
- Going from sound to structure

The problem

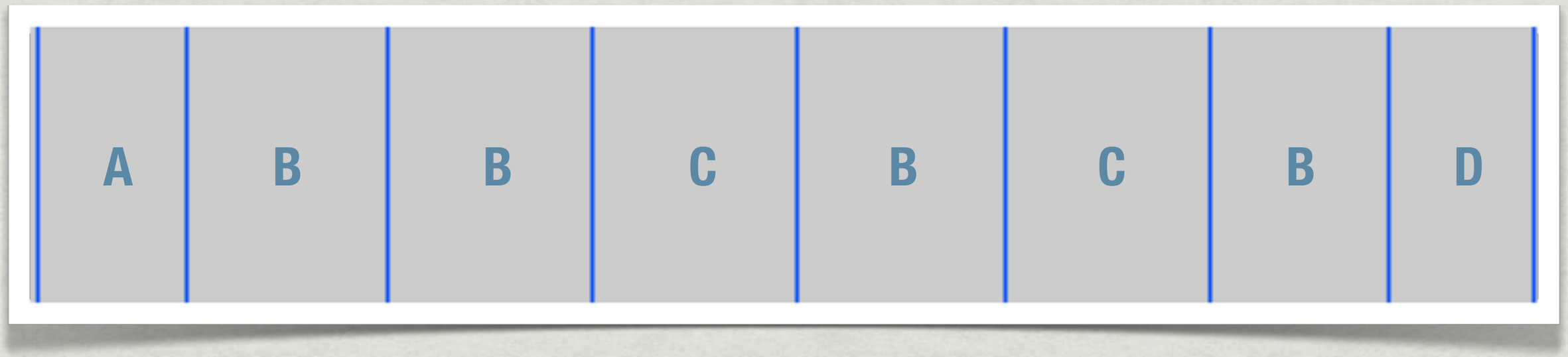


- Going from sound to structure

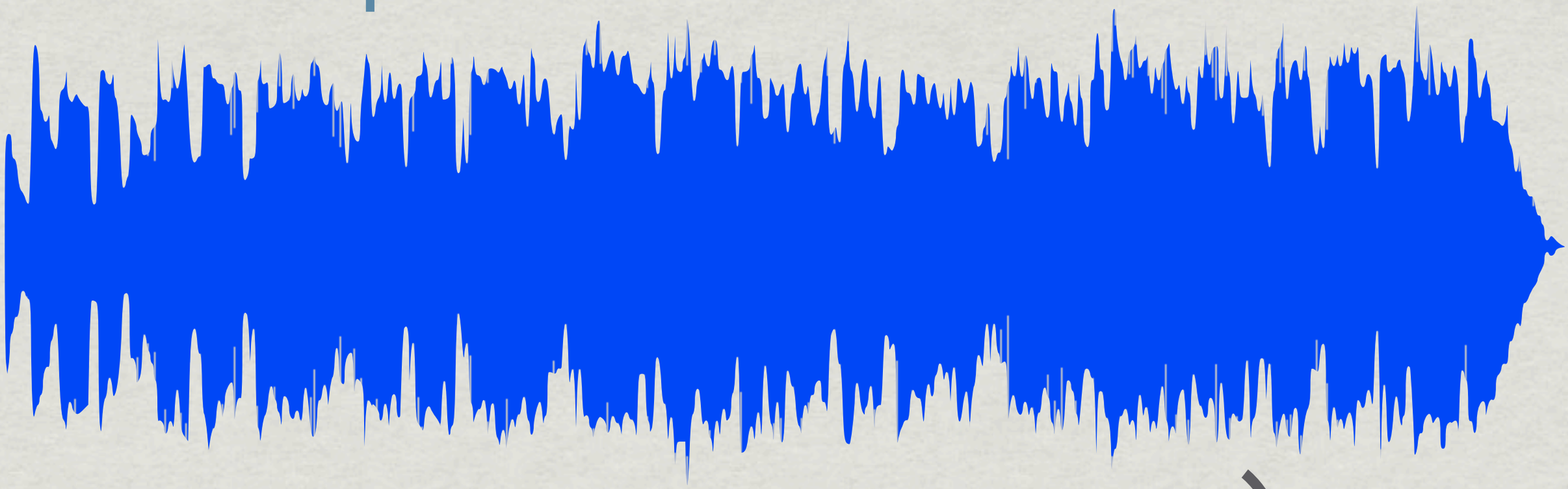
The problem



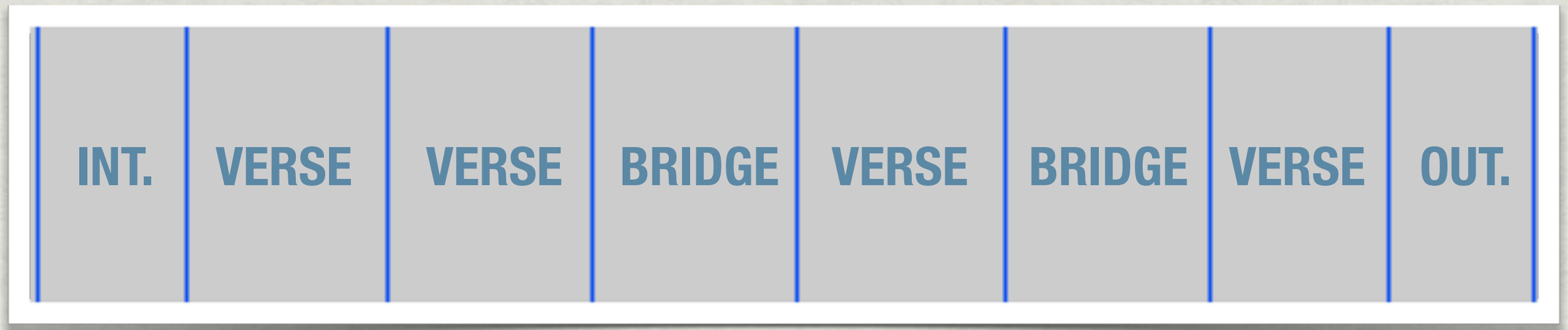
- Going from sound to structure



The problem



- Going from sound to structure



My objective today:

- To describe the variety of methods out there
- To illustrate three ways of subdividing the field

Three ways to look at the field:

- By hypotheses about what structure is → Sequences vs. states
- By hypotheses about how structure is expressed → Timbre vs. harmony (vs. rhythm vs. lyrics...)
- By techniques to do structure analysis → Similarity matrix vs. clustering

Outline

1. Two hypotheses:

- States
- Sequences

2. A word on features:

- Timbre, harmony, etc.
- Dynamic features

3. Two techniques:

- Similarity matrix
- Clustering models

Features (fast)

- Timbral features: instrumentation, vocal quality, etc.

MFCC,
MPEG-7 DESCRIPTORS

- Pitch features: what notes and chords are being played

CHROMA VECTOR,
FUNDAMENTAL FREQUENCY

- Rhythmic features: pulse periods

RHYTHMOGRAM

Features (fast)

- Timbral features: instrumentation, vocal quality, etc.

  MFCC, MPEG-7 DESCRIPTORS

- Pitch features: what notes and chords are being played

  CHROMA VECTOR, FUNDAMENTAL FREQUENCY

- Rhythmic features: pulse periods

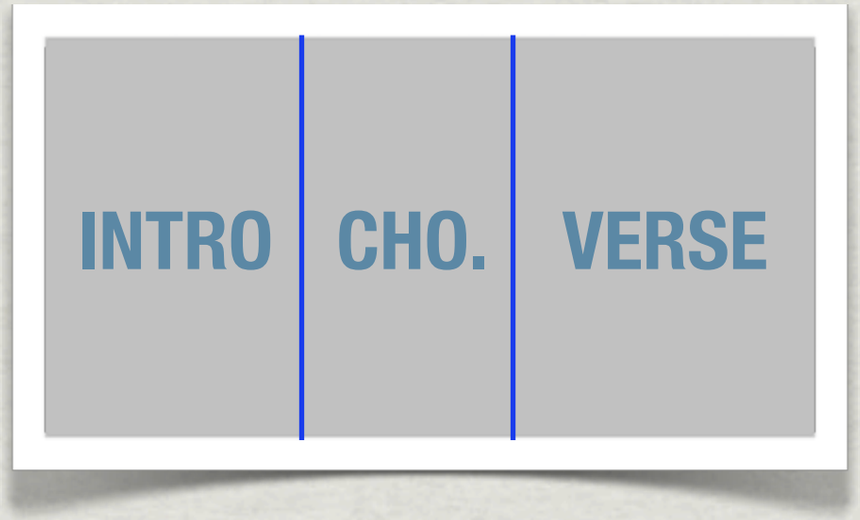
  RHYTHMOGRAM

- Lyrics

  LYRICS

Features

- What makes sections different?



PAUL SIMON: "CAN'T RUN BUT"

Features

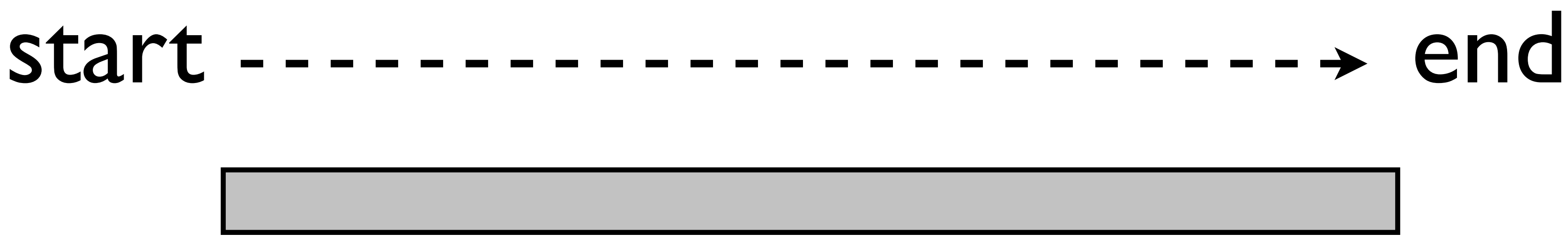
- What makes sections similar?

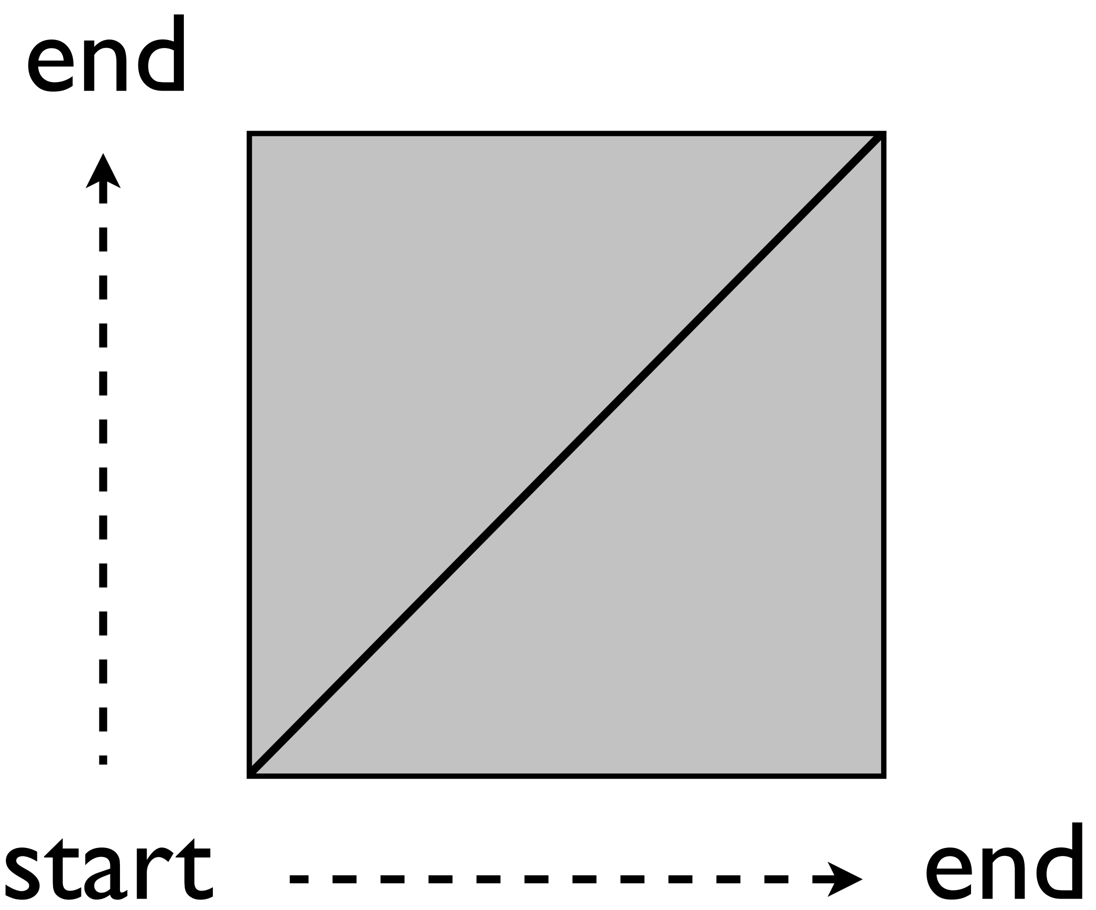
...

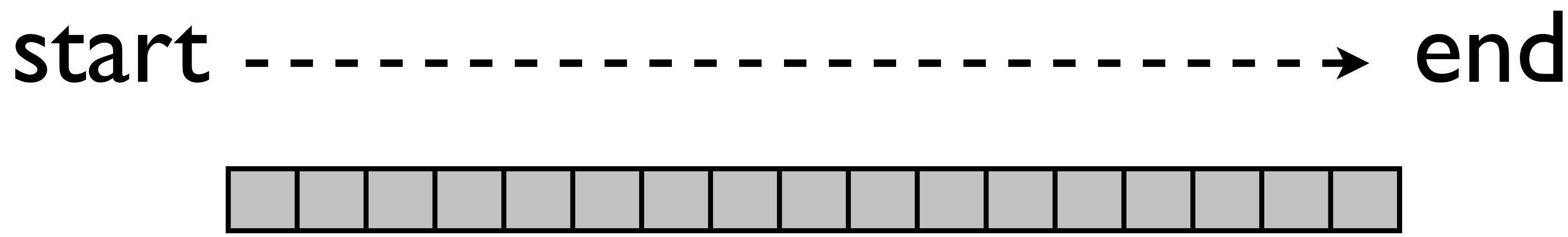
...

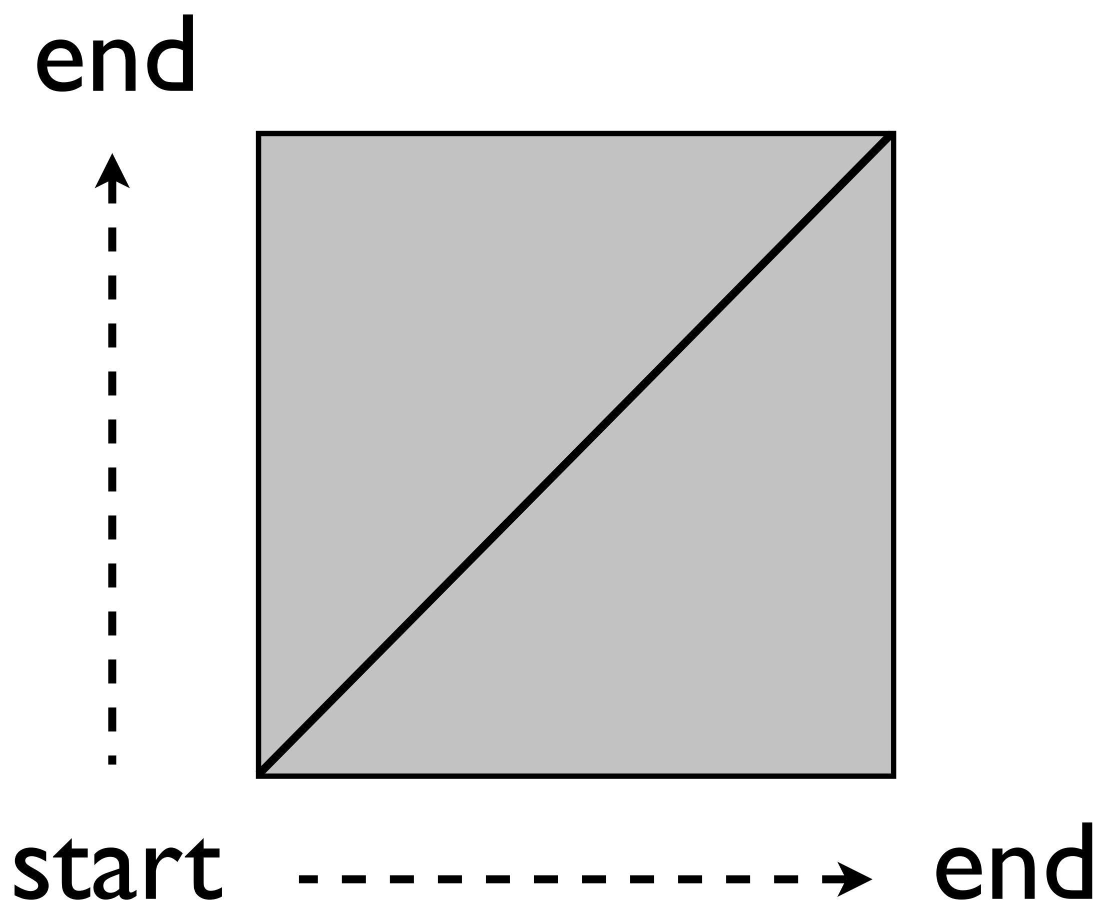
THE BEATLES: “BABY IT’S YOU”

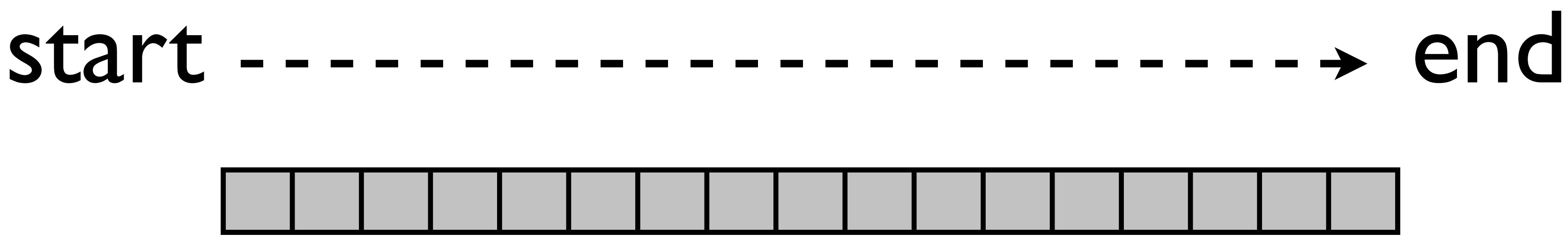
Technique 1:
Similarity Matrices

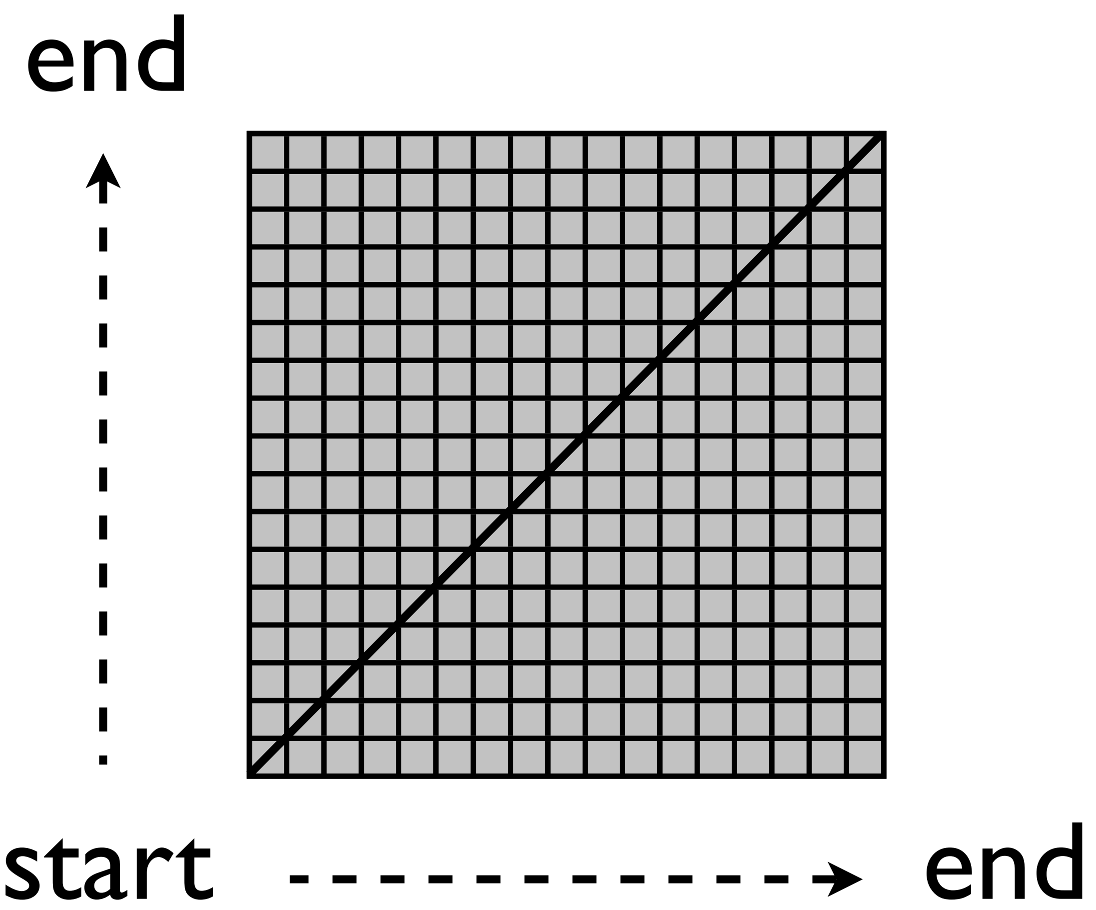

start
end

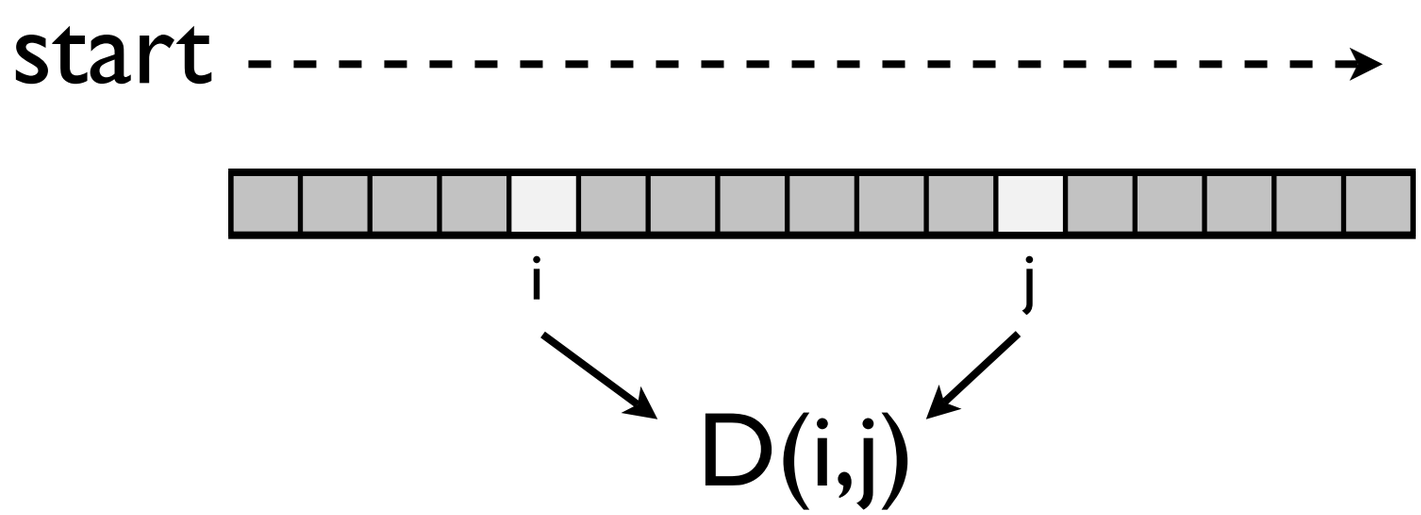

end
start
end

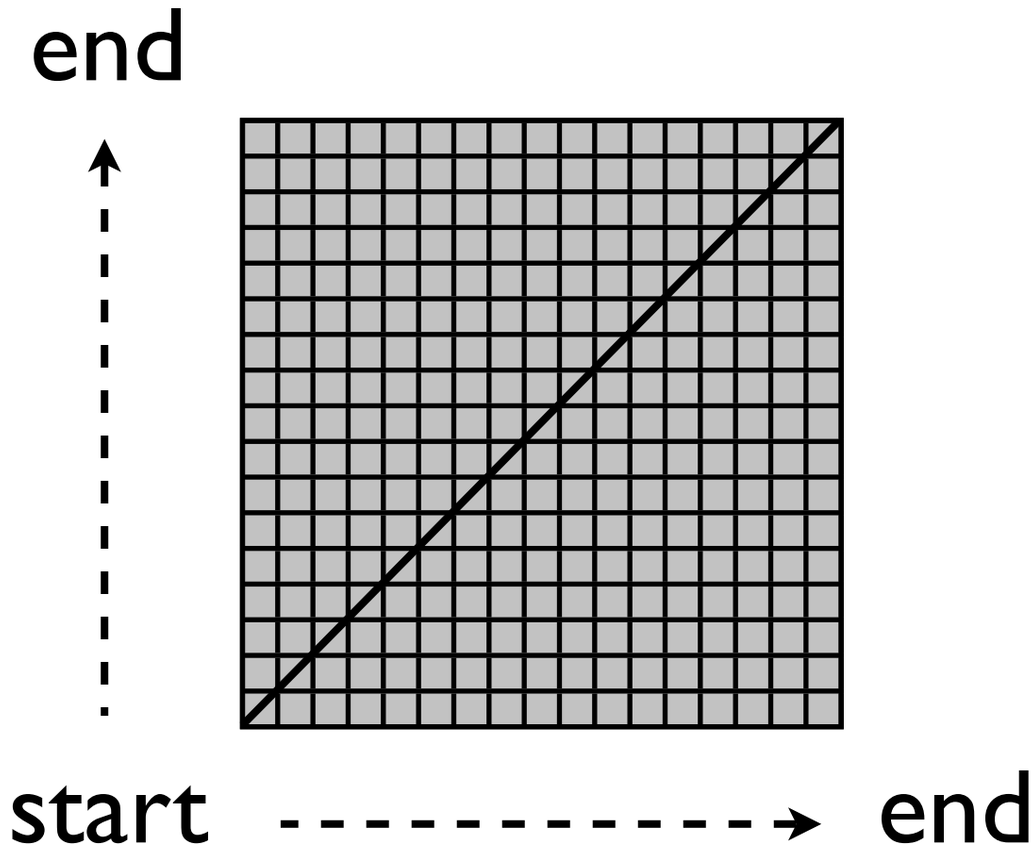

start
end

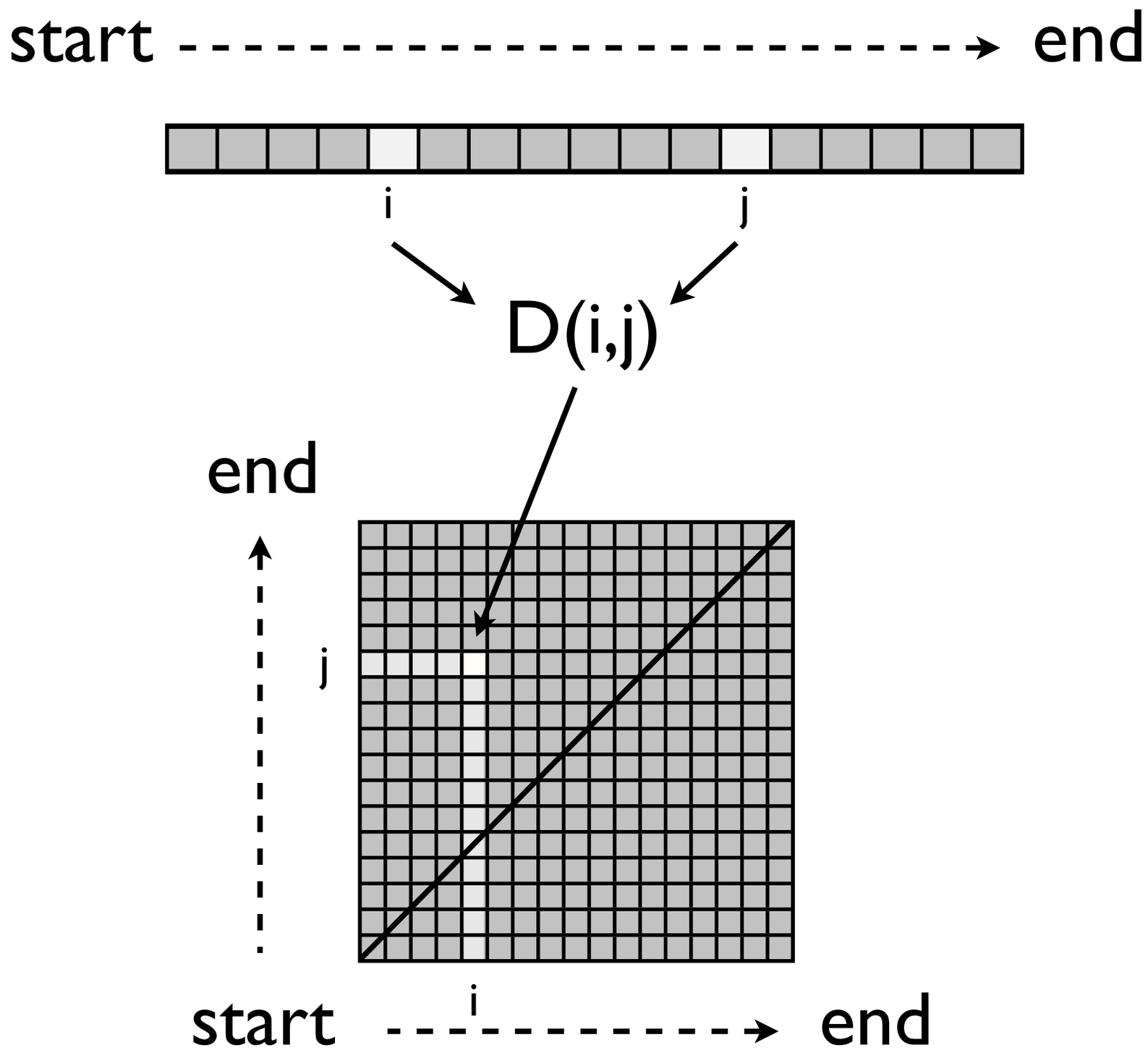

end
start
end

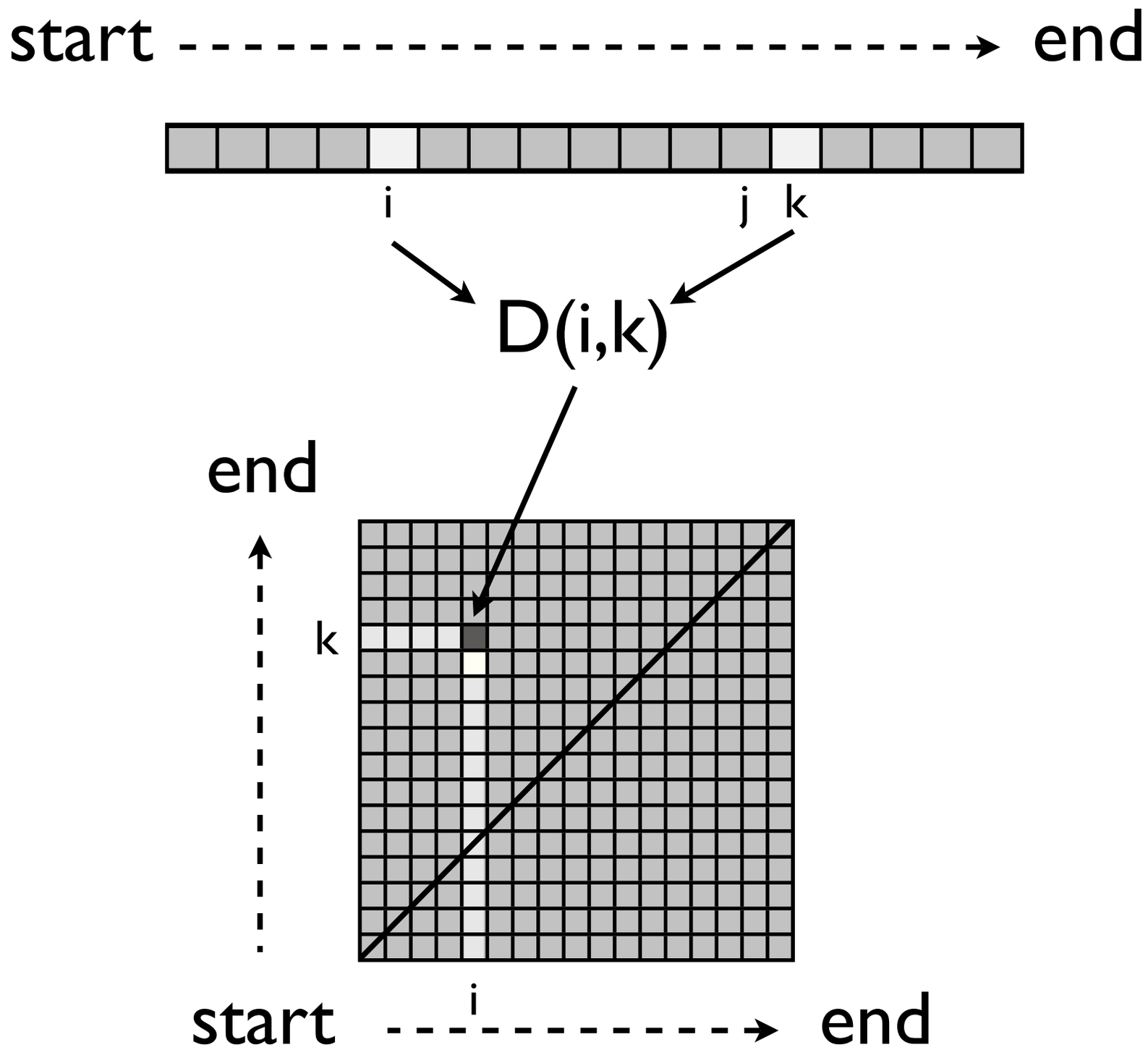

start
end


end
start
end


start
end
i
j
D(i,j)


end
start
end


start
end
i
j
D(i,j)
end
j
start
i
end


start
end
i
j
k
D(i,k)
end
k
i
start
end

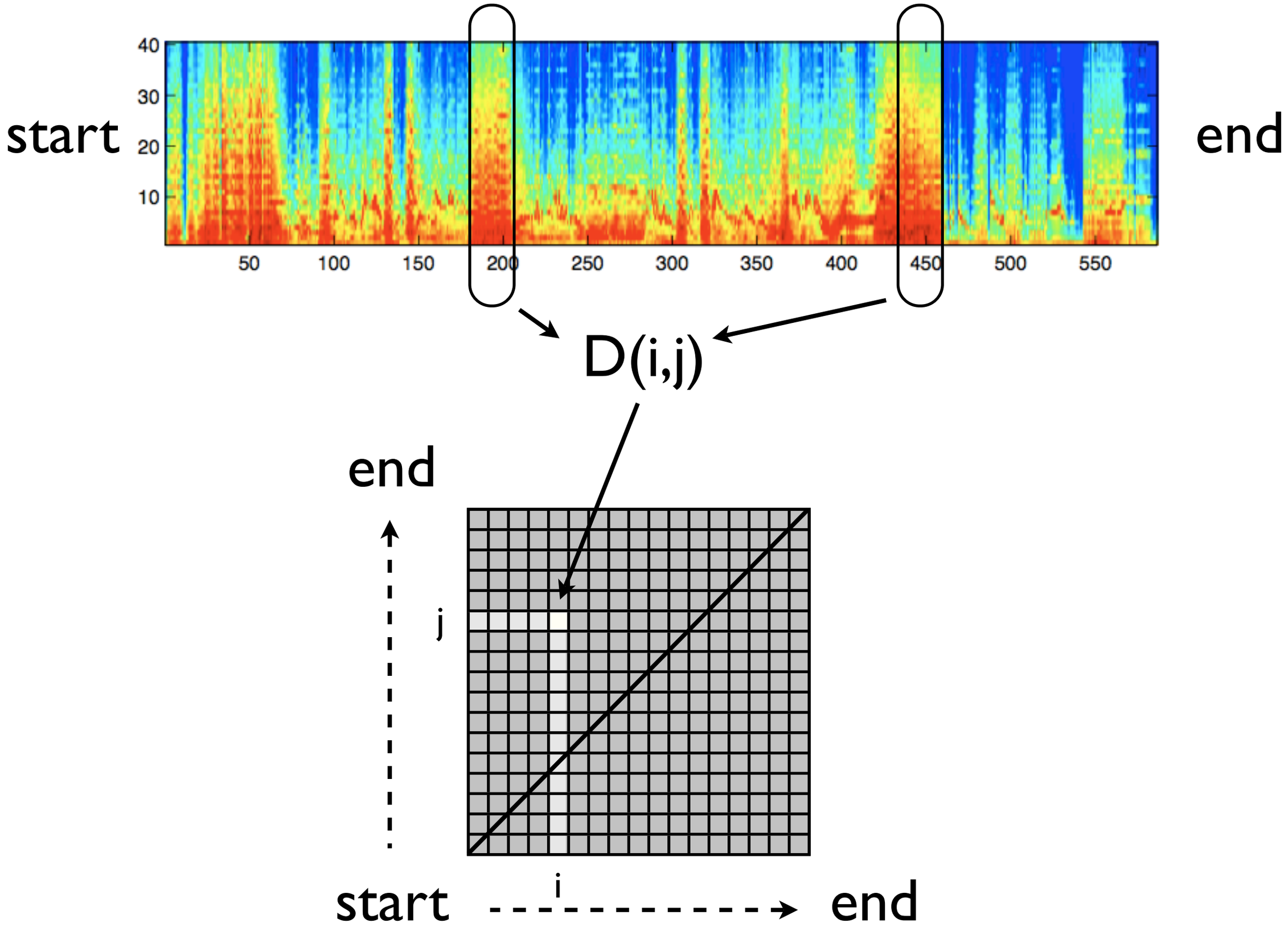


image: Foote 2000a

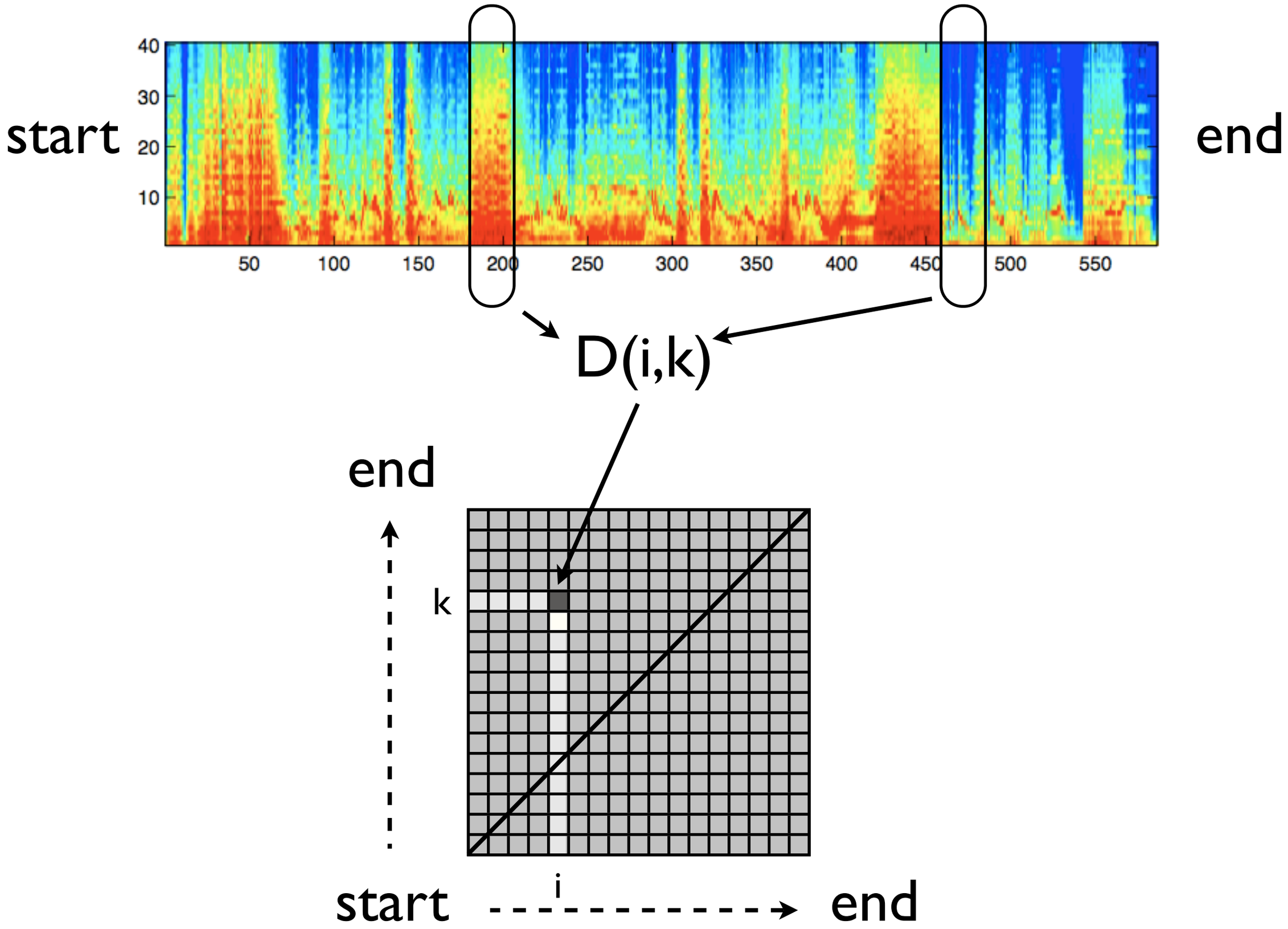


image: Foote 2000a

Similarity Matrices

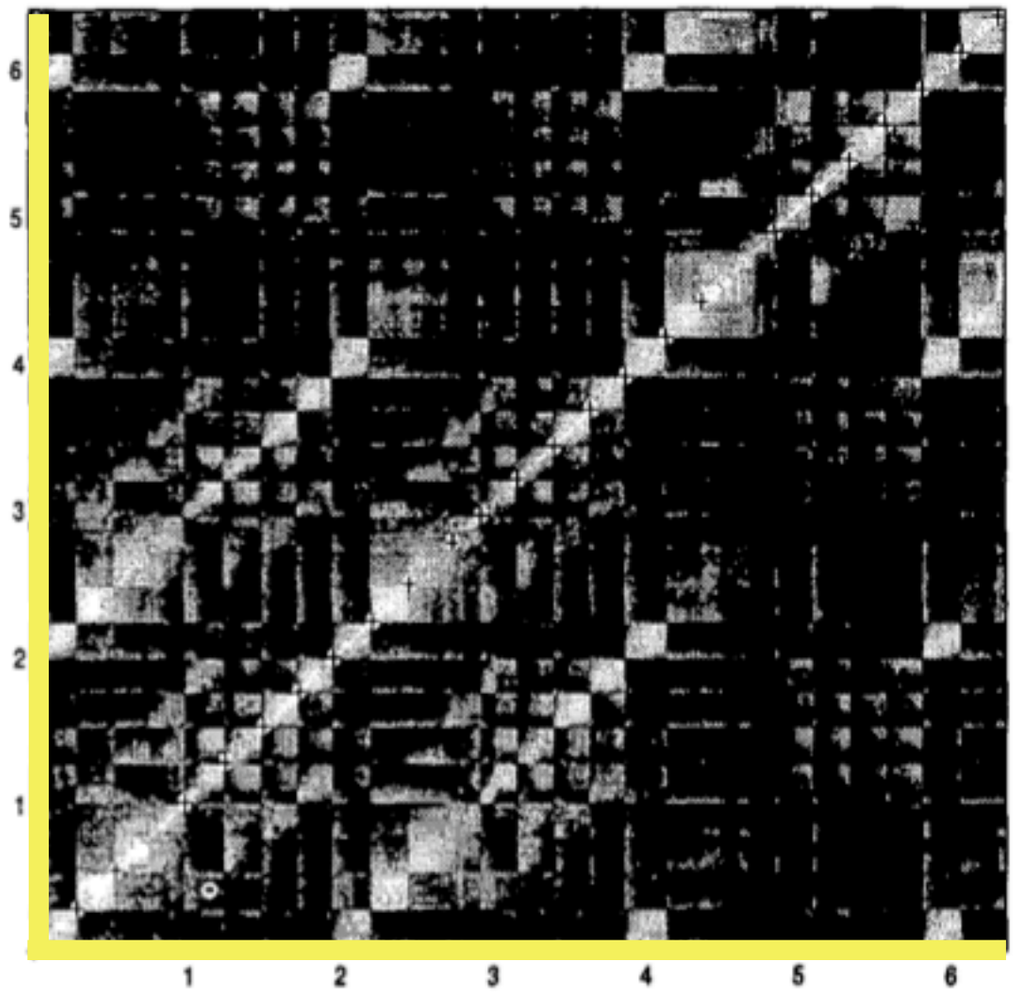


Figure 2. Gould performance showing note boundaries

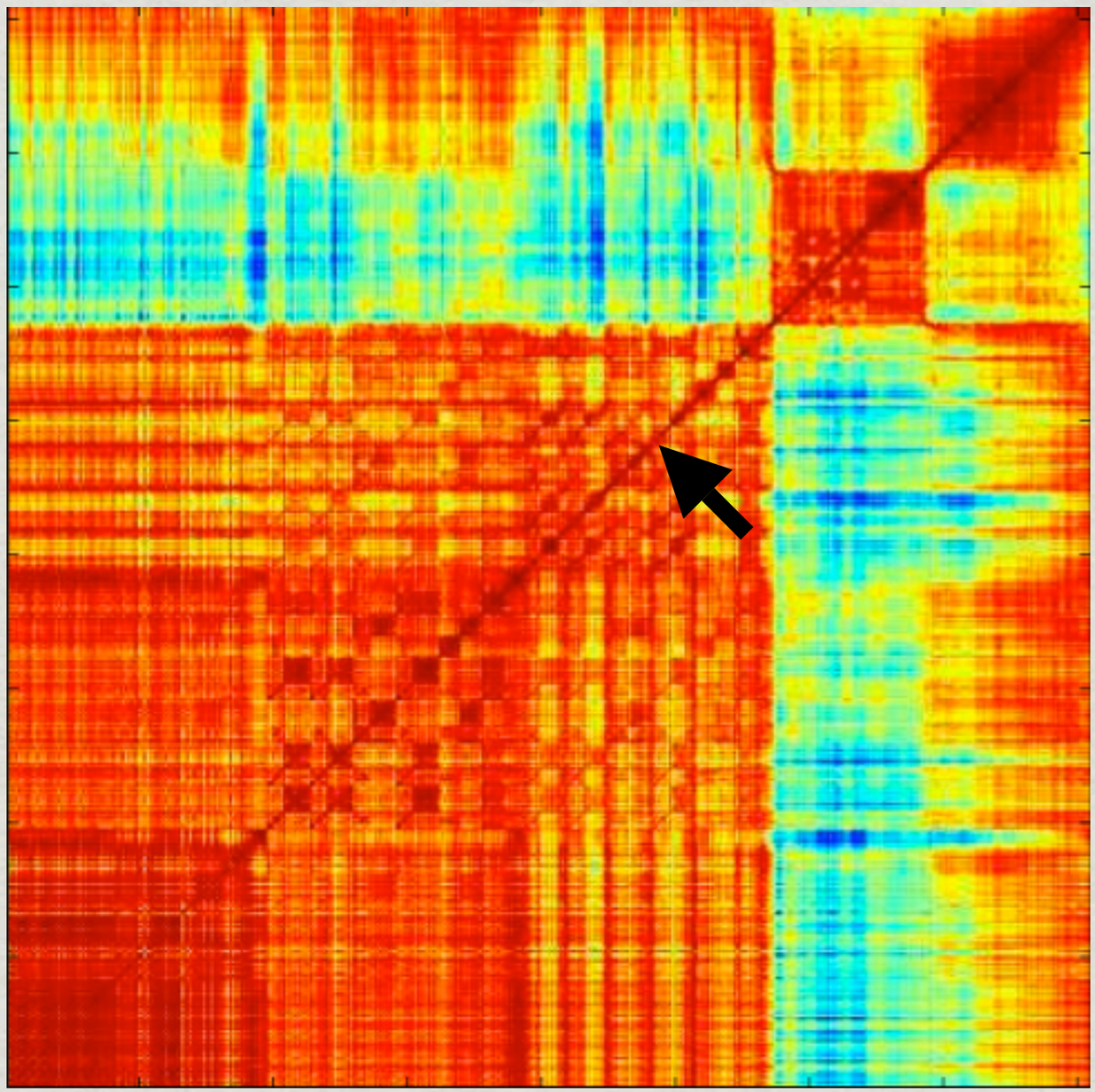
image: Foote 2000b

audio: http://youtube.com/watch?v=bOTW2VS4w5w

Similarity matrices

- They can show us stuff:

  Points of novelty



THE BEATLES: “FLYING”

Novelty detection

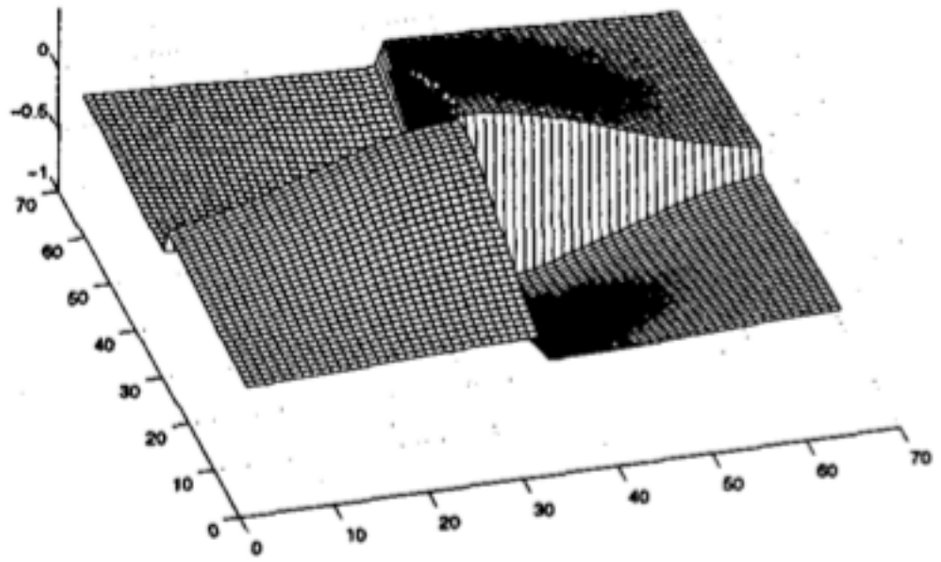
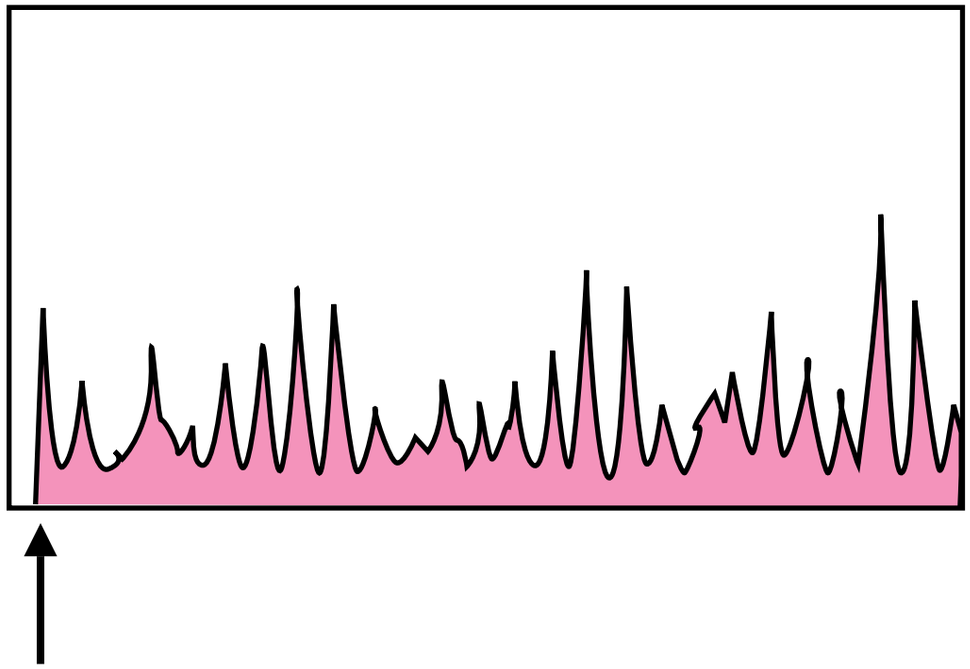


Figure 3. 64 x 64 checkerboard kernel with Gaussian taper

image: Foote 2000b

Novelty detection



Novelty scores

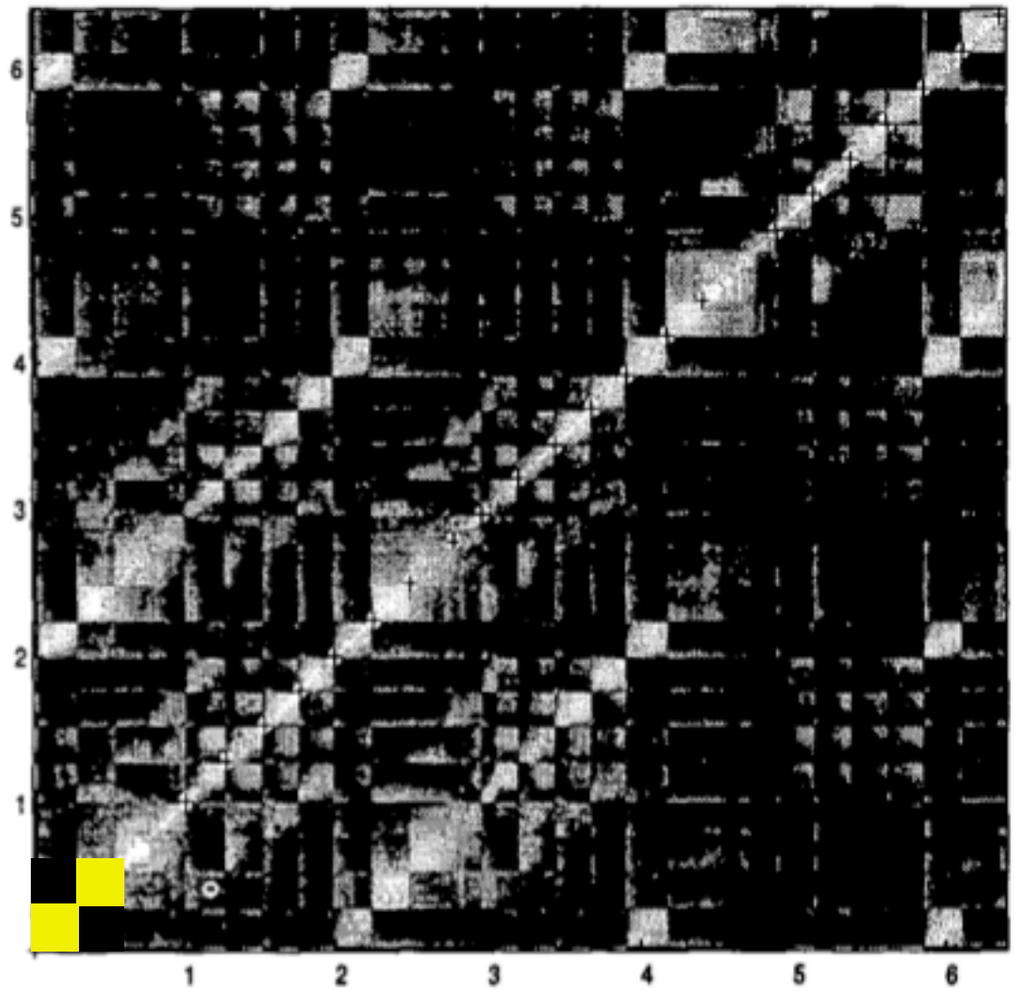
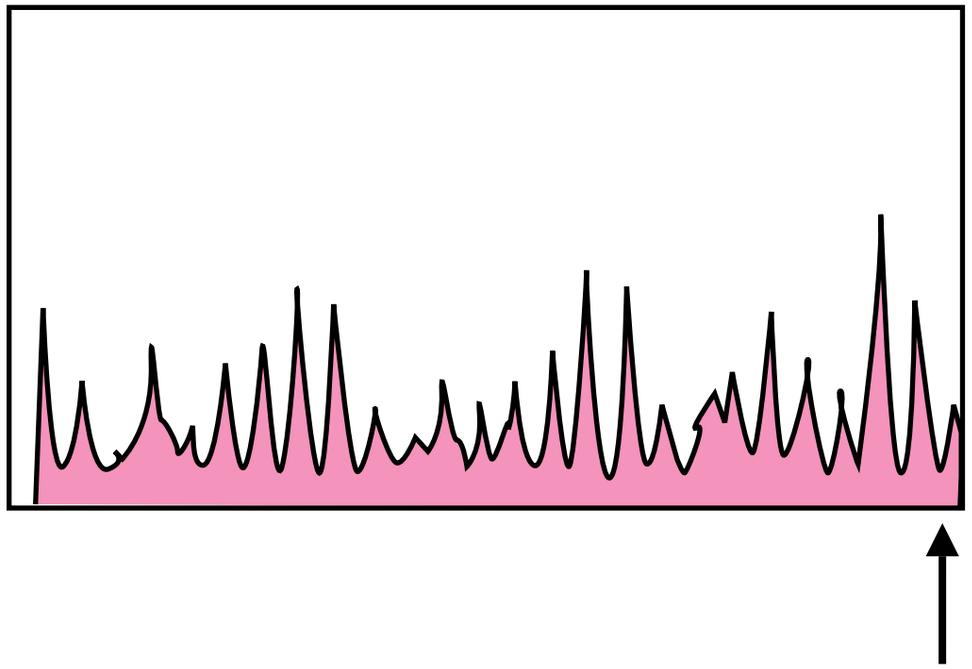


Figure 2. Gould performance showing note boundaries

images: Foote 2000b

Novelty detection



Novelty scores

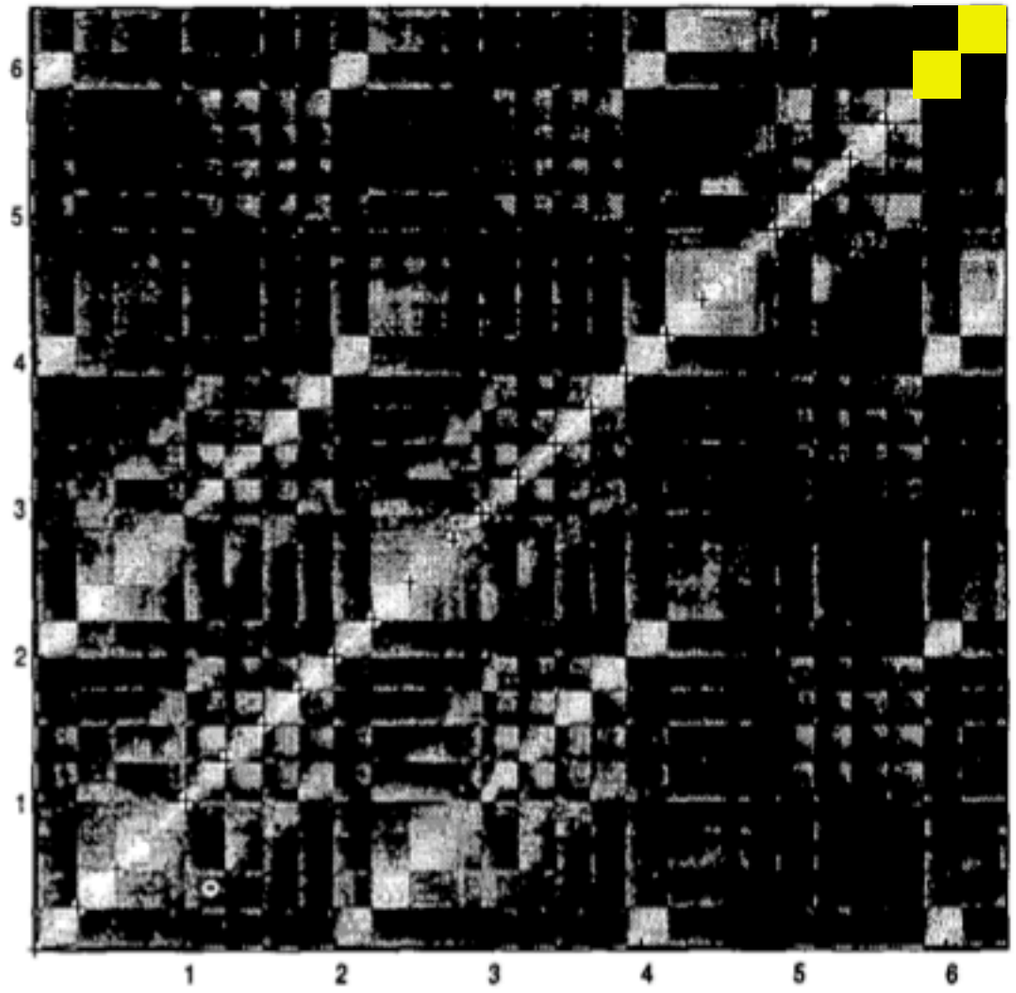
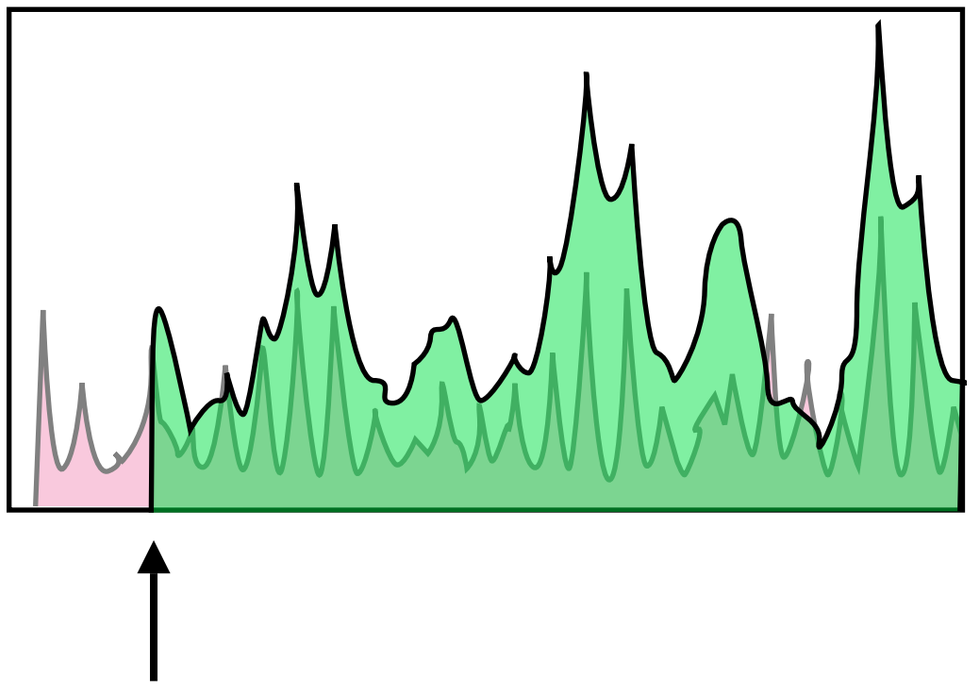


Figure 2. Gould performance showing note boundaries

images: Foote 2000b

Novelty detection



Novelty scores

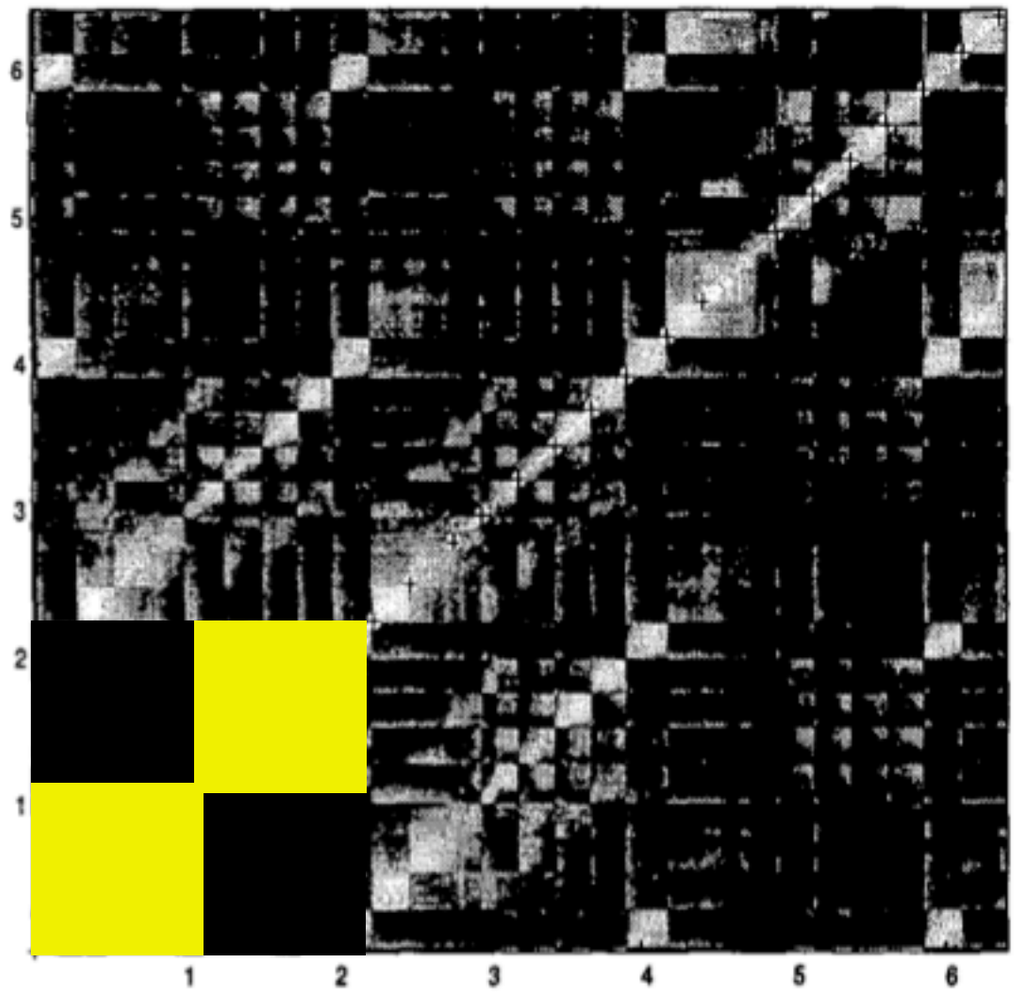
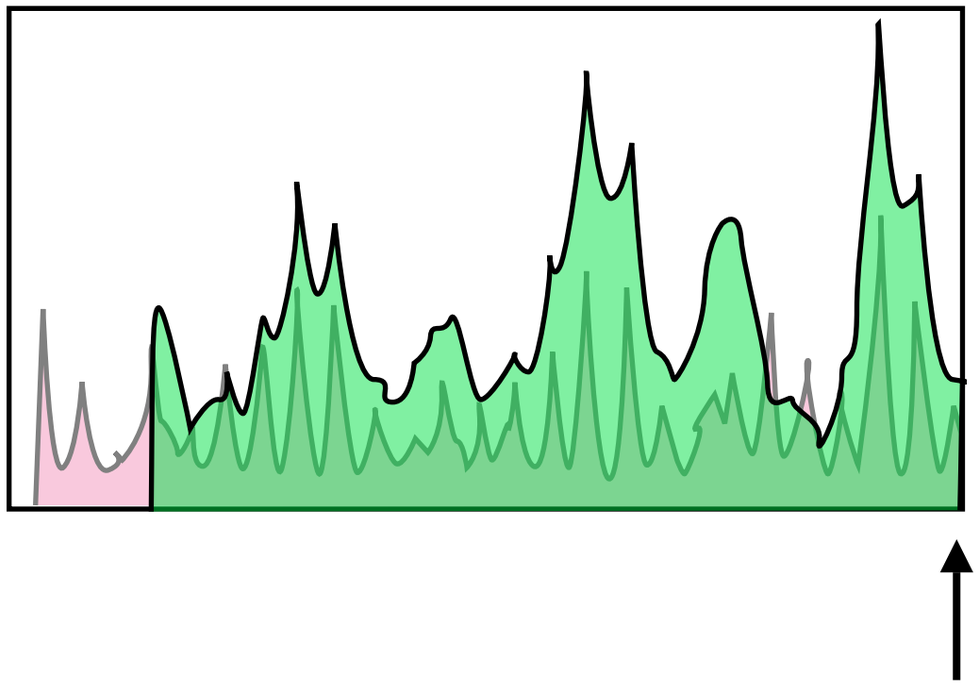


Figure 2. Gould performance showing note boundaries

images: Foote 2000b

Novelty detection



Novelty scores

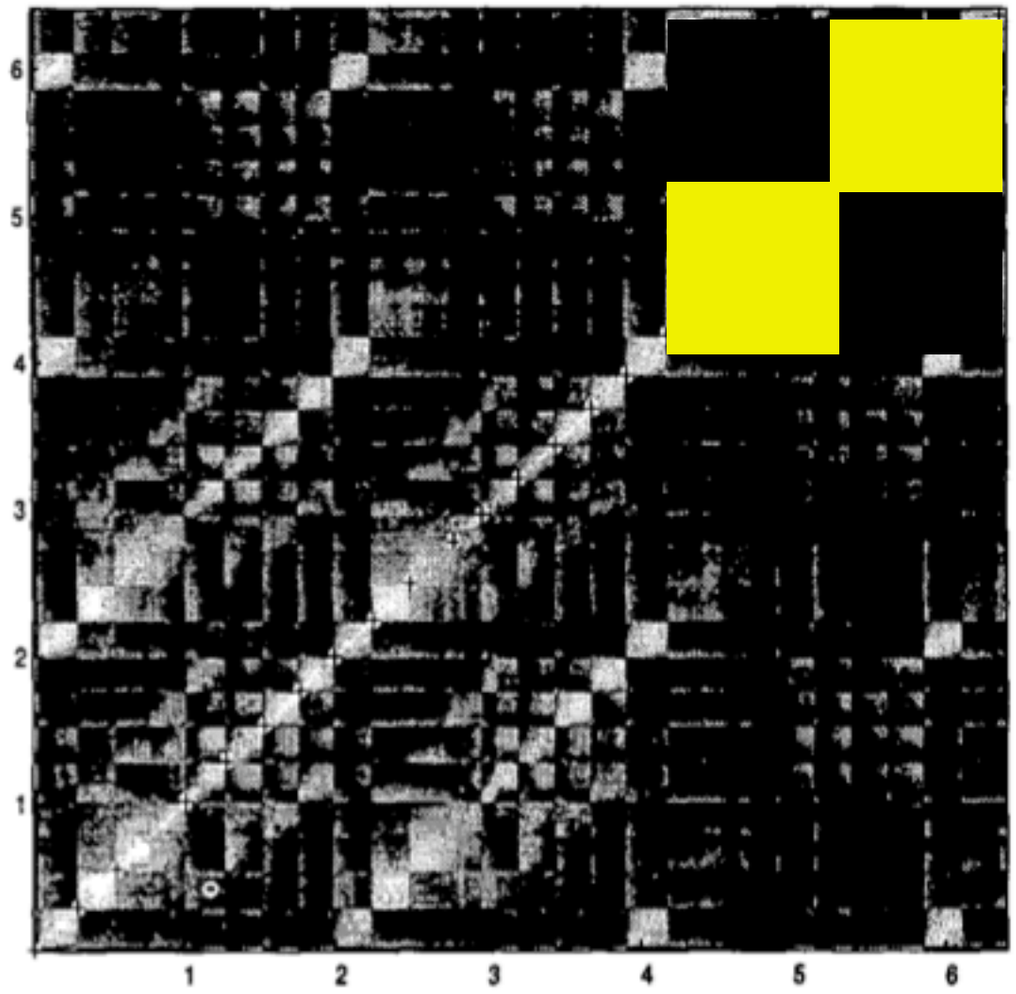
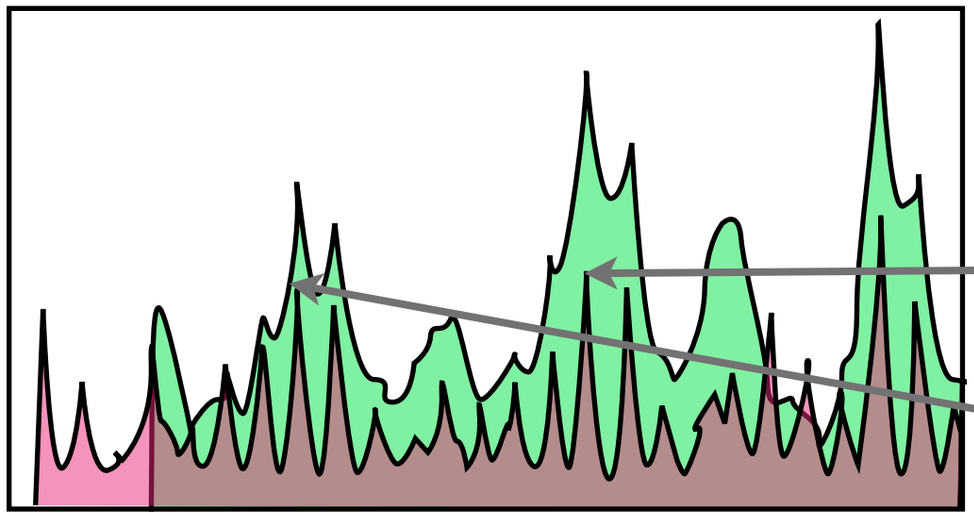


Figure 2. Gould performance showing note boundaries

images: Foote 2000b

Novelty detection



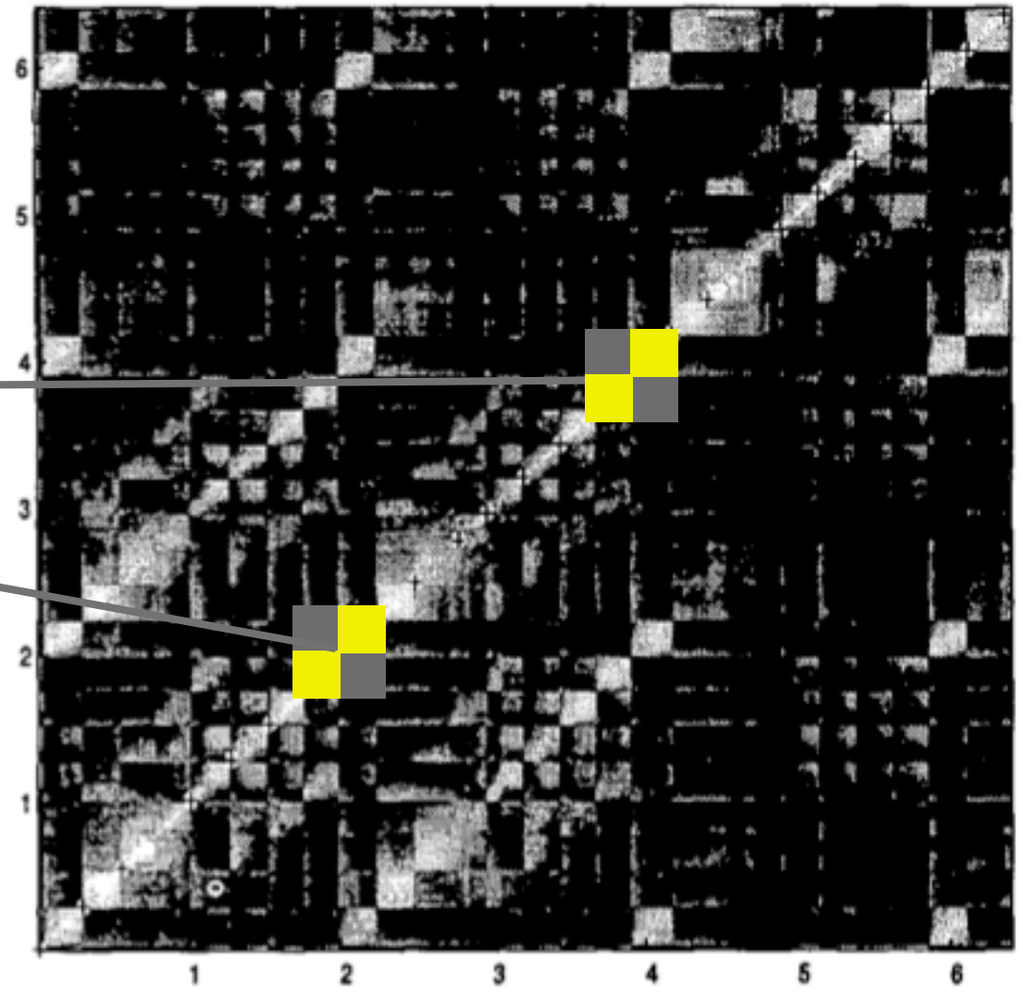
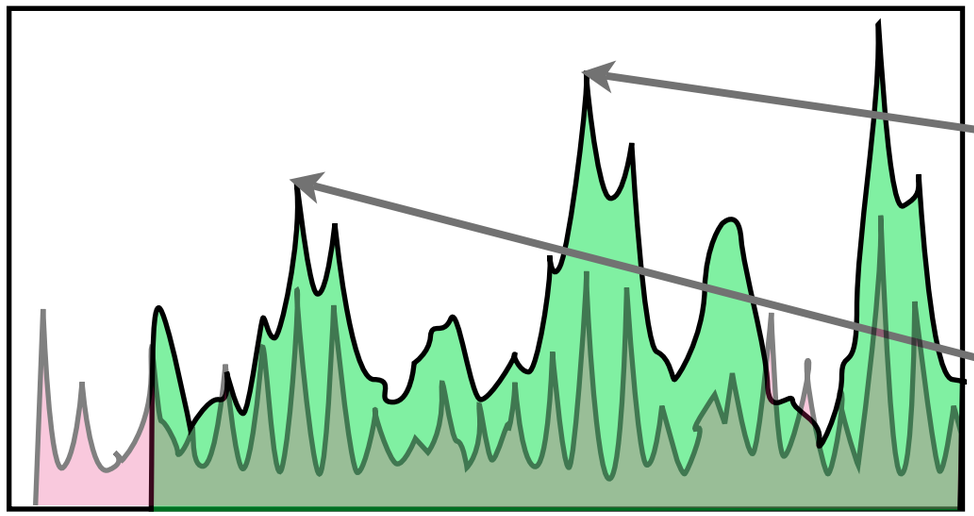


Figure 2. Gould performance showing note boundaries

Novelty scores

images: Foote 2000b

Novelty detection



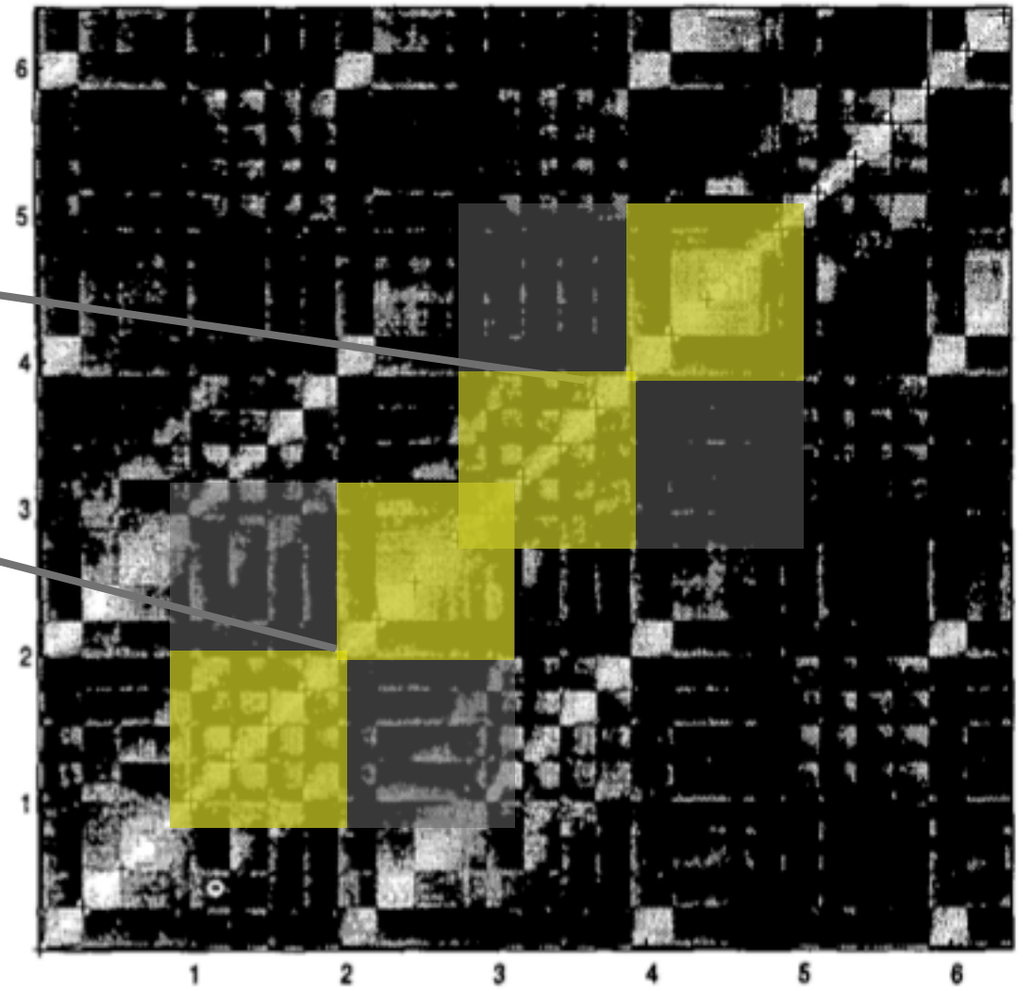
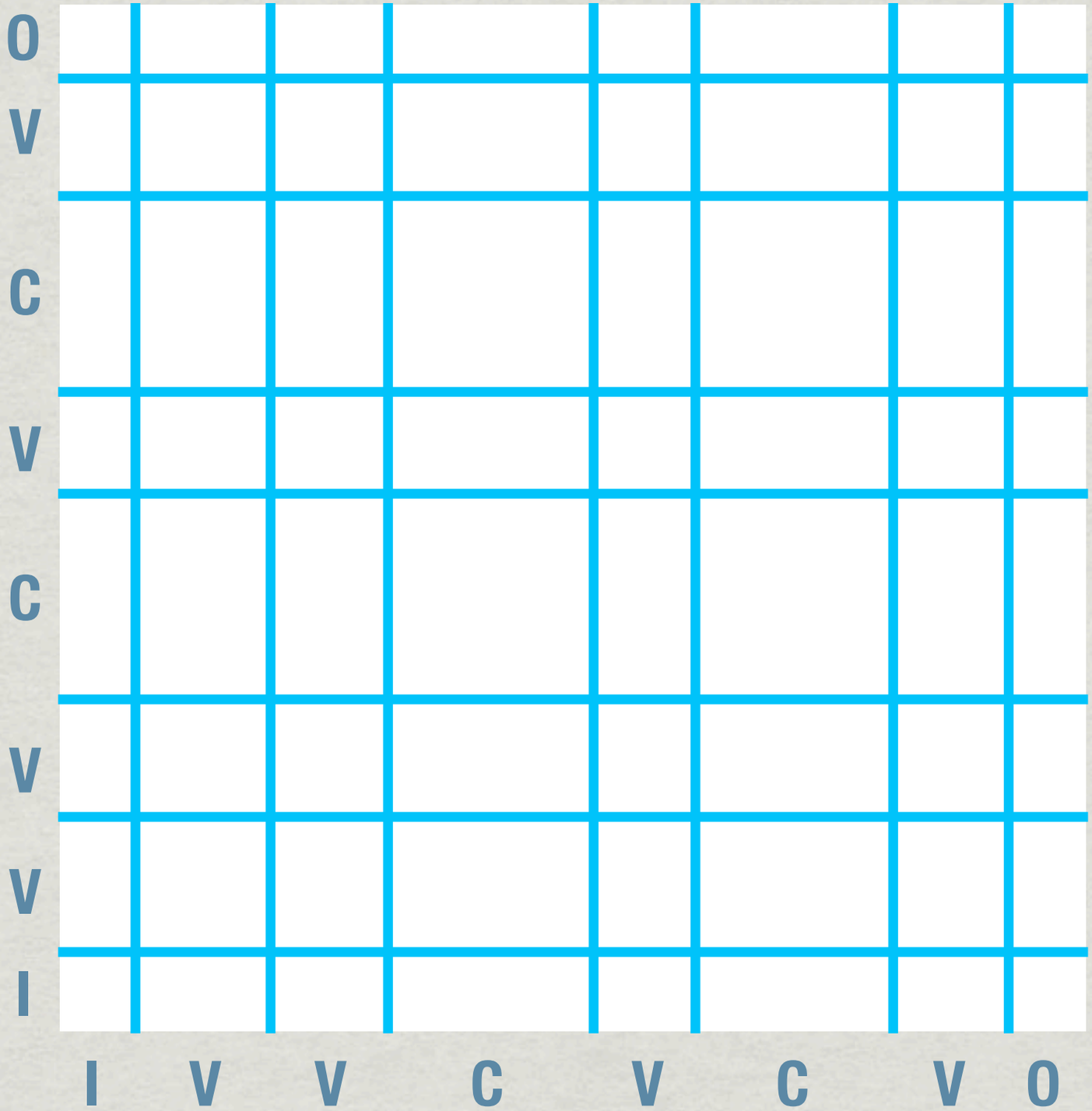


Figure 2. Gould performance showing note boundaries

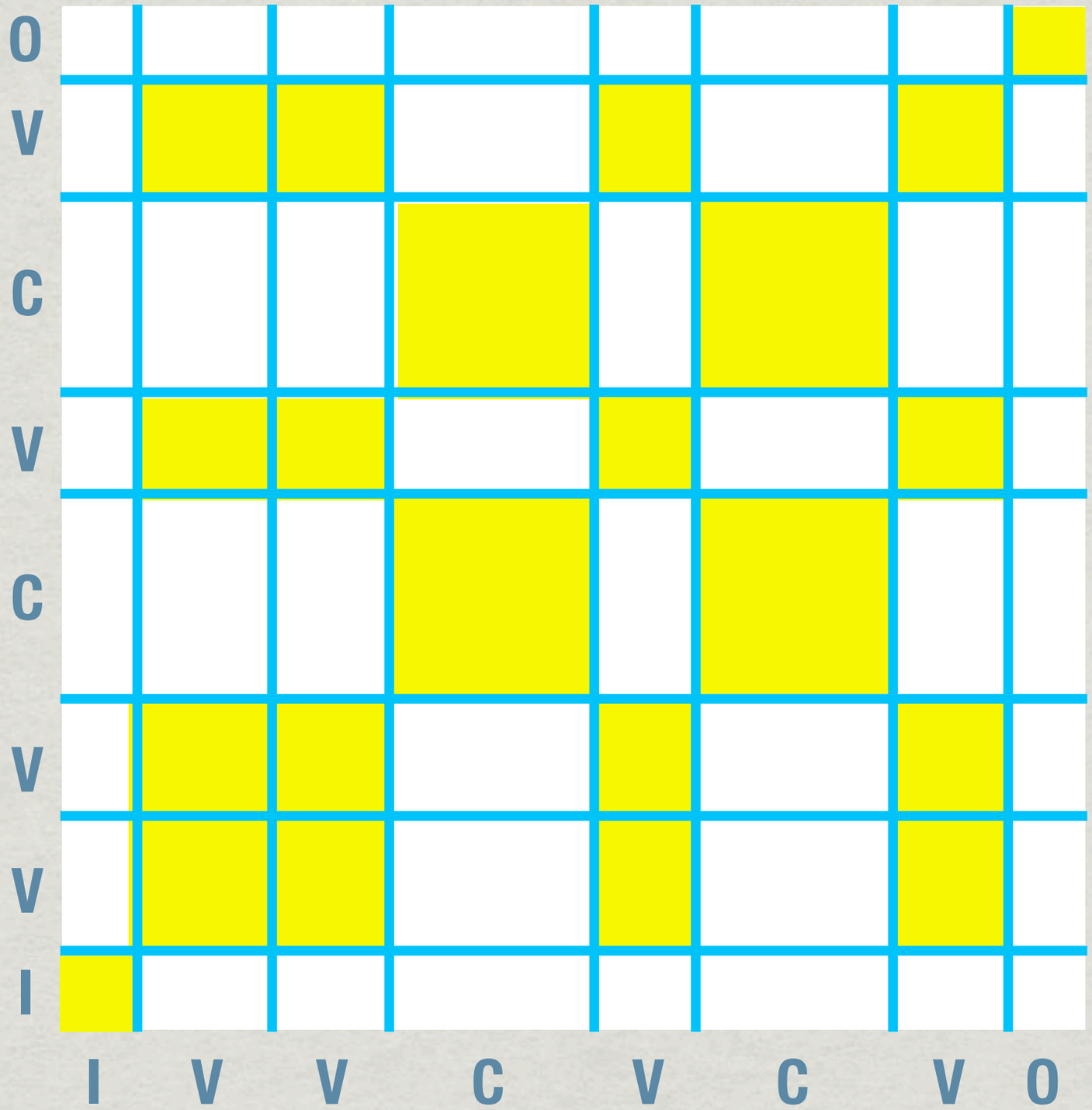
Novelty scores

images: Foote 2000b

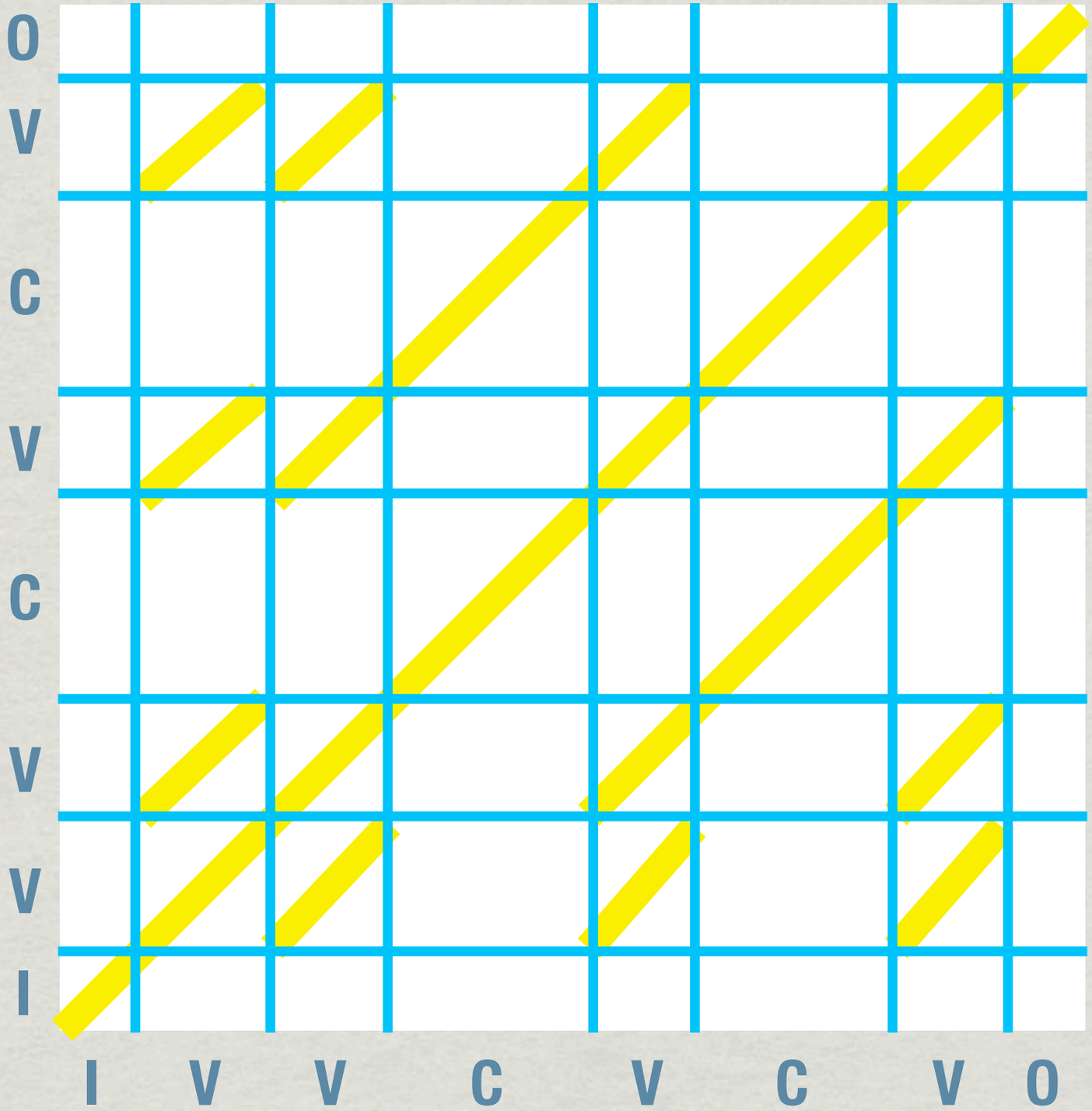
STATES OR SEQUENCES?



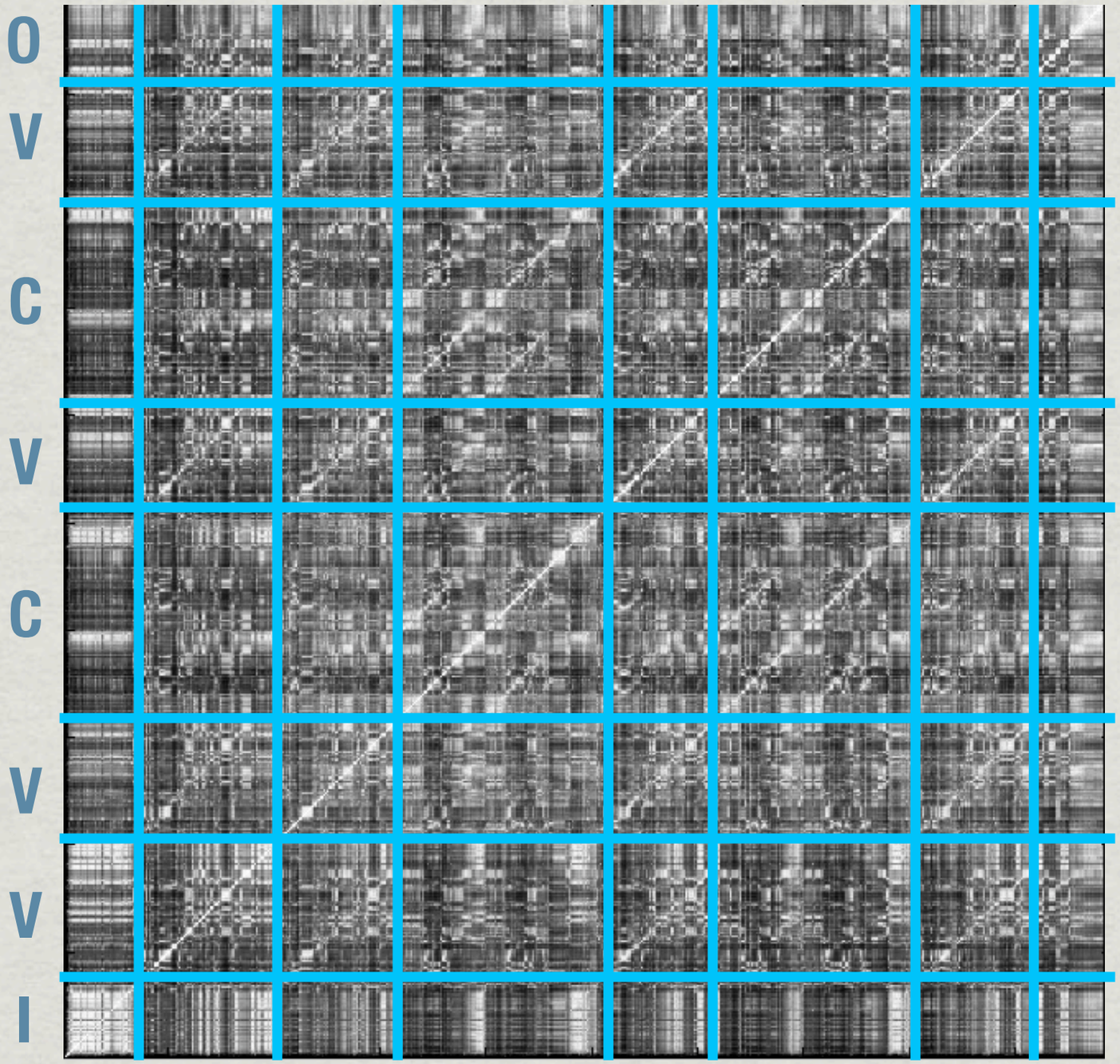
STATES VIEW



SEQUENCES VIEW



THE BEATLES: “YESTERDAY”



SEQUENCE WORKFLOW

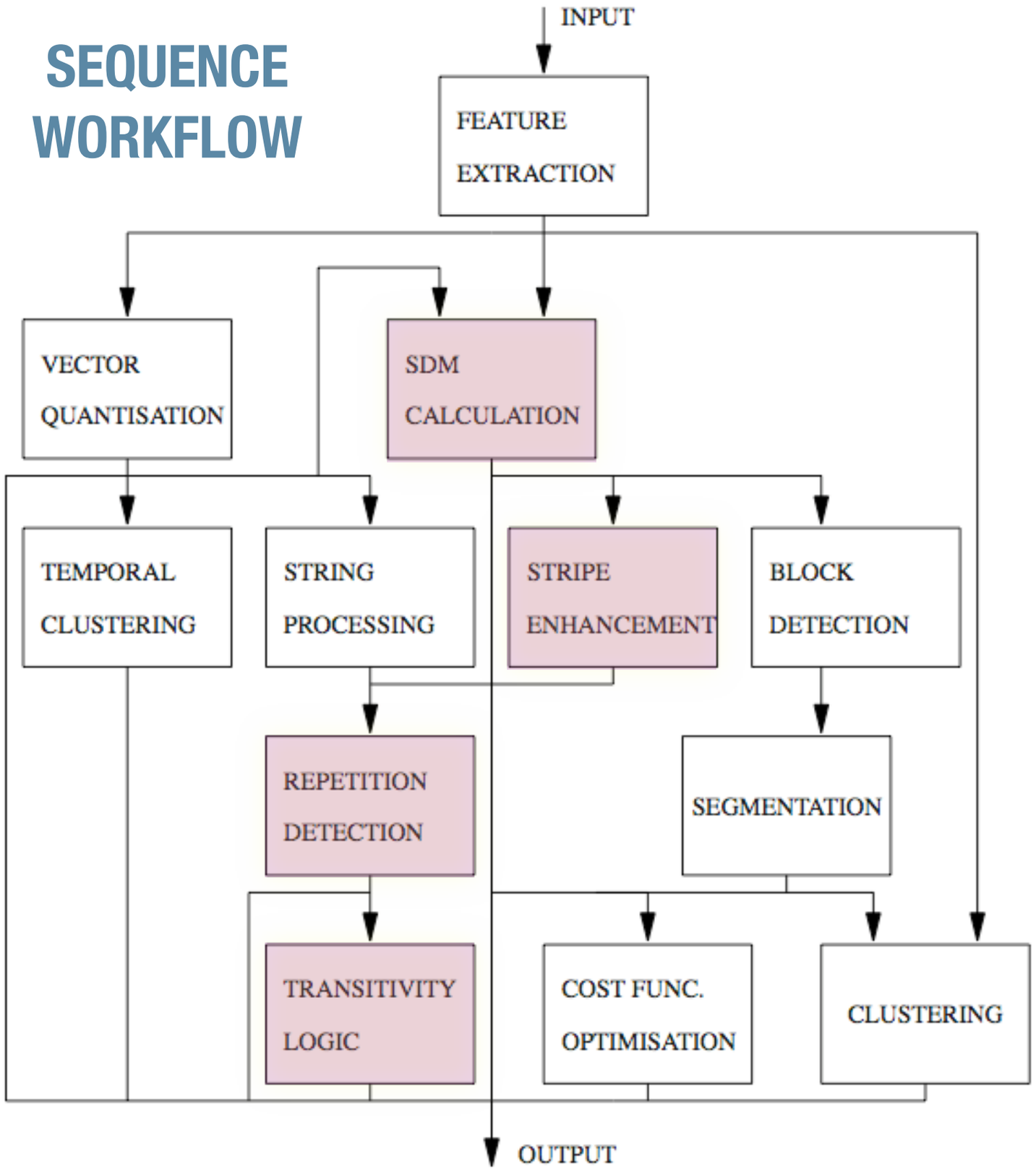


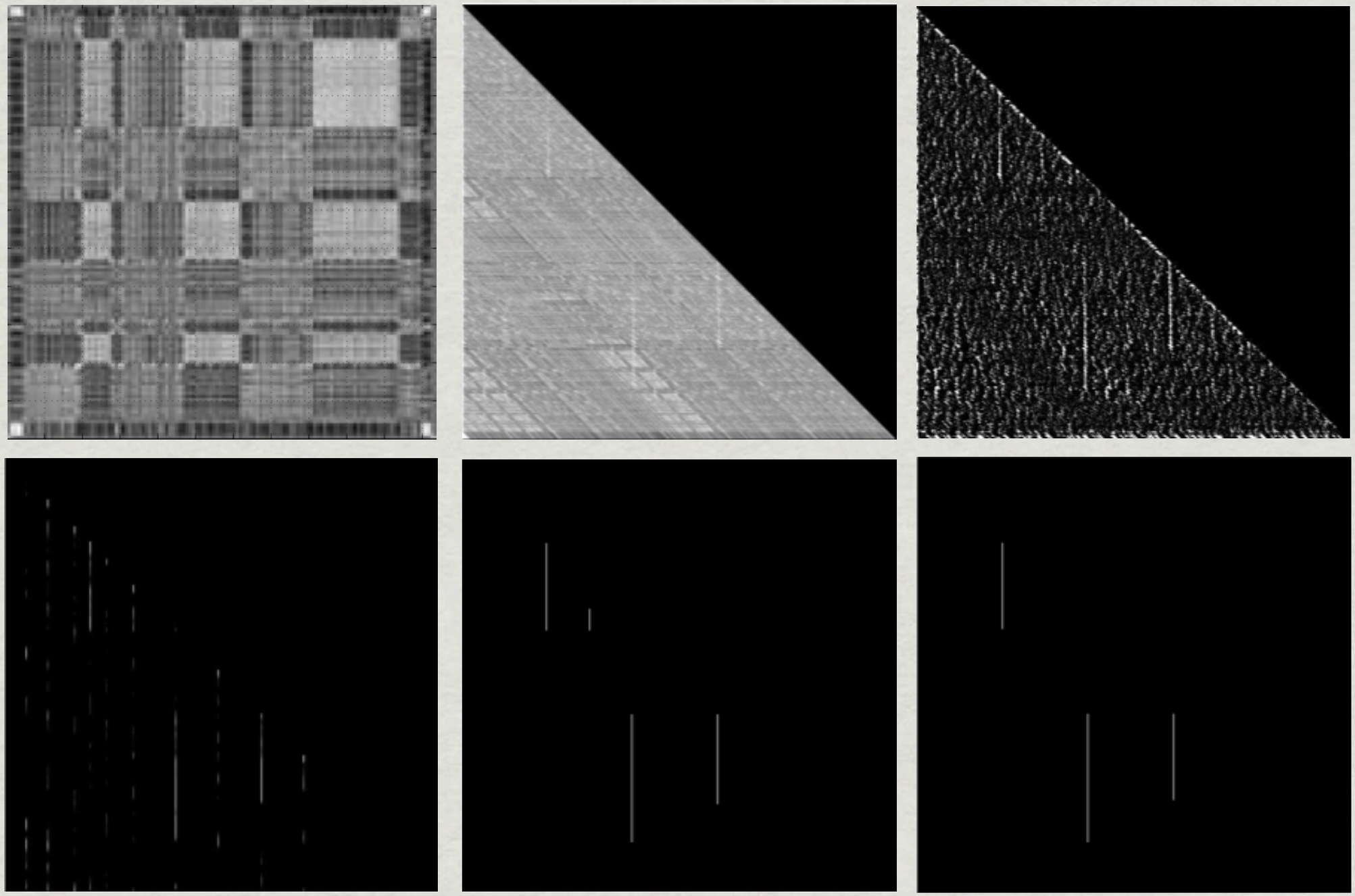
image: Paulus 2009

STRIPE SEARCH

TIME-LAG

FILTER

THRESHOLD



ERODE / DILATE

GROUND TRUTH

STATE WORKFLOW

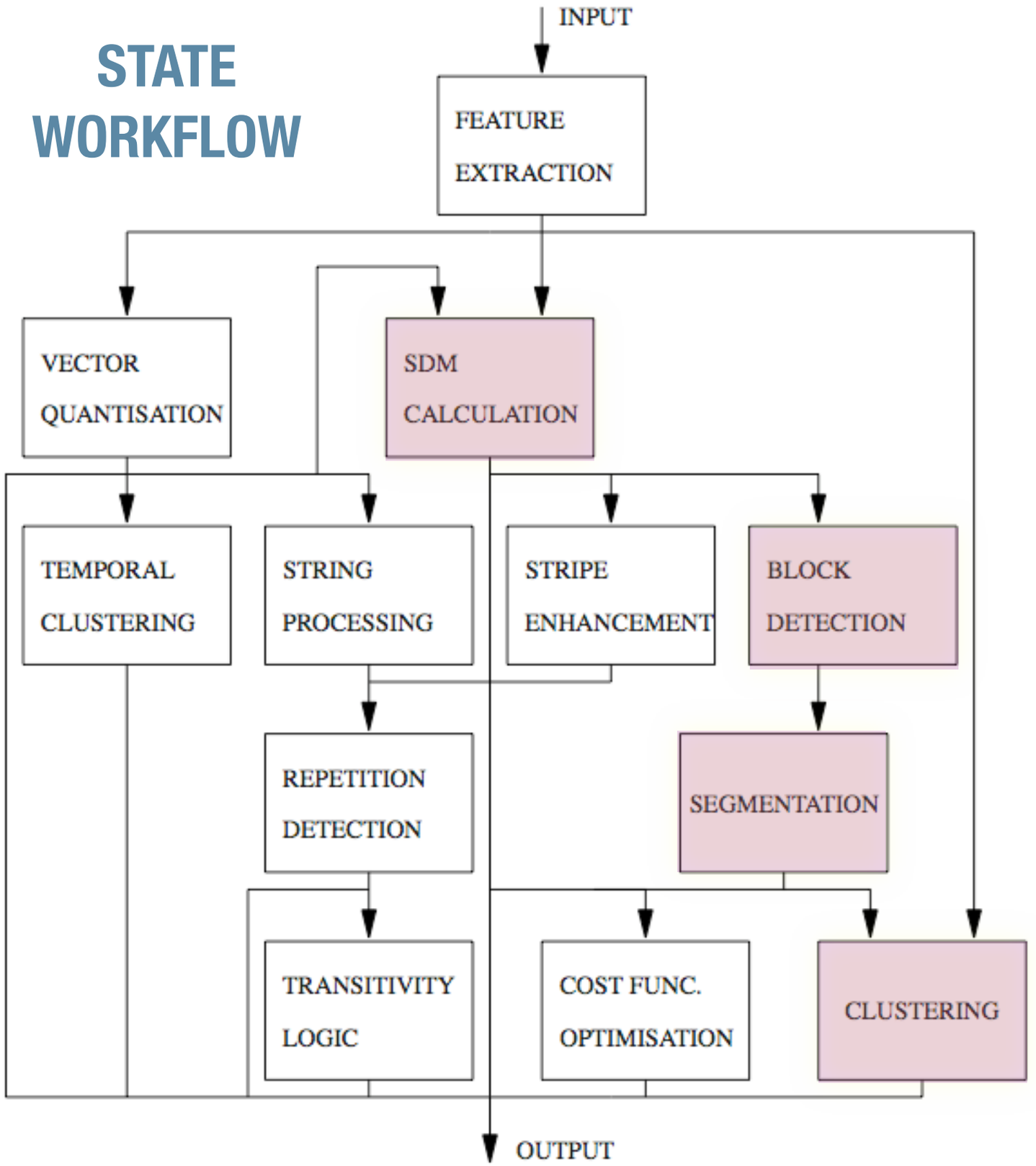
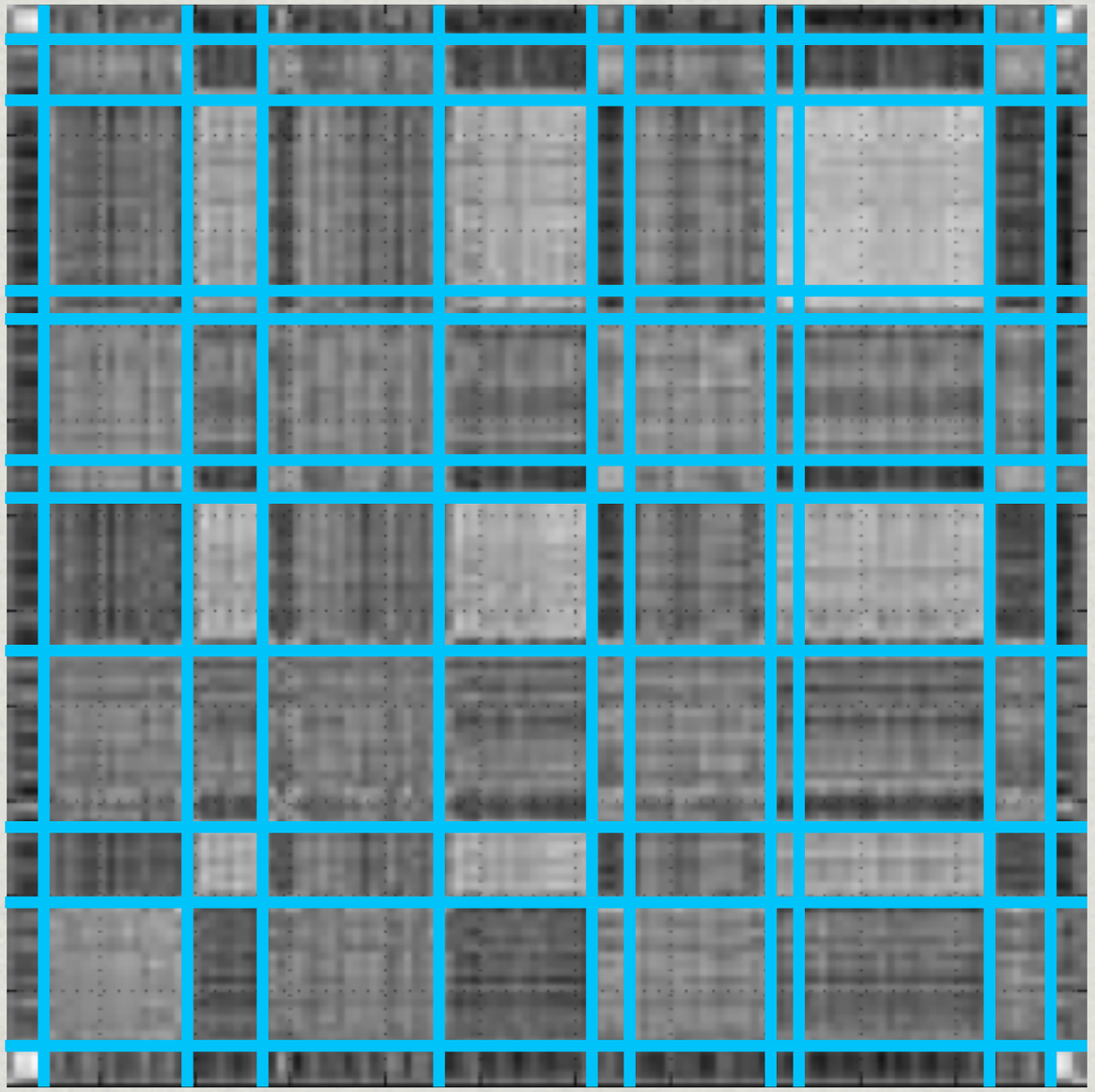
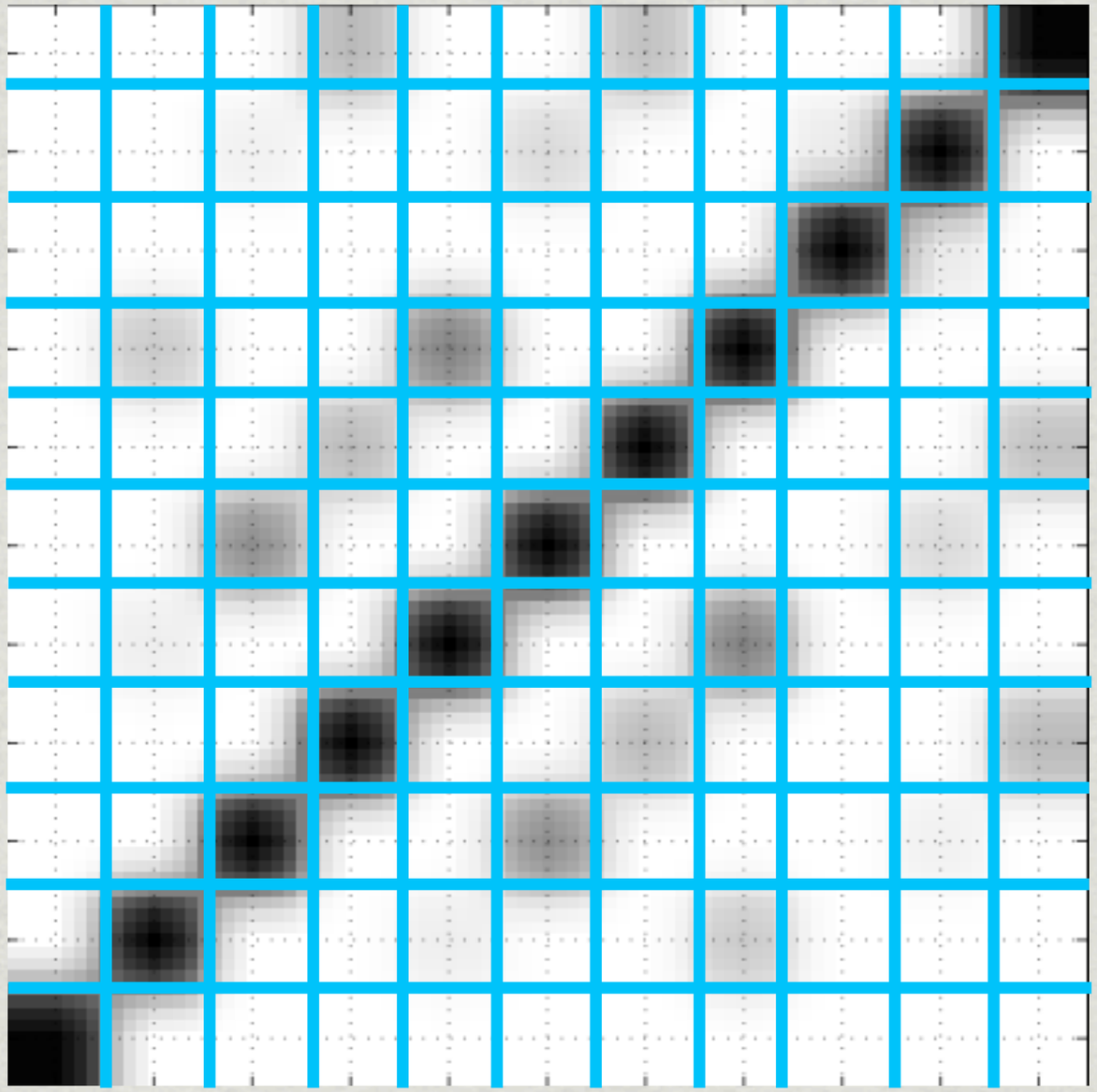


image: Paulus 2009

BLOCK SEARCH



BLOCK SEARCH



Outline

1. Two hypotheses:

- ✓ States
- ✓ Sequences

2. A word on features:

- ✓ Timbre, harmony, etc.
- Dynamic features

3. Two techniques:

- ✓ Similarity matrix
- Clustering models

Technique 2:
Clustering Models

Clustering

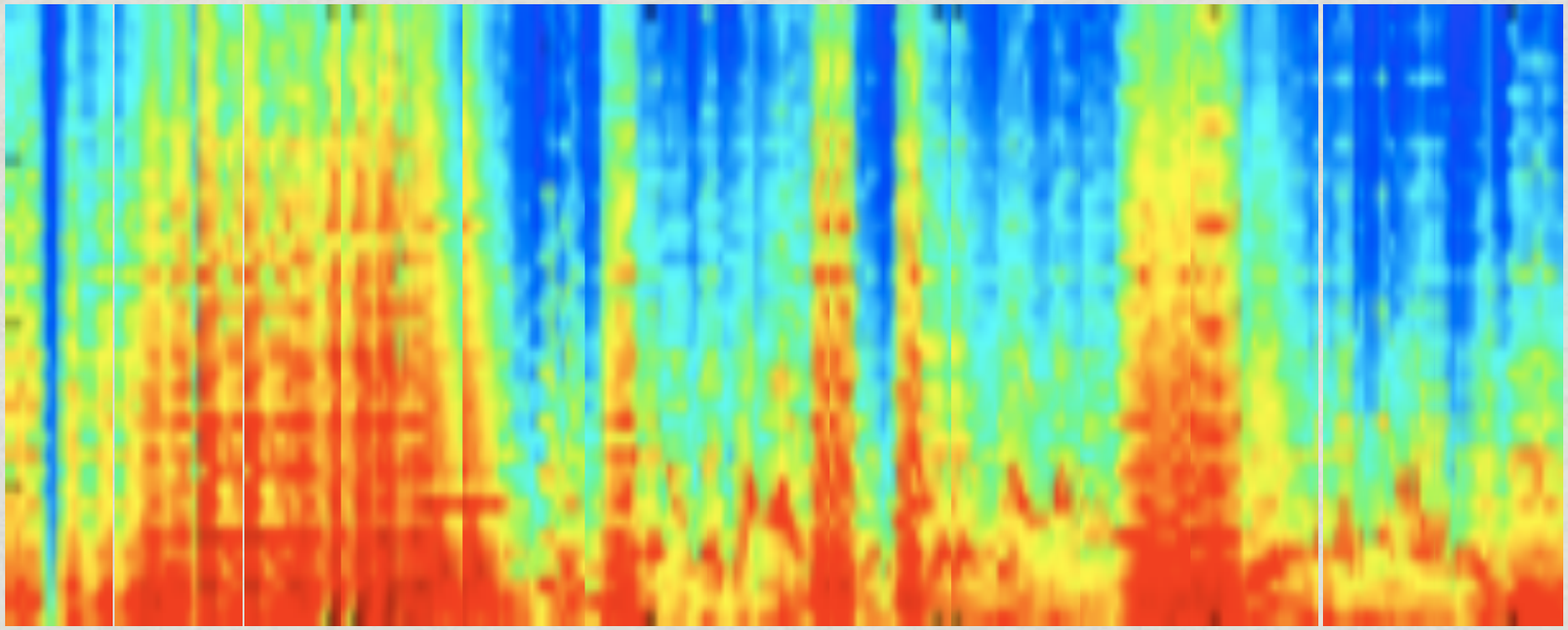


image: Foote 2000a

Clustering

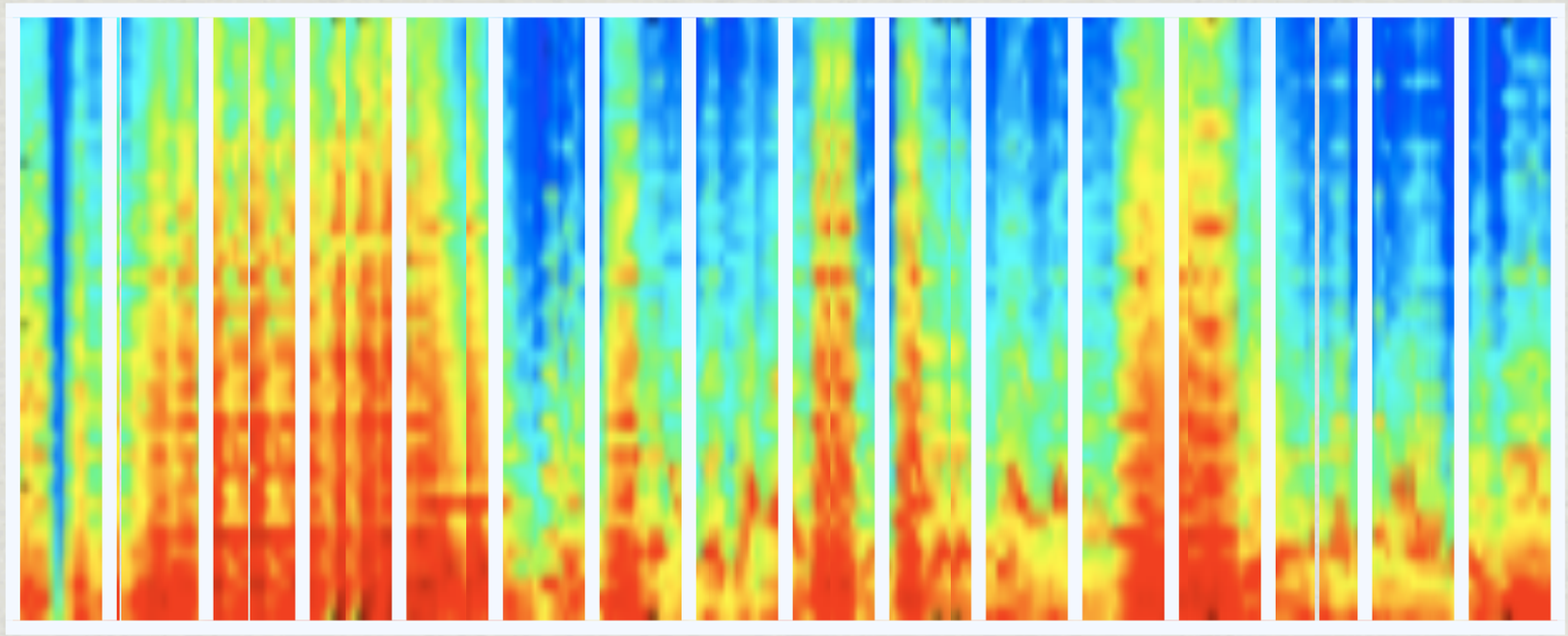
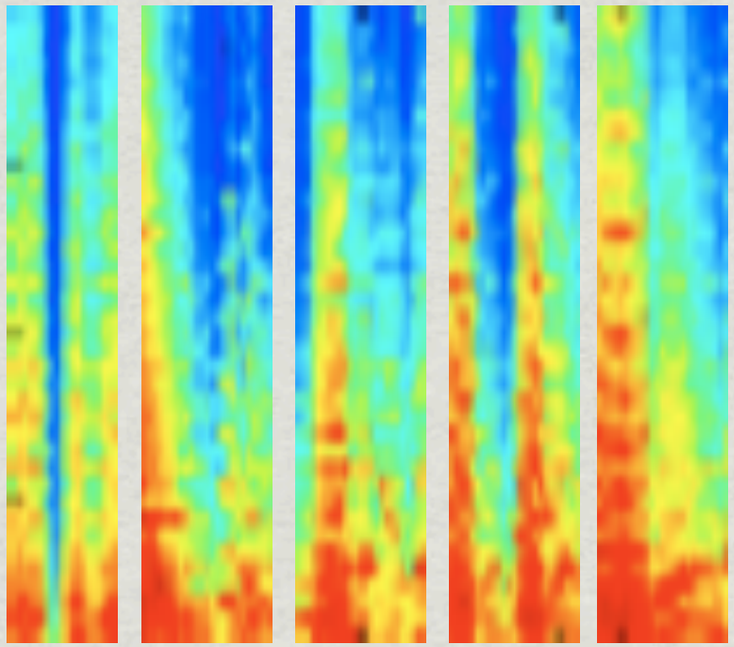
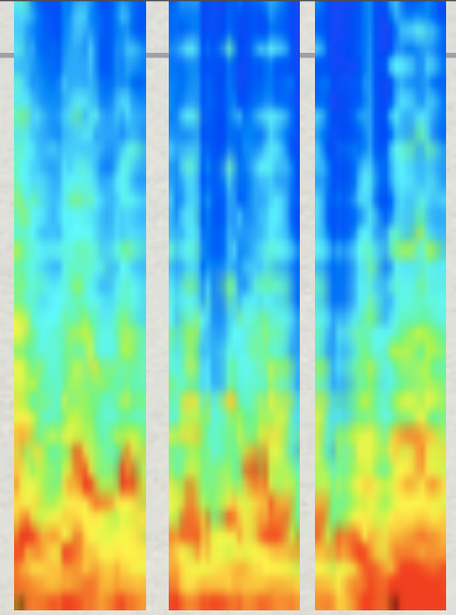
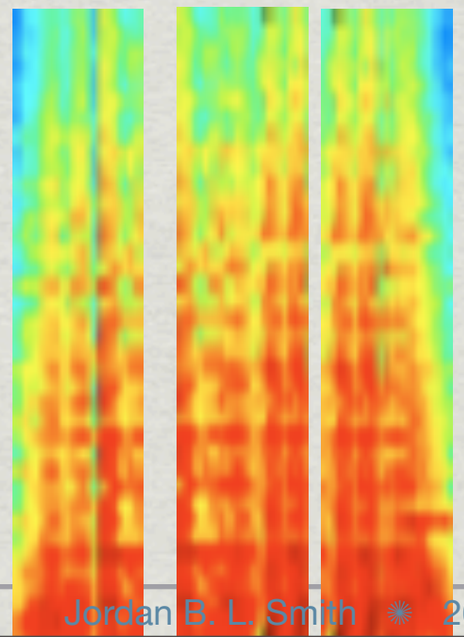


image: Foote 2000a

Clustering







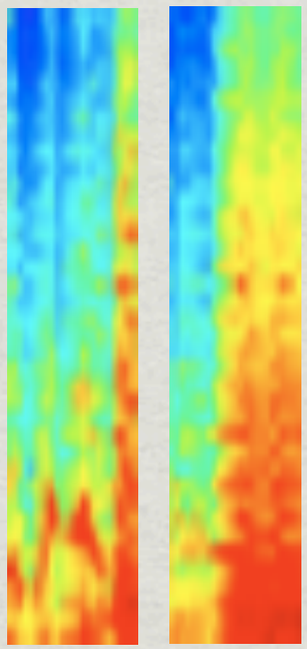
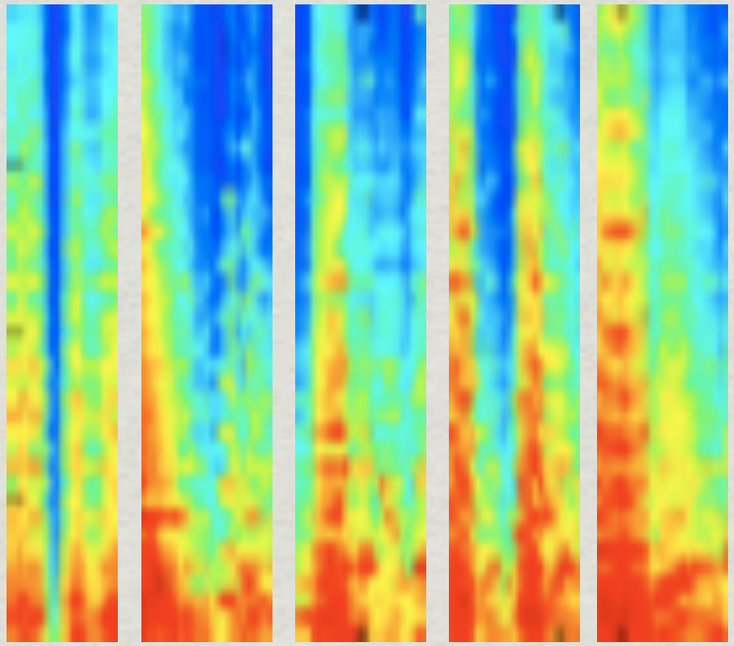
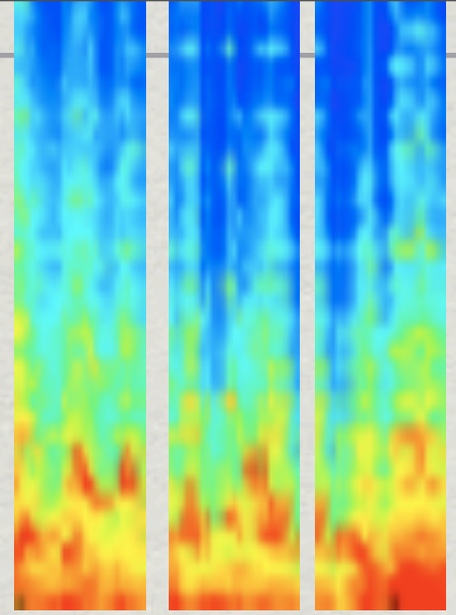


image: Foote 2000a

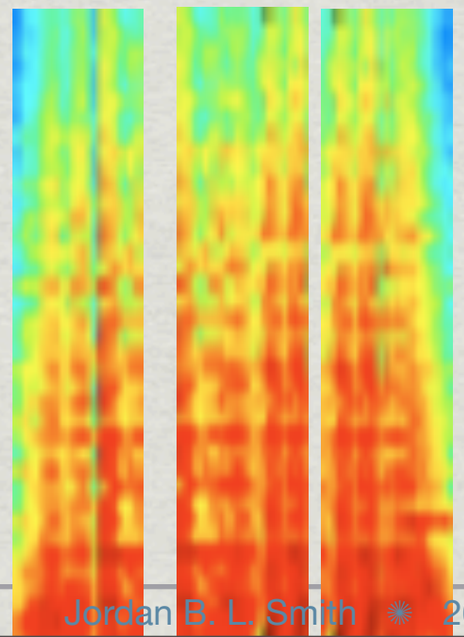
Clustering



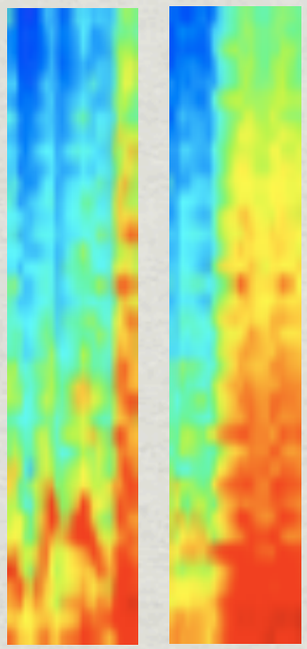
GROUP 1



GROUP 2



GROUP 3



GROUP 4

image: Foote 2000a

Clustering

1 3 3 3 1 1 4 1 2 4 1 2 2 ...

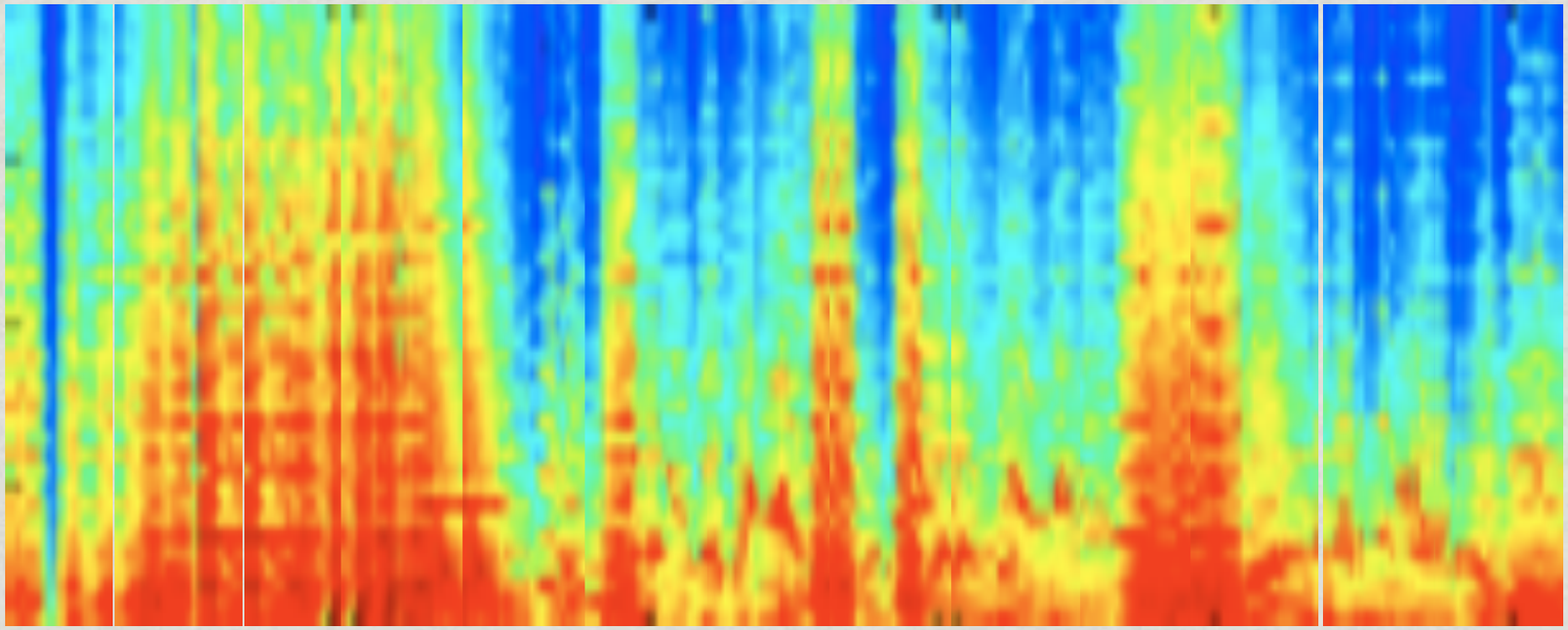


image: Foote 2000a

Clustering with HMM

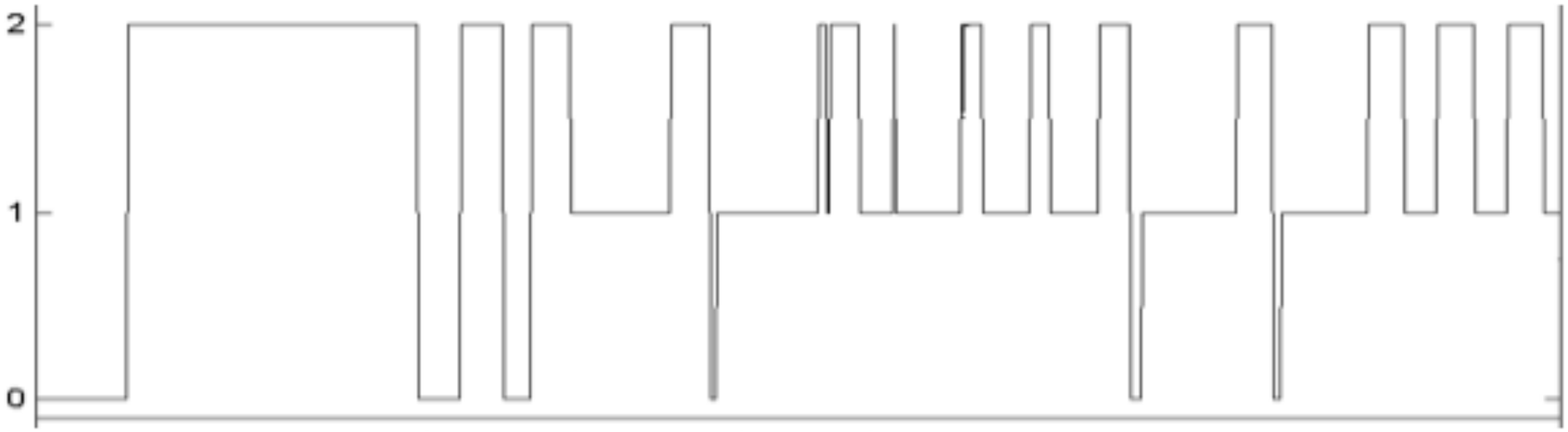


Figure 3-12: Segmentation of Bourvil's song. State 0 is {silence}, state 1 is {voice + accordion + accompaniment} and state 2 is {accordion + accompaniment}

image: Aucouturier 2001

Clustering with HMM

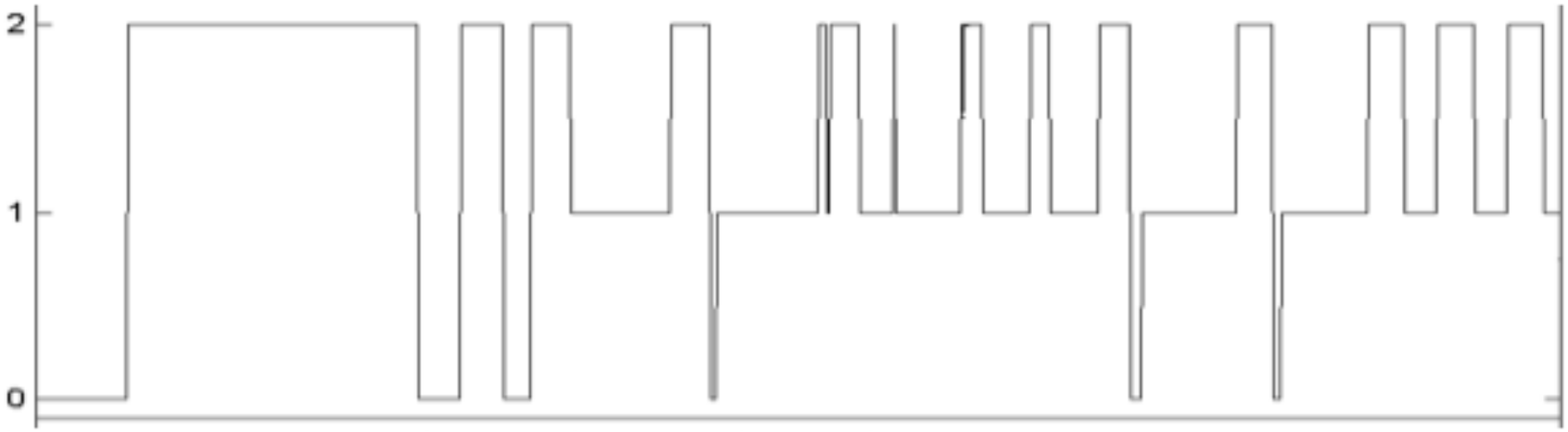


Figure 3-12: Segmentation of Bourvil's song. State 0 is {silence}, state 1 is {voice + accordion + accompaniment} and state 2 is {accordion + accompaniment}

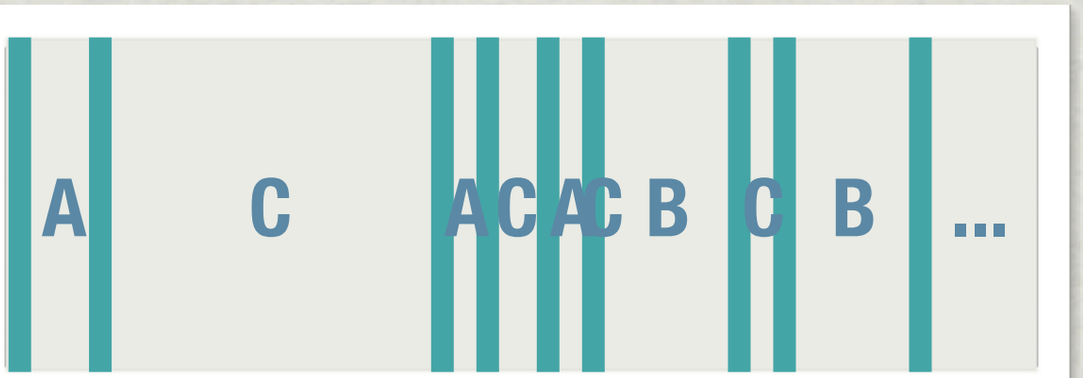


image: Aucouturier 2001

CLUSTERING

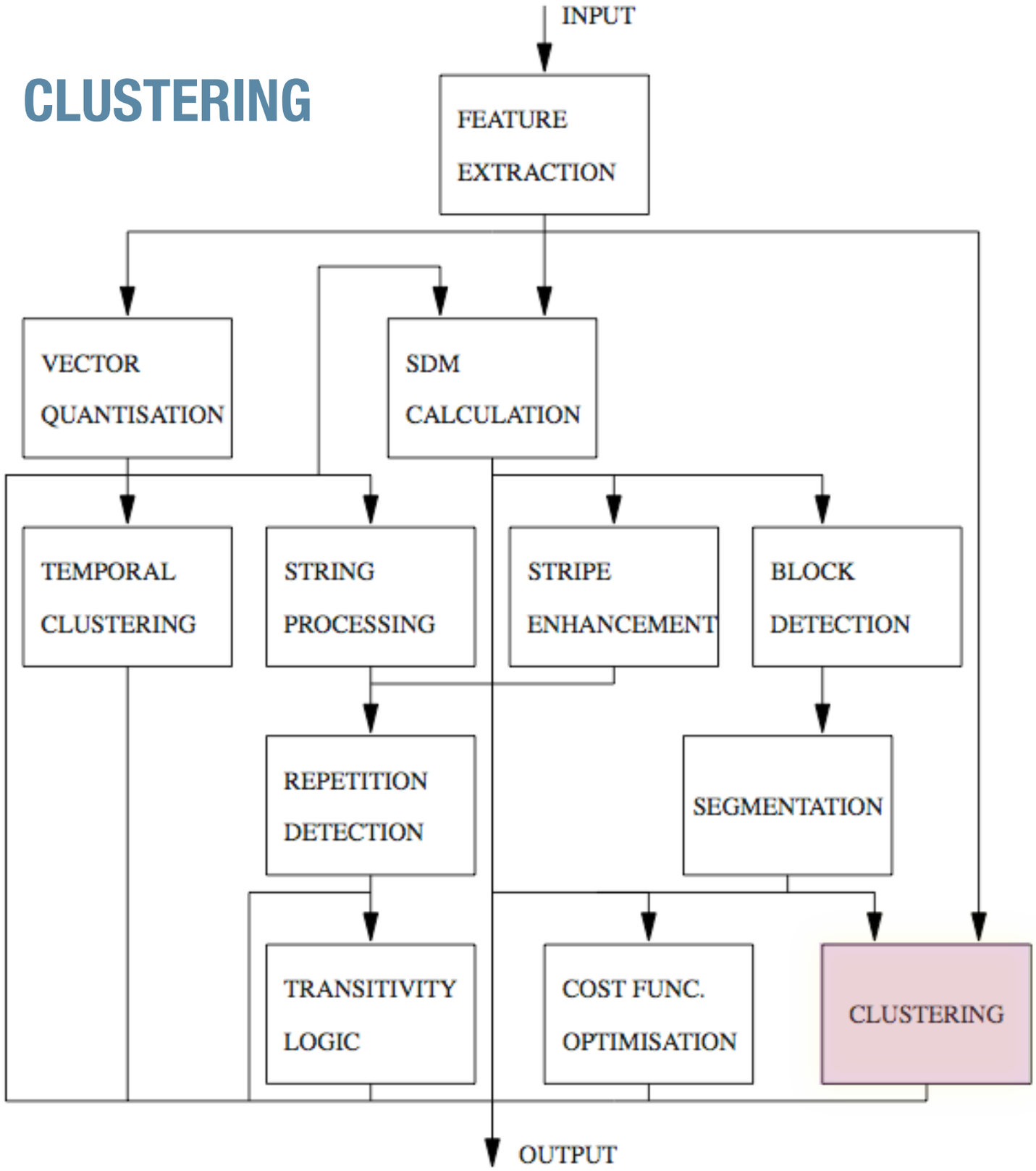


image: Paulus 2009

TEMPORAL CLUSTERING

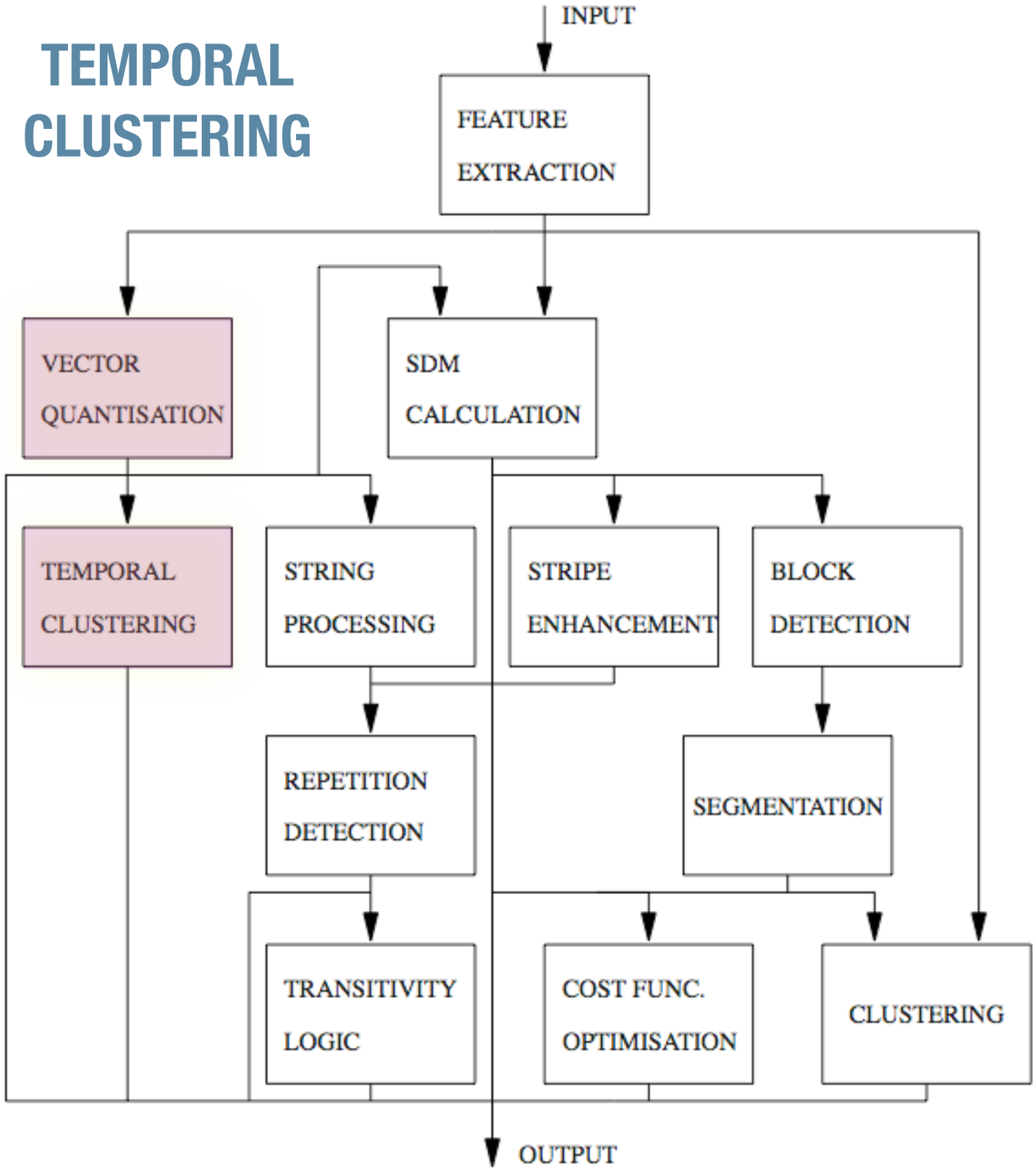
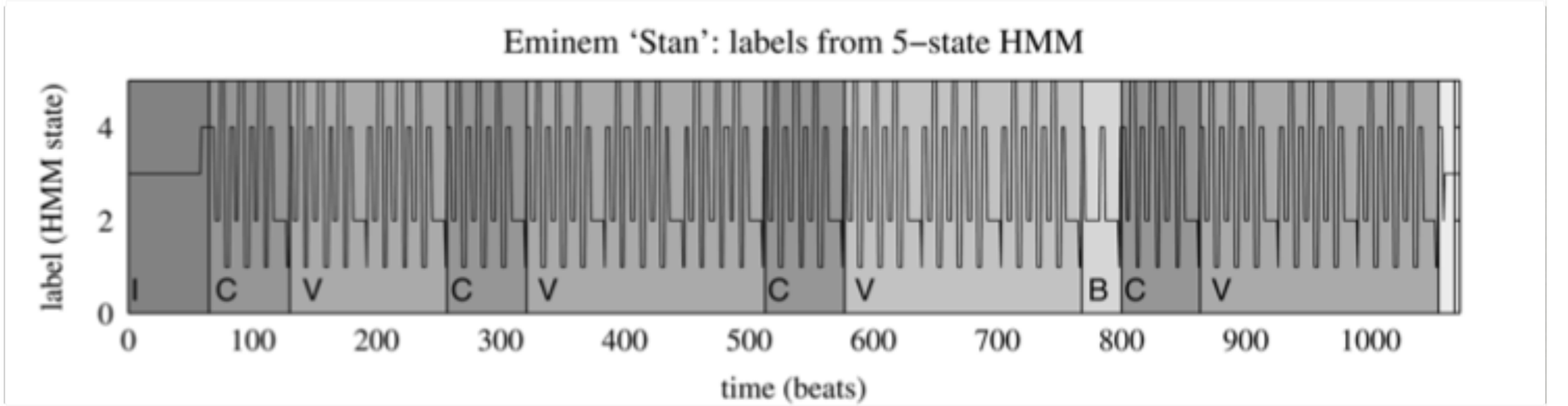
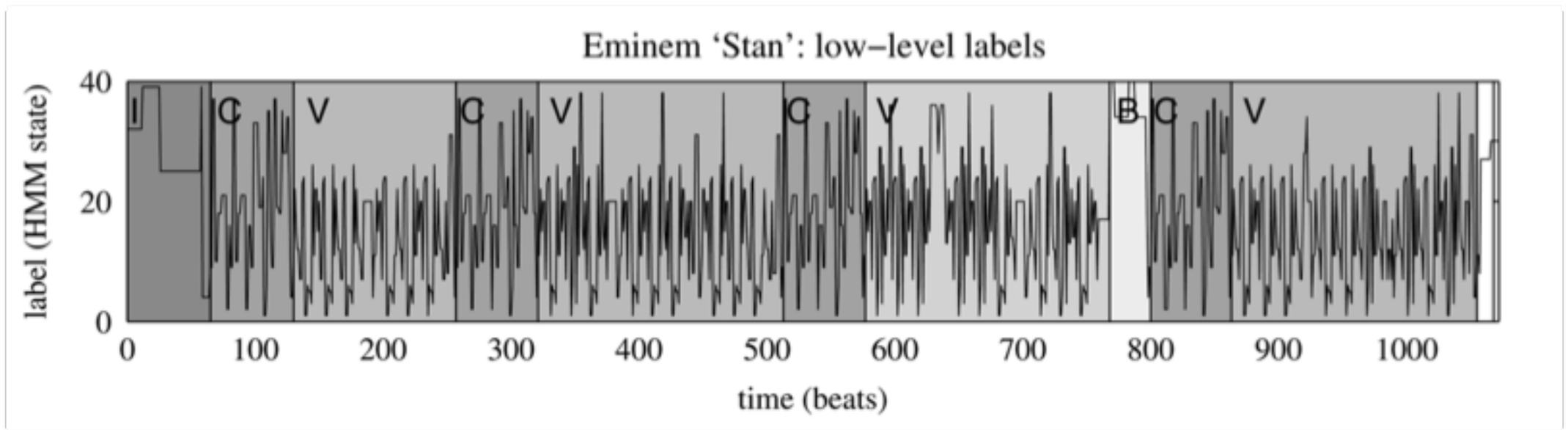


image: Paulus 2009

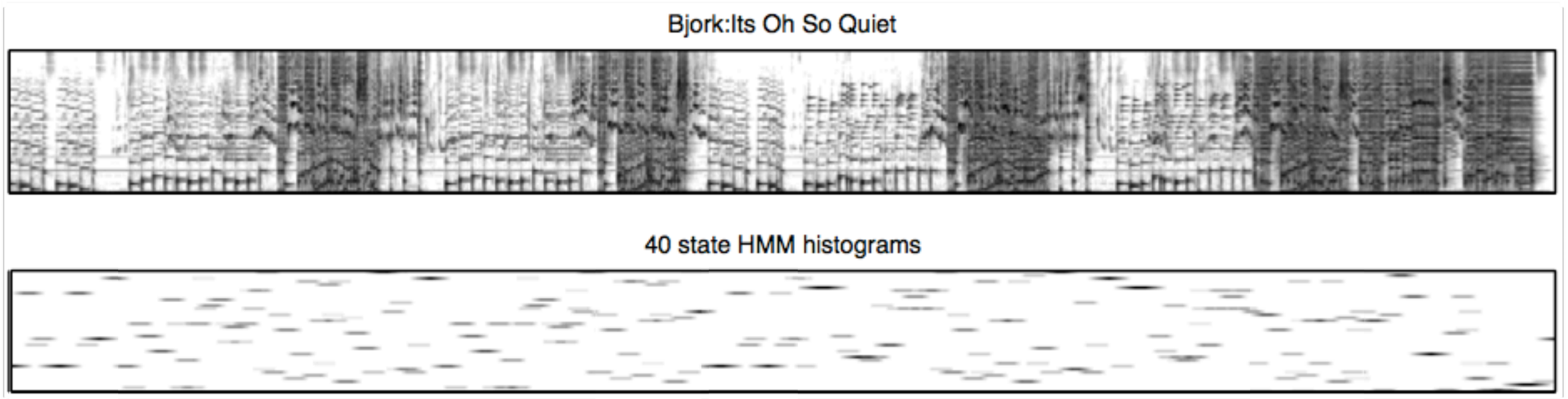
Clustering as mid-level representation

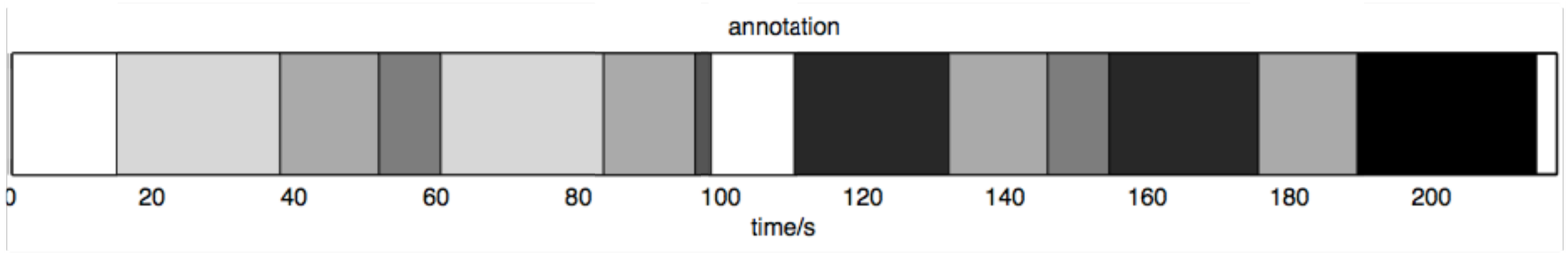




Images: Levy and Sandler 2008

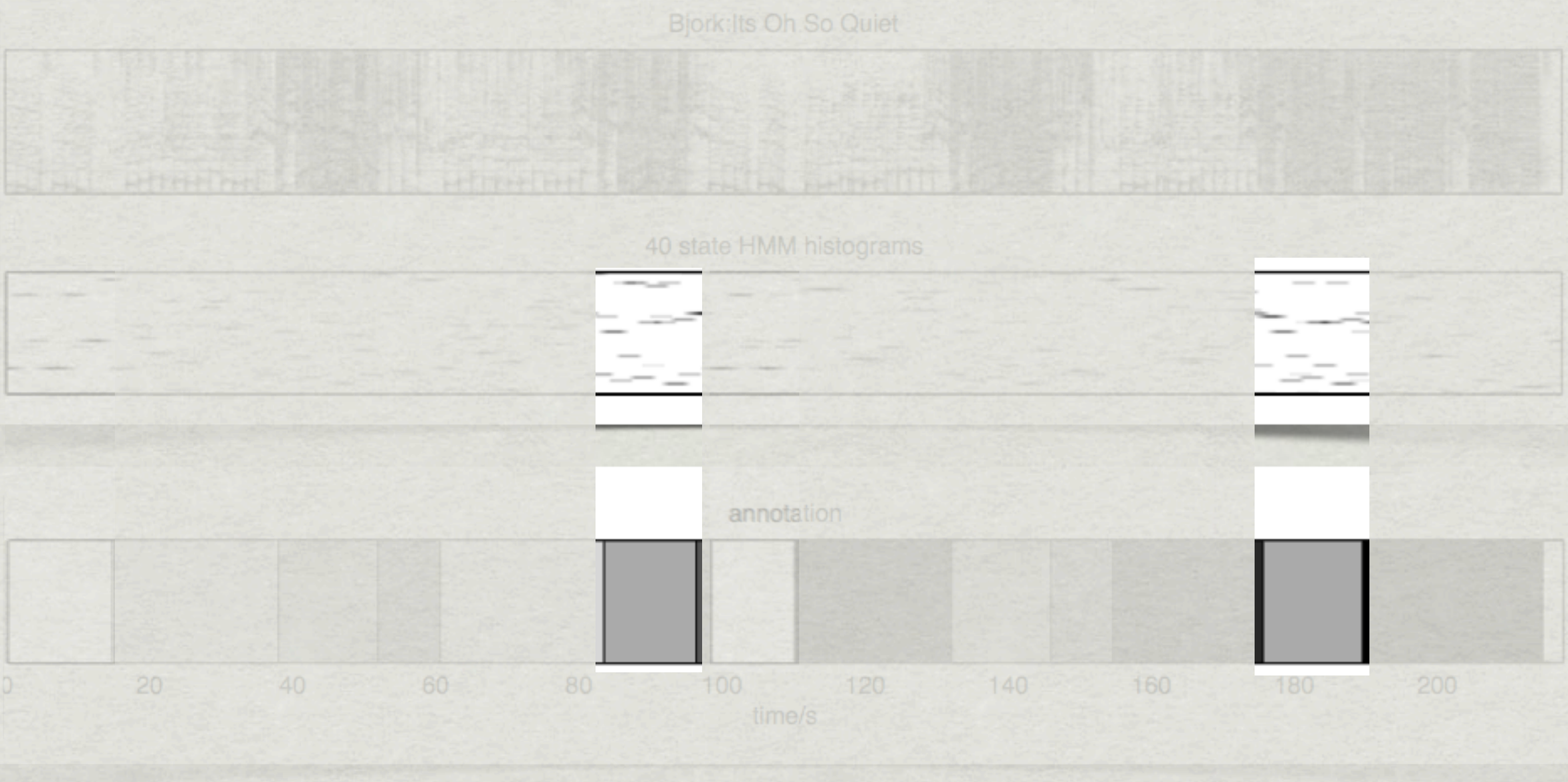
Clustering as mid-level representation





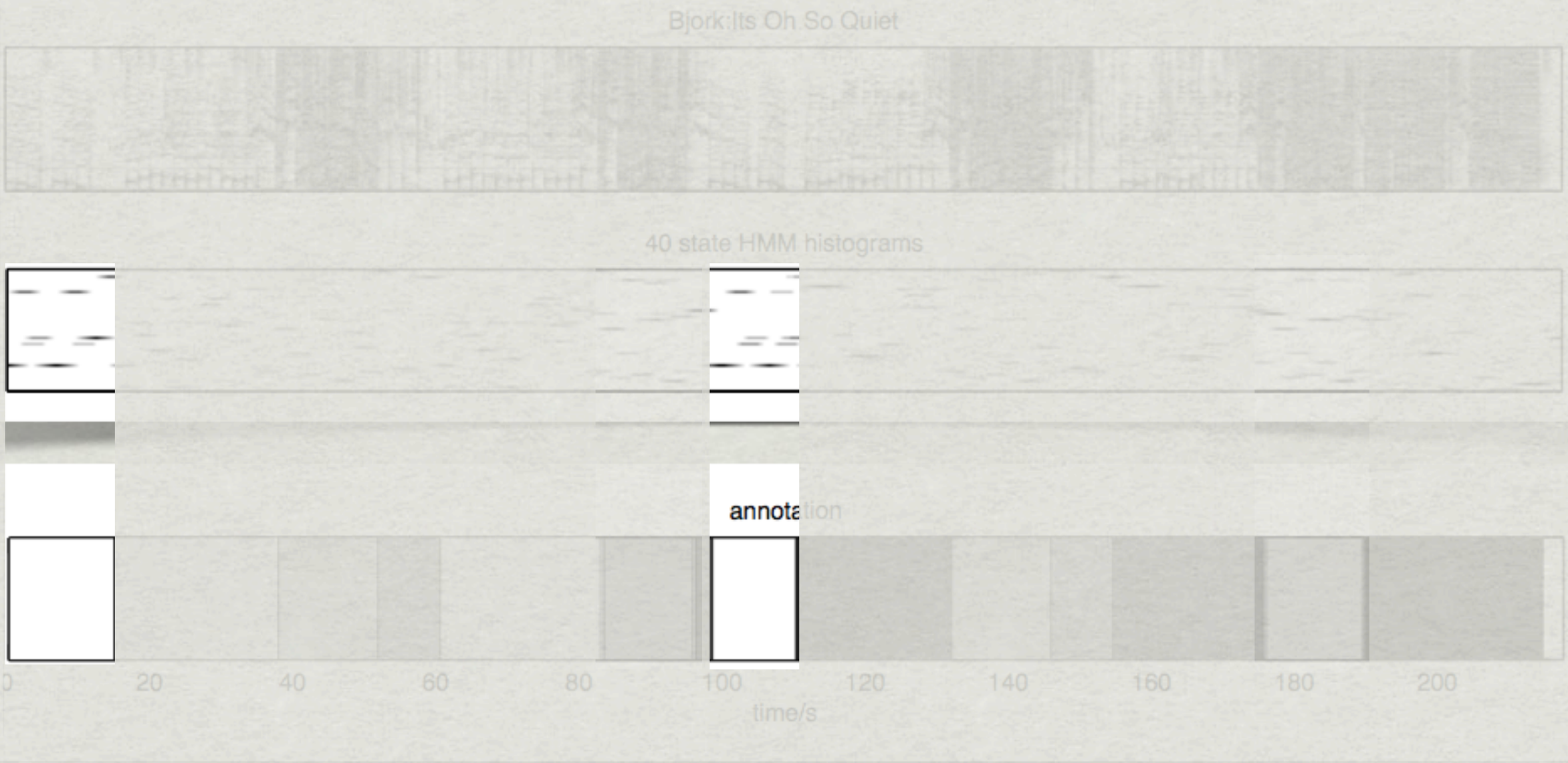
images: Abdallah et al. 2005

Clustering as mid-level representation



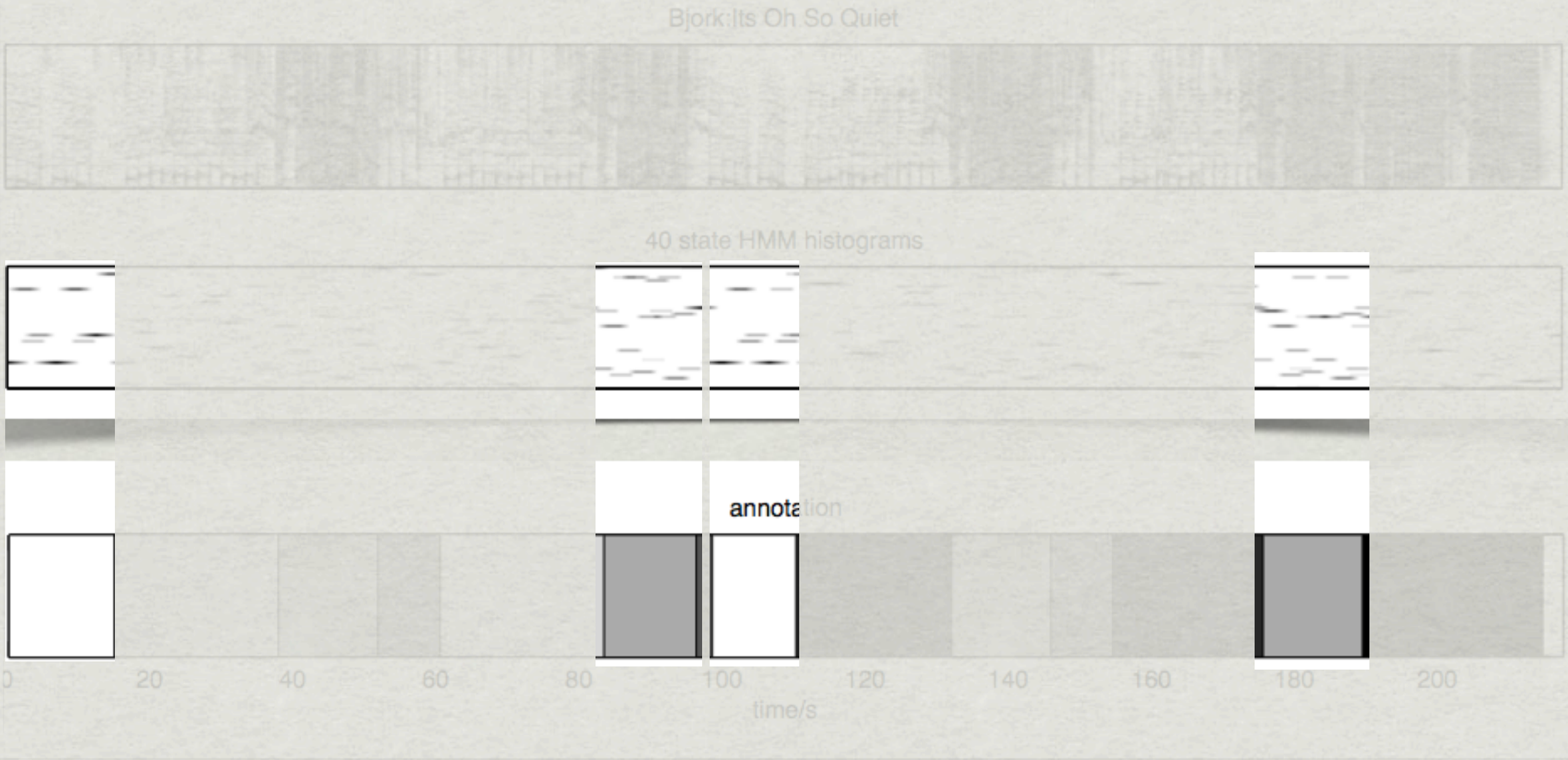
images: Abdallah et al. 2005

Clustering as mid-level representation



images: Abdallah et al. 2005

Clustering as mid-level representation



images: Abdallah et al. 2005

Clustering as mid-level representation

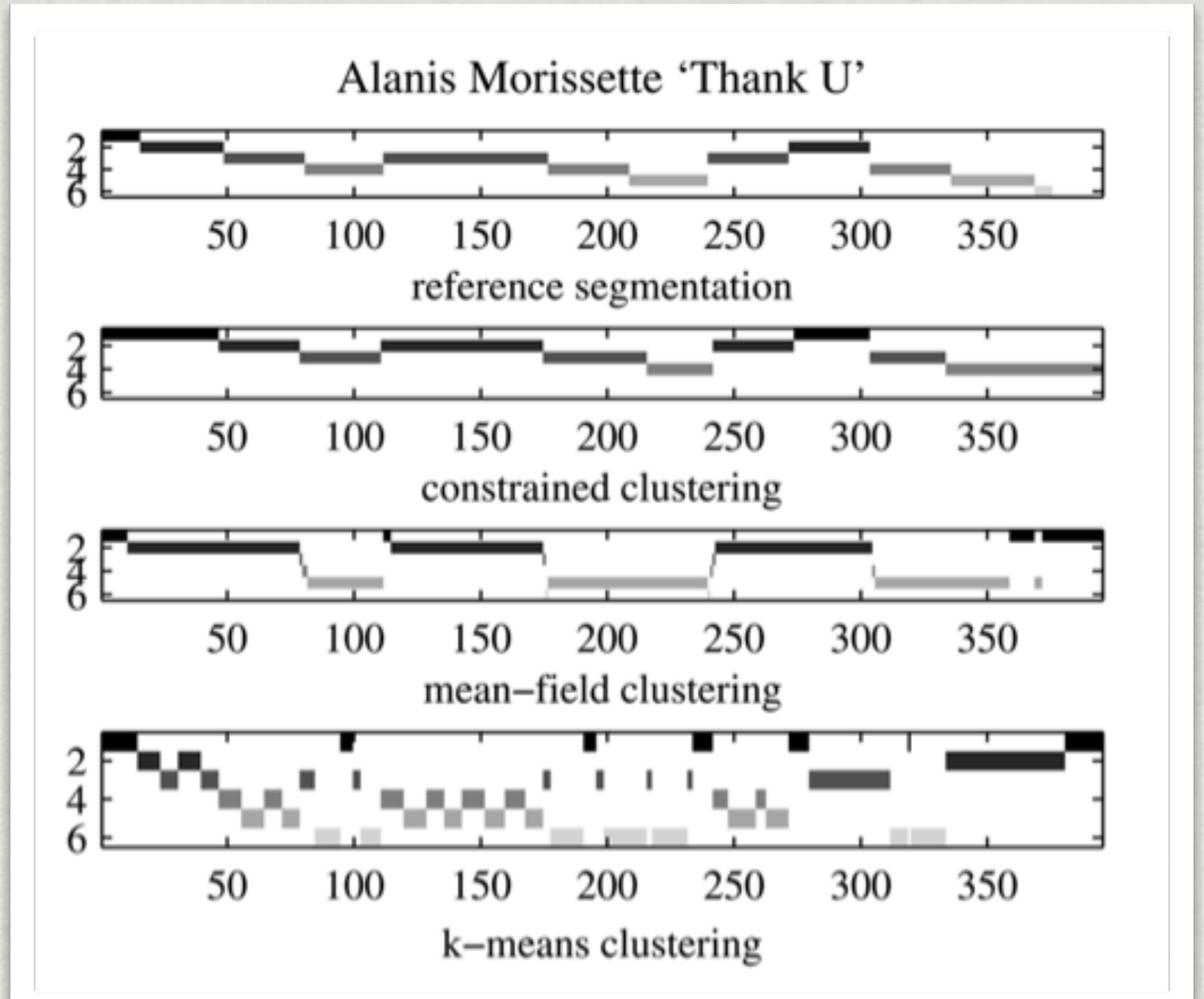


image: Levy and Sandler 2008

Features again

- Most features: static
  - each frame described by a vector
  - no information about temporal extent

- Solution: dynamic features

Dynamic features

- Information about timing or context:
  - Histograms (just saw)
  - Frame-wise derivatives (many)
  - Difference features (Turnbull et al. 2007)
  - FFTs of features (Peeters 2004)
  - Dynamic Texture Models (Barrington 2009)

~~Outline~~ Summary

1. Two hypotheses:

- ✓ States
- ✓ Sequences

2. A word on features:

- ✓ Timbre, harmony, etc.
- ✓ Dynamic features

3. Two techniques:

- ✓ Similarity matrix
- ✓ Clustering models

Discussion

- What can supervised learning do for structure analysis?
- Are either of the “states” or “sequences” hypotheses correct?
- Which of these methods can solve the Bohemian Rhapsody problem? (i.e., through-composed or ‘ABCD’ music)

Supervised learning

- Paulus & Klapuri 2010: applying semantic labels to analyses
- Turnbull et al. 2007: learning what boundaries look like

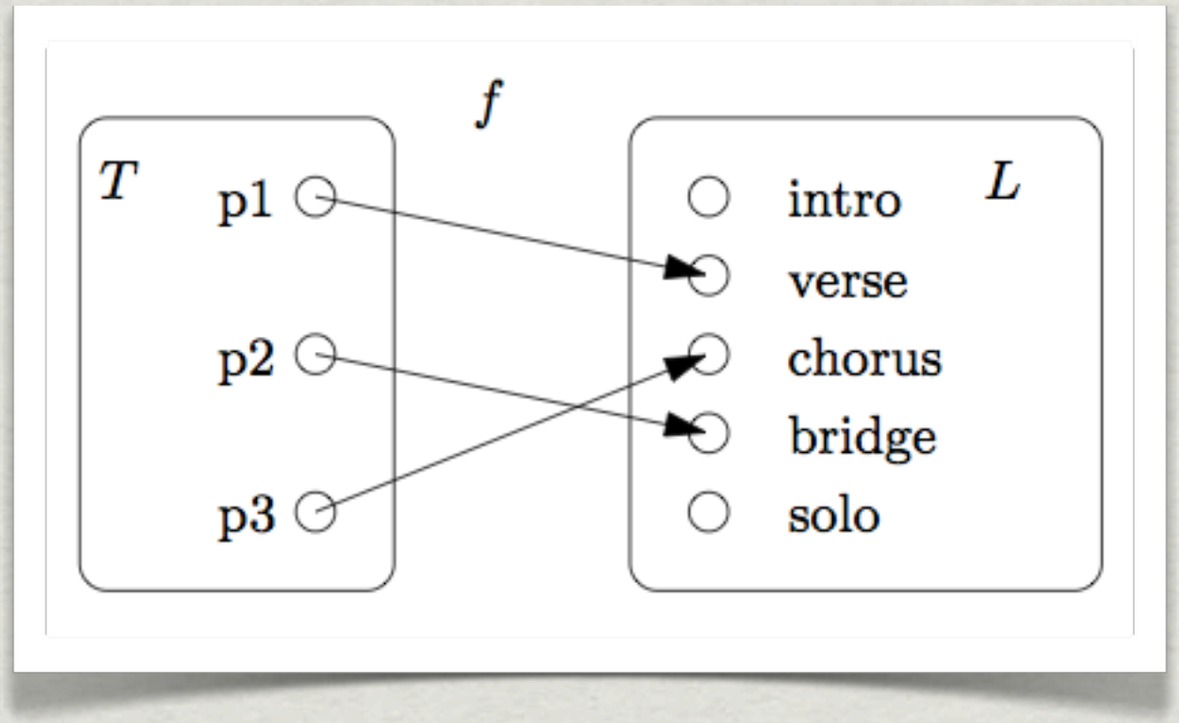


image: Paulus & Klapuri 2010

Thank you!

And thanks to:

McGill

Schulich School of Music
École de musique Schulich

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

DDMAL

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Image credits

- Abdallah, S., K. Noland, M. Sandler, M. Casey, and C. Rhodes. 2005. Theory and evaluation of a Bayesian music structure extractor. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), London, 420-5.
- Aucouturier, J.-J. 2001, July. Segmentation of musical signals, and applications to the analysis of musical structure. Master's thesis, Kings College, University of London.
- Foote, J. 2000a. Arthur: Retrieving orchestral music by long-term structure. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Plymouth, MA, USA.
- Foote, J. 2000b. Automatic audio segmentation using a measure of audio novelty. In Proceedings of the IEEE International Conference on Multimedia & Expo (ICME), 452-5.
- Levy, M., and M. Sandler. 2008, Feb. Structural segmentation of musical audio by constrained clustering. IEEE Transactions on Audio, Speech, and Language Processing 16 (2): 318-26.
- Paulus, J. 2009. Signal processing methods for drum transcription and music structure analysis. Ph.D. thesis, Tampere University of Technology, Tampere, Finland.

References

- Abdallah, S., M. Sandler, C. Rhodes, and M. Casey. 2006. Using duration models to reduce fragmentation in audio segmentation. Machine Learning 65 (2-3): 485-515.
- Aucouturier, J.-J. 2001, July. Segmentation of musical signals, and applications to the analysis of musical structure. Master's thesis, Kings College, University of London.
- Barrington, L., A. Chan, and G. Lanckriet. 2009. Dynamic texture models of music. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Washington, DC, USA, 1589-92. IEEE Computer Society.
- Chai, W. 2005, September. Automated analysis of musical structure. Ph. D. thesis, Massachusetts Institute of Technology, MA, USA.
- Foote, J. 2000a. Arthur: Retrieving orchestral music by long-term structure. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Plymouth, MA, USA.
- Foote, J. 2000b. Automatic audio segmentation using a measure of audio novelty. In Proceedings of the IEEE International Conference on Multimedia & Expo (ICME), 452-5.
- Foote, J., and M. Cooper. 2003. Media segmentation using self-similarity decomposition. In M. Yeung, R. Lienhart, and C.-S. Li (Eds.), Proceedings of the SPIE: Storage and Retrieval for Media Databases, Volume 5021, Santa Clara, CA, USA, 167-75. SPIE.
- Goto, M. 2003a. A chorus-section detecting method for musical audio signals. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Volume 5, 437-40.
- Jehan, T. 2005. Hierarchical multi-class self similarities. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, United States, 311-4.
- Levy, M., and M. Sandler. 2008, Feb. Structural segmentation of musical audio by constrained clustering. IEEE Transactions on Audio, Speech, and Language Processing 16 (2): 318-26.
- Logan, B., and S. Chu. 2000. Music summarization using key phrases. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Volume 2, Washington D.C., USA, 749-52. IEEE Computer Society.
- Maddage, N., C. Xu, M. Kankanhalli, and X. Shao. 2004. Content-based music structure analysis with applications to music semantics understanding. In Proceedings of the ACM International Conference on Multimedia, New York, NY, United States, 112-9.
- Paulus, J. 2009. Signal processing methods for drum transcription and music structure analysis. Ph.D. thesis, Tampere University of Technology, Tampere, Finland.
- Peeters, G. 2004. Deriving musical structures from signal analysis for music audio summary generation: “sequence” and “state” approach. In G. Goos, J. Hartmanis, and J. van Leeuwen (Eds.), Computer Music Modeling and Retrieval, Volume 2771, 169-85. Springer Berlin / Heidelberg.
- Shiu, Y., H. Jeong, and C.-C. J. Kuo. 2006b. Similarity matrix processing for music structure analysis. In Proceedings of the ACM Workshop on Audio and Music Computing Multimedia (AMCMM), New York, NY, USA, 69-76. ACM.
- Turnbull, D., G. Lanckriet, E. Pampalk, and M. Goto. 2007. A supervised approach for detecting boundaries in music using difference features and boosting. In Proceedings of the International Conference on Music Information Retrieval (ISMIR), Vienna, Austria, 51-4.