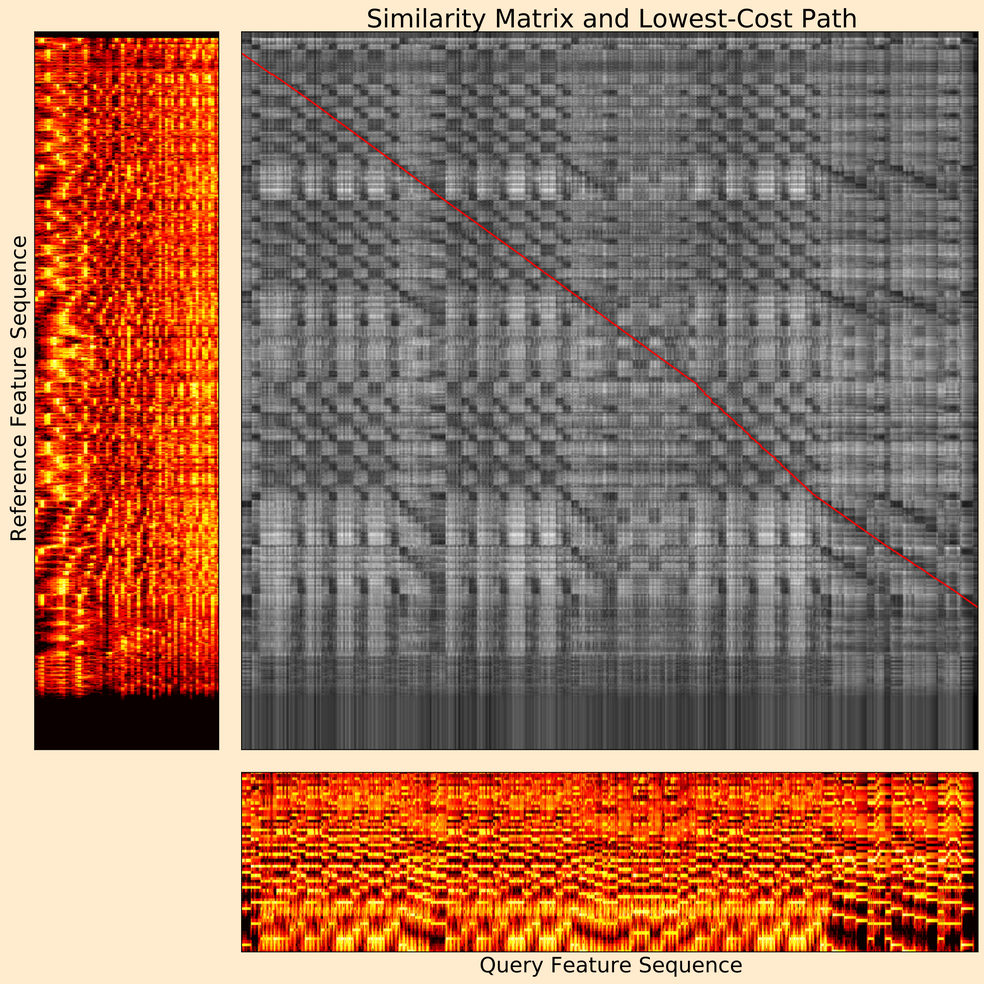
Lab ROSA
Laboratory for the Recognition and Organization of Speech and Audio

LEARNING EFFICIENT REPRESENTATIONS FOR SEQUENCE RETRIEVAL

Colin Raffel, LabROSA, Dept of Electrical Engineering, Columbia University

Baseline Method: Dynamic Time Warping

- Natural metric for comparing sequences
- Useful in retrieval, classification, and alignment
- $\mathcal{O}(NM)$-complex using dynamic programming
- Various "pruning methods" exist which approach linear time...
- However, most are not universally applicable
- Data dimensionality can cause expensive "local distance" calculations
- Quadratic penalty when the data is sampled too finely
- Inappropriate when sequences come from different modalities
- Relies on a non-learned metric for comparing feature vectors



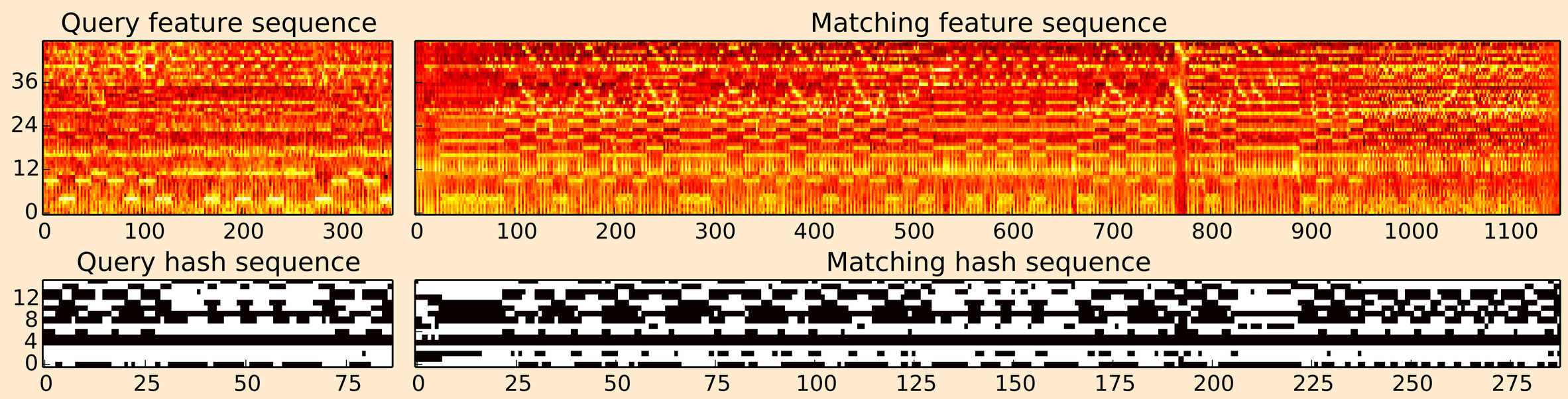
References

[1] Colin Raffel and Daniel P. W. Ellis. Large-scale content-based matching of MIDI and audio files. In *16th International Society for Music Information Retrieval Conference (to appear)*, 2015.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *arXiv preprint arXiv:1503.03832*, 2015.

Hash Feature Vectors and Implicitly Downsample Sequences for Faster DTW [1]



Map Entire Feature Vector Sequences to an Embedded Fixed-Size Euclidean Space [2, 3]