Audio Cover Song Identification: Beyond The Notes

Chris Tralie
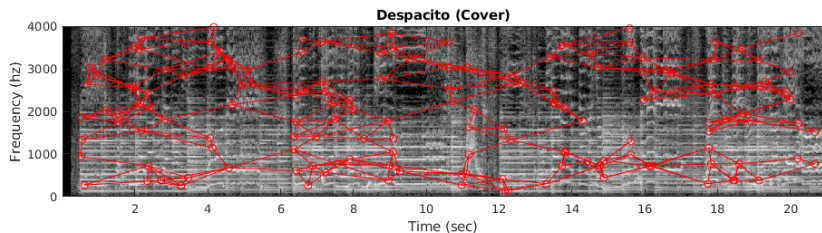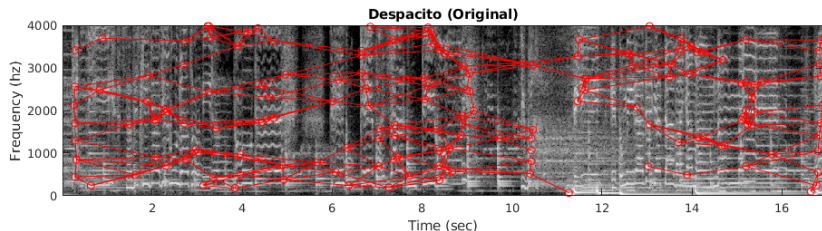
Duke University ECE / Math

John's Hopkins CBME

Ursinus College Math/CS (Fall 2019)

2/9/2018

Just Use Shazam! (?)

Despacito (Original)

Despacito (Cover)

▷ Traditional audio fingerprinting is abysmal on covers

Multi-Feature Cover Song Identification

▶ Small Scale Multi-Feature CSI

▷ Large Scale Multi-Feature CSI

[1] Christopher J Tralie and Paul Bendich. "Cover Song Identification with Timbral Shape". In: *16th International Society for Music Information Retrieval (ISMIR) Conference.* 2015

[2] Christopher J Tralie. "MFCC And HPCP Fusion for Robust Cover Song Identification". In: *18th International Society for Music Information Retrieval (ISMIR).* 2017

HPCP

HPCP Reo Speedwagon

HPCP Styx
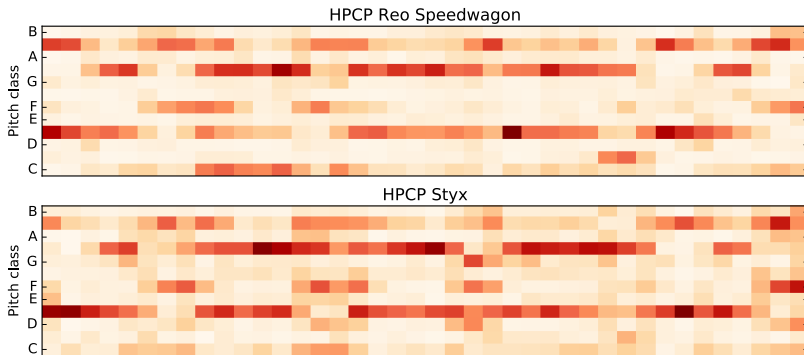
Emilia Gómez. "Tonal description of polyphonic audio for music content processing". In: *INFORMS Journal on Computing* 18.3 (2006), pp. 294–304

Daniel PW Ellis. "Identifying 'cover songs' with beat-synchronous chroma features". In: *MIREX 2006* (2006), pp. 1–4

Juan Pablo Bello. "Audio-Based Cover Song Retrieval Using Approximate Chord Sequences: Testing Shifts, Gaps, Swaps and Beats.". In: *ISMIR*. vol. 7. 2007, pp. 239–244

Joan Serra et al. "Chroma binary similarity and local alignment applied to cover song identification". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 16.6 (2008), pp. 1138–1151

Joan Serra, Xavier Serra, and Ralph G Andrzejak. "Cross recurrence quantification for cover song identification". In: *New Journal of Physics* 11.9 (2009), p. 093017

Chroma / HPCP

▷ Create cross-similarity matrix with cosine distance between beat-synchronous blocks of HPCP features between two songs

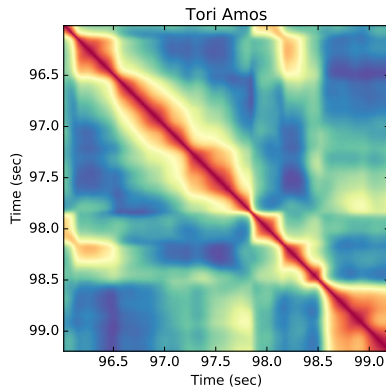▷ Find "fuzzy diagonals" some way (e.g. Smith Waterman on binary CSM)
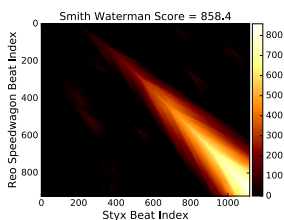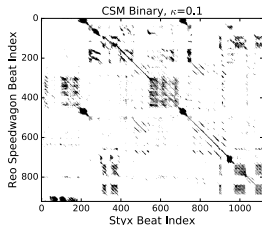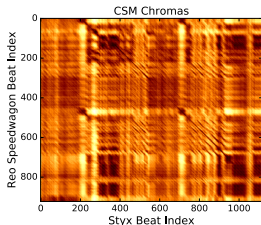
"Something So Right"

Annie Lennox   Paul Simon
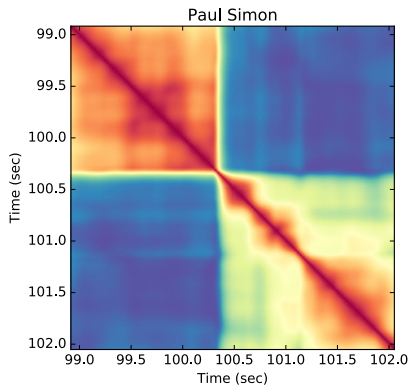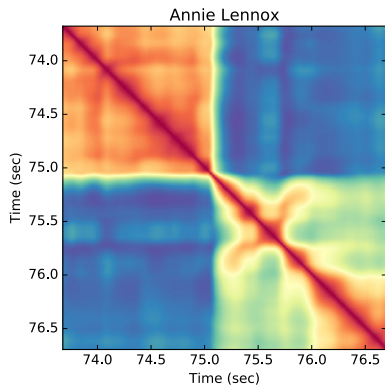
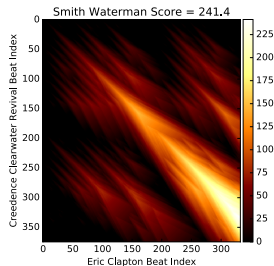"Time"
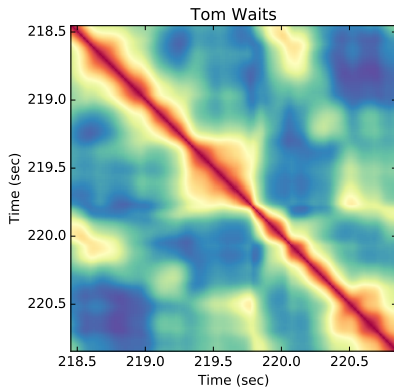
SSMs As Geometric Features

Joint work with Paul Bendich (Duke)

Resize all beat-synchronous SSMs to same resolution $d \times d$

$$\mathsf{CSM}_{ij} = ||\mathsf{SSMA}_i - \mathsf{SSMB}_j||_F$$

▷ True Cover Pair: "Before You Accuse Me"

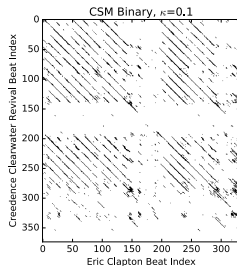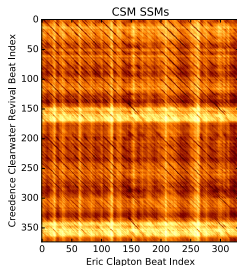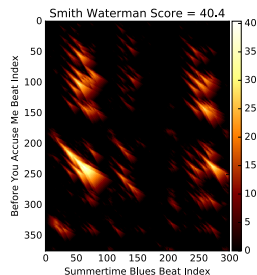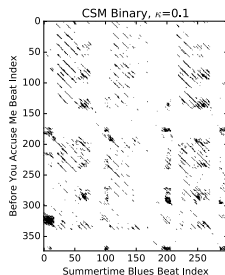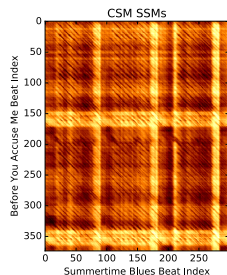# SSMs As Geometric Features

Resize all beat-synchronous SSMs to same resolution $d \times d$

$$\text{CSM}_{ij} = ||\text{SSMA}_i - \text{SSMB}_j||_F$$

▷ False Cover Pair: "Before You Accuse Me" vs "Summertime Blues"

Similarity Network Fusion

▷ Unsupervised similarity learning by cross-diffusion[1]

Figure 1. The first 10 retrieved shapes in MPEG7 by Shape Context (SC) [3] (first row), Inner Distance(IDSC) [13] (second row) and Cross Diffusion with SC & IDSC (last row).The first column shows the query shape.

[1] Bo Wang et al. "Unsupervised metric fusion by cross diffusion". In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE. 2012, pp. 2997–3004

[2] Bo Wang et al. "Similarity network fusion for aggregating data types on a genomic scale". In: *Nature methods* 11.3 (2014), pp. 333–337

[3] Ning Chen, Wei Li, and Haidong Xiao. "Fusing similarity functions for cover song

Similarity Network Fusion

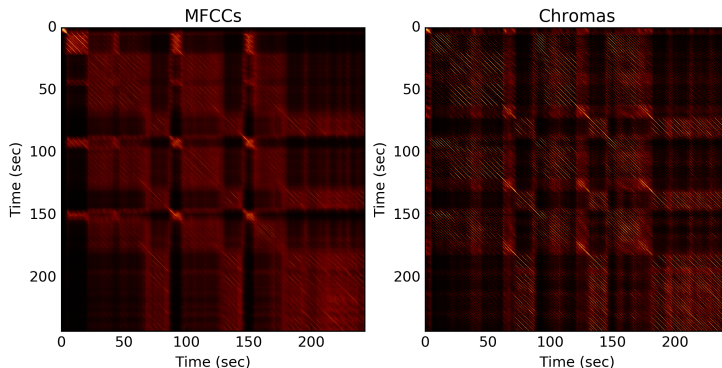▷ Unsupervised similarity learning by cross-diffusion[1]

[1] Bo Wang et al. "Unsupervised metric fusion by cross diffusion". In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE. 2012, pp. 2997–3004

[2] Bo Wang et al. "Similarity network fusion for aggregating data types on a genomic scale". In: *Nature methods* 11.3 (2014), pp. 333–337

[3] Ning Chen, Wei Li, and Haidong Xiao. "Fusing similarity functions for cover song identification". In: *Multimedia Tools and Applications* (2017), pp. 1–24. ISSN:

My Contribution: Cross Similarity Network Fusion
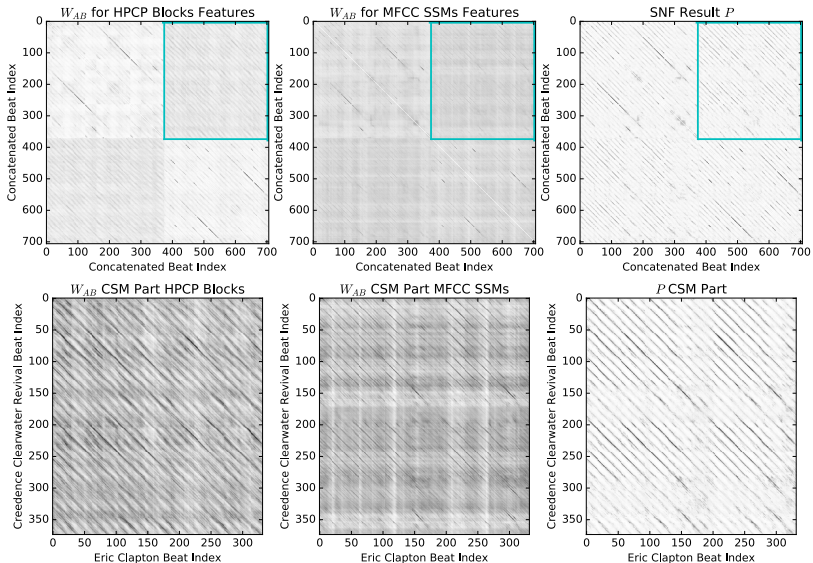
Covers 80 Results

Single Feature Results

——————-	MR	MRR	MDR	Top 1	Top 25	Top 50	Top 100	Score
SSM	15.14	0.615	1	91	130	144	155	48/80
MFCC	29.71	0.538	2	79	108	122	142	42/80
HPCP	16.14	0.669	1	100	130	140	150	52/80

Fusion Results

——————-	MR	MRR	MDR	Top 1	Top 25	Top 50	Top 100	Score
SSMs/MFCC	13.96	0.7	1	107	132	142	155	55/80
HPCP/SSMs 3 Iters	**7.52**	**0.849**	1	**131**	**150**	**152**	155	**68/80**
Chen 2017[1]	?	0.625	?	?	?	?	?	?

[1] Ning Chen, Wei Li, and Haidong Xiao. "Fusing similarity functions for cover song identification". In: *Multimedia Tools and Applications* (2017), pp. 1–24. ISSN: 1573-7721. DOI: 10.1007/s11042-017-4456-9. URL: http://dx.doi.org/10.1007/s11042-017-4456-9

	MR	MRR	Top-01	Top-10
MFCCs	83.3	0.618	583	679
SSMs	72.5	0.623	581	698
HPCPs	44.4	0.757	727	809
Late	19.8	0.875	855	931
Early	22.5	0.829	798	884
Early + Late	**14**	**0.904**	**884**	**950**

Table: Results of different features and fusion techniques on the Covers 1000 dataset.

Why does this work so well?

Javascript CSM Viewer

▷ Small Scale Multi-Feature CSI

► Large Scale Multi-Feature CSI

[1] Christopher J Tralie. "GraphDitty: A Software Suite for Geometric Music Structure Visualization". In: *19th International Society for Music Information Retrieval (ISMIR), Late Breaking Session*. 2018
[2] Christopher J Tralie and Brian McFee. "Enhanced Hierarchical Music Structure Annotations via Feature Level Similarity Fusion". In: *ICASSP*. 2019
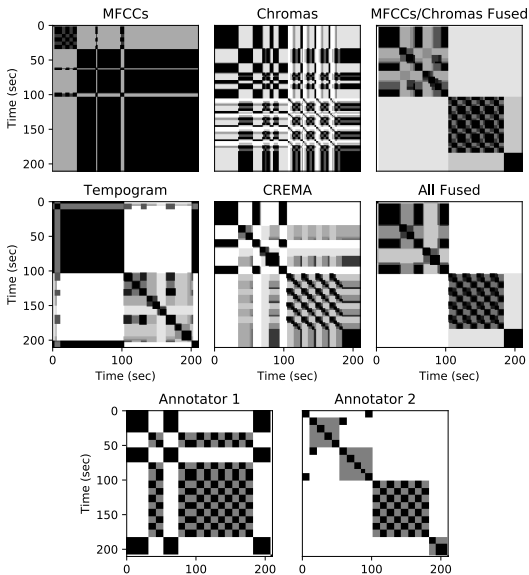
▷ `http://www.covers1000.net/GraphDitty`
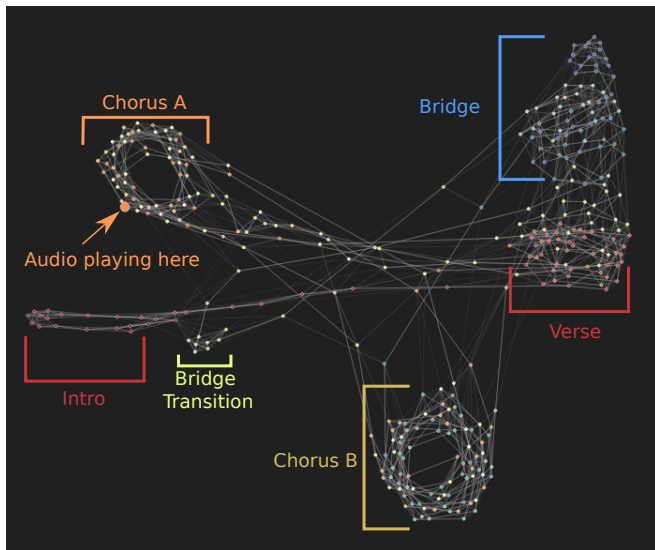
Joint work with Brian McFee

Thank You!

Contact: chris.tralie@gmail.com

Supplementary slides

▷ "Parent SSM": SSM on song A concatenated to song B
▷ Learning similarity functions for parent SSMs fusing different features