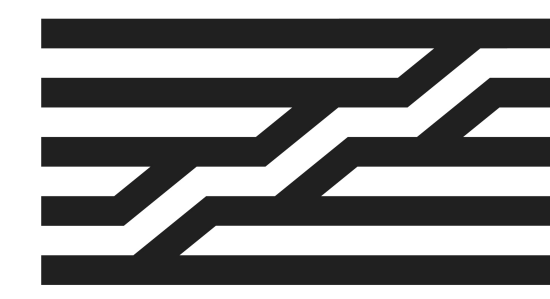


And what if two musical versions don't share melody, harmony, rhythm or lyrics ?

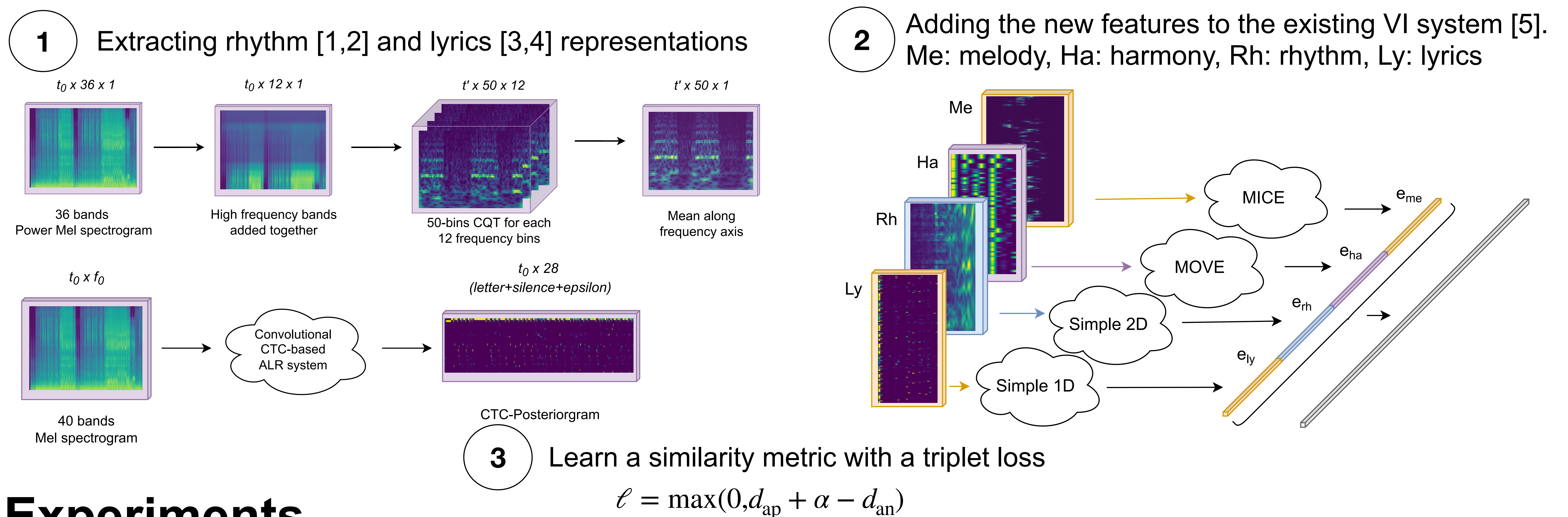
Mathilde Abrassart, Guillaume Doras



Problem

1. establish a similarity relationship $S_{ij} = g(f(A_i), f(B_j))$ between a query track A_i and a reference track B_j
2. f is a feature extraction function and shall preserve common musical facets between versions
3. g is a comparison function and shall allow fast lookup into large music corpora

Proposed method



Experiments

✓ Training : Dali (for lyrics representation only), SHS₅₊ (62k versions of ~7.5k works) and SHS_{5+/4-}

✓ Evaluating with SHS₄₋ (50k covers of ~20k works) and Da-Tacos (13k versions of 1k works & 2k confusing tracks)

(datasets are pruned to avoid overlapping samples - see <https://ircam-anasynth.github.io/papers/2022/abrassart>)

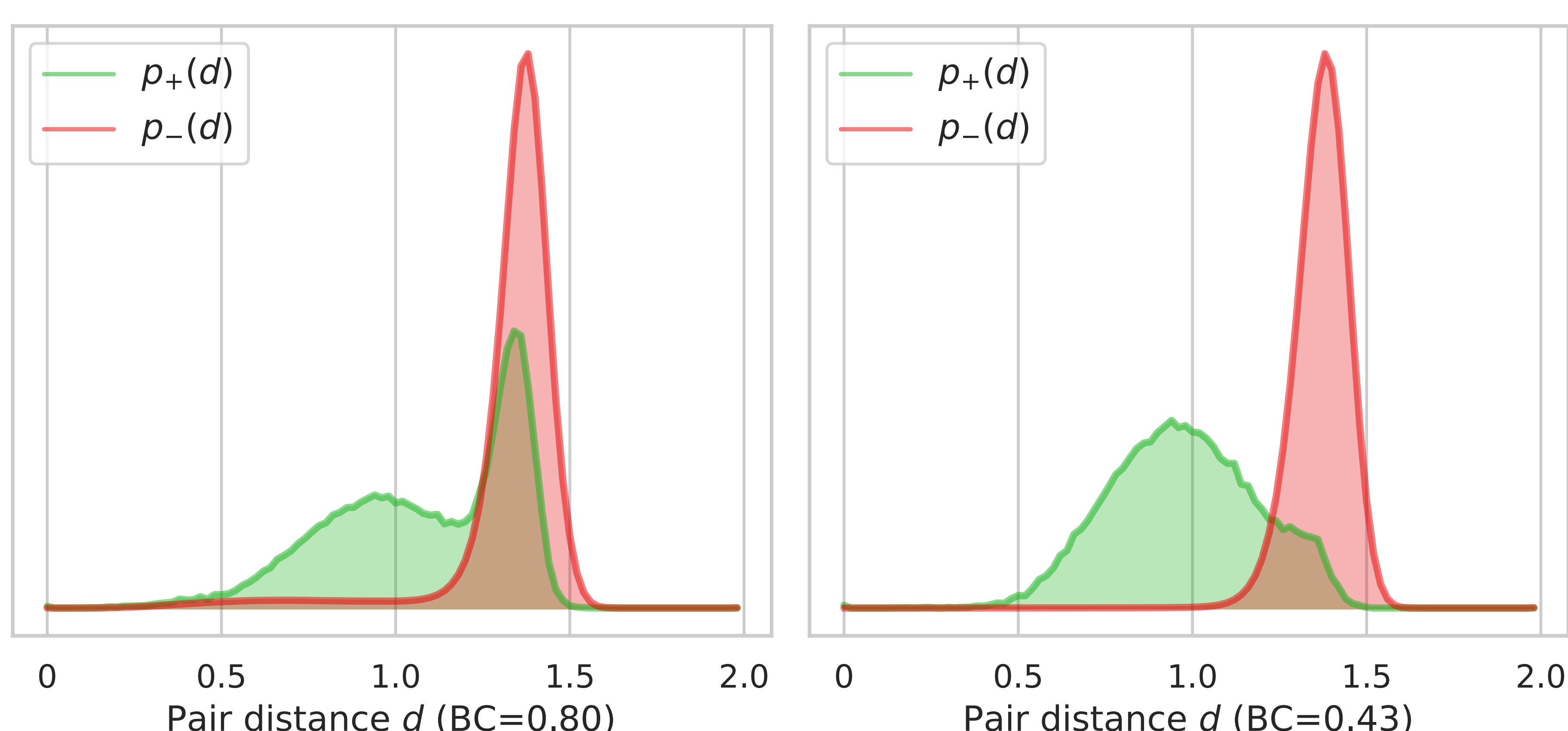
✓ Comparing with baseline model results [5]

*results on Da-Tacos-Vocals (w/o instrumental tracks)

Test set	Pruned SHS ₄₋			Pruned Da-Tacos		
	MAP	MT@10	MR1	MAP	MT@10	MR1
Me+Ha (baseline)	0.693	1.256	453	0.626 0.717*	6.668 6.290*	32 21*
Me+Ha+Ly	0.800	1.396	291	0.602 0.818*	6.480 7.205*	33 16*
Me+Ha+Rh+Ly	0.785	1.378	286	0.560 0.765*	6.054 6.714*	33 14*

✓ Comparing distribution for lyrics with Da-Tacos and Da-Tacos-Vocals (w/o instrumental tracks)

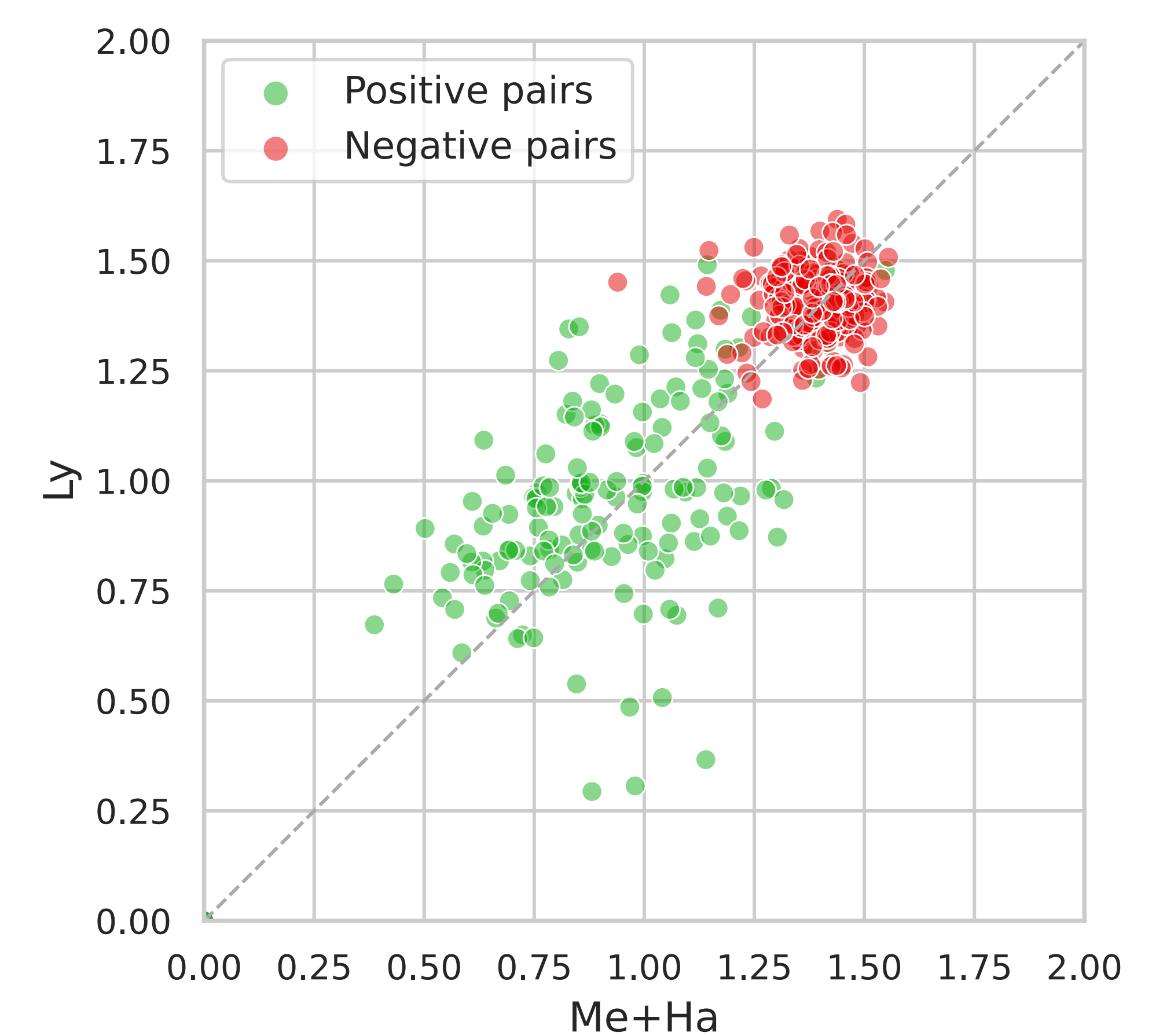
- Disappearance of a bump depicting the presence of false negative pairs



✓ Pairwise distances for Me+Ha vs. Ly for 500 pairs from SHS₄₋

Positive pairs: samples which are versions, negative pairs: samples which are not versions

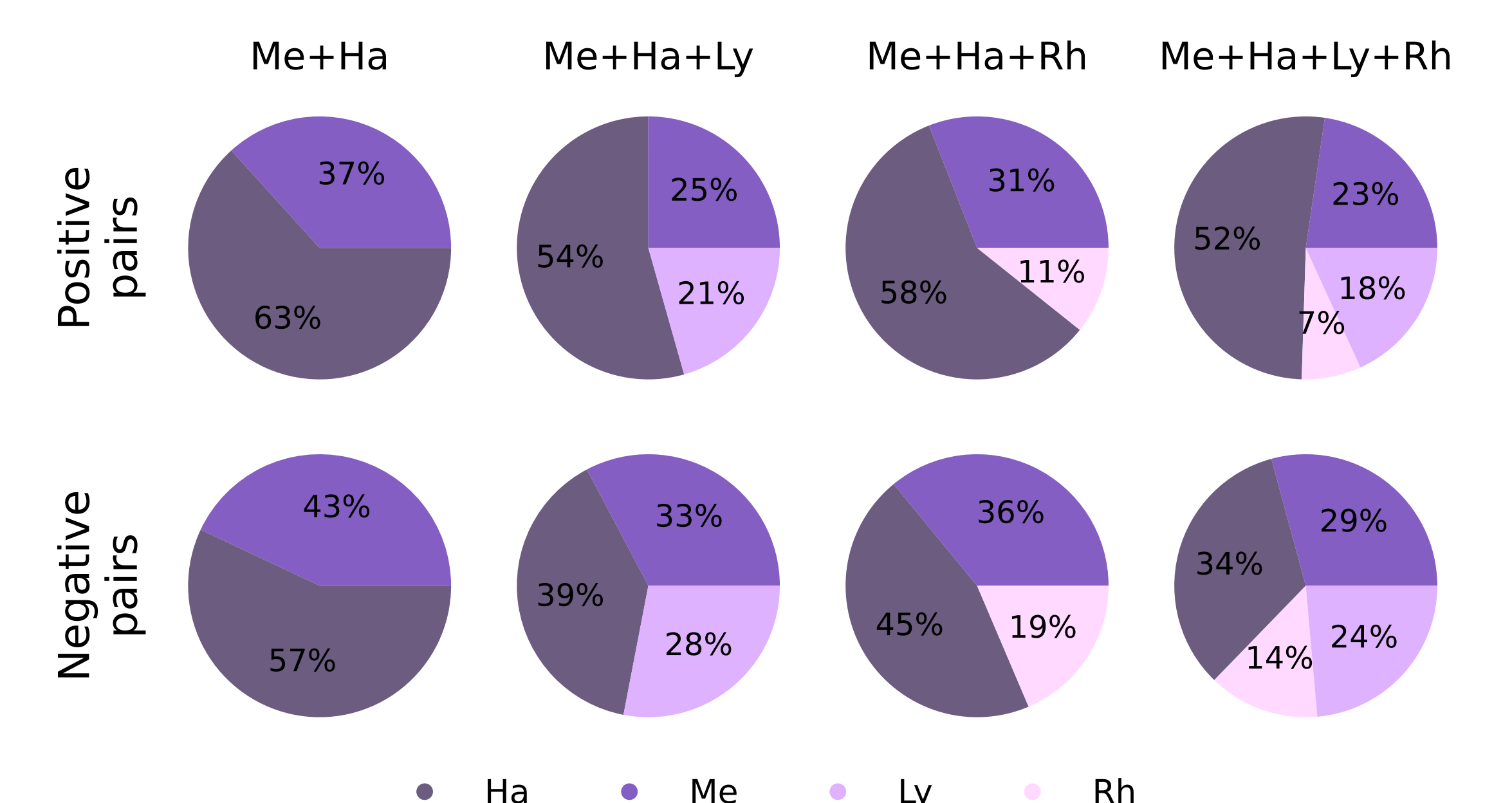
- Evidence of complementarity between Me+Ha and Ly



✓ Observing oracle results (considers the best performing feature to compute each pairwise distance)

Test set	Pruned SHS ₄₋			Pruned Da-Tacos		
	MAP	MT@10	MR1	MAP	MT@10	MR1
Me+Ha (baseline)	0.879	1.521	97	0.837	8.709	4
Me+Ha+Ly	0.963	1.637	14	0.918	9.398	1
Me+Ha+Rh+Ly	0.978	1.658	4	0.951	9.657	1

✓ Observing most relevant feature proportions to identify positive and negative pairs on SHS₄₋



✓ Comparing SoA results on SHS₄₋ and Da-Tacos

*results on Da-Tacos-Vocals (w/o instrumental tracks)

Test set		SHS ₄₋			Da-Tacos		
Model	Emb.	MAP	MT@10	MR1	MAP	MT@10	MR1
Doras et al. [5]	512	0.660	1.080	657	0.635	6.744	30
Vaglio et al. [6]	n/a	n/a	n/a	n/a	0.804*	n/a	n/a
Du et al. [7]	1536	n/a	n/a	n/a	0.791	n/a	19.2
Me+Ha+Ly	1536	0.800	1.396	291	0.818*	7.205*	16*
					0.602	6.480	33

[1] Pohle et al. On Rhythm and General Music Similarity. ISMIR 2009

[2] Foughmand et al. Deep-rhythm for tempo estimation and rhythm pattern recognition. ISMIR 2019

[3] Zeghidour et al. Fully convolutional speech recognition. arXiv:1812.06864, 2018

[4] Collobert et al. Wav2Letter: an End-to-End ConvNet-based Speech Recognition System. arXiv:1609.03193, 2016

[5] Doras et al. Combining musical features for cover detection. ISMIR, 2020

[6] Vaglio et al. The Words remain the Same: Cover Detection with Lyrics Transcription ISMIR, 2021

[7] Du et al. Bytecover2: Towards dimensionality reduction of latent embedding for efficient cover song identification. ICASSP, 2022