



Evaluation (audio)			
genre	recall_at_k	mrr_at_k	n_queries
hoerspiel	0.81	0.51	150
vocal jazz	0.29	0.14	150
lounge	0.25	0.08	150
latin	0.23	0.12	150
album rock	0.22	0.11	150
hard rock	0.22	0.11	150
soul	0.2	0.09	150
adult standards	0.19	0.1	150
art rock	0.19	0.1	150
rock	0.19	0.09	150
classic rock	0.17	0.07	150
mellow gold	0.15	0.07	150
soft rock	0.13	0.08	150
folk rock	0.11	0.04	150
country rock	0.1	0.05	150

Evaluation (audio & audio+lyrics)								
genre	baseline_available	recall_audio	recall_audio_lyrics	delta_recall	mrr_audio	mrr_audio_lyrics	delta_mrr	alpha_audio_weight
folk rock	1	0.11	0.14	0.03	0.04	0.02	-0.03	0.8
classic rock	1	0.17	0.16	-0	0.07	0.09	0.01	0.8
art rock	1	0.19	0.15	-0.04	0.1	0.1	-0	0.8
soft rock	1	0.13	0.08	-0.05	0.08	0.02	-0.06	0.8
rock	1	0.19	0.13	-0.06	0.09	0.06	-0.02	0.8
mellow gold	1	0.15	0.07	-0.09	0.07	0.01	-0.06	0.8
album rock	1	0.22	0.13	-0.09	0.11	0.06	-0.05	0.8
hard rock	1	0.22	0.1	-0.12	0.11	0.02	-0.09	0.8
psychedelic rock	0	null	0.42	null	null	0.2	null	0.8
symphonic rock	0	null	0.37	null	null	0.13	null	0.8

Outcomes (Genre-Based Similarity + Lyrics Experiment)

- **Genre segmentation:** Partitioned the catalog into 4,406 genres (\approx 432K tracks), enabling genre-specific retrieval and evaluation.
- **Top predictors:** Feature importance indicates similarity is driven primarily by tempo and loudness, with secondary effects from energy/acousticness/valence, varying by genre.
- **Effectiveness metric:** Measured ranking quality using Recall@10 and MRR@10 (higher is better), providing an objective retrieval benchmark per genre.
- **Lyrics impact ($\alpha=0.8$):** Adding lyrics produced mixed-to-negative deltas in Recall@10 and MRR@10 across comparable genres, suggesting the current text signal (API lyrics + baseline weighting) is not consistently additive and likely needs tuning (α , text representation, and sampling).