A reliable AFP dataset with real broadcasts and production music

ISMIR 2022 BENGALURU

BAF: An Audio Fingerprinting Dataset for Broadcast Monitoring

Guillem Cortès [♩♪] Alex Ciurana [♩] Emilio Molina [♩] Marius Miron [♪] Owen Meyers [♯] Joren Six [♭] Xavier Serra [♪]

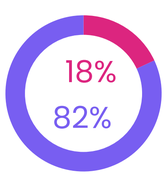
[♩] BMAT Licensing S.L. [♪] MTG, Universitat Pompeu Fabra [♯] Epidemic Sound [♭] IPEM, Ghent University

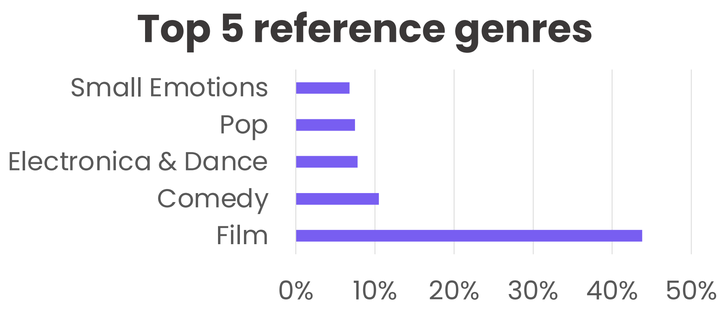
1. Problem context

Identify music in a broadcast stream giving exact *start* and *end* matching times. Music has a variable SNR and is usually in background with speech and SFXs on top.

2. BAF Dataset

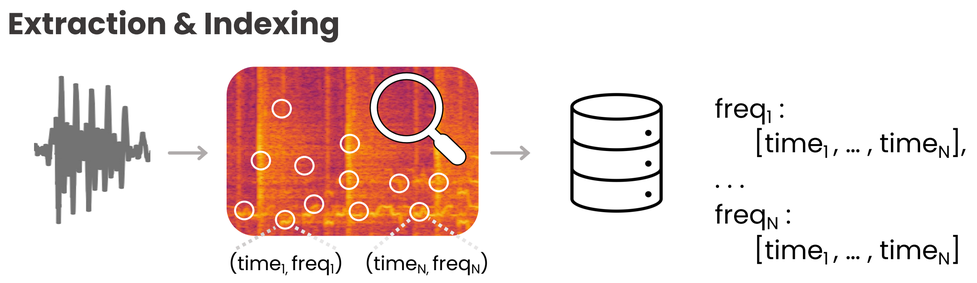
- The only publicly available broadcast monitoring dataset.
- Self-contained dataset, open for non-commercial research.
- Background music with variable SNR.
- Low sample rate, monaural audio.
  - 74 hours production music tracks from Epidemic Sound catalogue (2,000 entries).
  - 57 hours of 1-minute TV recordings (3,425 entries) from 203 TV channels across 23 countries.
  - +37 hours of matched query-reference.
  - Cross-annotated exact matching times by 6 annotators with high reliability annotation: >0.93 Fleiss' Kappa factor.

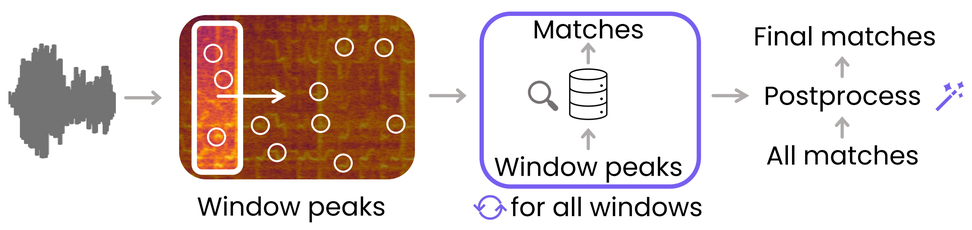




3. Proposed baseline: PeakFP

- Slow, simple, no-scalable, no-optimized AFP baseline. No usable on an industrial scale.
- Single-peak matching.





4. Benchmark

Algorithm	Match Ratio*	# matches Prec.	# matches GT Rec.	seconds Prec.	seconds Rec.	seconds F1
PeakFP (baseline)	1.64	.96	**.72**	.96	**.32**	**.47**
Panako2.0 [1]	1.85	**.98**	.21	**.98**	.06	.12
Panako2.0 (x1.5) [1]	2.12	.70	.41	.69	.15	.25
Olaf [2]	1.95	**.98**	.14	**.98**	.06	.11
NeuralFP [3]	1.39	.22	.23	.37	.10	.15
NeuralFP-spcm1510 [3]	1.56	.23	.45	.38	.22	.28
NeuralFP-spc3000 [3]	1.40	.69	.31	.83	.13	.22
Audfprint [4]	N/A△	.76	.05	.86	.02	.04
Audfprint (x2) [4]	N/A△	.71	.10	.81	.04	.08

* Match Ratio = # TP segments identified / # TP segments Groundtruth

△ Not Applicable. Audfprint reports only 1 match per query.

Algorithm	Extraction & Indexing	Matching	Index Size
Olaf [2]	53m	3h 30m	349 MB
Panako 2.0 [1]	2h 17m	5h 24m	273 MB
NeuralFP [3]	49h 34m	9h 30m	37 MB
Audfprint [4]	9h 50m	23h 01m	19 MB
PeakFP (baseline)	50m	98h 39m	160 MB

* Benchmarked in a 98GB RAM server with two 16-cores CPUs at 2.6GHz

- NeuralFP extraction is the slowest due to neural model complexity. It can be optimized with a GPU.
- PeakFP single peak matching produces a high collision of matches. High computational cost.

- Dataset is not challenging to False Positives, hence the high Precision results.
- Low Recall values indicate that algorithms have not been studied to work with background music.
- Recall in seconds decreases because identifications are partial and don't cover the full annotation groundtruth.
- Precision increases in seconds identified since False Positives are shorter in time than True Positives.
- Increasing FP density (Panako, Audfprint) boosts Recall at expense of Precision.

Paper #228

Audio Fingerprinting is not solved for Broadcast Monitoring

[1] Six, J., & Leman, M. (2014). Panako: a scalable acoustic fingerprinting system handling time-scale and pitch modification. In *15th International Society for Music Information Retrieval Conference (ISMIR-2014)*.

[2] Six, J. (2020). Olaf: Overly lightweight acoustic fingerprinting. Late-Breaking Demo In *21st International Society for Music Information Retrieval Conference (LBD ISMIR 2020)*.

[3] Chang, S., Lee, D., Park, J., Lim, H., Lee, K., Ko, K., & Han, Y. (2021). Neural audio fingerprint for high-specific audio retrieval based on contrastive learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP-2021).

[4] Ellis, D. (2014). The 2014 labrosa audio fingerprint system. In *15th International Society for Music Information Retrieval Conference (ISMIR-2014)*.

La Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación (The Research National Agency of the Ministry of Science and Innovation) supports the call *Retos-Colaboración 2019*, project *NextCore – New generation of music monitoring technology*, RTC2019-007248-7. Also, with the support of Industrial Doctorates plan from *Secretaria d'universitats i Recerca*, *Departament d'Empresa i Coneixement* de la *Generalitat de Catalunya*. DI46-2020.