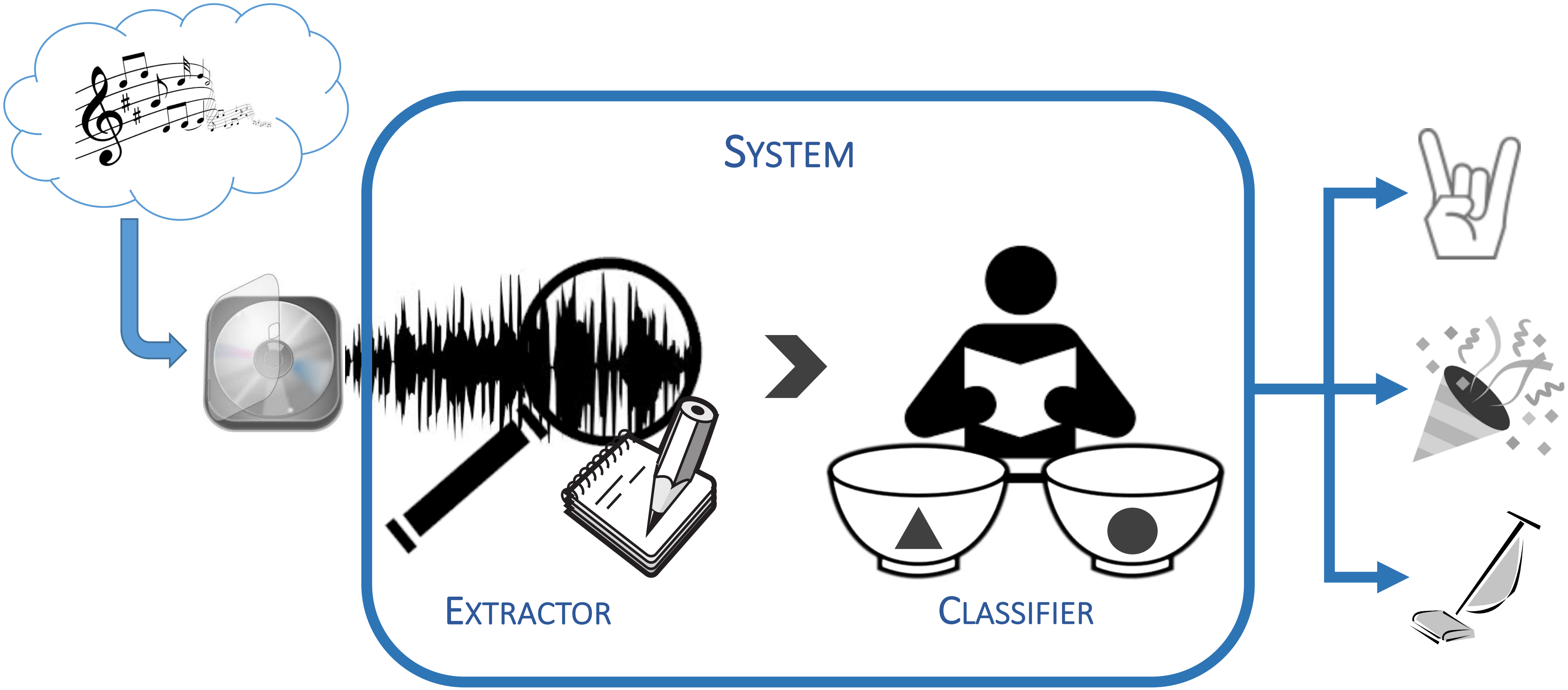
You Don’t Hear a Thing…
But My Horse Knows It’s Rock!

Francisco Rodríguez-Algarra

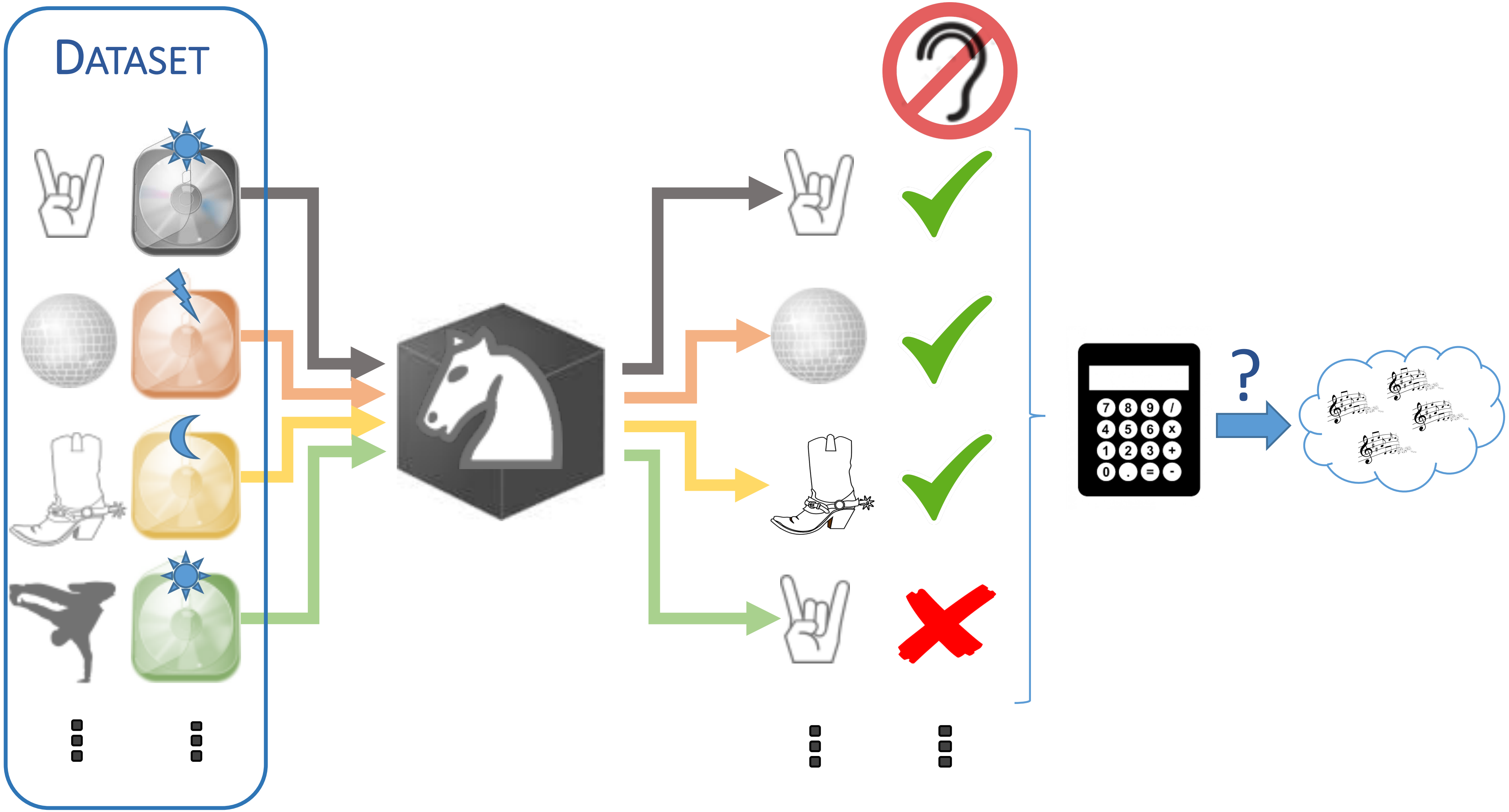
September 19th, 2016

Aims of MIR Research

A Music Listening Machine



EVALUATING MUSIC LISTENING MACHINES



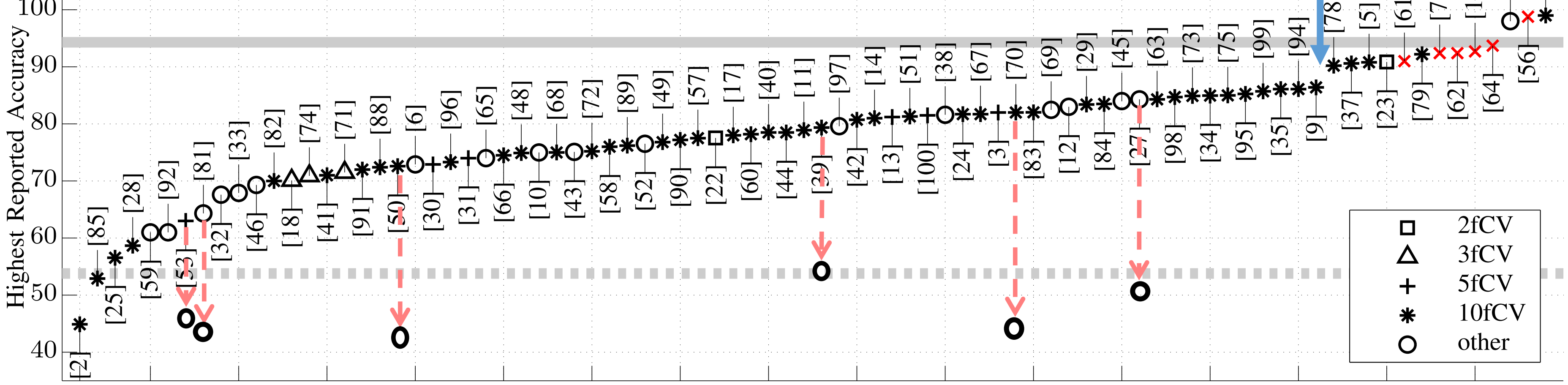
GTZAN Dataset [1]

Contents

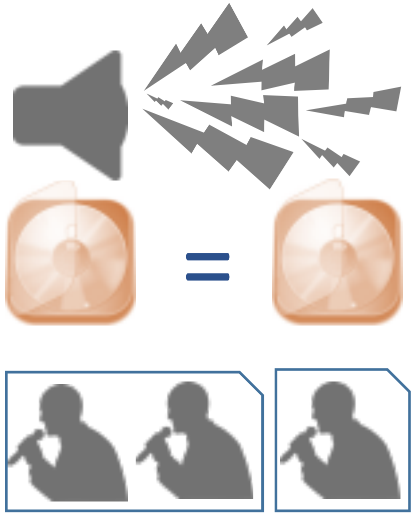
100 30-s excerpts of each of 10 classes

Benchmarking Results

Scattering Features + SVM [2]



But…



[1] Tzanetakis, G., & Cook, P. (2002). Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing 10(5)*, 293-302
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. IEEE Transactions on Signal Processing 62(16), 4114-4128

CASE STUDY

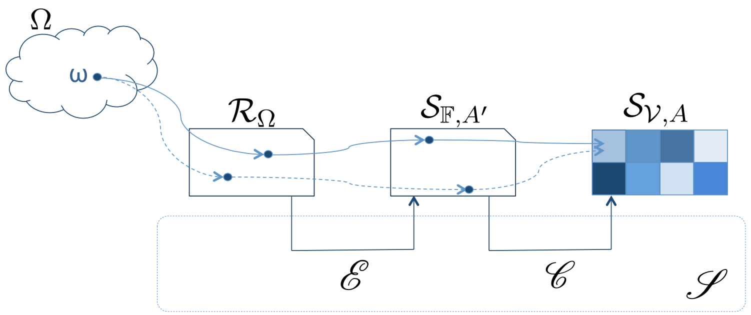
OR

Why do scattering-based music content analysis systems reproduce so many GTZAN annotations?

[3] Rodríguez-Algarra, F., Sturm, B. L., and Maruri-Haguilar, H. (2016). Analysing Scattering-based Music Content Analysis Systems: Where's the Music? In Proc. 17th *International Society for Music Information Retrievla Conference (ISMIR'16)*, New York City, NY, USA (Aug. 2016)

APPROACH

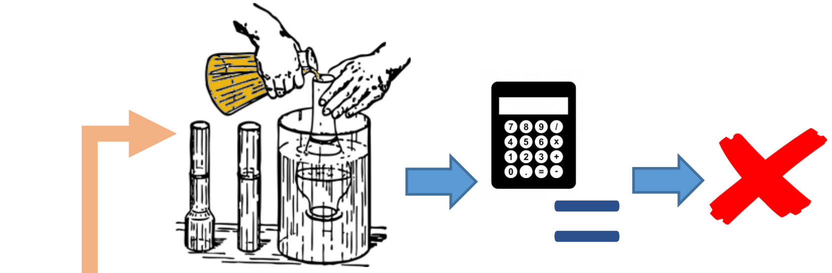
SYSTEM ANALYSIS

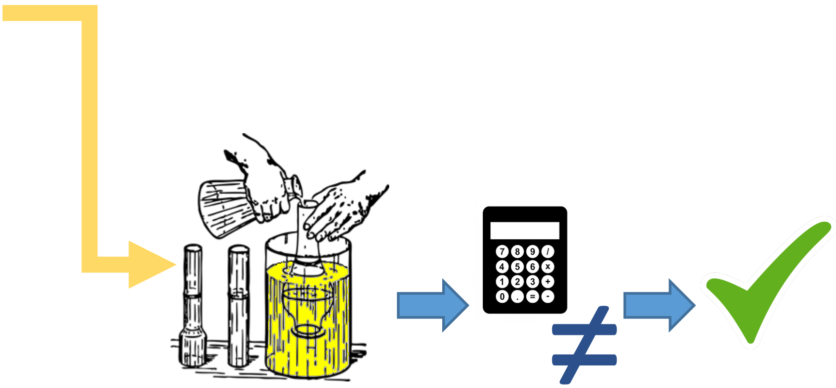


[+ GTZAN FAULTS]

INTERVENTION EXPERIMENTS

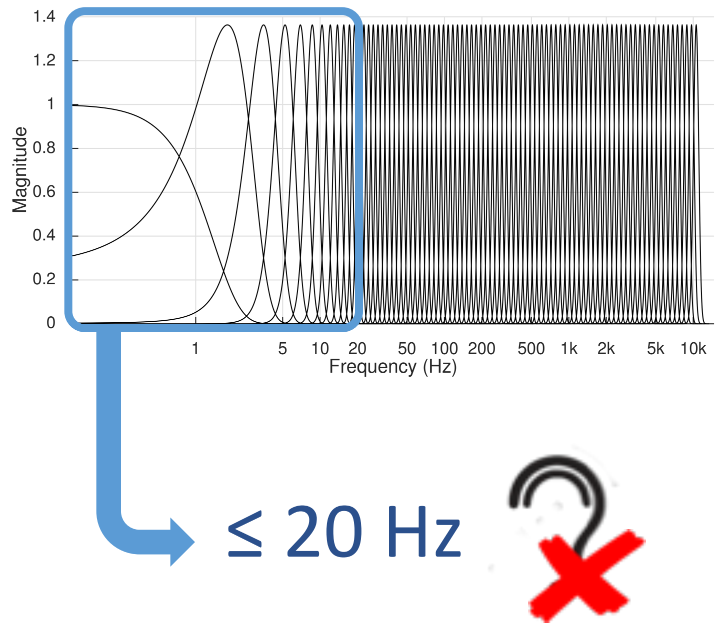
ID	Normalised accuracies [2]
a	82.0 ± 4.2
b	80.9 ± 4.5
c	89.3 ± 3.1
d	90.7 ± 2.4
e	91.4 ± 2.2
f	89.4 ± 2.5

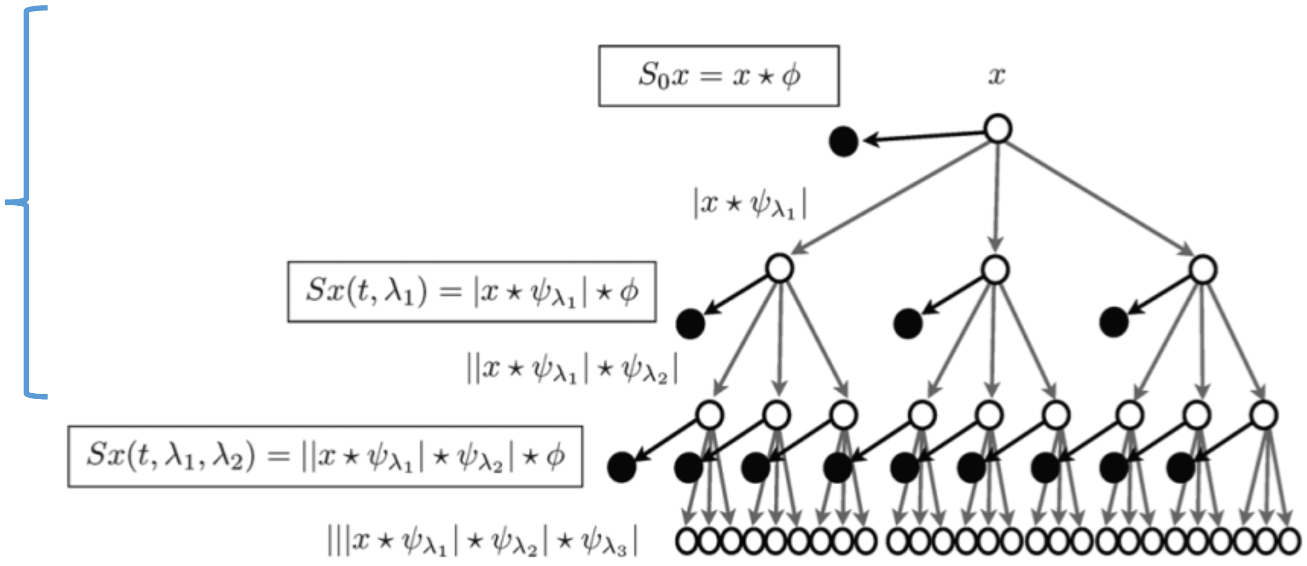




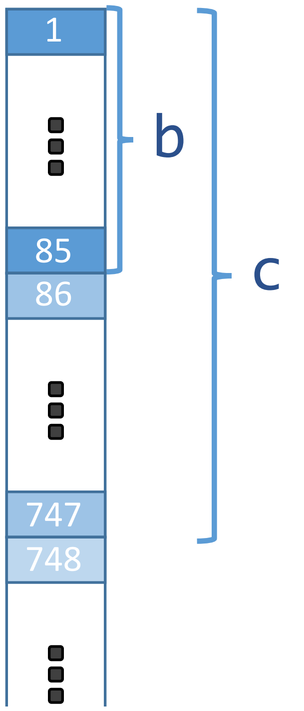
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

System Analysis: The Extractor



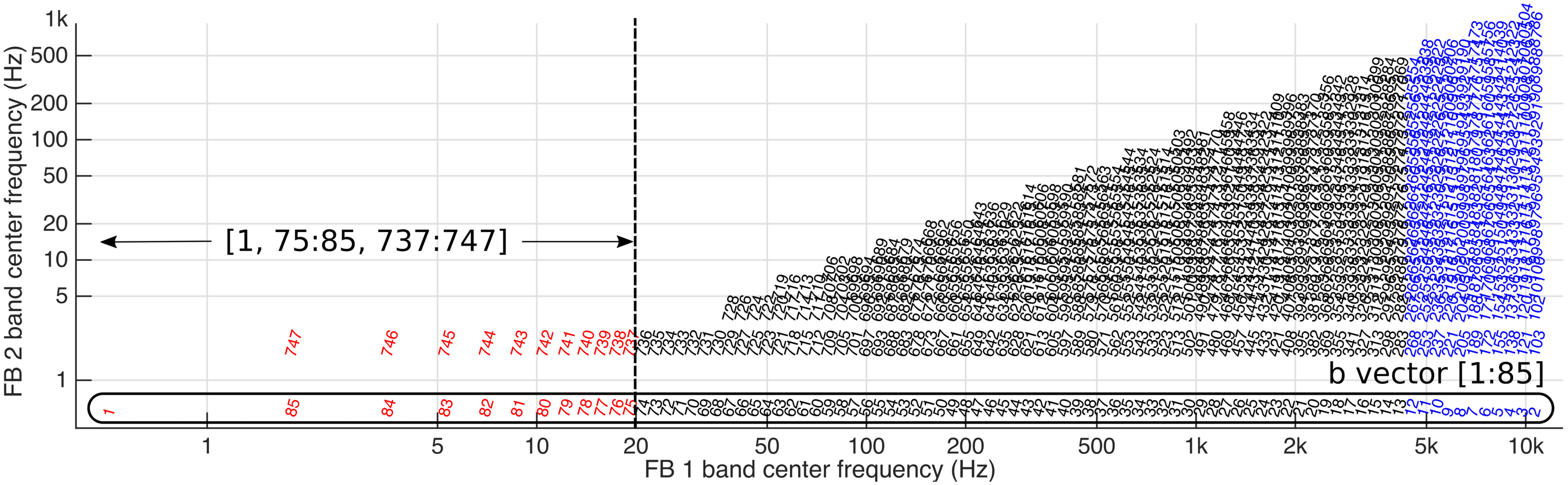


Convolutional structure of the Scattering Transform [2]



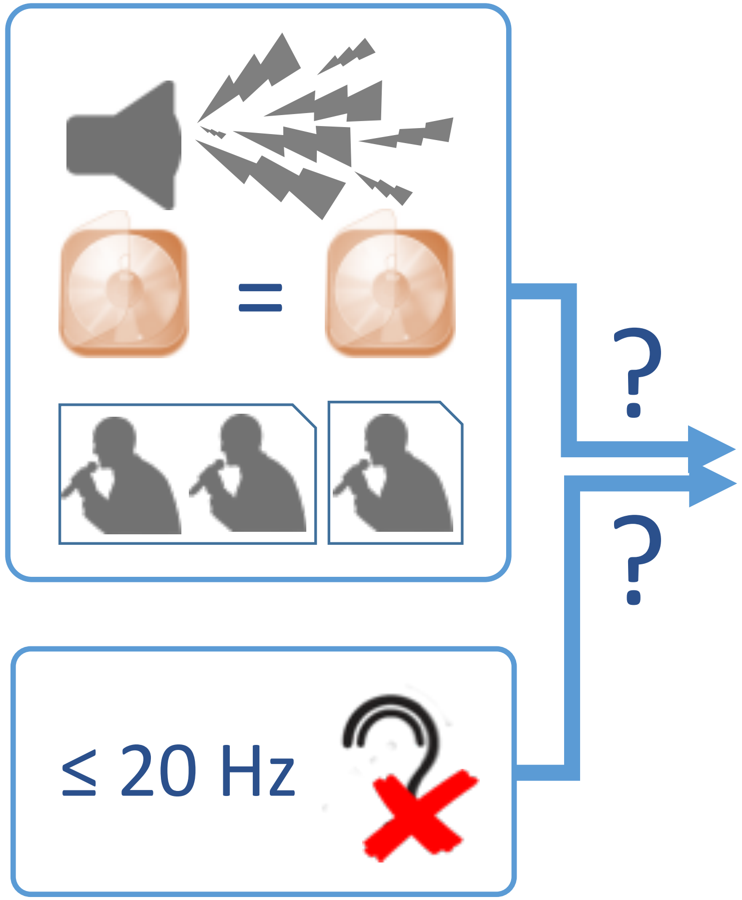
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

EXTRACTOR: FILTER CENTRE FREQUENCY VS. FEATURE DIMENSION



Information below 20 Hz could be captured

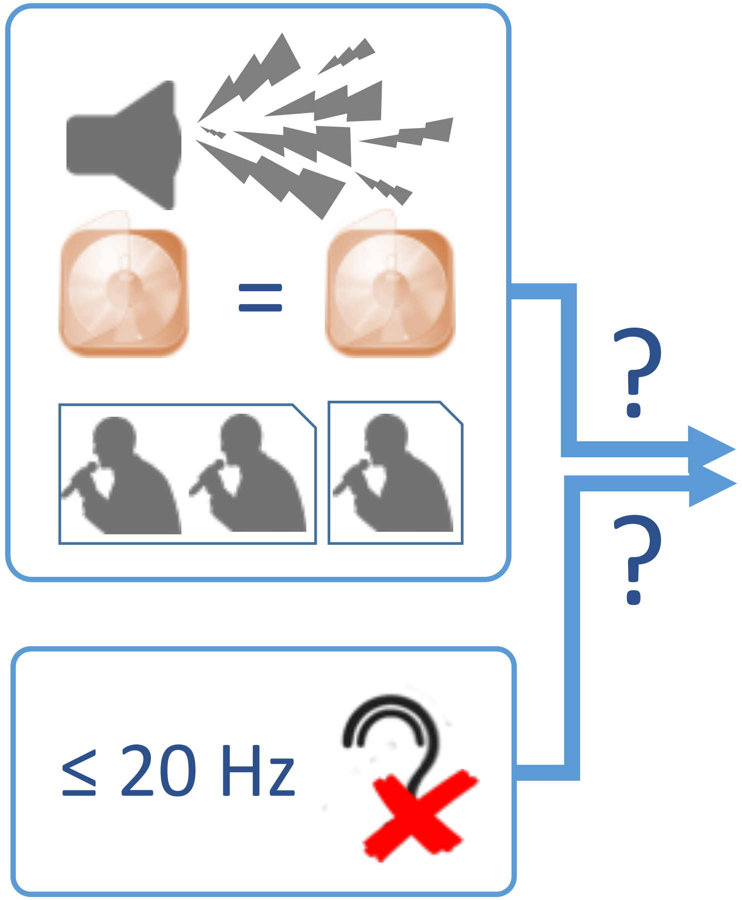
Intervention Experiments



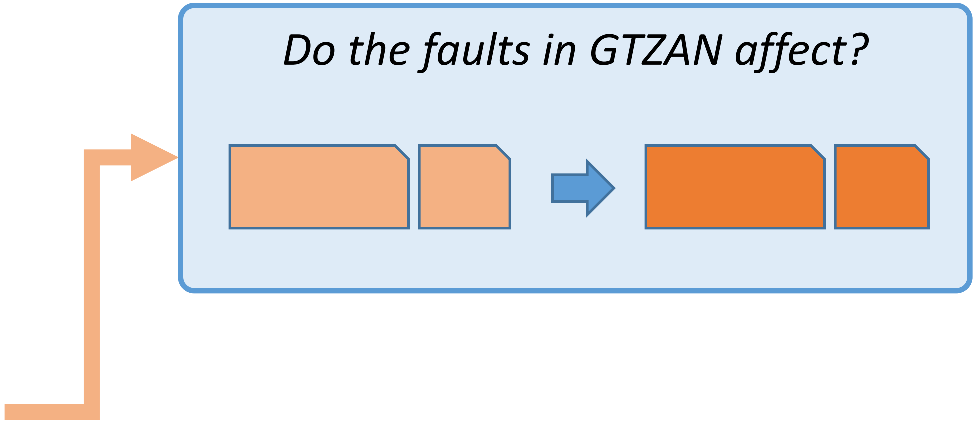
ID	Normalised accuracies [2]
a	82.0 ± 4.2
b	80.9 ± 4.5
c	89.3 ± 3.1
d	90.7 ± 2.4
e	91.4 ± 2.2
f	89.4 ± 2.5

[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

Partitioning Intervention



ID	Normalised accuracies [2]
a	82.0 $\pm$ 4.2
b	80.9 $\pm$ 4.5
c	89.3 $\pm$ 3.1
d	90.7 $\pm$ 2.4
e	91.4 $\pm$ 2.2
f	89.4 $\pm$ 2.5



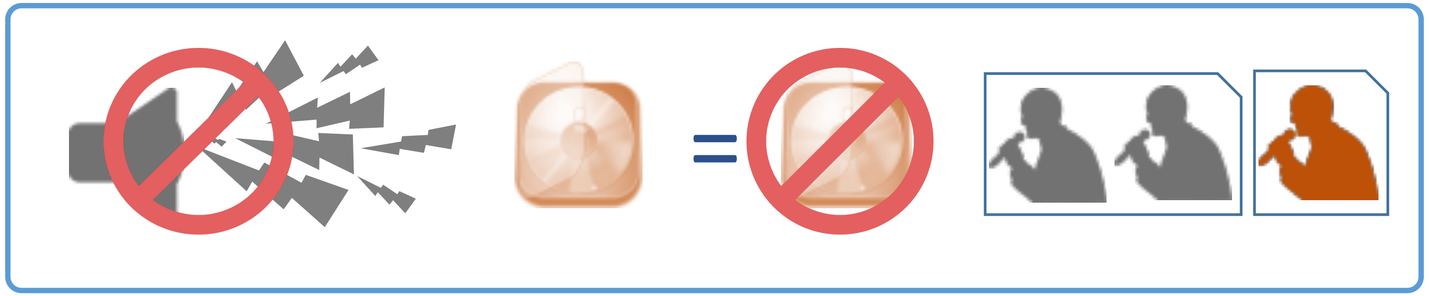
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

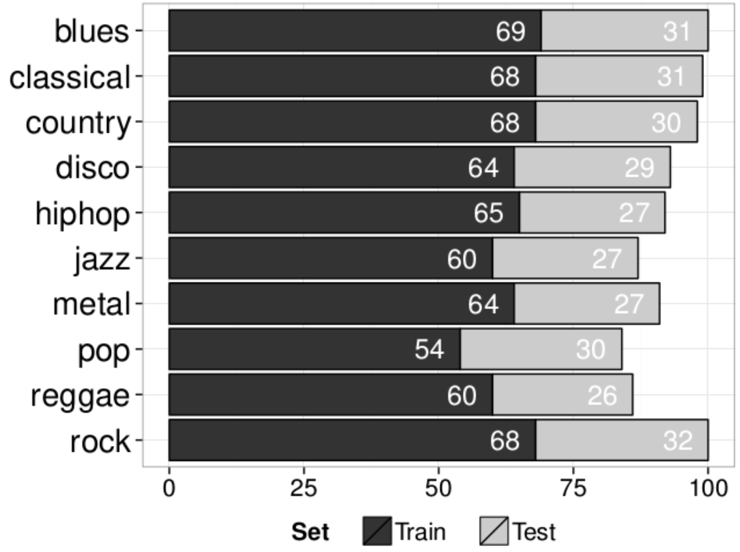
PARTITIONING INTERVENTION: METHOD

We compare two train/test partitioning conditions of *GTZAN*:

i. RANDOM:
750/250 stratified random selection
(same number of recordings per class);

ii. FAULT:
640/290 fault-filtered selection



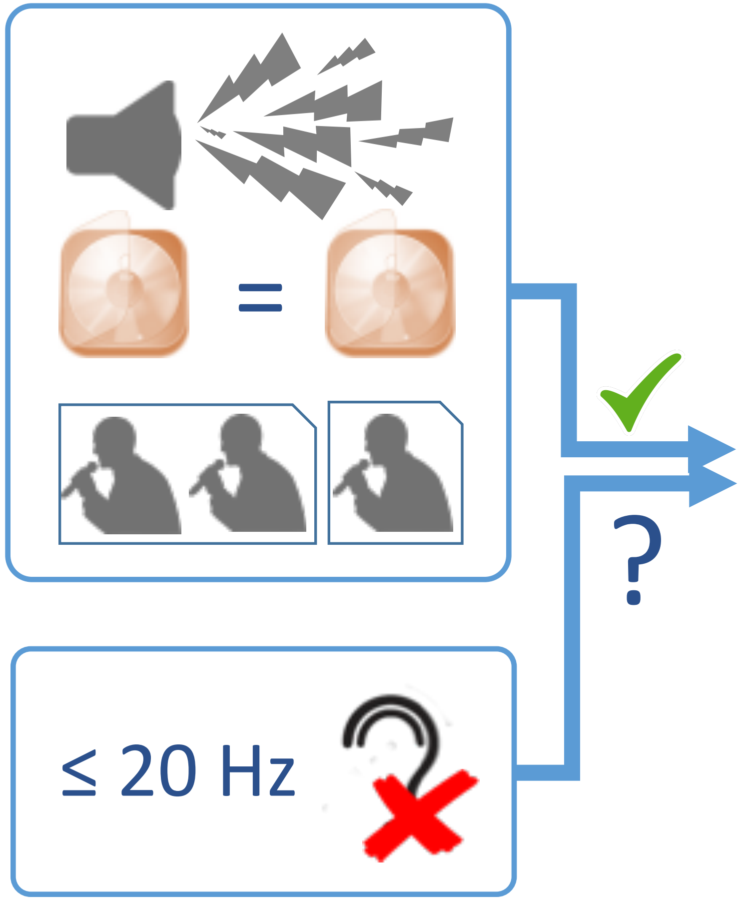


Partitioning Intervention: Results

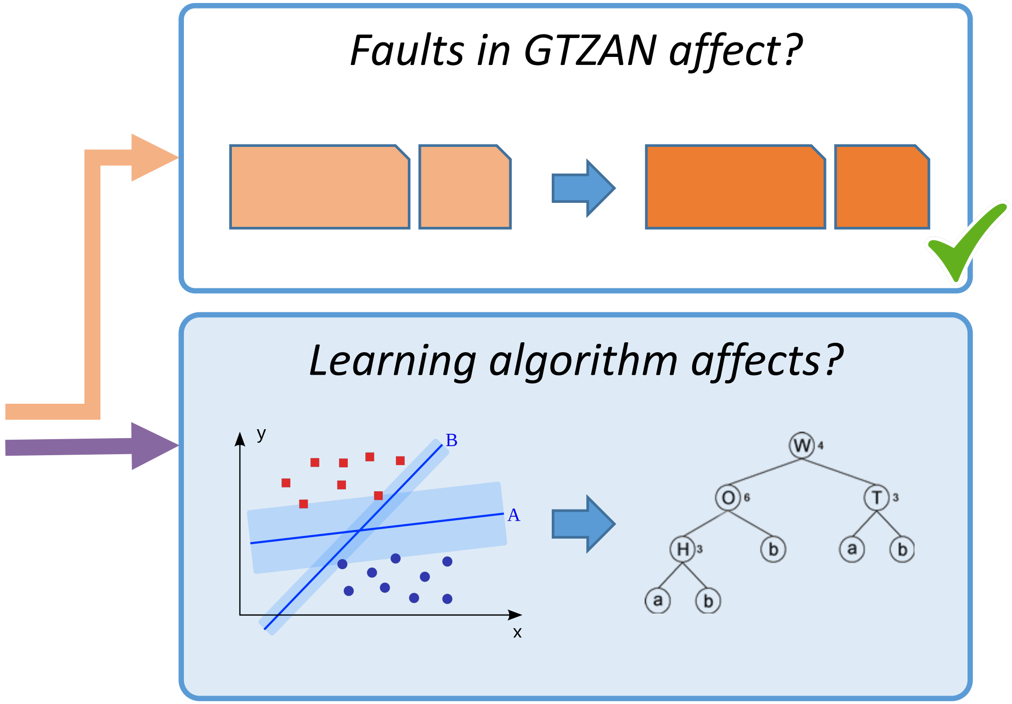
ID	Reported in [2]	*RANDOM*	*FAULT*
a	82.0 ± 4.2	78.00	53.29
b	80.9 ± 4.5	79.20	54.96
c	89.3 ± 3.1	88.00	66.46
d	90.7 ± 2.4	87.20	68.49
e	91.4 ± 2.2	85.60	68.61
f	89.4 ± 2.5	83.60	68.32

[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

Classifier Intervention



ID	Normalised accuracies [2]
a	82.0 ± 4.2
b	80.9 ± 4.5
c	89.3 ± 3.1
d	90.7 ± 2.4
e	91.4 ± 2.2
f	89.4 ± 2.5



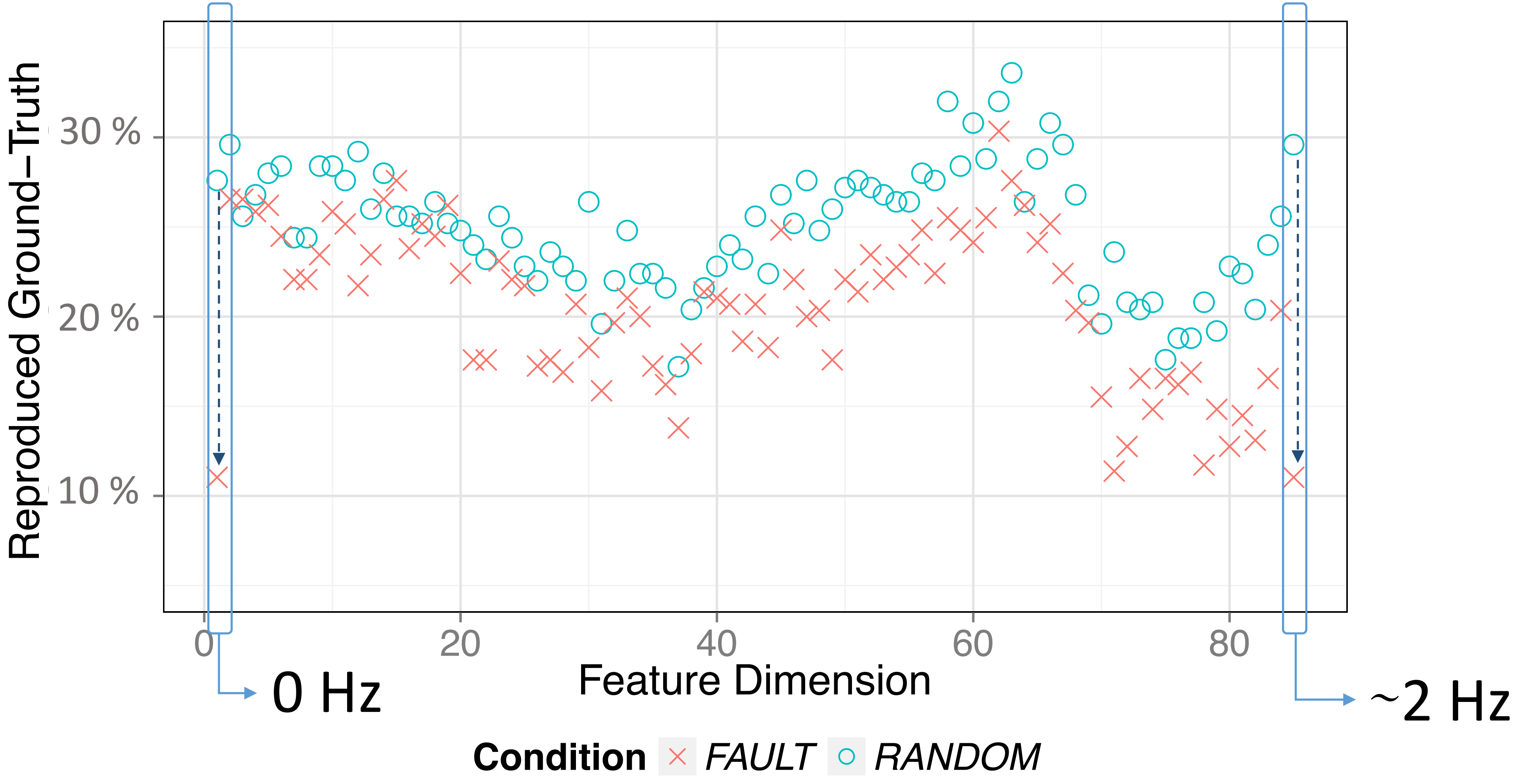
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

Classifier Intervention: Full Feature Vectors

	SVM			BDT	
ID	**Reported in [2]**	***RANDOM***	***FAULT***	***RANDOM***	***FAULT***
a	82.0 ± 4.2	78.00	53.29	72.80	45.70
b	80.9 ± 4.5	79.20	54.96	71.60	42.35
c	89.3 ± 3.1	88.00	66.46	80.00	49.91
d	90.7 ± 2.4	87.20	68.49	79.20	46.81
e	91.4 ± 2.2	85.60	68.61	79.60	44.77
f	89.4 ± 2.5	83.60	68.32	79.20	46.48

[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

Classifier Intervention: Single Dimensions



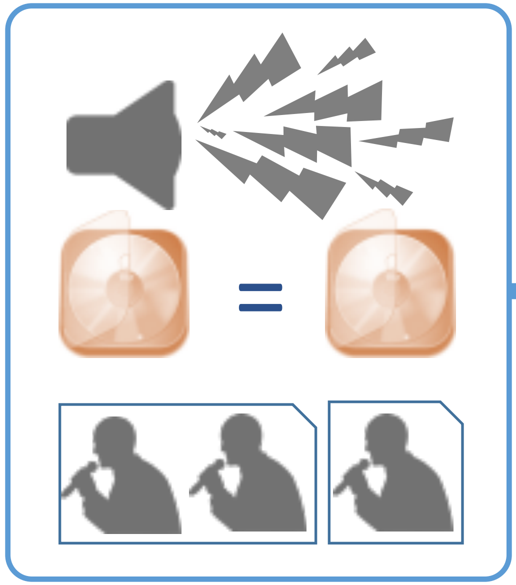
Classifier Intervention: Only Dimensions Below 20 Hz

BDT in RANDOM with only dimensions **[1, 75-85]** …

… ≈ originally reported for GTZAN! [1]

[1] Tzanetakis, G., & Cook, P. (2002). Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing 10(5),* 293-302

FILTERING INTERVENTION

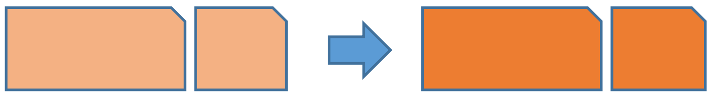


≤ 20 Hz

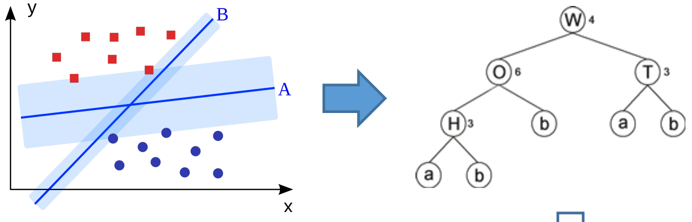
?

ID	Normalised accuracies [2]
a	82.0 ± 4.2
b	80.9 ± 4.5
c	89.3 ± 3.1
d	90.7 ± 2.4
e	91.4 ± 2.2
f	89.4 ± 2.5

Faults in GTZAN affect?



Learning algorithm affects?



≈

Information ≤ 20 Hz affects?

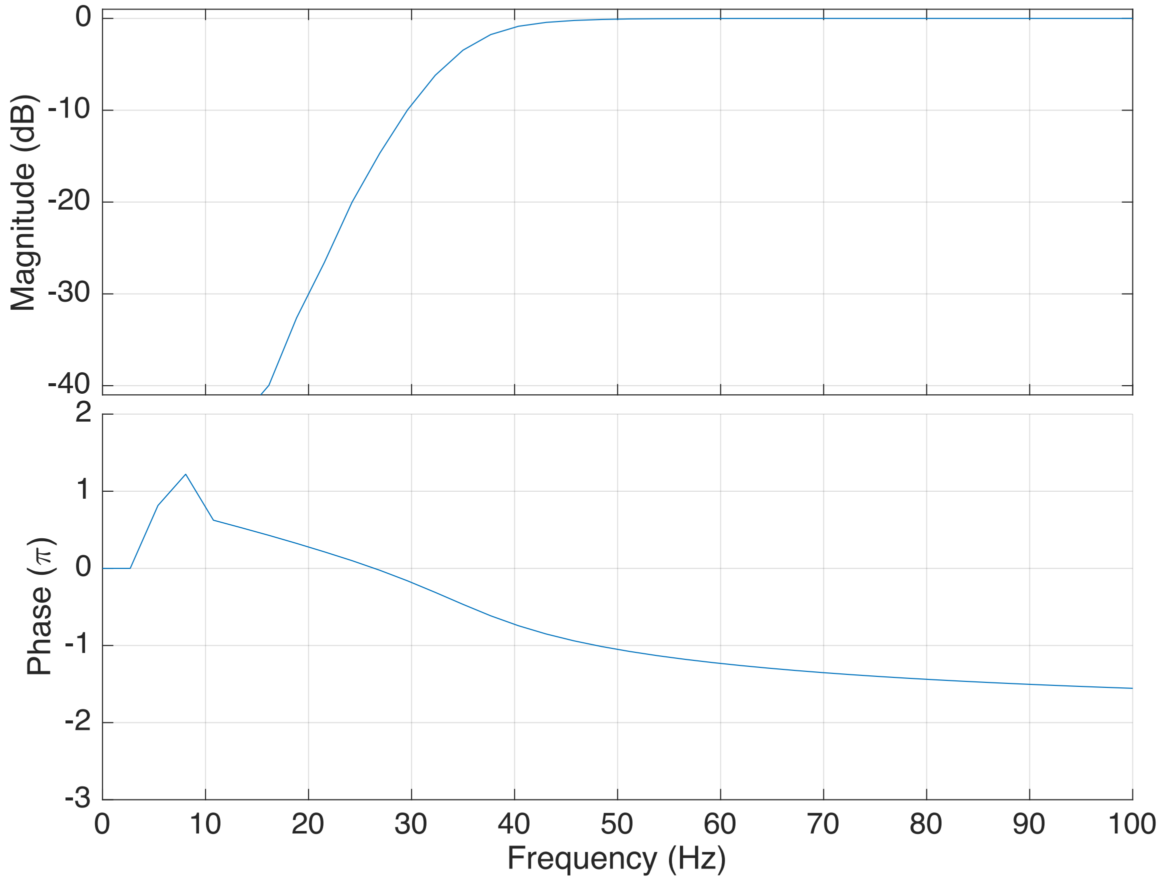
[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16)*, 4114-4128

FILTERING INTERVENTION: METHOD

1. We apply a 5th order Butterworth HPF to test recordings
   - Power at frequencies below 20 Hz attenuated at least 30 dB

2. We check we do not perceive differences

3. We test SVM systems trained in the partitioning intervention with filtered recordings

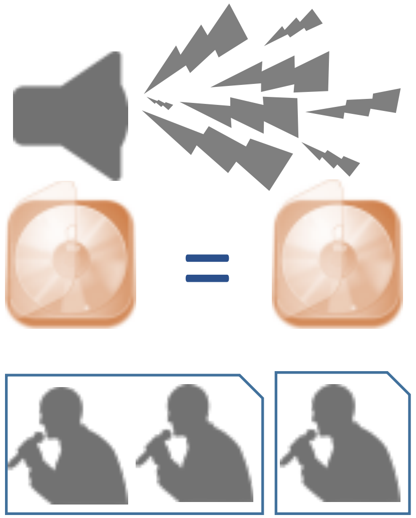


Filtering Intervention: Results

	Original *GTZAN* recordings			Attenuated [0, 20] Hz	
ID	Reported in [2]	*RANDOM*	*FAULT*	*RANDOM*	*FAULT*
a	82.0 ± 4.2	78.00	53.29	39.20	30.09
b	80.9 ± 4.5	79.20	54.96	31.60	22.42
c	89.3 ± 3.1	88.00	66.46	50.80	44.47
d	90.7 ± 2.4	87.20	68.49	62.40	55.11
e	91.4 ± 2.2	85.60	68.61	64.80	44.52
f	89.4 ± 2.5	83.60	68.32	64.80	53.16

[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

Conclusions: From the Case Study

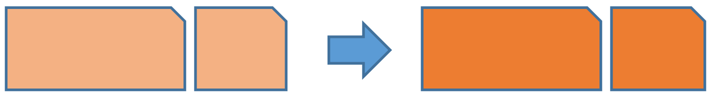


≤ 20 Hz

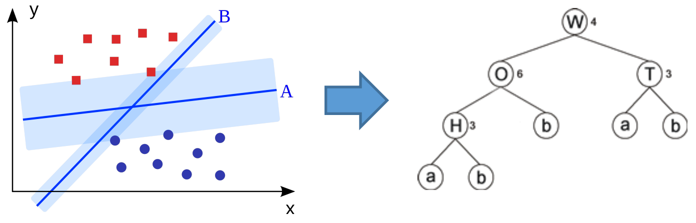
ID	Normalised accuracies [2]
a	82.0 ± 4.2
b	80.9 ± 4.5
c	89.3 ± 3.1
d	90.7 ± 2.4
e	91.4 ± 2.2
f	89.4 ± 2.5

What if we trained discarding ≤ 20 Hz dimensions?

Faults in GTZAN affect?



Learning algorithm affects?



≈

Information ≤ 20 Hz affects?

[2] Andén, J., & Mallat, S. (2014). Deep Scattering Spectrum. *IEEE Transactions on Signal Processing 62(16),* 4114-4128

CONCLUSIONS: BEYOND THE CASE STUDY

System Analysis + Intervention Experiments

- ✓ Improves Relevance of Evaluation Information
- ✓ Highlights Validity Issues
- ✗ Systematic?

Can we control for any possible confounder?

Can Experimental Design help us?

Thank YOU!!!

Any questions/criticisms/suggestions?

BONUS: Horse hearing range ≈ 50-33500 Hz