Parameter Sensitivity of Deep-Feature based Evaluation Metrics for Audio Textures

Chitralekha Gupta, Yize Wei, Zequn Gong, Purnima Kamath, Zhuoyao Li, Lonce Wyse
National University of Singapore

1. Introduction

- An audio texture is defined by statistical parameters that remain constant over time, for example the sound of wind blowing at a certain constant strength is an audio texture [1]
- Existing objective metrics have been tested for evaluating the quality of the generated audio, but not if the intended control parameter is faithfully preserved in the generated audio
- This is a systematic study of the **parameter variation sensitivity** of existing **standard metrics** and **potential Gram matrix and cochlear-model based metrics** for **audio textures** validated with subjective evaluation

2. Metrics

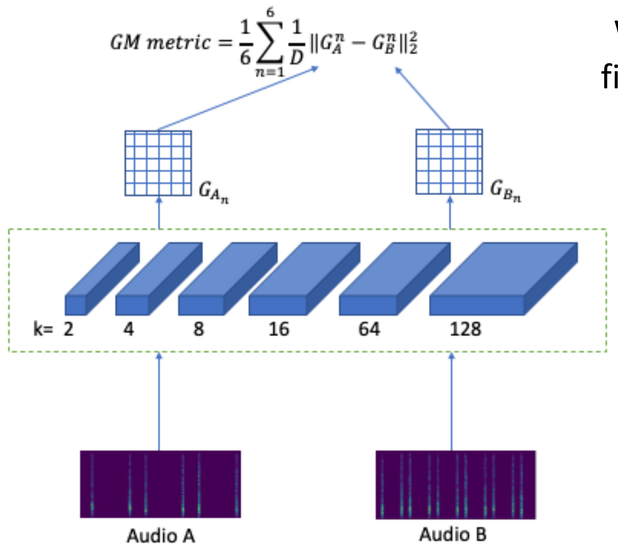
Existing Methods

- **Fréchet Audio Distance (FAD) [2]:** computes the distance between the Gaussian distributions of the embeddings of train and test set audio data extracted from a pre-trained VGGish audio classifier
- **L2 Distance:** Euclidean distance between the spectrograms of the two input audio signals to be compared

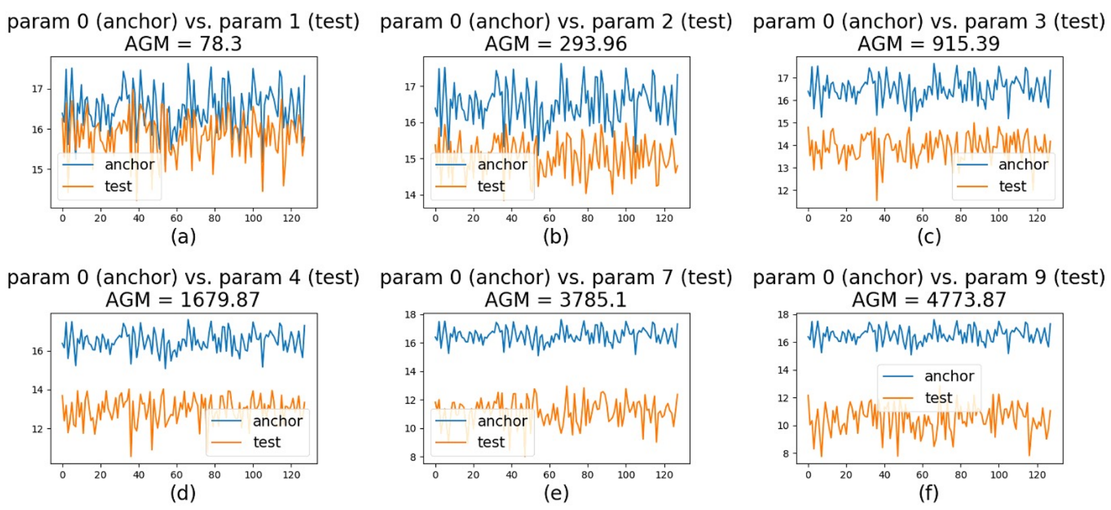
New Potential Methods

- **Gram Matrix Metric (GM):** Gram matrix is computed as the correlation between feature maps of CNN which provides a summary statistic of the audio texture [3]. GM is the mean- squared error between the Gram matrices of two textures

$$GM\ metric = \frac{1}{6}\sum_{n=1}^{6}\frac{1}{D}\|G_A^n - G_B^n\|_2^2$$



- **Gram Matrix Cosine Metric (GMcos):** cosine distance between the Gram matrices of the two textures
- **Accumulated Gram Metric (AGM):** we compute a summarizing Gram vector by aggregating the values from its six Gram matrices. AGM is the dot product of the difference between gram vectors of two audio textures



- **Cochlear Param-Metric (CPM):** A filterbank with a set of cochlear filters represent statistics of an audio texture [1]. CPM is the cosine distance between the respective statistics of the two audio textures

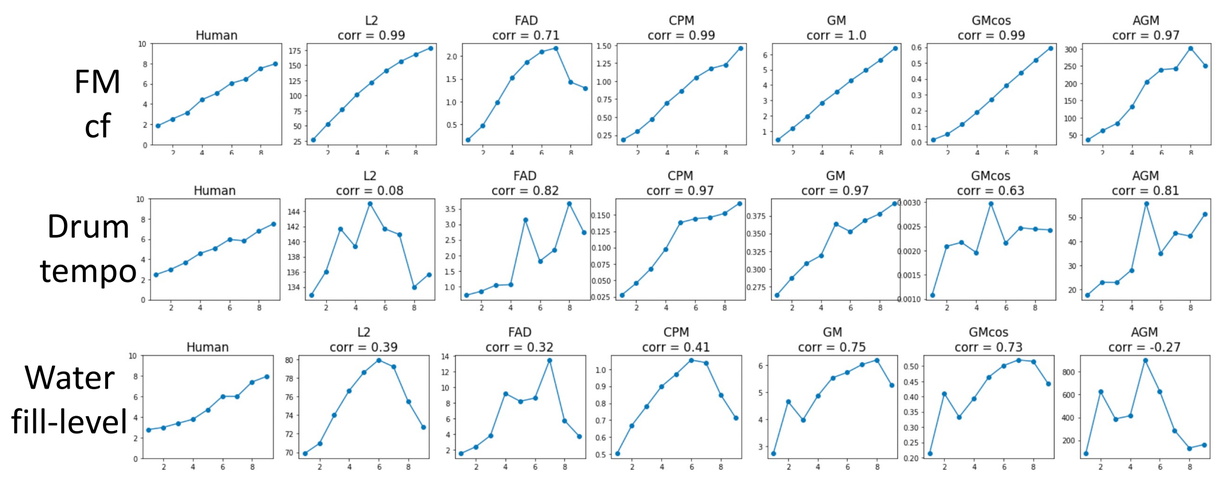
3. Dataset

- **Pitched** (FM, windchimes, brass etc.)**, Rhythmic** (tapping, drumbreak etc)**, Others** (wind, waterfill, bees etc) [4]

4. Results

Pearson's correlation between the avg human perception rank orders and the objective metric distance of test examples from the anchor (low param value) for a subset of data

Texture Param	L2	FAD	CPM	GM	GMcos	AGM
FM-cf	**0.99**	0.71	**0.99**	**1.00**	**0.99**	0.97
Windchimes-strength	**0.97**	**0.97**	0.82	0.95	0.95	0.62
Tapping-rate	**1.00**	0.90	0.91	0.64	**0.94**	0.76
Drum-tempo	0.08	0.82	**0.97**	**0.97**	0.63	0.81
Wind-strength	**0.96**	0.86	**0.88**	0.87	**0.88**	0.56
Water-fill level	0.39	0.32	0.41	**0.75**	**0.73**	-0.27
Applause-rate	0.66	0.83	**0.86**	0.66	0.78	**0.90**



5. Discussion

- A comprehensive study on the parameter change sensing property of existing audio evaluation metrics as well as three potential audio statistical and deep-feature based metrics
- CPM and FAD emerge as the best metrics, while GM based metrics show promising results
- Further investigation is required to understand how metrics behave in complex realistic situations where multiple params vary simultaneously (eg. Water-fill)
- Future work is needed to understand the behavior of these metrics for cross texture-type comparisons

6. References

[1] McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron, 71(5), 926-940.
[2] Kilgour, K., Zuluaga, M., Roblek, D., & Sharifi, M. (2019). Fréchet Audio Distance: A Reference-Free Metric for Evaluating Music Enhancement Algorithms. In INTERSPEECH (pp. 2350-2354).
[3] Antognini, J. M., Hoffman, M., & Weiss, R. J. (2019). Audio texture synthesis with random neural networks: Improving diversity and quality. In ICASSP 2019 (pp. 3587-3591). IEEE.
[4] Wyse, L., & Ravikumar, P. T. (2022, April). Syntex: parametric audio texture datasets for conditional training of instrumental interfaces. In NIME 2022.

Listening samples, dataset description, supplementary material, code

ISMIR 2022 BENGALURU