Towards an Acoustic-Semantic Space of Extreme Metal Vocal Styles

Isabella Czedik-Eysenberg[1], Eric Smialek[2], Jan-Peter Herbst[2]
[1]SInES, Department of Musicology, University of Vienna, Austria; [2]University of Huddersfield, United Kingdom

This research was generously funded by EU Horizon / the Engineering and Physical Sciences Research Council (EPSRC) in the United Kingdom under grant EP/X02590X/1.

Background

Growled vocals in **extreme metal** are characterized by **low harmonicity** and **high roughness** and are often associated with expressive traits like "aggressiveness" (Tsai et al., 2010). Audio features can help classify these vocals into broad style categories (Nieto, 2013; Kato & Ito, 2013; Kalbag & Lerch, 2022).
Despite this awareness of vocal effects specific to individual subgenres, the **perceptual** organization of these styles has not yet been empirically demonstrated via participant responses and linked to relevant **audio features**.

Aims

We aim to provide empirical evidence on how listeners interpret subgenres of extreme metal vocals. We synthesize **acoustic** and **verbal** evidence via a **semantically meaningful** space of verbal associations correlated with audio features.

Methods

We extracted short phrases from **115 professional metal vocal tracks** provided via a partnership with *Unstoppable Recording Machine*. These excerpts were used in perceptual experiments and analyzed acoustically by extracting audio features using `PRAAT`/`Parselmouth` (Boersma, 2001; Jadoul et al., 2018), `Librosa` (McFee et al., 2015), and `Essentia` (Bogdanov et al., 2013).

Experiment 1: Similarity Rating

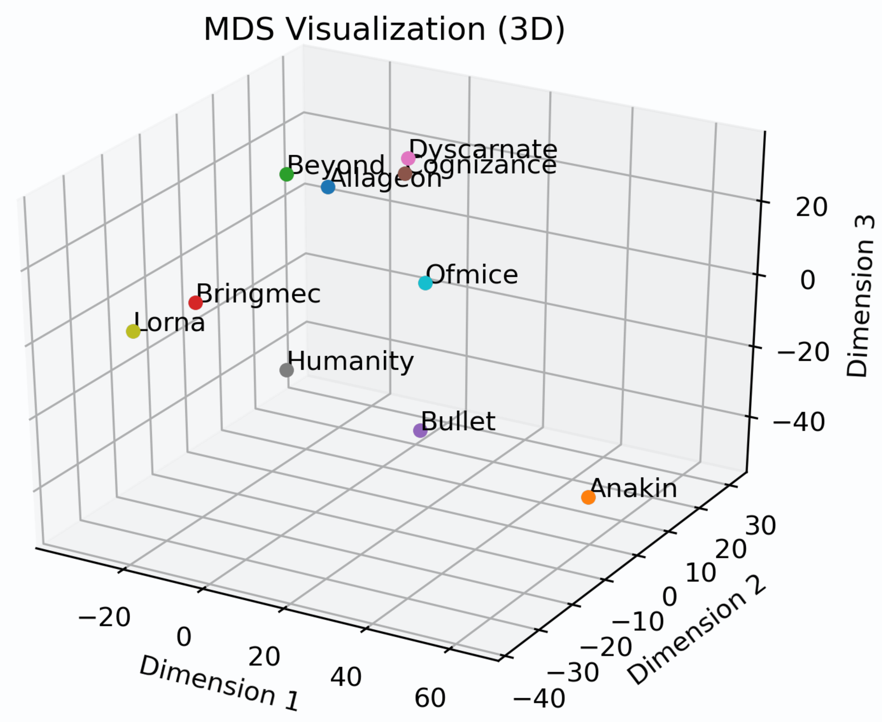
In order to identify the main **perceptual dimensions** of different metal vocal styles, 14 subjects rated a subset of 10 excerpts on a slider for **pairwise similarity** (45 comparisons). The resulting mean similarity matrix forms the basis for a perceptual similarity space computed using **multidimensional scaling** (MDS).

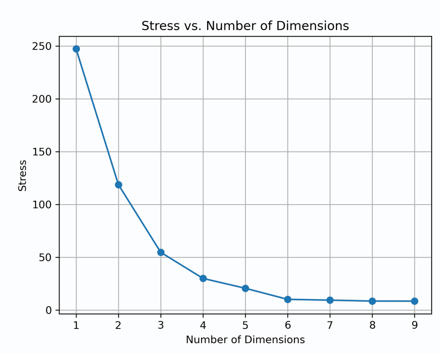
Experiment 2: Verbal Associations

In a second experiment, vocal excerpts were played to participants across the entire dataset on a self-developed web platform to collect **verbal descriptions** of the vocals.
Participants responded both by typing **free associations** and using **preselected tags**.

67 people participated in the task, providing 6,073 descriptive adjectives in total (4,493 tags and 1,580 free associations).

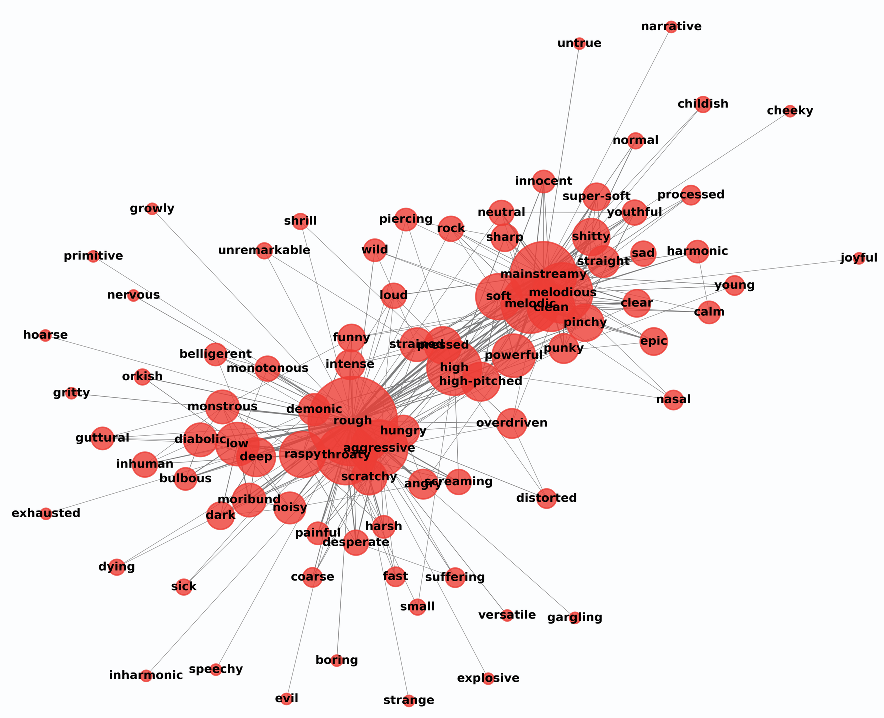
Similarity Space





MDS reveals a **three-dimensional similarity space**, with the first major axis contrasting **harmonic** vs. **inharmonic** vocals (*Harmonic-to-Noise Ratio* (HNR): $r = 0.837$, $p = 0.005$; *Spectral Complexity*: $r = -0.959$, $p < 0.001$). The second perceptual dimension shows no linear correlations with extracted sound features, while the third dimension is related to the position of the higher **formants** (e.g., *F2*: $r = -0.855$, $p = 0.003$).

Co-occurrence Network of Free Associations

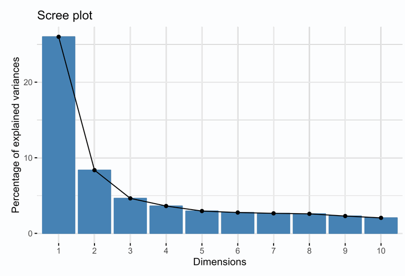


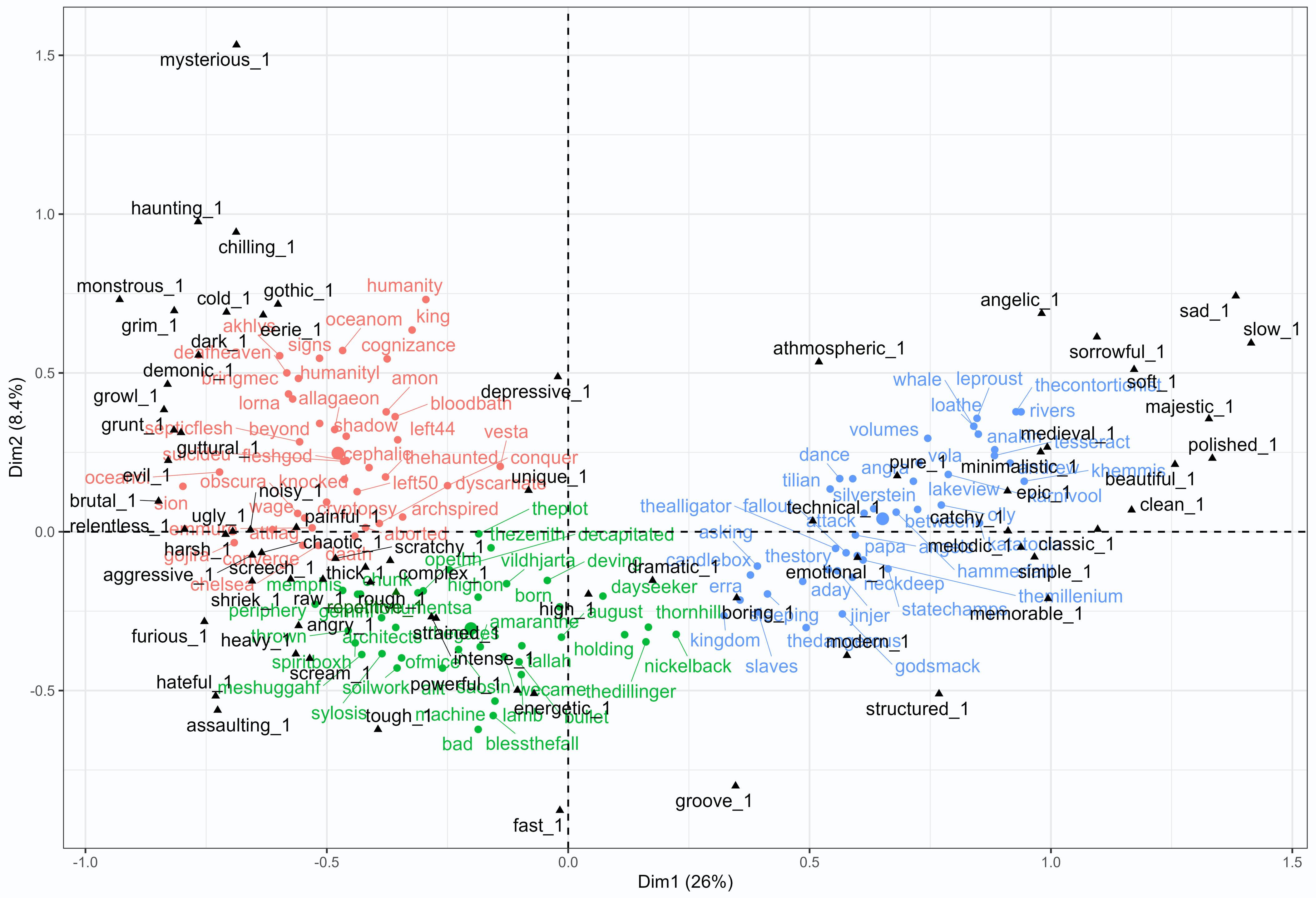
From the **free verbal associations**, we computationally constructed a **semantic network** via **co-occurrence** analysis. Words (*nodes*) were considered as co-occurring (*edges*) if they were used—by any participant—to refer to the same stimulus. In order to exclude overly idiosyncratic associations, only co-occurrences appearing in at least three stimuli were considered for constructing the graph.

Overall, a dichotomy between ***rough/raspy*** and ***clean/melodious*** vocals constitutes the predominant axis of the semantic network of verbal associations, which also shows subclusters of more fine-grained descriptions.

Multiple Correspondence Analysis of Verbal Tags

Starting with the **4,493 selected tags**, we conducted a **multiple correspondence analysis** (MCA) based on whether particular tags occurred for particular stimuli. We opted for a two-dimensional configuration, with the first dimension explaining 26% of the variance and the second dimension explaining 8.4%.





Acoustically, the **first dimension** of descriptions according to the MCA, shows a very strong correlation with *Harmonic-to-Noise Ratio* (HNR) and other related descriptors referring to aspects of **inharmonicity**, **noisiness**, and **roughness** (see table).

Dimension 1

Audio Feature	r	p
Harmonic-to-Noise Ratio	0.932	<0.001
Shimmer	-0.929	<0.001
Spectral Contrast (400-800 Hz)	0.916	<0.001
Sensory Dissonance	-0.825	<0.001

Dimension 2

Audio Feature	r	p
Valence (Model)	-0.468	<0.001
Arousal (Model)	-0.449	<0.001
Minimum Frequency (5%)	-0.391	<0.001
Formant 1	-0.263	0.004

The **second dimension** of the MCA, however, demonstrates a much less clear relationship with audio features (see table). The strongest relations are found with audio models for predicting perceived **valence** and **arousal**. It is characterized by a contrast between two groups of associations: *fast/groove/tough/energetic/assaulting* vs. *mysterious/haunting/chilling/angelic/atmospheric*.

Discussion and Conclusion

The three analytical approaches all indicate that **Harmonicity** is the most important perceptual axis for evaluating different styles of metal vocals. With the MCA, the second axis may further represent a broad dichotomy of aesthetic tropes related to ***"quotidian human toughness"*** vs. the **supernatural**. Smialek (2023) argues that this distinction sets apart traditional metal genres from more controversial, newer forms like metalcore.
Our findings can be **explored interactively** through a **web application**, allowing users to experience them both aurally and visually.

Participate in the experiment:

Explore the interactive metal vocals space:

References

Bogdanov, D., Wack, N., Gómez Gutiérrez, E., Gulati, S., Boyer, H., Mayor, O., Roma, G., Salamon, J., Zapata, J. & Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. In *14th Conference of the International Society for Music Information Retrieval (ISMIR)*, pp. 493–8. | **Boersma, P. (2001).** PRAAT, a system for doing phonetics by computer. *Glot International, 5*, 341–345. | **Jadoul, Y., Thompson, B., & De Boer, B. (2018).** Introducing Parselmouth: A Python interface to praat. *Journal of Phonetics, 71*, 1–15. | **Kalbag, V., & Lerch, A. (2022).** Scream detection in heavy metal music. *arXiv preprint* arXiv:2205.05580. | **Kato, K., & Ito, A. (2013).** Acoustic features and auditory impressions of death growl and screaming voice. In *Ninth International Conference on IIHMSP* (pp. 460–463). IEEE. | **McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E. & Nieto, O. (2015).** librosa: Audio and music signal analysis in python. *Proceedings of the 14th python in science conference* (Vol. 8, 18–25). | **Nieto, O. (2013).** Unsupervised clustering of extreme vocal effects. In *Proc. 10th Int. Conf. Advances in Quantitative Laryngology* (p. 115–116). | **Smialek, E. (2023).** Contempt-of-Core: A Reception History of Metalcore Subgenres as Abject Genres. In J.-P. Herbst (Ed.), *The Cambridge Companion to Metal Music* (pp. 281–298). Cambridge University Press. | **Tsai, C. G., Wang, L. C., Wang, S. F., Shau, Y. W., Hsiao, T. Y., & Auhagen, W. (2010).** Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception, 27(3)*, 209–222.