Background
Growing 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; Olsen et al., 2018). Audio features can help classify these vocals into broad style categories (Nieto, 2013; Kato & Ito, 2013; Kalbarg & 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).

Results 1: Similarity Space
MDS reveals a three-dimensional similarity space, with the first major axis contrasting harmonic vs. inharmonic vocals (Harmonic-to-Noise Ratio (HNR): $r = 0.837$, $p = 0.003$). Spectral Complexity: $r = -0.959$, $p < 0.001$.

The second perceptual dimension shows no significant correlations with extracted sound features, while the third dimension is related to the position of the higherformants (e.g., $F2: r = -0.855$, $p = 0.003$).

Results 2: 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%.

Discussion and Conclusion
Both experimental approaches indicate that Harmonicity is the most important perceptual axis for evaluating different styles of metal vocals. Comparing the results of the two experiments, it can be found that not only the first MCA dimension significantly correlates with the respective MDS dimension ($r = 0.901$, $p < 0.001$), but also the second MCA dimension corresponds to the (inverted) second dimension of the MDS relatively well ($r = -0.813$, $p = 0.004$).

This second dimension, however, seems to demonstrate a less clear relationship with audio features. Moderate relations are found with audio models for predicting perceived valence and arousal. The dimension is characterized by a contrast between two groups of associations: fast/groove/tough/energetic/assaulting vs. mysterious/haunting/chilling/angry/atmospheric and 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.