

Additional information

Competing interests

While this manuscript was being drafted, CL, DM, and RD were employed by and/or held equity in COMPASS Pathways, a company involved in psychedelic-assisted therapies. They are no longer affiliated with the company, and this affiliation did not influence playlist selection, data analysis, or manuscript preparation. DLB is co-founder of Spiritune, a company involved in music production for psychedelic-assisted therapy. This relationship did not influence playlist selection, data analysis, or manuscript preparation.

Section A: Musical selection

Imperial 1¹ (48 tracks; playlist name: Psychedelic Therapy Playlist 1; Author: Mendel Kaelen) and Imperial 2² (36 tracks; playlist name: Psychedelic Therapy Playlist 2; Author: Mendel Kaelen) were created for psilocybin therapy research at Imperial College London, consisting primarily of neoclassical and ambient music, with elements of jazz, classical, and non-Western styles²¹. JH Classical³ (49 tracks; playlist name: Psilocybin Research: Johns Hopkins; Author: William A. Richards, note: that was the title at the time of our analysis, the title has since been removed) and JH Overtone⁴ (31 tracks; playlist name: Johns Hopkins Psilocybin Playlist: Overtone-based music; Author: Justin Strickland; note: that was the title at the time of our analysis, the title has since been removed) were developed at Johns Hopkins University. The former consists of Western classical music; the latter contains music emphasizing instruments with particularly strong overtone signatures, such as Tibetan singing bowls, gongs, didgeridoo, chimes, bells, and sitar. These two playlists have significant overlap, with JH Overtone containing about 25 percent of the tracks in JH Classical^{39,59}. Chacruna Baldwin⁵ (39 tracks; playlist name: Psilocybin Playlist for Chacruna; Author: Matt Baldwin) and Chacruna Kelan Thomas 2⁶ (51 tracks; playlist name: 2 Kelan Thomas - A playlist for psilocybin; Author: Matt Baldwin) were assembled for psilocybin research at the Chacruna Institute in San Francisco, incorporating a mix of indie, new wave, and post-rock selections to reflect a broader contemporary aesthetic. Copenhagen⁷ (60 tracks; playlist name: Copenhagen Music Program for Psilocybin; Author: Catharina Messell) was designed as part of an academic initiative at Copenhagen University, informed by existing therapeutic playlists and Guided Imagery and Music (GIM) practices³⁸. Compass v2⁸ (68 tracks; playlist name: COMPASS Playlist v2.0 20181309; Author: George Goldsmith; note: this playlist has since been deleted or made private) was curated by COMPASS (London) as part of their clinical trials of COMP360—a synthetic form of psilocybin, and is largely comprised of new age, ambient, electronic music with elements of classical, rock, and jazz.

Several tracks appeared across multiple playlists, with Górecki's Symphony No. 3 included in four playlists (Copenhagen, JH Classical, JH Overtone, Imperial 1), and Om Namah Shivaaya in three playlists (JH Classical, JH Overtone and Copenhagen). Seventeen tracks were repeated twice, mainly among the Johns Hopkins playlists. For 6 tracks (13 repetitions) we found diverging labels from the time-based convention classification (e.g., a track would be labeled as return in Compass v2, and peak in Imperial 1)—such tracks were removed from the analysis. In contrast, 14 tracks repeated across playlists occurred in the same phase, supporting the validity of the time-based convention. All phase labels were then manually reviewed to ensure continuity within each playlist. In three cases where time-based rules split consecutive tracks from the same artist and album across phases, labels were adjusted to preserve coherence. For example, Górecki's Symphony No. 3 was retained as an “onset” track across playlists, resulting in a larger onset phase for Imperial 1.

After removal of repeated tracks, the final sample comprised 369 tracks across the eight playlists, with 528 minutes of audio in the onset phase (playlist mean = 66 minutes, range = 53–91), 961 minutes in the peak phase (mean = 120, range = 99–141), and 1217 minutes in the return phase (mean = 152, range = 127–200).

Section B: Alternate time-based boundary analysis

To address potential artifacts from our predefined phase boundaries (at 60 and 180 minutes) and class imbalance in the dataset, we tested several variations: (a) excluding tracks spanning phase boundaries, (b) clipping the return phase duration to match the shortest playlist (Imperial 2; 126.9 minutes), (c) applying SMOTE (Synthetic Minority Oversampling Technique,⁶⁰), and (d) random undersampling to balance class sizes.

¹<https://open.spotify.com/playlist/2mT6IpOU4ipJ0BkoCigAiw>

²<https://open.spotify.com/playlist/6q0rnAIIxokcRlu6vRJPNX>

³<https://open.spotify.com/playlist/5KWf8H2pM0tlVd7niMtqeU>

⁴<https://open.spotify.com/playlist/400OW1CBqejT3j0r56FDTS>

⁵<https://open.spotify.com/playlist/2KrqZEhp2qUrWD6UJlFett>

⁶<https://open.spotify.com/playlist/2szSJ3rB6oQUn6ywhEGzO>

⁷<https://open.spotify.com/playlist/6QqLlJMtGAlw40kcMtBGDr>

⁸<https://open.spotify.com/playlist/53Z2M4AQxO4s2R1MUMVdEO>

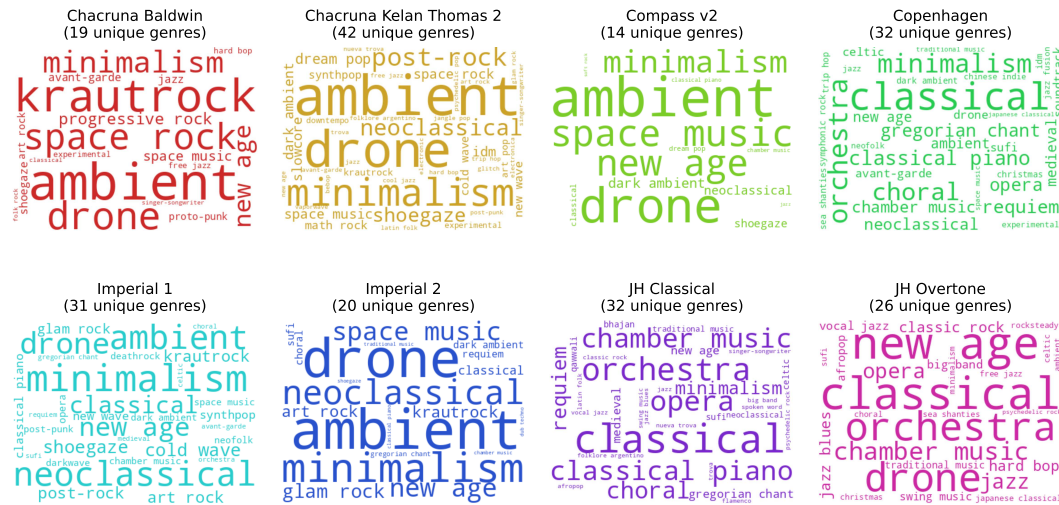


Figure 8. Genre composition of the psilocybin therapy music playlists. Word clouds representing the most common genres in each of the eight playlists. Font size indicates genre frequency. Playlist titles are shown above each word cloud.

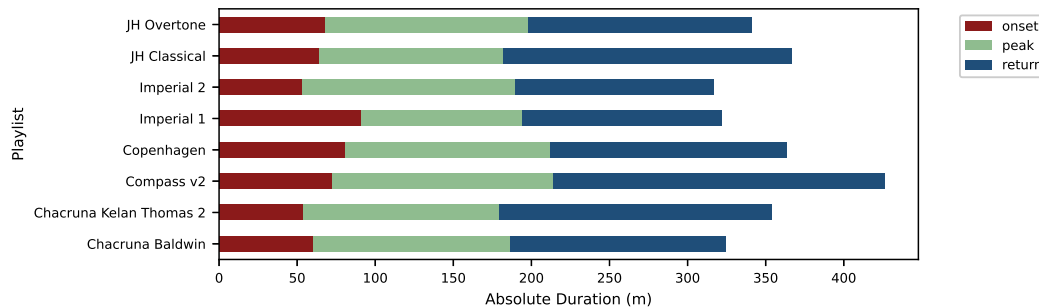


Figure 9. Absolute playlist duration. Stacked bar plot showing the distribution of playlist durations (in minutes) across the three phases (onset, peak, return).

Supervised classification

At the track level, the default random forest model significantly outperformed the majority-class baseline (macro $F_1 = 0.55$, $t(4) = 7.43$, $p = 0.002$), with SMOTE and undersampling maintaining similar performance (macro $F_1 = 0.55$ and 0.51 , both $p < 0.001$). At the segment level, only SMOTE and undersampling configurations achieved statistical significance (macro $F_1 = 0.50$, $p = 0.038$ and macro $F_1 = 0.48$, $p = 0.001$, respectively), while the default model did not (macro $F_1 = 0.48$, $p = 0.073$). Notably, excluding boundary-spanning tracks reduced performance to nonsignificance at both track and segment levels ($p = 0.056$ and $p = 0.985$, respectively), while clipping the return phase similarly eliminated significant effects ($p = 0.191$ and $p = 0.081$). These patterns confirm that low classification accuracy reflects limited acoustic separability between phases rather than artifacts of boundary placement or class imbalance. The strongest evidence for phase differentiation emerged when the full dataset with class balancing was used; however, the classification metrics were similar to the default setting, suggesting that whatever acoustic differences exist are subtle.

Unsupervised clustering

Since SMOTE and random undersampling are not applicable to unsupervised learning, we tested only two preprocessing variations for clustering: clipping return phase duration to the shortest playlist and removing boundary-spanning tracks.

Clustering performance remained poor across all conditions. For the default configuration, silhouette scores were 0.140 (segment-level) and 0.159 (track-level), with all agreement metrics (ARI, NMI, V-measure) near zero (< 0.01), indicating essentially random cluster-label correspondence. Clipping the return phase modestly improved some metrics—particularly at the segment level, where V-measure increased to 0.094—but values remained well below thresholds indicative of meaningful structure⁵⁵. Removing boundary-spanning tracks produced no improvement, with metrics nearly identical to the default condition (all V-measures < 0.013). These results confirm that the failure to recover phase structure through unsupervised

Table 6. Low-level audio feature-based classification statistics. The results of paired t-tests comparing cross-validations between trained models and corresponding majority-class baseline models. Each model type was calculated at the level of 30-second segments as well as tracks. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Level	Sampling	Macro F1 (mean $\pm$ SD)	Model acc. (mean $\pm$ SD)	Baseline Acc. (mean $\pm$ SD)	t	p
Segment	Default	0.483 $\pm$ 0.089	0.507 $\pm$ 0.056	0.449 $\pm$ 0.045	2.419	0.073
Segment	SMOTE	0.502 $\pm$ 0.054	0.509 $\pm$ 0.039	0.449 $\pm$ 0.045	3.063	0.038*
Segment	UNDER	0.478 $\pm$ 0.022	0.482 $\pm$ 0.018	0.196 $\pm$ 0.056	8.719	0.001***
Segment	Clipping	0.376 $\pm$ 0.179	0.417 $\pm$ 0.155	0.211 $\pm$ 0.122	2.327	0.081
Segment	Boundaries	0.430 $\pm$ 0.030	0.480 $\pm$ 0.020	0.481 $\pm$ 0.077	-0.020	0.985
Track	Default	0.554 $\pm$ 0.032	0.583 $\pm$ 0.054	0.434 $\pm$ 0.017	7.426	0.002**
Track	SMOTE	0.553 $\pm$ 0.043	0.569 $\pm$ 0.057	0.203 $\pm$ 0.012	11.356	0.000***
Track	UNDER	0.514 $\pm$ 0.039	0.531 $\pm$ 0.051	0.203 $\pm$ 0.012	12.399	0.000***
Track	Clipping	0.498 $\pm$ 0.174	0.580 $\pm$ 0.133	0.427 $\pm$ 0.090	1.572	0.191
Track	Boundaries	0.520 $\pm$ 0.062	0.568 $\pm$ 0.066	0.452 $\pm$ 0.030	2.661	0.056

clustering is robust across data preprocessing approaches and reflects the absence of intrinsic acoustic organization aligned with therapeutic phases.

Table 7. Clustering performance. Sil = silhouette score; ARI = adjusted rand index; NMI = normalized mutual information; H = homogeneity; C = completeness; V = V-measure.

Level	Sampling	Sil	ARI	NMI	H	C	V
Segment	Default	0.140	0.004	0.002	0.002	0.002	0.002
Segment	Clipping	0.179	0.098	0.094	0.095	0.093	0.094
Segment	Boundaries	0.139	0.008	0.005	0.004	0.006	0.005
Track	Default	0.159	-0.002	0.009	0.009	0.009	0.009
Track	Clipping	0.210	-0.016	0.057	0.056	0.057	0.057
Track	Boundaries	0.155	-0.001	0.013	0.013	0.013	0.013

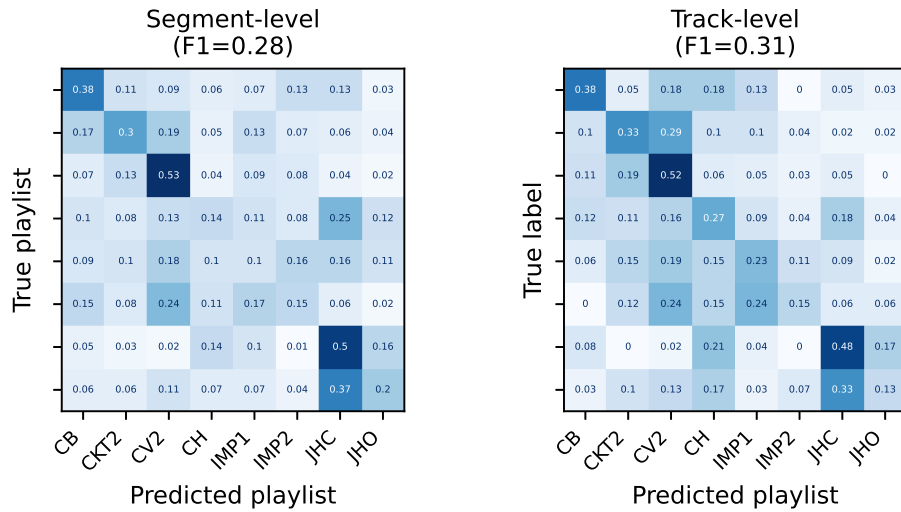


Figure 10. Confusion matrices. The playlist confusion matrices from Figure 7A shown with numerical values.

Section C: Principal component structure

Clustering was performed on the principal components that retained 99% of the variance. See Table 8 for numbers of retained components and explained variance in each analysis. Figures 5 and 7B share the same components. The first two components explain between 47% and 62% of the variance, so the projections in Figures 5, 6B and 7B leave out a large part of the space in which clustering took place.

PC1 and PC2 had similar loadings in all four analyses (Table 9). PC1 loaded positively and evenly on the RASTA-filtered auditory spectrum bands, which we interpret as broadband spectral energy. PC2 loaded on psychoacoustic sharpness, spectral centroid and spectral roll-off, which we interpret as spectral brightness. PC3 varied across analyses, loading on (1) spectral variance and entropy at the track level, (2) loudness and spectral flux at the segment level, and (3) voice-quality measures (shimmer, jitter and harmonics-to-noise ratio) in JH Classical. These labels are provisional and have not been validated perceptually. The analysis code on GitHub reproduces the full loadings.

Table 8. Principal component analysis of low-level audio features. Number of components that retain 99% of the variance, and eigenvalue (explained variance, %) of the first three components.

Level	Playlists	Components	PC1	PC2	PC3
Segment	All	39	20.59 (31.7)	10.14 (15.6)	6.87 (10.6)
Track	All	34	21.90 (33.6)	11.15 (17.1)	7.49 (11.5)
Track	Imperial 2	20	29.83 (44.5)	11.65 (17.4)	8.06 (12.0)
Track	JH Classical	22	26.87 (40.5)	10.31 (15.5)	10.20 (15.4)

Table 9. Highest principal component loadings. Five highest-magnitude loadings per component, across all playlists. RASTA = RASTA-filtered auditory spectrum.

	Segment level	Track level
PC1	RASTA spectrum L1 norm (.22); RASTA bands 13, 15, 16, 17 (.20)	RASTA spectrum L1 norm (.21); RASTA bands 12, 13, 15, 16 (.20)
PC2	Psychoacoustic sharpness (.30); spectral centroid (.29); roll-off 50% (.29); roll-off 75% (.29); roll-off 25% (.28)	Roll-off 50% (.28); psychoacoustic sharpness (.28); roll-off 25% (.28); spectral centroid (.26); roll-off 75% (.26)
PC3	Auditory spectrum L1 norm (.30); spectral flux (.27); RMS energy (.24); spectral variance (.23); 1–4 kHz band energy (.23)	Spectral variance (.27); spectral entropy (.27); roll-off 90% (.25); auditory spectrum L1 norm (.23); shimmer (–.23)