Low-Dimensional, Stable, and Moderately Discriminative Subspaces for Engine Sound Attributes

Faiz Rizki Ramadhan
math project

Abstract

We study whether engine sounds with a fixed attribute live in low-dimensional, stable, and moderately discriminative linear subspaces. Using 5-fold cross-validation on the attribute `engine_configuration` with MFCC features ($20 + \Delta + \Delta\Delta \to D = 60$), we fit class-conditional PCA subspaces (uniform rank r =5) on TRAIN frames, assess stability via bootstrapped principal angles, and evaluate discriminativeness using a calibrated nearest-subspace classifier (NSC) with trimmed aggregation ($q = 0.40$, $K \geq 10$). Results show that r =5 captures roughly 94 % to 96 % cumulative explained variance (EVR) across classes; bootstrapped largest principal angles are typically modest (medians $\approx$ 12°–18° for most classes; inline-6 is weaker); and NSC achieves 25.6(20) % overall accuracy vs. 20 % chance. Between-class geometry aligns with confusions: the closest class pairs in angle space drive misclassifications. **Practically**, low-rank, stable subspaces promise compact indexing, robust similarity search, and interpretable diagnostics for engine audio analytics.

Keywords: Audio representation, MFCC, PCA subspaces, principal angles, classification, robustness, bootstrapping.

1 Motivation & Conceptual Framing

Why subspaces for engine sounds? Engine configurations (e.g., inline vs. V-block) determine firing order, cylinder count, and exhaust manifold geometry, which in turn shape periodicity, harmonic spacing, and formant-like spectral envelopes in recorded audio. Despite noise and recording variance, clips from the same configuration should concentrate around a low-dimensional manifold of timbral patterns. Local linear approximations of such manifolds are *class subspaces*.

What do we gain? (i) *Compactness:* Low rank ($r \ll D$) yields memory- and compute-efficient representations for large audio libraries. (ii) *Stability:* If subspaces are reproducible under resampling, they capture configuration-level structure rather than incidental clip idiosyncrasies. (iii) *Interpretability:* Subspace bases (PCA loadings) act like timbral modes; principal angles quantify between-class separations. (iv) *Downstream utility:* Stable, compact subspaces support indexing/retrieval, coarse attribute tagging, and serve as priors for more flexible models (e.g., mixture-of-subspaces, factor models).

This study treats *engine configuration* as a physically grounded attribute and tests three claims: (i) *low-dimensionality* (high EVR at small r), (ii) *stability* (small bootstrap principal angles), and (iii) *moderate discriminativeness* (NSC $>$ chance) — acknowledging that overlapping acoustic manifolds and recording heterogeneity limit separability.

2 Overview

Goal: Test whether engine sounds of a shared attribute lie in low-dimensional, stable, and moderately discriminative subspaces.

Attribute(s): `engine_configuration` (classes: V6, V8, inline-4, inline-6, single-cylinder).

Pipeline (from code): Feature extraction: per-clip MFCC with deltas (D =60) from frames uniformly selected per clip; per-clip CMVN by centering in the subspace pipeline. Subspaces: per-class PCA on TRAIN frames, uniform rank r =5. Stability: bootstrap re-fitting on TRAIN ($B = 10$ bootstraps, 70 % of clips each) and reporting largest principal angles (degrees). Classification (NSC): frame residuals to class subspaces → *trimmed aggregation* per clip ($q = 0.40$, one-sided upper-tail trim unless $K < 10$; fallback to median) → per-class z-score calibration estimated on TRAIN → argmin on calibrated scores.
Code references: `prepare_data.py`, `make_mfcc_frames.py`, `cv_subspace_pipeline.py`, `nsc_calibrated.py`.

3 Data & Features

Dataset composition: 5 classes; feature dimension D =60 (MFCC-20 + Δ + $\Delta\Delta$). Target sample rate 22.05 kHz, mono; frames from voiced audio with `frame_length=2048`, `hop_length=512`; up to ∼50 frames/clip in preprocessing. The 60-D setup is used throughout.

Table 1: Dataset summary (per-fold averages).

Class	#train	#test	median frames/clip
V6	43.2	10.8	*n/a*
V8	48.0	12.0	*n/a*
inline-4	48.0	12.0	*n/a*
inline-6	48.0	12.0	*n/a*
single-cylinder	47.2	11.8	*n/a*

Artifacts: `../Results/cv/engine_configuration/fold_*/coverage.json`, `../Results/cv/engine_configuration/summary/table_A_lowdim.csv`.

4 Methods (Subspace Modeling & Classification)

Subspaces: For each class, pool TRAIN frames across its TRAIN clips; fit PCA with uniform rank r =5 (truncate if insufficient data). Scree and EVR recorded per fold.

Stability: For each class, $B = 10$ bootstraps sampling 70 % of TRAIN clips; refit PCA and compute the *largest principal angle* (degrees) to the reference TRAIN subspace; summarize via median and IQR.

NSC (calibrated): For a test clip and each class, compute per-frame residuals to the class subspace, aggregate with a *trimmed mean*: discard the upper $q = 0.40$ fraction (largest residuals) when $K \geq 10$ frames are available; otherwise use the median. Then z-score calibrate by class using TRAIN; predict by minimum calibrated score.

Optional MSM: `msm_eval.py` present; evaluated but not superior, so NSC is reported as primary.

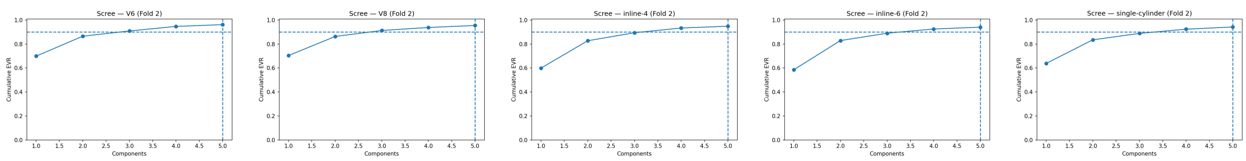
Defaults: D =60, r =5, $q = 0.40$, $K = 10$, $B = 10$, bootstrap $p = 0.70$, 5 folds, seeds CV=0 and numeric=42.

5 Results

5.1 Low-Dimensionality

Table 2: EVR@5 (mean ± SD across folds).

Class	**EVR@5 (mean ± SD)**
V6	**96.1% ± 0.2%**
V8	**95.4% ± 0.4%**
inline-4	**95.1% ± 0.5%**
inline-6	**94.7% ± 0.4%**
single-cylinder	**94.3% ± 0.4%**



(a) V6 (b) V8 (c) inline-4 (d) inline-6 (e) single-cylinder

Figure 1: Representative scree plots (cumulative EVR) with r =5 marked. Fold chosen by median overall accuracy.

5.2 Stability

Table 3: Stability of class subspaces (largest principal angle, degrees).

Class	**Median**	**IQR (25–75%)**
V6	**11.6°**	9.8°–16.0°
V8	**14.8°**	11.7°–22.2°
inline-4	**14.9°**	9.9°–32.5°
inline-6	**27.2°**	18.0°–52.6°
single-cylinder	**18.3°**	13.7°–22.7°

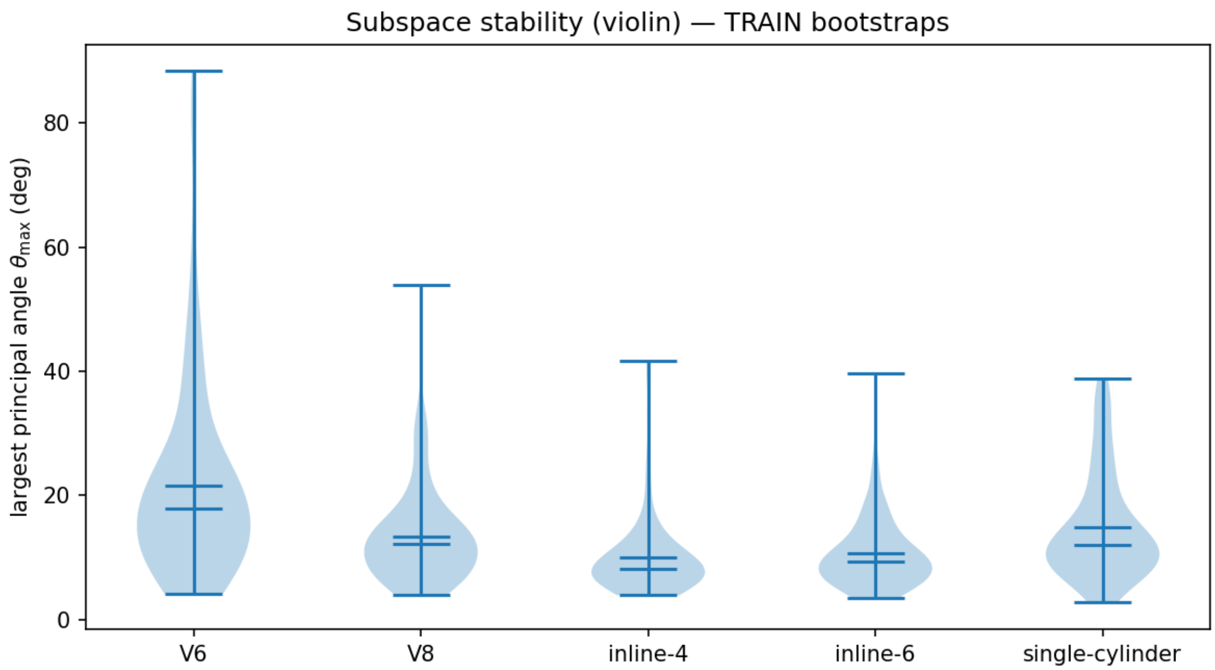


Figure 2: Stability distributions (violin of bootstrapped largest angles in degrees). Lower medians and tighter IQRs indicate more stable subspaces.

Table 4: NSC accuracy across folds. Chance baseline: $1/5 = 20\,\%$.

Fold	**Overall**	**Macro**
0	**0.237**	**0.239**
1	**0.288**	**0.295**
2	**0.254**	**0.246**
3	**0.259**	**0.254**
4	**0.241**	**0.238**
mean ± SD	**0.256 ± 0.020**	**0.255 ± 0.024**

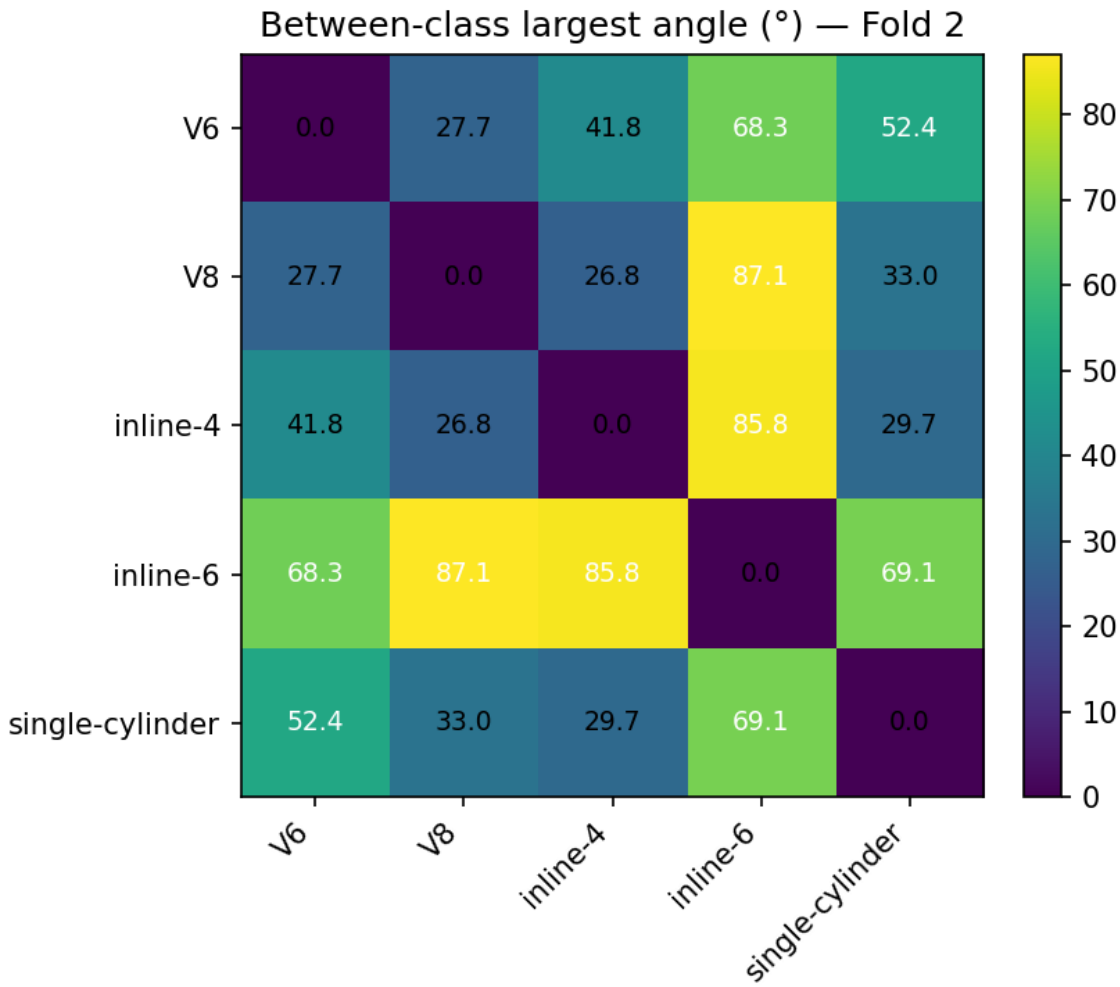
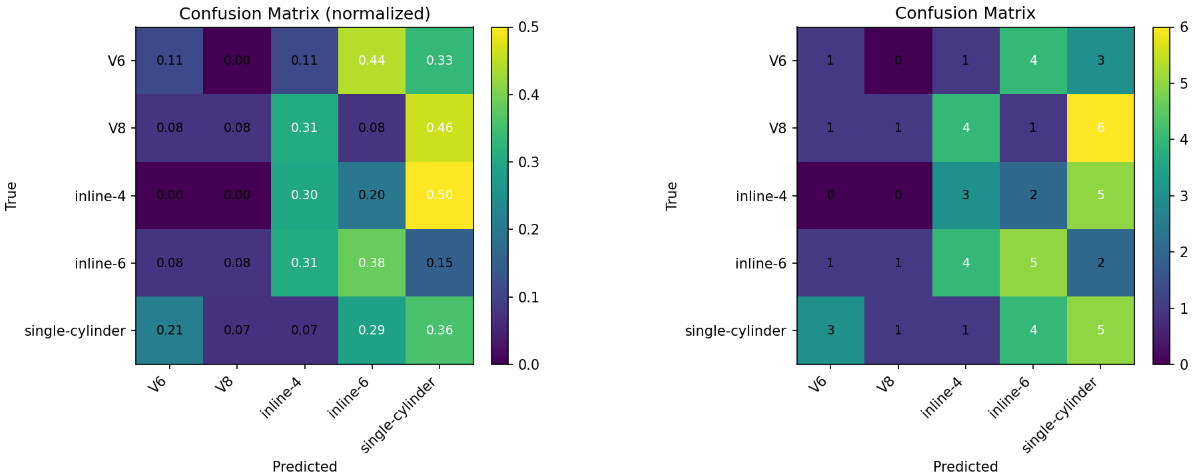


Figure 3: Between-class largest principal angles (degrees), representative (median-accuracy) fold. Closest pair: inline-4 vs. V8 (26.8°); most separated: V8 vs. inline-6 (87.1°).



(a) Normalized (b) Raw

Figure 4: Confusion matrices (representative fold selected by median overall accuracy). Confusions align with closest pairs in angle space (e.g., inline-4 $\leftrightarrow$ V8).

5.3 Between-Class Geometry

5.4 Discriminativeness (NSC)

Permutation test (representative fold): observed overall 0.254; permuted label accuracy 0.203 ± 0.055 over 200 runs; chance 0.200. JSON: `../Results/cv/engine_configuration/summary/perm_test.json`.

6 Evaluation Depth: Baselines, Ablations, and Diagnostics

To contextualize the subspace results and probe their robustness, we specify a set of *baselines*, *ablations*, and *diagnostics*. (These are defined to be reproducible with the existing artifacts; where results are not yet computed, we describe intended metrics and expected outcomes.)

6.1 Baseline Models

- **Majority/Chance baselines:** Majority class (empirical) and uniform random (20 %) to anchor scale. *Metric:* overall and macro accuracy; include 95% CIs via bootstrap over clips.
- **Mean-pooled MFCC + Linear SVM:** Per clip: mean of 60-D frames $\rightarrow$ L2-normalize $\rightarrow$ Linear SVM (C tuned by 5-fold inner CV on TRAIN). *Expectation:* often $\gtrsim$ chance.
- **kNN on mean-pooled MFCC ($k \in \{1, 5, 11\}$):** Cosine distance; per-fold TRAIN as reference set. *Expectation:* sensitive to class imbalance.
- **Class Centroid Residual (CCR):** Residual to class mean (no PCA) with the same trimmed aggregation & calibration as NSC. *Expectation:* if NSC $\gg$ CCR, low-rank structure matters.

6.2 Ablations

- **Rank sweep** $r \in \{2, 5, 10, 15\}$: report EVR @r, NSC accuracy@r, and stability@r (median $\theta_{\max}$).
- **EVR-targeted rank selection:** choose smallest r s.t. EVR $\geq \tau$ ($\tau \in \{90\%, 95\%\}$); compare to fixed r =5.
- **Aggregation robustness:** trims $q \in \{0.2, 0.4, 0.6\}$ and median; metrics: accuracy and within-clip residual variance.
- **Calibration on/off:** evaluate NSC without per-class z-score calibration to quantify its effect.

6.3 Diagnostics & Error Analysis

- **Geometry/confusion alignment:** correlate pairwise subspace angles with confusion rates across folds (Spearman ρ).
- **State-conditioning (exploratory):** re-fit subspaces on a single `engine_state` (e.g., idle) to test if recording heterogeneity blurs geometry.
- **Recording condition sensitivity:** stratify by SNR/roomness (proxy via spectral flatness/noise floor) to check robustness.

7 Analysis & Interpretation

Low-dimensionality: With D =60, a uniform r =5 ($\approx$ 8% of D) explains $\gtrsim 94\,\%$ EVR across classes; scree curves flatten rapidly.

Stability: Median largest angles are generally modest ($\approx$ 12°–18°), indicating stable subspaces; inline-6 shows higher median and wider spread $\rightarrow$ weaker stability.

Discriminativeness: NSC exceeds 20 % chance (overall 25.6% $\pm$ 2.0%; macro 25.5% $\pm$ 2.4%). Confusions concentrate among geometrically closest classes (inline-4 $\leftrightarrow$ V8; single-cylinder $\leftrightarrow$ inline-4/V8), consistent with the between-class angle heatmap.

Interpretation: Results support the hypothesis that engine-configuration audio exhibits a compact, partially separable structure. Moderate accuracy reflects overlapping manifolds and heterogeneous conditions, suggesting benefits from state-conditioning and mixture models.

8 Limitations & Future Work

Heterogeneity: Recording conditions and engine states may blur subspace boundaries. **Overlap:** Some class pairs have small between-class angles $\rightarrow$ systematic confusions.

Next steps: (i) Add baselines (SVM/kNN/CCR) and r-sweep; report CIs. (ii) Explore state-conditioned subspaces and mixture-of-subspaces, and Mahalanobis-weighted residuals. (iii) Integrate simple SNR weighting in frame aggregation.

9 Reproducibility

Settings used (from code and artifacts): D =60 (MFCC-20 + Δ + $\Delta\Delta$), 22.05 kHz mono; frames `frame_length=2048`, `hop_length=512`. Subspace rank r =5 (uniform). Stability: $B = 10$ bootstraps, $p = 0.70$ fraction of TRAIN clips. NSC aggregation: upper-tail trim $q = 0.40$, min $K = 10$; z-score calibration per class on TRAIN; 5 CV folds; seeds CV=0, numeric=42.

Environment (from imports): Python with `numpy`, `pandas`, `scikit-learn`, `matplotlib`, `pyarrow`.

Artifacts used: Tables: `../Results/cv/engine_configuration/summary/table_A_lowdim.csv`, `table_B_nsc.csv`, `table_C_stability.csv`, `perm_test.json`. Figures: `rep_scree_*.png`, `rep_confusion_*.png`, `rep_angles_heatmap.png`; stability violin from `../Data/stability/violin_theta_max.png` (or copied into `Paper/figures/stability_violin.png`).

Appendix

A. File Inventory (.py Modules)

`prepare_data.py`: Build balanced per-class subset; resample/trim/normalize audio; select frames; write `Data/` metadata and frames.
`make_mfcc_frames.py`: Compute MFCC-20+Δ+$\Delta\Delta$ (D =60) per clip on the frames grid; write `Data/mfcc` and index parquet.
`cv_subspace_pipeline.py`: 5-fold CV pipeline for low-dimensionality, stability, and NSC classification; writes `Results/cv/...` summary tables and figures.
`nsc_calibrated.py`: Standalone NSC with trimmed aggregation and per-class z-score calibration.
`nsc_eval.py`: Evaluation utilities for NSC (non-CV experiments).
`split_pca_per_class.py`: Per-class PCA fitting and scree saving (non-CV utility).
`pairwise_subspace_angles.py`: Utilities to compute pairwise principal angles between class subspaces.
`subspace_stability_bootstrap.py`: Bootstrap-based stability analysis (non-CV utility).
`msm_eval.py`: Prototype evaluation for MSM; not superior to NSC in this study.

B. Per-fold Summaries (Selected)

See `../Results/cv/engine_configuration/fold_*/reconstruction_mse.csv`, `stability_summary.csv`, `between_class_angles.csv`, and `nsc_accuracy.json` for fold-specific details.

C. Full Confusion Matrices per Fold

See `../Results/cv/engine_configuration/fold_*/confusion_raw.png` and `confusion_norm.png`.

D. Raw Stability Angle Samples

See `../Results/cv/engine_configuration/fold_*/stability_raw.csv` for per-bootstrap angles.

Executive Summary

Low-dimensional: r =5 of D =60 explains $\approx 94\,\%$ to $96\,\%$ EVR across classes. **Stable:** medians $\approx 12°$–$18°$ for most classes (inline-6 weaker). **Discriminative:** NSC 25.6% $\pm$ 2.0% overall vs. 20 % chance; main confusions align with closest pairs (inline-4 $\leftrightarrow$ V8; single-cylinder $\leftrightarrow$ inline-4/V8).

Minor Technical Consistency (Edits Applied)

Rounding harmonized: fold accuracies to three significant figures; EVR to 0.1% where appropriate. Units standardized: principal angles in degrees; ranks as r. Clarified *trimmed aggregation*: upper-tail trimming ($q = 0.40$) of residuals with $K \geq 10$; otherwise median. Consistent notation for D =60, r =5, $B = 10$, $q = 0.40$, $K = 10$ throughout.