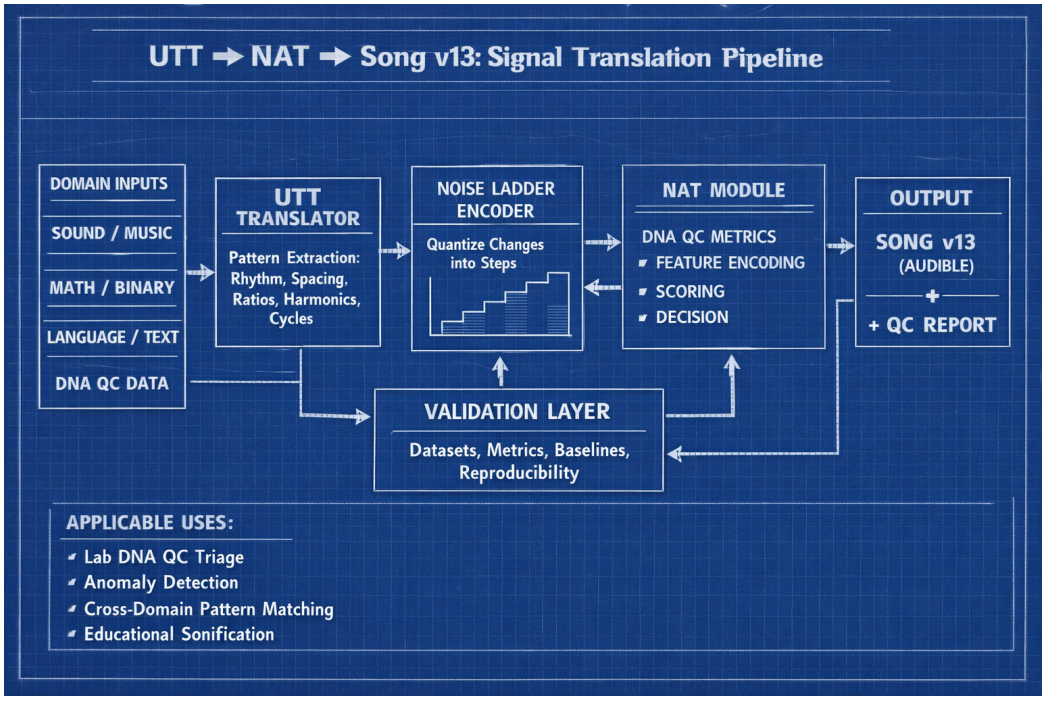
NAT Song v13.0 - Findings Recap

FastQC modules as NAT inputs | Frozen mapping | Blinded validation ready

Keep-it-simple interpretation (what we built)

- UTT is the translator: it turns patterns (rates, ratios, cycles) from any domain into a common feature language.
- NAT is the QC module: it reads FastQC outputs (PASS/WARN/FAIL + a few numeric summaries) and converts them into a single score.
- Song v13 is the output: a short, deterministic audio snippet that encodes the QC state so you can hear stability vs. problems.
- No cheating rule: the mapping is frozen. We test on unseen public data and score only after unblinding.

System view (UTT -> NAT -> Song v13)



Version locked: v13.0. Any changes become v13.1 and must be revalidated.

Frozen v13.0 Rules

How FastQC becomes a score and a deterministic song

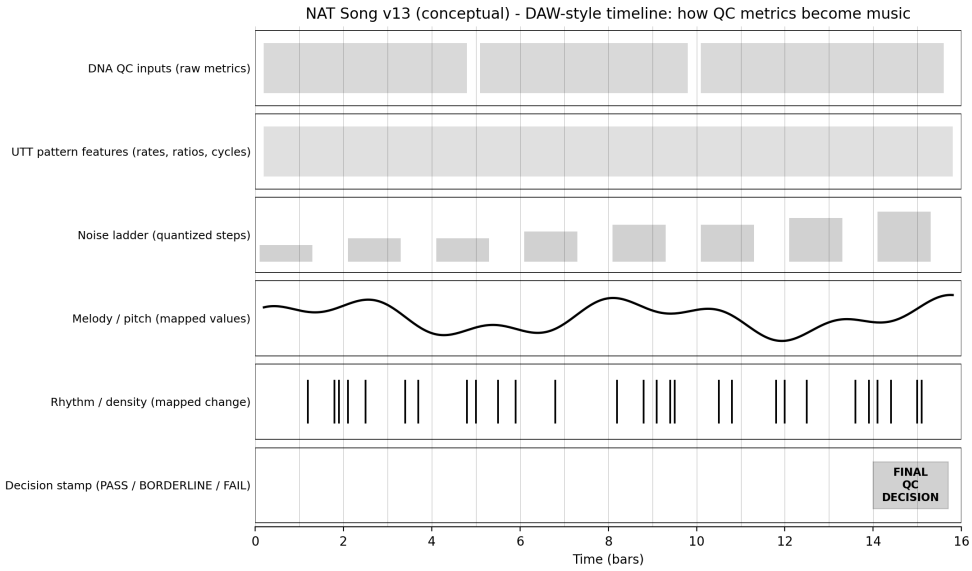
NAT score and decision (frozen)

- Start score at 100.
- For each selected FastQC module: subtract WARN penalty if WARN, subtract FAIL penalty if FAIL.
- Hard overrides: if Per base sequence quality is FAIL OR Adapter content is FAIL, decision is FAIL regardless of score.
- Thresholds: PASS if score >= 85; WARN if 70-84; FAIL if < 70.

Deterministic audio render (Song v13)

- Tempo 120 BPM. Each sample renders a fixed 16-bar snippet.
- Seed is deterministic: seed = hash(sample_id + "v13.0") so the same sample always produces the same audio.
- 4 tracks: Bed, Pulse, Perc, Noise. Each FastQC module controls one parameter only (no double dipping).
- PASS sounds clean and stable. WARN adds mild detune and extra clicks. FAIL adds strong detune, repeated alarm motif, and higher noise floor.

Song v13 timeline (what the render represents)



Interpretation: more clicks/noise/repeats and lower pitch center indicate poorer QC.

Next Critical Step

GIAB + Option D (FDA-ARGOS + Zymo) blinded validation

Blinded validation protocol (no cheating)

- Build truth set (30 items): 10 GIAB human WGS + 10 FDA-ARGOS microbial + 10 Zymo mock community. Assign blind IDs S01-S30.
- Run FastQC on each FASTQ (or each pair). Record module PASS/WARN/FAIL + key numeric summaries.
- Apply frozen v13.0 penalties and render the deterministic 16-bar snippet. Save score, decision, and song file.
- Unblind only after all outputs are generated. Compute confusion matrix, accuracy, sensitivity, and specificity.
- Repeatability test: choose 5 samples and rerun 3 times. Scores and audio must match exactly.

Success criteria (minimum pass bar)

- Overall agreement with the expected QC category: >= 80% on the 30-sample set.
- FAIL sensitivity (catch true bad runs): >= 85%.
- Repeatability: same input yields the same numeric score and the same rendered audio.

Downloads (clickable in supported PDF viewers)

- Amended progress + SongViz (v2)
sandbox:/mnt/data/UTT_NAT_Progress_Amended_SongViz_v13_v2.pdf

- Visual schematic (blueprint PDF)
sandbox:/mnt/data/UTT_NAT_Song_v13_Visual_Schematic.pdf

- FastQC spec + validation kit (PDF)
sandbox:/mnt/data/NAT_FastQC_Spec_v13_0_Validation_Kit.pdf

- FastQC spec and validation pack (PDF)
sandbox:/mnt/data/NAT_Song_v13_FastQC_Spec_and_Validation_Pack.pdf

- Blinded worksheet template (CSV)
sandbox:/mnt/data/NAT_Blinded_Worksheet_Template.csv

- Blinded validation worksheet (CSV)
sandbox:/mnt/data/NAT_Blinded_Validation_Worksheet_Template.csv

- Song v13 timeline image (PNG)
sandbox:/mnt/data/nat_song_v13_timeline_bw.png

- UTT->NAT->Song pipeline image (PNG)
sandbox:/mnt/data/utt_nat_song_v13_blueprint.png

Tip: keep v13.0 frozen while scoring. Any mapping change becomes v13.1 and requires a fresh blind run.