Transparent Triage: A White Paper on Honest File-Risk Detection for the Post-Digging DJ Workflow

Why audio software should show its work instead of scoring your taste

DigWash research track

2026-07-04

Contents

1 Abstract

A DJ comes home from a digging session with a folder of tracks. Some came from Bandcamp, some from a promo pool, some from a YouTube rip, some from a friend's USB stick three re-encodes ago. The filenames say `FLAC`, `WAV`, `320kbps`. The folder does not say which of those claims are true, which files are hot enough to clip a club system, which ones lose their bottom end the moment a booth sums them to mono, or which ones are simply worth a closer listen before they earn a place in the library.

This white paper describes a different way to build audio software for that moment. Instead of asking a machine to judge whether a track is good — a question no algorithm can honestly answer from a single file — DigWash asks a narrower, answerable question: *what can be measured from this file alone, and which of those measurements are worth a human's attention before the track reaches a set?* The answer is a transparent triage layer, not a quality score: file facts, graded by evidence strength, surfaced as reasons a DJ can act on.

We present the diagnostic model as three layers a working DJ or audio engineer will recognize on sight — provenance and integrity, rig and dynamics safety, and spatial/spectral translation — instead of the six research families the underlying evidence review uses internally. We show why each layer's strongest claims are standards-backed or peer-reviewed, and we are explicit about where the evidence is thinner. We close with a staged blueprint any DJ-software team can use to build this kind of tool honestly, and with the user-facing language that keeps a technical fact from turning into a verdict on someone's taste.

2 The DJ Problem

Digging is the fun part. What comes after digging is not.

A serious DJ's collected folder is not a curated library — it is raw material. It mixes containers (`FLAC`, `WAV`, `AIFF`, `MP3`), provenance (store purchase, promo pool, stream rip, USB hand-off, a rip of a rip), and reliability (some files are pristine, some are quietly damaged, some are not what their extension claims). None of this is visible by looking at a file list. All of it becomes visible the moment a track hits a large PA at volume, in front of a crowd, where a clipped peak, a mono-collapsed bassline, or a lossy transcode disguised as a lossless file stops being an abstraction and starts being an audible problem.

Software that only asks *"will this file open?"* solves the wrong problem. A file can decode perfectly and still be the wrong file — a 128 kbps download rewrapped in a lossless container, a master limited so hard it has no punch left, a stereo mix that disappears the instant a booth sums it to mono. None of that shows up in a format check. All of it is knowable, to a defensible degree, from the audio itself.

That is the practical question this white paper's research answers:

> Given only a collected folder of unknown-provenance audio files — with no access to the original master, the venue, the playback system, or the DJ's own gain structure — how can a DJ turn that folder into a useful archive while knowing which files deserve a closer listen before they reach a large system?

The honest answer has a hard boundary built into it. A distributed audio file does not contain the room it will be played in, the system it will be played on, the gain the DJ will apply, or the crowd that will hear it. Software that pretends otherwise — that promises a track is "club-ready" or predicts how a room will react — is selling certainty it cannot possess. Software that refuses to look at the file at all is leaving useful, computable evidence on the table. The right answer sits between those two failures: compute what the file can honestly tell you, and say plainly what it cannot.

3 The Core Philosophy: Facts, Risks, and Outcomes Are Not the Same Thing

Every claim an audio-analysis tool can make about a file falls into exactly one of three categories, and confusing them is the single most common way this kind of software becomes dishonest.

File facts are measurements computable directly from the decoded signal: true peak, integrated loudness, a clipping count, stereo correlation, bandwidth, a codec-grid trace. These are not opinions. Given the same file and the same method, two implementations should agree.

File risks are calibrated, advisory interpretations of those facts: *this track is unusually hot for your library*, *this file's lossless container looks like it is hiding a lossy source*, *this track loses significant energy when summed to mono.* A risk is still honest, but it depends on a reference point — a threshold, a library baseline, a calibration corpus — that must be disclosed, not hidden inside a single confident-sounding number.

Venue or taste outcomes are the claims a file cannot support on its own: that a track will sound good in a specific club, that it will fatigue a specific crowd on a specific night, that it is simply good or bad music. No file-side measurement — however sophisticated — contains the room, the system, the gain staging, the audience, or the artistic intent that would be required to make that claim honestly.

This is the paper's central rule, and we state it as plainly as we can:

$$\text{file descriptor} \neq \text{venue outcome} \neq \text{musical value}.$$

We do not offer this as a limitation to apologize for. We offer it as a manifesto for how audio software should be built. A growing category of tools promises to measure something like "vibe" or "hit potential" from an audio file — a single opaque score standing in for a judgement no algorithm has the standing to make. That approach photographs confidence, not truth: it hides the reasoning that produced the number, so neither the listener nor the software's own maker can audit whether the score means anything for a particular track, genre, or room. The correct failure mode for that kind of system is silence, not a wrong answer dressed as certainty.

The alternative is transparent triage. Every warning this white paper's model produces must be traceable to a specific measured fact, a specific evidence source, and a specific limit on what that fact does and does not prove. A DJ should be able to ask *"why is this file flagged?"* and get a real answer — "this file's true peak is 1.8 dB above digital full scale" — not "the model said so." That discipline is slower to build and less impressive to demo than a single glowing number. It is also the only version of this product that deserves to be trusted with a DJ's library.

4 Reading This Paper

The next three chapters present the diagnostic model as three layers a working audio professional will recognize: what a file *is* (provenance and integrity), what a file will physically *do* to a system

(rig and dynamics safety), and what a file will *become* once it is summed, played loud, or moved off a sweet spot (spatial translation and spectral fatigue). The chapter after that turns the model into a staged implementation blueprint and the user-facing language that keeps every warning honest. The detailed formulas, per-source evidence tables, and calibration mathematics behind every claim in the main text are collected in the appendix, so a reader can verify the derivation behind any claim without wading through it to read the argument.

5 Provenance and Integrity: Is the File What It Claims to Be?

5.1 The problem this layer solves

A container tells you almost nothing about what happened before the file reached it. `FLAC` means "encoded losslessly," not "sourced losslessly." A file can be decoded from a 128 kbps MP3 and then re-encoded as FLAC, and every ordinary check — file extension, decoder, checksum, even a casual listen on small speakers — will report a clean lossless file. On a large system, at volume, the missing high end, the quantization artifacts, or the phantom "holes" left by a lossy codec's psychoacoustic model stop being subtle. By then the DJ has already trusted the wrong file.

This is a different failure mode from ordinary format incompatibility. Converting a file to a more compatible container can fix a player's refusal to load it. No conversion can restore information a lossy codec already threw away. So the question this layer answers is not "can I open this file?" It is:

> Given only one distributed audio file, with no access to the original master, is there evidence that the signal has already passed through a lossy codec — regardless of what the container claims?

5.2 Why this is answerable without a reference

Every mainstream lossy codec — MP3, AAC, Vorbis, WMA — follows the same abstract recipe: transform the signal into a time-frequency representation, use a model of human hearing to decide what can be represented coarsely or dropped entirely, quantize what remains, and losslessly pack the result (Hennequin et al. 2017). The only stage that actually removes information is the quantization step. That step leaves fingerprints: coefficients pushed to zero because the encoder decided they were inaudible, and — more distinctively — values that sit unnaturally close to the specific numerical grid the codec's transform uses. A detector that knows how to look for that grid does not need the original master. It needs only the suspect file and a model of what quantization tends to leave behind.

The strongest published method for this problem, from Derrien, targets exactly DigWash's scenario: a file that decodes as lossless but was actually produced by decoding a lossy codec (commonly AAC) and re-encoding losslessly (Derrien 2019). The intuition is straightforward even though the implementation is precise: a coefficient that has already been quantized by a lossy encoder sits closer to that encoder's rounding grid than a coefficient from genuine, never-compressed audio. Counting how often a file's coefficients land suspiciously close to that grid, across enough frames, separates transcoded material from genuine lossless material with a low error rate in the paper's own evaluation of over 1,500 files. A related method from Kim and Rafii searches more broadly over candidate codec parameters and framings to find the configuration that best exposes this kind of structure, and reports correctly identifying source format in the large majority of tested cases, including at high bitrates (Kim 2018).

Other groups have approached the same problem from the signal-statistics side. Luo and colleagues show that simple statistics of a decoded WAV signal's transform-domain coefficients

— how many land exactly on zero, how their magnitudes distribute across frequency bands — reliably distinguish never-compressed audio from decompressed audio, and can further estimate which lossy codec and roughly which bitrate produced a file (Luo 2012, 2013). Bianchi and colleagues extend this to a further practical problem — a file compressed more than once, at different bitrates — and show the resulting statistical fingerprint is detectable, particularly when the second compression pass used a higher bitrate than the first (Bianchi 2014).

A data-driven alternative exists: train a classifier directly on spectrograms to recognize lossy-codec marks without hand-designing the feature (Hennequin et al. 2017). This approach is fast and codec-agnostic, but it has a specific, consequential weakness: its accuracy collapses at high bitrates, precisely where a DJ is most likely to be fooled by a transcode that "sounds fine." A learned detector is also only as good as the codec configurations it was trained on — a validated robustness concern, not a hypothetical one (Koops 2024). That combination argues for leading with the deterministic, grid-based methods and treating a learned classifier as an accelerator to add later, not the first source of truth.

5.3 The bottom line

$$\text{playable container} \not\Rightarrow \text{genuine source quality.}$$

A file that opens cleanly has told you nothing about where it came from. The honest product behavior is to report both facts side by side — the declared container and the inferred history — and to let the DJ decide whether to keep, replace, or simply flag a suspect file. No detector in this family recovers lost audio; the value is entirely in surfacing the truth before the DJ trusts a file that isn't what it claims to be.

5.4 What the warning should say

What was detected	What the file gets told
Lossless container, lossy-codec quantization trace	"This looks like a lossy file rewrapped as lossless."
Strong high-frequency sparsity plus a codec-grid match	"This file carries lossy compression traces."
Estimated prior low bitrate	"The real source quality may be lower than the filename or container suggests."
Evidence of more than one compression pass	"This file may have been encoded more than once."
Weak or unsupported evidence	"Codec history could not be determined reliably."

Every one of these is a statement about the file, not a verdict on the track. None of them says the track is unusable — only that it deserves a second look before it is trusted. The full evidence table, the rounding-error and grid-search formulas behind these detectors, and their documented limits are in the appendix's Provenance and Integrity section.

6 Rig and Dynamics Safety: What Will This File Do to a System?

6.1 The problem this layer solves

A track can sound perfectly fine on headphones and still be the file that clips a club system, trips a limiter, or embarrasses a DJ the moment the volume comes up. The signal contains real,

measurable clues about that risk — how close its peaks sit to full scale, how loud it is on average, how consistently it has been compressed or limited. None of those clues require knowing the room. They are properties of the file, and they are exactly the properties a DJ needs surfaced before a track reaches a large rig.

This layer deliberately stops short of predicting what will actually happen in a venue. A distributed file does not contain the PA's gain structure, the limiter settings, the subwoofer layout, the room's modal behavior, or the engineer's mix decisions. What it does contain is enough to say, honestly: *this file is hot, or clipped, or unusually flat, and here is the number that says so.*

6.2 The standards-backed core: true peak and loudness

The strongest ground in this entire white paper sits here, because it rests on a published international standard rather than a proposed method. ITU-R BS.1770-5 defines exactly how to measure a programme's loudness and true peak (ITU 2023), and EBU R 128 and Tech 3342 extend the same foundation with production targets and a loudness-range descriptor (EBU 2023a, 2023b). These are not audio-quality heuristics; they are the same measurement family broadcasters and streaming platforms already use to normalize programme loudness worldwide.

True peak matters because the peak visible in a file's raw samples is not the actual peak of the reconstructed analog waveform — real peaks routinely fall *between* samples, and a meter that only checks sample values can under-read a transient by several decibels (ITU 2023). A true-peak meter oversamples the signal to catch what a naive peak meter misses. For a DJ, this is the difference between a file that looks safe on a basic meter and one that will actually clip the instant it passes through a converter, a gain stage, or a format change.

Integrated loudness is computed with frequency weighting that approximates how the ear perceives loudness across the spectrum, energy averaging over time, and a gating procedure that excludes silence and very quiet passages so they don't distort the result. The standard's formula for the gated, K-weighted loudness of a signal is:

$$L_K = -0.691 + 10 \log_{10} \left(\sum_{i \in I} G_i z_i \right) \quad \text{LKFS},$$

where each channel i contributes its mean-square energy z_i weighted by a standard channel coefficient G_i. This single number tells a DJ, reliably and comparably across files, whether a track is unusually loud relative to their library — useful for spotting a hot master before it embarrasses a transition, and for understanding why a very loud file might get turned down automatically by a streaming platform's own loudness normalization.

Neither of these numbers is a rig-stress predictor by itself. A loud integrated-loudness value indicates a hot master, not a known venue SPL — the final level in the room depends on the DJ's gain staging, the amplifier, and the room, none of which the file contains. What the standard gives DigWash is a trustworthy, reproducible measurement; what the product adds on top is the judgment of when that measurement is worth a warning.

6.3 Dynamics: flat, hot, and unusually jumpy

A file's *dynamics* — how its level moves over time, and how close its peaks sit to its average — say something different from its raw loudness. Two tracks can share the same integrated loudness while one breathes and the other has been limited into a flat wall. The relevant descriptors here are crest factor (peak level minus average level), peak-to-loudness ratio, and EBU Tech 3342's Loudness Range (LRA), which quantifies how much a track's loudness moves over its duration

using a percentile spread of short-term loudness values, deliberately excluding quiet fades and isolated loud events from dominating the number (EBU 2023b).

These descriptors should never be read as a verdict on mastering taste. Compression and limiting are normal, often deliberate, production tools — modest compression can even be preferred by listeners in some contexts. What the evidence supports is a *direction*, not a threshold: heavier limiting measurably reduces crest factor and reshapes the amplitude envelope, and listener studies show that once loudness differences are equalized, heavy compression tends to reduce perceived quality and preference rather than improve it (Croghan 2012). A large-scale analysis of real-world mixes and masters confirms this is not a rare edge case: loudness and dynamics issues, including overcompression and clipping, are common in mastered commercial audio, and overcompression measurably correlates with tonal-balance changes such as reduced high-frequency energy (Mourgela 2024). So the honest product behavior is comparative, not absolute: flag a file as unusually flat *relative to a DJ's own library or a genre baseline*, never as flat in some universal sense that a techno tool and a classical recording are both supposed to satisfy.

6.4 Low-frequency and mono-sum risk

Large systems are unusually sensitive to low frequencies, because long wavelengths interact strongly with room geometry and subwoofer placement. Cinema research on wide-area low-frequency reproduction shows this sensitivity produces real spatial variance across a room — a bass response that is fine at one position and cancelled at another (Hill 2016). That finding comes from cinema rooms, not clubs, and DigWash cannot compute a room's spatial response from a file; but the file *can* tell you how much low-frequency energy is concentrated in a track, and — more usefully — how much of that low end survives when the stereo signal is summed to mono, which is exactly what many subwoofer arrays, phones, livestream encoders, and mono booth zones do to a signal in practice. A stereo file whose bass content partially cancels under mono summation is a concrete, file-computable risk worth flagging before it reaches a system that sums it.

6.5 What the warning should say

What was measured	What the file gets told
True peak near or above full scale	"This file has little peak headroom and may clip after conversion, gain, or processing."
Integrated loudness far above library baseline	"This master is unusually loud; check gain before playing it loud."
Repeated flat-top or clipped samples	"This file appears clipped."
Low loudness range and low peak-to-loudness ratio vs. baseline	"This master is very flat compared with your library."
High low-frequency energy concentration	"This track is very sub-heavy; check it on the system before relying on it."
Bass content that weakens under mono summation	"Low bass may change or weaken if the system sums bass to mono."

Every one of these pairs a specific, reproducible measurement with a specific, reversible action — check the gain, listen before trusting the low end. None of them predicts what will happen in any particular room. The gating-block mathematics, the worst-case peak-sample derivation, and the full evidence tables behind this chapter are in the appendix's Rig and Dynamics Safety section.

7 Spatial Translation and Spectral Fatigue: Listen-Check Prompts, Not Verdicts

7.1 The problem this layer solves

Some file problems only reveal themselves once a track leaves the studio environment it was mixed in: a stereo image that collapses when summed to mono, a hi-hat or synth that turns piercing at club volume, a resonance that was invisible on headphones and unmistakable through a horn-loaded PA. None of these are format problems. All of them are real, and all of them are hard to predict from a casual listen at home.

This is the layer where the evidence is least settled, and we say so plainly rather than dress it up. The strongest claim we can defend here is narrower than it sounds: DigWash can compute file-side facts — stereo correlation, mono-sum behavior, high-frequency energy concentration, tonal peak persistence, modulation characteristics — that correlate with known perceptual risks documented in psychoacoustics and spatial-audio research. What the file cannot tell you is whether a specific listener, in a specific room, at a specific volume, will actually find a given track fatiguing or find its stereo image unstable. The correct output of this layer is therefore a *listen-check prompt* — "this file is worth hearing on loud monitors before you trust it" — never a verdict.

7.2 Why mono-sum loss and stereo translation matter

A stereo file genuinely is two independent channels, left and right — that part of the format is exactly what it sounds like. But the same two channels can also be read a second way, purely by arithmetic: their average (the content common to both channels, which is what survives when the two are summed to mono) and their difference (the content that makes the left and right channels unlike each other, which is what carries stereo width). Audio engineers call these the *mid* and *side* signals. Computing them needs nothing beyond the file itself — no reference, no second version of the track — and doing so is what lets DigWash measure mono compatibility directly instead of guessing at it. It matters because a large share of real playback chains do not preserve a file's full stereo image: subwoofer arrays are commonly summed or heavily coupled, phones and many streaming and livestream paths fold to mono, and some booth zones route to mono outright. A track whose bass content partially cancels when its left and right channels are summed will lose exactly that content in every one of those contexts — and the listener will simply experience it as bass that "isn't there," with no obvious cause.

The perceptual anchor for treating stereo correlation as a width/translation signal comes from a classic loudspeaker-listening study: perceived image width tracks the magnitude of the cross-correlation between two channels, and perceived distance tracks its sign (Kurozumi 1983). That study measured room acoustics, not distributed music files, so we treat its transfer to a file-side correlation measurement as a reasoned analogy, not as directly validated evidence — a distinction this white paper insists on holding even where it would be more convenient to blur it. Two things about a file's own left/right correlation, however, are not analogy but direct signal mathematics: channels that are nearly identical correlate near +1, and channels that partially cancel under summation correlate below zero. A file with sustained negative correlation combined with high side-channel energy is therefore a legitimate candidate for a mono/stereo listen-check, even though the strength of that recommendation is still a DigWash hypothesis awaiting calibration against real listening data, not an imported law.

Nor is stereo correlation the whole story. Real production datasets confirm that mono-compatibility problems and phase issues are not theoretical edge cases — they show up regularly in both mixes and finished masters (Mourgela 2024). And the older assumption that phase differences are simply inaudible does not hold up under listening tests: phase

changes have been shown to measurably affect perception for both synthetic and real signals (Stepankova, n.d.). Phase should therefore be treated as a diagnostic feature worth surfacing, not as a detail files-side software can safely ignore — while remembering that many wide-stereo effects are deliberate, so a phase-difference measurement is evidence for a listen-check, not proof of a mistake.

7.3 Why spectral outliers deserve a second listen

Loudness perception is not flat across frequency, and it changes shape with playback level: the balance of frequencies that sounds fine on a laptop at home is not guaranteed to sound the same once the same file is reproduced much louder on a large system (Zwicker 2007). That single fact is the reason a spectral-fatigue layer exists at all — amplified venues genuinely do operate at levels most home listening never reaches, and managing those levels is an established practical concern for the industry (Mulder 2016).

Psychoacoustics gives this problem well-defined vocabulary, not just intuition. **Sharpness** describes how much a signal's perceived intensity is weighted toward high frequencies. **Roughness** describes the perception of fast amplitude modulation — the buzzy or grainy quality of certain textures. **Fluctuation strength** describes slower modulation, the kind that produces a sense of rhythmic throb rather than buzz. **Tonality** describes how strongly a signal is dominated by narrow, whistle-like tonal components rather than broadband texture. Each of these has a rigorous textbook definition (Zwicker 2007) and each is computable from a decoded file without any reference signal.

What the literature does *not* supply is a validated threshold for when any of these measurements should be called fatiguing for a DJ, on a club system, over the course of a set. The published just-noticeable-difference values for these metrics come from a controlled study of refrigerator noise — real numbers, but from a different domain entirely, useful only as scale intuition (You and Jeon 2008). The correct, honest response to that gap is not to invent club thresholds and present them as findings. It is to make every warning **corpus-relative**: flag a file as a sharpness, tonality, or roughness *outlier compared with the DJ's own library or a genre reference group*, not as absolutely, universally too bright or too harsh. A four-on-the floor kick pattern or a rolling percussion loop should never be flagged for having exactly the rhythmic modulation it was built to have.

7.4 Why the explanation layer matters as much as the measurements

A file can trip several of these warnings at once — a suspect codec history, a hot true peak, a mono-sum problem, a sharpness outlier — and a DJ does not need five disconnected numbers. They need to know which risk is driving the warning and why. That is the role of an explanation layer sitting on top of every measurement in this white paper: it does not invent a sixth kind of judgment about the file. It takes the already-computed, already-graded risks from every other layer and ranks them, so the product can say *"the strongest warning here comes from codec-provenance evidence"* or *"the strongest warning here comes from low-frequency mono loss"* instead of hiding both behind one opaque number.

This is also the point in the diagnostic model where the temptation toward false authority is strongest, so we name it directly. Perceived audio quality is not a single property waiting to be extracted from a file — it depends on context, listener history, and expectation that a file simply does not contain (Raake and Blauert 2013), and objective audio-quality measures do not transfer cleanly between domains they were not built and validated for (Torcoli 2021). The one peer-reviewed anchor for weighting distortions by how much they actually dominate a listener's attention comes from a full-reference, codec-quality-evaluation setting that has a reference signal to compare against — a setting DigWash's no-reference problem does not share (Delgado and

Herre 2022). Adapting that salience idea to a no-reference DJ-triage tool is therefore this white paper's own hypothesis, not an imported, validated finding — and it should be labelled that way for as long as it remains uncalibrated against real listening outcomes.

The failure mode this explanation layer exists to prevent is a single black-box score standing in for all of this nuance:

$$\text{file} \rightarrow \text{black-box quality score} \rightarrow \text{false authority.}$$

The alternative this white paper argues for is slower to explain and harder to market, but it is the only version that is actually honest:

$$\text{measured facts} \rightarrow \text{graded risks} \rightarrow \text{ranked, explained warning} \rightarrow \text{human decision.}$$

7.5 What the warning should say

What was measured	What the file gets told
Sustained negative L/R correlation with high side energy	"Wide/phasey stereo: listen-check before trusting it in a set."
Bass content that weakens under mono summation	"Mono compatibility warning: this file loses energy when summed to mono."
One channel carrying substantially more level or peak energy	"Channel imbalance: one channel is louder than the other."
Sharpness, tonality, or roughness outlier vs. reference set	"Spectral fatigue check: this file is an outlier for your library. Listen-check before relying on it at high level."
Several risks present at once	"The strongest warning here comes from [named risk]." Every other measured risk stays visible underneath.

None of this language claims the track will fail in a club, fatigue a crowd, or measures anything resembling "vibe." That is deliberate: those are exactly the claims a file cannot support, and this white paper's entire argument is that saying so plainly is a feature, not an admission of weakness. The grounding-status notices, the full evidence tables, and the derivations for every proxy measurement in this chapter are in the appendix's Spatial Translation and Spectral Fatigue section.

8 A Blueprint for Honest File-Risk Triage

The three layers above are a diagnostic model, not a build order. This section turns them into a staged blueprint that any team building this kind of tool — DigWash or otherwise — can follow, and into the plain-language discipline that keeps a technical measurement from turning into a moral judgment about someone's mastering choices.

8.1 A staged blueprint, not one monolithic feature

The evidence reviewed across this paper does not support building an "audio-quality AI" as a single feature. It supports a staged rollout, where each tier is only as confident as its underlying evidence, and nothing ships ahead of the confidence it deserves.

Tier 1 — Deterministic and standards-backed. Decode every file to a canonical analysis signal and record its container, codec, duration, sample rate, bit depth, and channel facts. Compute BS.1770-compatible true peak and integrated loudness. Compute clipping, flat-top, crest factor, peak-to-loudness ratio, loudness range, and level-envelope summaries. Compute left/right balance, stereo correlation, side/mid ratio, and mono-sum loss, including its low-frequency-specific form. Every one of these is explainable and useful even before a validation corpus exists, because each rests on a published standard or on signal mathematics rather than a threshold borrowed from another domain.

Tier 2 — Codec provenance. Add bandwidth and high-frequency sparsity analysis, deterministic codec-grid and quantization-trace detection, and fake-lossless checks for lossy-origin audio hiding in a lossless container. Add double-compression indicators only alongside a confidence label, since the evidence for detecting a second compression pass is direction-dependent — strongest when the second pass used a higher bitrate than the first. This tier answers the single most product-specific question a DJ actually asks: *is this file what it claims to be?*

Tier 3 — Calibrated perceptual advisories. Add corpus-relative sharpness, tonality, roughness, fluctuation, and spectral-outlier descriptors. Add library-relative dynamics and transient-movement indicators. Add a salience-weighted explanation layer that ranks warnings and names the leading cause, only after real labelled examples exist to calibrate it against. Keep every individual component visible even once a summary ranking is shown — a summary is a convenience for reading the warnings faster, never a replacement for the facts underneath it.

The ordering is deliberate: it moves from the claims with the strongest evidence to the claims that require the most calibration, so a product built this way is never asserting more confidence than its current tier can support.

8.2 The language discipline that keeps this honest

A staged blueprint only stays honest if the product copy respects the same boundary the evidence does. Every warning should read as a fact plus a consequence, phrased so a DJ can act on it — never as a judgment about mastering taste, artistic intent, or musical worth.

Defensible:

- "This looks like a lossy file rewrapped as lossless."
- "This file carries lossy compression traces."
- "This file may clip after gain, conversion, or processing."
- "This master is very flat compared with your library."
- "Low bass may change if the system sums bass to mono."
- "This file has stereo/phase conditions that deserve a mono check."
- "This file is an outlier for sharpness, tonal peaks, or rough modulation compared with your reference set."
- "This is a calibration-based warning summary, not a music-quality score."

Never defensible, regardless of how confident the underlying model appears:

- "This track is bad."
- "This is guaranteed club-ready."
- "This will fatigue the crowd."
- "This predicts the vibe."
- "This replaces listening."
- "This should be deleted."

The pattern underneath both lists is the same rule this paper opened with: name the measured fact, name the reasonable action, and stop exactly there. A tool that follows this discipline can

be wrong about a threshold and still be honest about what it is doing. A tool that collapses everything into one confident score cannot be corrected, because there is nothing underneath the number to correct.

8.3 The call to action

None of this is a case against building smarter audio tools. It is a case for building them so their reasoning survives contact with a real file, a real DJ, and a real room. The industry does not need software that claims to know whether a track will move a crowd — no file contains that information, and pretending otherwise erodes the trust the tool needs to be useful at all. It needs software that tells a DJ, plainly and with a reason attached, which files in a freshly dug folder deserve a closer listen before they earn a place in the set.

Audio tools should assist human listening. They should not replace it.

9 Appendix: Evidence Tables and Derivations

This appendix holds the material kept out of the main text so the body reads as an argument rather than a derivation-by-derivation review: the per-source evidence tables and the formulas behind every measurement this paper names. Each section below carries exactly the same rigor and grounding-status labelling as its corresponding main-text chapter — nothing here is a looser or less-checked version of the argument, only a different organization of it, kept together so a reader can verify any claim without wading through the derivation to follow the main text.

9.1 Evidence Grading, and How to Read the Grounding-Status Labels

Every source cited in this paper is graded by evidence type: a **standard** (a published measurement standard, such as an ITU or EBU recommendation), a **peer-reviewed** paper, a **textbook**, a **synthesis** (a review or overview that explains context but establishes no new threshold or formula on its own), or **grey** literature (blogs, forums, vendor material — useful for background, never a threshold source). Standards, peer-reviewed papers, and textbooks ground claims within their actual scope; synthesis and grey sources never do.

The three chapters draw on evidence of different strengths, and this paper says so plainly rather than presenting every claim as equally settled:

Chapter	Topic	Grounding status
Provenance and Integrity	Codec-history and fake-lossless detection	Strongest grounding in this paper: direct peer-reviewed no-reference detection methods.
Rig and Dynamics Safety	Physical-rig measurement (true peak, loudness, clipping)	Standards-backed: ITU-R BS.1770-5 for true peak and loudness.
Rig and Dynamics Safety	Dynamics measurement (loudness range, crest factor, compression)	Standards-backed descriptors (EBU R128/Tech 3342); interpretive warnings still require library-relative calibration.

Chapter	Topic	Grounding status
Spatial Translation and Spectral Fatigue	Psychoacoustic descriptors (sharpness, roughness, tonality)	Method-grounded, threshold-empirical: textbook and peer-reviewed descriptors, but no published DJ or club threshold exists for any of them.
Spatial Translation and Spectral Fatigue	Stereo and spatial measurement	Ongoing research, borrowed-domain: its central perceptual anchor is transported from a 1983 loudspeaker/room-acoustics study by analogy, not re-validated on distributed music files. Establishes no thresholds.
Spatial Translation and Spectral Fatigue	Salience and explanation layer	Ongoing research, hypothesis-grounded: the salience principle has one peer-reviewed anchor, obtained in a full-reference codec-quality setting; every construction built on it here (risk normalization, salience weights, aggregation) is a proposed hypothesis requiring calibration. Establishes no thresholds.

9.2 Provenance and Integrity — Evidence Table and Derivations

9.2.1 Evidence table

Source	Finding relied on
Derrien 2019 (Derrien 2019)	AAC fake-lossless detection from time-frequency quantization error; 1,576-file evaluation, null false positives, very low false negatives under reported settings.
Kim and Rafii 2018 (Kim 2018)	Lossy-format identification by searching codec parameters/framing where zeroed coefficients become visible; reported accuracy 0.96.
Hennequin et al. 2017 (Hennequin et al. 2017)	Codec-independent CNN detection from PCM spectrograms; very low error under 192 kbps, severe weakness at AAC 256/320 kbps and MP3 320 kbps (error $\approx$ 98% at AAC 320 kbps).
Luo et al. 2012, 2013 (Luo 2012, 2013)	MDCT/MFCC statistics identify prior MP3/WMA/AAC compression and bitrate from a decoded WAV signal.

Source	Finding relied on
Bianchi et al. 2014 (Bianchi 2014)	Double-compressed MP3 detection via quantized-MDCT histogram chi-square distance; strong when the second bitrate exceeds the first, weaker otherwise.
Koops et al. 2024 (Koops 2024)	Learned lossy/lossless classifiers can look near-perfect on matched test data yet degrade under unseen codec-parameter variation — a robustness warning, not a detector.

9.2.2 Mechanism

For a block of samples $x[n]$, an MDCT-like analysis coefficient is:

$$X_k = \sum_{n=0}^{N-1} x[n]w[n] \cos\left[\frac{2\pi}{N}\left(n + \frac{1}{2} + \frac{N}{4}\right)\left(k + \frac{1}{2}\right)\right].$$

9.2.3 Derrien's fake-lossless detector

With a scaled MDCT magnitude $X_{sc}(k)$ and rounding operator $\mathrm{Rd}(\cdot)$, the local rounding error is:

$$\epsilon(k) = \mathrm{Rd}(|X_{sc}(k)|) - |X_{sc}(k)|.$$

Subband error energy over subband s:

$$E(s) = \sum_{k=k_{\min}(s)}^{k_{\max}(s)} \epsilon^2(k).$$

The score is the normalized rate at which $E(s)$ falls below a threshold $\tau(s)$:

$$L = \frac{1}{N_f N_{sf} N_s} \sum_{i_f} \sum_{i_{sf}} \sum_{s} \mathbf{1}\{E_{i_f, i_{sf}}(s) < \tau(s)\},$$

with decision rule $L > \lambda \Rightarrow$ transcoded, $L \leq \lambda \Rightarrow$ genuine lossless. λ and $\tau(s)$ are the paper's fitted operating points, not universal constants — any implementation must validate them against its own corpus.

9.2.4 Kim and Rafii's grid search

$$S_\theta(i) = \Delta(\,\mathrm{Energy}(T_\theta(x_{i:i+L-1}))), \qquad \hat{\theta} = \arg\max_{\theta \in \Theta} \mathrm{Combine}_i\, S_\theta(i),$$

where T_θ is a time-frequency transform under candidate codec parameters θ (transform type, window, length, hop).

9.2.5 Compression-history statistics (Luo et al.)

Zero-coefficient ratio over MDCT coefficients $C_{m,k}$:

$$z = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} \mathbf{1}\{C_{m,k} = 0\}.$$

Combined with a 20-band mean-magnitude feature μ_b into a feature vector $\mathbf{v} = [z, \mu_1, \ldots, \mu_{20}]$ for classification.

9.2.6 Double-compression (Bianchi et al.)

Chi-square distance between the observed quantized-MDCT histogram X and a simulated singly-compressed histogram Y:

$$D_\chi(X, Y) = \sum_{i=1}^{N} \frac{(X_i - Y_i)^2}{2(X_i + Y_i)}.$$

9.2.7 Implementable detector shape

$$\mathbf{q}(x) = [q_{\text{bandwidth}}, q_{\text{sparsity}}, q_{\text{grid}}, q_{\text{history}}, q_{\text{container}}], \qquad R_{\text{codec}} = w^\top \mathbf{q}(x),$$

mapped to `clean` / `suspect` / `likely lossy history` / `unknown` bands whose cut points α, β and weights w are calibration data, not literature constants.

9.3 Rig and Dynamics Safety — Evidence Table and Derivations

9.3.1 Evidence table

Source	Finding relied on
ITU-R BS.1770-5 (ITU 2023)	Standard algorithm for programme loudness and true-peak level: K-weighting, mean-square energy, channel weighting, gating, true-peak oversampling rationale.
EBU R 128 (EBU 2023a)	Target broadcast loudness −23 LUFS; production true peak ≤ −1 dBTP; Maximum Momentary/Short-term Loudness as supplementary descriptors.
EBU Tech 3342 (EBU 2023b)	Loudness Range (LRA): 3 s short-term loudness, −70 LUFS absolute gate, −20 LU relative gate, 10th–95th percentile span; explicitly not the same as dynamic range or crest factor.
Hill et al. 2016 (Hill 2016)	Cinema wide-area low-frequency reproduction is prone to spatial/temporal variance from room modes and comb-filtering; mechanism transfers to large rooms by analogy, source domain is cinema, not clubs.
Mulder 2016 (Mulder 2016)	Amplified-music level management is socio-technical and venue-dependent; not a file-only SPL threshold.

Source	Finding relied on
Croghan et al. 2012 (Croghan 2012)	Listener study: modest compression can be preferred in some conditions; high compression generally reduces quality/preference once loudness is equalized; decreasing crest factor and envelope modification track increasing compression.
Coretto and Giordano 2017 (Coretto 2017)	Popular dynamic-range descriptors have weak statistical grounding; a transient-power-distribution statistic detects compression more consistently.
Mourgela et al. 2024 (Mourgela 2024)	218,109-entry dataset: loudness/dynamics issues prevalent in mastered audio; masters often exceed streaming loudness references; overcompression associates with tonal-balance change.

9.3.2 Peak measurements

Sample peak: $p_s = \max_n |x[n]|$, in dBFS: $P_s = 20\log_{10}(p_s)$.

True peak, from an oversampled reconstruction $\tilde{x}[m]$: $p_{tp} = \max_m |\tilde{x}[m]|$, $P_{tp} = 20\log_{10}(p_{tp})$. ITU-R BS.1770-5's worst-case peak-sample under-read, for oversampling ratio n and max normalized frequency f_{norm}:

$$\Delta_{\text{under}} = 20\log_{10}\left[\cos\left(\frac{\pi f_{\text{norm}}}{n}\right)\right].$$

9.3.3 Integrated loudness (full gating procedure)

For 400 ms blocks with 75% overlap and per-channel mean square z_{ij} in block j:

$$l_j = -0.691 + 10\log_{10}\left(\sum_{i\in I} G_i z_{ij}\right).$$

Blocks pass an absolute gate $\Gamma_a = -70$ LKFS, then a relative gate 10 dB below the loudness computed after absolute gating. For the surviving block set J_g:

$$L_{KG} = -0.691 + 10\log_{10}\left(\sum_{i\in I} G_i \frac{1}{|J_g|}\sum_{j\in J_g} z_{ij}\right) \quad \text{LKFS/LUFS.}$$

Standard channel weights: $G_L = G_R = G_C = 1.0$, $G_{Ls} = G_{Rs} = 1.41$, LFE excluded from the main summation.

9.3.4 Crest factor, PLR, LRA

$$C = P_{tp} - P_{\text{RMS}}, \qquad \text{PLR} = P_{tp} - L_{KG}.$$

Loudness Range, from the gated short-term loudness sequence ℓ_g and percentile function Q_p:

$$\mathrm{LRA} = Q_{95}(\ell_g) - Q_{10}(\ell_g) \quad \mathrm{LU}.$$

Library-relative standardization used for flat/hot warnings:

$$z_{\mathrm{PLR}} = \frac{\mathrm{PLR}(x) - \mathrm{median}_{y \in \mathcal{L}_g} \mathrm{PLR}(y)}{\mathrm{MAD}_{y \in \mathcal{L}_g} \mathrm{PLR}(y) + \epsilon}.$$

9.3.5 Low-frequency and mono-sum measurements

Low-frequency energy ratio over sub-band $[f_1, f_2]$ against full-band energy:

$$R_{\mathrm{LF}} = \frac{\int_{f_1}^{f_2} |X(f)|^2 df}{\int_0^{f_N} |X(f)|^2 df}.$$

Mid/side decomposition and low-frequency side ratio:

$$m[n] = \frac{x_L[n] + x_R[n]}{2}, \quad s[n] = \frac{x_L[n] - x_R[n]}{2}, \quad R_{\mathrm{side,LF}} = \frac{\int_{f_1}^{f_2} |S(f)|^2 df}{\int_{f_1}^{f_2} |M(f)|^2 df + \epsilon}.$$

Stereo correlation: $\rho_{LR} = \dfrac{\sum_n x_L[n] x_R[n]}{\sqrt{\sum_n x_L^2[n]} \sqrt{\sum_n x_R^2[n]}}$.

9.3.6 Transient/movement proxy

Band envelope $e_b[j] = 20 \log_{10}\left(\sqrt{\frac{1}{|W_j|} \sum_{n \in W_j} x_b^2[n]} + \epsilon \right)$, inter-band movement score:

$$M_{\mathrm{band}} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{std}_j(\Delta e_b[j]), \qquad \Delta e_b[j] = e_b[j] - e_b[j-1].$$

This is a file-computable proxy for envelope movement, not a validated "punch" metric; the underlying AES sources for punch/clarity (Fenton and Wakefield 2012; Fenton et al. 2014) are downgraded context because their complete manuscripts were not peer reviewed.

9.4 Spatial Translation and Spectral Fatigue — Evidence Tables and Derivations

9.4.1 Evidence table — spectral fatigue

Source	Finding relied on
Zwicker and Fastl 2007 (Zwicker 2007)	Textbook grounding for critical bands, specific loudness, sharpness, roughness, fluctuation strength, psychoacoustic annoyance.
Osses et al. 2023 (Osses 2023)	Psychoacoustic metrics require calibration, model-implementation, and sound-class awareness before interpretation.

Source	Finding relied on
Lopez-Ballester et al. 2019, 2020 (Lopez-Ballester et al. 2019; Lopez-Ballester 2020)	Zwicker-style psychoacoustic-annoyance formula and constants; real-time computation feasibility.
You and Jeon 2008 (You and Jeon 2008)	JND values for loudness/sharpness/roughness/fluctuation from a refrigerator-noise study — scale intuition only, not a music or DJ threshold.
Pospischil et al. 2025 (Pospischil 2025)	Binaural sharpness channel-combination is nontrivial; single-channel maximum shortcuts can fail.
Mourgela et al. 2024 (Mourgela 2024)	Tonal-profile issues and overcompression/high-frequency-profile links in real mixes/masters.

An NRLFC synthesis document proposing specific club-level psychoacoustic ceilings (sharpness 1.35 acum, roughness 0.80 asper, fluctuation strength 0.06 vacil) was reviewed and explicitly excluded: its bibliography rests on blogs, forums, and unrelated-domain noise-engineering papers, and its numbers are unvalidated hypotheses dressed as findings. No figure from that document appears anywhere in this paper.

9.4.2 Psychoacoustic descriptor formulas

Specific loudness integrated over critical-band rate z: $N(m) = \int_0^{24} N'(z,m)\,dz$.

Sharpness, as a loudness-weighted first moment favoring high critical bands:

$$S(m) = C_S \frac{\int_0^{24} N'(z,m)\, g_S(z)\, z\, dz}{N(m)}.$$

Roughness and fluctuation-strength proxies, from a band-envelope modulation spectrum $A_z(f_\mu)$:

$$R_{\text{proxy}} = \sum_z w_R(z) \int_{15}^{300} A_z(f_\mu) h_R(f_\mu)\, df_\mu, \qquad F_{\text{proxy}} = \sum_z w_F(z) \int_{0.5}^{20} A_z(f_\mu) h_F(f_\mu)\, df_\mu.$$

Tonal-peak prominence: $T(k,m) = 10 \log_{10} \dfrac{P(k,m)}{\text{median}_{r \in \mathcal{N}(k)} P(r,m) + \epsilon}$.

Zwicker-style psychoacoustic annoyance:

$$PA = N\left(1 + \sqrt{w_S^2 + w_{FR}^2}\right), \quad w_S = 0.25(S - 1.75)\log(N + 10), \quad w_{FR} = \frac{2.18(0.4F + 0.6R)}{N^{0.4}}.$$

The 1.75 acum pivot is legitimate inside this canonical formula and nowhere else; it is not a club sharpness ceiling and must never be quoted as one.

You and Jeon's refrigerator-noise JNDs, quoted only as scale intuition: $\Delta N \approx 0.5$ sone, $\Delta S \approx 0.08$ acum, $\Delta R \approx 0.04$ asper, $\Delta F \approx 0.012$ vacil.

Library-relative standardization used for every spectral-fatigue warning:

$$z_q(f) = \frac{f_q - \mu_{g,q}}{\sigma_{g,q} + \epsilon}, \qquad \text{warning}_q = \mathbf{1}\{z_q(f) > \theta_q\}.$$

9.4.3 Evidence table — spatial

Source	Finding relied on
Kurozumi and Ohgushi 1983 (Kurozumi 1983)	Perceptual loudspeaker study (IACC/apparent-source-width lineage): image width tracks |cross-correlation|, distance tracks its sign. Transported to file-side correlation by analogy only.
Stepankova (Stepankova, n.d.)	Phase-spectrum changes measurably affect perception for tested signals — rejects the "phase never matters" simplification.
Flessner et al. 2019 (Flessner 2019)	Overall binaural quality can be dominated by the weaker of monaural/binaural aspects; model is reference-based, so DigWash uses only the conceptual split.
Mourgela et al. 2024 (Mourgela 2024)	Mono compatibility and phase issues occur in both mixes and masters in real production data.

9.4.4 Spatial formulas

Mid/side transform as above. Level imbalance: $\Delta_{LR} = 10\log_{10}\dfrac{E_L + \epsilon}{E_R + \epsilon}$.

Zero-lag cross-correlation: $\rho_{LR} = \dfrac{\sum_n L[n]R[n]}{\sqrt{\sum_n L^2[n]\sum_n R^2[n]} + \epsilon}$, with short-time variants $\rho_{\min}$, $\rho_{05} = Q_5\{\rho_{LR}[m]\}$, and $p_{\rho<0}$ (fraction of windows with negative correlation).

Side-to-mid ratio: $R_{SM} = 10\log_{10}\dfrac{\sum_n S^2[n] + \epsilon}{\sum_n M^2[n] + \epsilon}$.

Whole-file mono-sum measure: $\Delta_{\text{mono}} = 10\log_{10}\dfrac{2\sum_n M^2[n] + \epsilon}{\sum_n L^2[n] + \sum_n R^2[n] + \epsilon}$, and its band-specific/low-frequency form $\Delta_{\text{sub}} = \Delta_{\text{mono}}$(20–120 Hz).

Energy-weighted inter-channel phase-difference feature, from $\Delta\phi(k,m) = \text{wrap}(\phi_L(k,m) - \phi_R(k,m))$ and weight $w(k,m) = |X_L(k,m)|^2 + |X_R(k,m)|^2$:

$$\Phi_b = \frac{\sum_{k\in b,m} w(k,m)|\Delta\phi(k,m)|}{\sum_{k\in b,m} w(k,m) + \epsilon}.$$

Proposed (uncalibrated) warning rules: mono_loss_flag $= \mathbf{1}\{\Delta_{\text{mono}} < \theta_{\text{mono}}\}$; sub_mono_flag $= \mathbf{1}\{\Delta_{\text{sub}} < \theta_{\text{sub}}\}$; wide_phase_flag $= \mathbf{1}\{R_{SM} > \theta_{SM}\} \wedge \mathbf{1}\{p_{\rho<0} > \theta_\rho\}$. No source in the corpus validates these three rules or their thresholds; they ship, if at all, as uncalibrated listen-check prompts.

9.4.5 Evidence table — perception salience

Source	Finding relied on
Raake and Blauert 2013 (Raake and Blauert 2013)	Sound-quality formation combines bottom-up signal processing with top-down cognition, context, and expectation.
Torcoli et al. 2021, 2018 (Torcoli 2021, 2018)	Objective audio-quality measures are domain-dependent; strong correlation for some artifact types does not generalize to others.
Delgado and Herre 2022 (Delgado and Herre 2022)	Peer-reviewed anchor: cognitive salience modeled as interaction between distortion metrics and cognitive-effect metrics, in a full-reference codec-quality setting.
Delgado and Herre 2023, 2024 (Delgado and Herre 2023, 2024)	Preprint extensions (informational masking, multidimensional salience weighting); method direction, not peer-reviewed authority.
Wilson and Fazenda 2014 (Wilson 2014)	In mastered commercial music, listeners distinguish and rate distortion character (clean/hard-clipped/soft) differently.

9.4.6 Salience-layer formulas (DigWash's own proposed machinery — none of it is a literature constant)

Per-family risk normalization from a robust library-relative statistic:

$$z_i = \frac{x_i - \text{median}(X_i)}{1.4826\ \text{MAD}(X_i) + \epsilon}, \qquad r_i = \sigma(\alpha_i(z_i - \tau_i)), \qquad \sigma(u) = \frac{1}{1 + e^{-u}}.$$

Salience weighting and weighted-sum summary:

$$w_i = \frac{\exp(\beta r_i)}{\sum_j \exp(\beta r_j)}, \qquad R_{\text{sal}} = \sum_i w_i r_i.$$

Alternative "is anything seriously wrong" aggregation (noisy-OR):

$$R_{\text{any}} = 1 - \prod_i (1 - r_i).$$

Agreement-aware confidence from semi-independent evidence sources c_m:

$$c_{\text{family}} = 1 - \prod_m (1 - c_m).$$

None of α_i, τ_i, or β come from literature; all belong to a calibration corpus that must pair, for each file, its descriptor vector, its derived risks, a practical outcome label, and a short reason (e.g. "played fine," "bass vanished," "sounds like a bad transcode").

9.5 Full User-Facing "Do Not Say" Register

Consolidated across every chapter, for a single audit point: no version of this product, at any tier, should say a track is objectively bad, is guaranteed club-ready, will fatigue the crowd, predicts the vibe, measures musical worth, replaces listening, or should be deleted. Every one of these fails the paper's central rule — file descriptor $\neq$ venue outcome $\neq$ musical value — regardless of how confident any underlying model appears.

Bianchi. 2014. "Detection and Localization of Double Compression in MP3 Audio Tracks." *EURASIP Journal on Information Security*.

Coretto. 2017. "Nonparametric Estimation of the Dynamic Range of Music Signals." *Australian & New Zealand Journal of Statistics*.

Croghan. 2012. "Quality and Loudness Judgments for Music Subjected to Compression Limiting." *JASA 132(2):1177-1188*.

Delgado, and Herre. 2022. "A Data-Driven Cognitive Salience Model for Objective Perceptual Audio Quality Assessment." *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 986–90.

Delgado, and Herre. 2023. *An Improved Metric of Informational Masking for Perceptual Audio Quality Measurement*. arXiv / AudioLabs manuscript.

Delgado, and Herre. 2024. *Towards Improved Objective Perceptual Audio Quality Assessment - Part 1: A Novel Data-Driven Cognitive Model*. arXiv / submitted IEEE-ACM TASLP manuscript.

Derrien. 2019. "Detection of Genuine Lossless Audio Files. Application to the MPEG-AAC Codec." *JAES / HAL*.

EBU. 2023a. *Loudness Normalisation and Permitted Maximum Level of Audio Signals*. EBU Recommendation R 128.

EBU. 2023b. *Loudness Range: A Measure to Supplement EBU r 128 Loudness Normalisation*. EBU Tech 3342.

Fenton, Steven, Hyunkook Lee, and Jonathan P. Wakefield. 2014. "Elicitation and Objective Grading of Punch Within Produced Music." *AES 136th Convention*.

Fenton, Steven, and Jonathan P. Wakefield. 2012. "Objective Profiling of Perceived Punch and Clarity in Produced Music." *132nd Audio Engineering Society Convention*.

Flessner. 2019. "Subjective and Objective Assessment of Monaural and Binaural Aspects of Audio Quality." *IEEE/ACM TASLP*.

Hennequin, Romain, Jimena Royo-Letelier, and Manuel Moussallam. 2017. "Codec-Independent Lossy Audio Compression Detection." *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 726–30. https://doi.org/10.1109/ICASSP.2017.7952251.

Hill. 2016. "Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas." *JAES*.

ITU. 2023. *Recommendation ITU-r BS.1770-5: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level*. ITU-R.

Kim. 2018. "Lossy Audio Compression Identification." *EUSIPCO*.

Koops. 2024. "Robust Lossy Audio Compression Identification." *arXiv / UMG*.

Kurozumi. 1983. "The Relationship Between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality." *JASA 74(6):1726-1733*.

Lopez-Ballester. 2020. "Enabling Real-Time Computation of Psycho-Acoustic Parameters in Acoustic Sensors Using CNNs." *IEEE Sensors Journal*.

Lopez-Ballester, Pastor-Aparicio, Segura-Garcia, Felici-Castell, and Cobos. 2019. "Computation of Psycho-Acoustic Annoyance Using Deep Neural Networks." *Applied Sciences*.

Luo. 2012. "Identifying Compression History of Wave Audio and Its Applications." *ACM TOMCCAP*.

Luo. 2013. "Compression History Identification for Digital Audio Signal." *ICASSP / Conference Paper*.

Mourgela. 2024. "Exploring Trends in Audio Mixes and Masters: Insights from a Dataset Analysis." *AES 157th Convention*.

Mulder. 2016. "Amplified Music and Sound Level Management." *JAES*.

Osses. 2023. "Considerations for the Perceptual Evaluation of Steady-State and Time-Varying Sounds Using Psychoacoustic Metrics." *Forum Acusticum 2023 / SQAT*.

Pospischil. 2025. "Directional Sharpness Perception Under Different Listening Conditions." *Acta Acustica*.

Raake, and Blauert. 2013. "Comprehensive Modeling of the Formation Process of Sound-Quality." *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*.

Stepankova. n.d. "Sensitivity of Auditory Perception to Changes in Phase Spectrum." *Original Research*.

Torcoli. 2018. "Comparing the Effect of Audio Coding Artifacts on Objective Quality Measures and on Subjective Ratings." *AES 144th Convention*.

Torcoli. 2021. *Objective Measures of Perceptual Audio Quality Reviewed*. IEEE/ACM TASLP.

Wilson. 2014. "Characterisation of Distortion Profiles in Relation to Audio Quality." *DAFx-14*.

You, and Jeon. 2008. "Just Noticeable Differences in Sound Quality Metrics for Refrigerator Noise." *Noise Control Engineering Journal 56(6):414-424*.

Zwicker. 2007. *Psychoacoustics: Facts and Models*. Springer textbook.