Engineering Sound for the Retail Floor

A programmable, verifiable music system for sales performance
and staff alertness in sealed commercial environments

Hakan Emren
Founder, notFM

August 2026
Version 1.0

Abstract

Retail environments manage light, temperature, layout, scent and price with engineering discipline. Sound — the one variable that reaches every person in the room, continuously, for the entire working day — is almost universally left to chance. This paper argues that background music in commercial space should be treated as an engineered system with a specification, a control loop and a verification method, and describes such a system.

We summarise the published evidence linking musical tempo to shopper pace, dwell time and purchase behaviour, and the separate literature on cognitive performance in sealed, mechanically ventilated interiors — the conditions under which most retail staff actually work. We then describe notFM: a catalogue built with measured tempo, a deterministic selection engine, and a telemetry layer that reports what each room actually heard.

The engine's behaviour is stated as a closed-form selection rule and validated against 8,784 real broadcast events. The central result is that selection is 34 to 333 times more regular than random assignment; in a 261-track pool, *every* track had been played either 8 or 9 times.

Our stated objective is to raise commercial performance and staff alertness through controlled sound. This paper does not report that we have achieved it. It reports the instrument we built in order to pursue it, and the evidence that the instrument does what it is specified to do. Sound delivery in retail is usually unmeasurable. We show that it does not have to be.

1 The last unmanaged variable

A modern retail floor is an instrumented space. Lighting is specified in lux and colour temperature. HVAC is specified in air changes per hour. Planograms are tested. Queue design is modelled. Price is optimised continuously.

Sound is specified as "play something nice."

This is not a small omission. Music is the only environmental variable that is *simultaneously* experienced by every customer and every member of staff, that persists for the entire trading day, that cannot be opted out of, and that acts on physiology rather than on attention. A shopper can ignore a poster. They cannot ignore tempo.

The gap is not one of belief — most retailers accept that music matters — but one of *control*. The typical in-store solution is a playlist, a radio stream or a streaming account. None of these can answer the three questions an operations director would ask of any other system:

1. What exactly is playing in each location, right now?
2. Can I change it deliberately, by hour and by day, and be sure the change took effect?
3. Can I prove after the fact that it did what I asked?

A playlist answers none of them. This paper is about a system that answers all three, and about why the answers matter commercially.

Statement of objective

We should be exact about what is being claimed, because this field is crowded with vendors who are not. notFM's objective is:

> *to make the acoustic environment of a commercial space a deliberate, scheduled and verifiable input — so that its effect on customer behaviour and on staff alertness can be pursued as an engineering problem rather than left to taste.*

That is an aim, not a result. Everything in §5 is a measurement of *delivery* — proof that the system broadcasts what it was instructed to broadcast. Outcome studies are described in §8 as work in preparation. We draw the line between the two deliberately and repeat it in §7.

2 What the evidence does and does not establish

The claim that background music affects commercial behaviour is repeated so often by vendors that it has stopped being examined. It deserves examining, because the version usually sold — *slow music increases sales* — is a forty-four-year-old result from a single store, and the conditions that produced it no longer exist.

2.1 The founding result, and its limits

Milliman's supermarket field experiment [1] is the origin of the entire field. Slow-tempo background music slowed the rate at which shoppers moved through the store, and was associated with a large increase in gross sales volume relative to fast-tempo conditions. Milliman reported the same direction in a restaurant [2], where slow tempo increased time at table.

Three things must be said about it plainly.

It was **one store**. A single supermarket, a single chain, a single market, in 1982. The reported effect size has never been cleanly replicated at that magnitude.

It measured **pace**, and inferred sales. The robust finding is that tempo changed how fast people moved. The sales figure followed from that in one setting; it is not an independently established law.

It was a **captive environment**. The 1982 supermarket shopper had no phone, no price comparison, no way to shorten the visit, and no competing audio of their own. Ambient sound had the room to itself.

2.2 What the meta-analyses actually found

The aggregated evidence is more careful than the industry's summary of it. Garlin and Owen [6] reviewed the retail music literature and found that tempo and volume reliably shift *arousal*, *time perception* and *approach behaviour* — but that effects on actual purchase behaviour are smaller, more conditional, and strongly dependent on the fit between the music and the setting. Kämpfe et al. [7], across a wider domain, reach the same shape of conclusion: background music reliably moves arousal and mood, and the direction depends on the match between music and task.

Where the evidence is strongest is where the mechanism is physiological rather than cultural. Musical fit shifts choice: French or German music in a wine aisle moved purchases toward the matching nationality, while shoppers denied any influence [4]; classical rather than top-forty music raised the value of items selected without changing their number [3]. Tempo alters perceived duration, which matters wherever there is a queue [5]. And in the exercise-science literature, where tempo is used deliberately to modulate arousal and perceived exertion, the entrainment mechanism is well characterised [14].

2.3 The context has changed, measurably

The strongest argument against transplanting a 1982 prescription into a 2026 store is not theoretical. The listening environment has changed in ways that can be counted.

Recorded music itself has compressed. On the Billboard year-end charts, average track length has fallen every decade since its peak: 4:25 in the 1990s (4:39 in 1992), 4:03 in the 2000s, 3:42 in the 2010s, and 3:18 in the 2020s [10, 11]. Tracks charting on streaming platforms now average close to three minutes — roughly thirty seconds shorter than in 2019 [12]. That is a reduction of about thirty percent from the 1990s peak.

The cause is economic, not aesthetic. Streaming pays per play, and a play is counted after roughly thirty seconds of listening. The model therefore rewards the *number* of plays rather than total listening time [13]. Songs did not shorten because audiences lost patience; they shortened because the payout structure rewarded shortening. Intros compressed for the same reason: a track that has not reached its hook within thirty seconds risks being skipped before it is paid for.

The listener has an exit. In 1982, a shopper could not skip a track, could not put on their own headphones without effort, and had no screen competing for attention. All three are now standard. Ambient sound no longer has the room to itself; it competes.

The room is not the same room. A shopping-centre fashion unit is not a 1982 supermarket. Dwell time, circulation pattern, staff-to-customer ratio and the moment of decision all differ. A prescription derived from one is not transferable to the other by assertion.

2.4 A practitioner's dissent

The following is stated in the first person because it is a professional judgement rather than a finding, and it should be attributable.

I have spent thirty years in radio and fifteen of them building music systems for commercial spaces, and I hold a degree in communications. On the strength of that, notFM's position — and mine — is that the Milliman prescription should no longer be repeated as guidance. I set out three reasons, none of which requires disputing that his 1982 measurements were correctly taken.

The unit of programming is the sequence, not the track. Radio learned this decades before retail research did. An audience does not respond to the tempo of a song; it responds to the relationship between consecutive songs, to the transition, and to whether the hour has a shape. Two identical playlists in different orders produce different rooms. A rule expressed as a single scalar — "slow" — cannot describe the thing that actually does the work.

The failure mode is asymmetric. Slow music does not make people stay. Wrong music makes them leave. In three decades I have never seen a tempo choice create an audience, and I have repeatedly seen one destroy the room in under a minute. Programming is therefore a discipline of avoiding the disqualifying error, not of finding a magic setting — which is precisely why a system needs constraints, exclusions and verification rather than a target BPM.

Monotony is punished faster than tempo is rewarded. The listener who hears the same forty songs on rotation stops hearing the music and starts hearing the repetition. This is the oldest lesson in the craft, and it is the one that survives every change of generation. It is also the reason this paper spends more space on rotation fairness (§6.1) than on tempo: the measurable failure in commercial audio is almost never that the tempo was wrong. It is that the catalogue collapsed onto a handful of tracks and nobody could see it happening.

Milliman remains in the bibliography because the field began there and because honest citation includes the work one disagrees with. It is cited as history, not as instruction.

2.5 What survives

Discarding the prescription is not the same as discarding the variable. What does not survive is the rule — *play slow music and sales will rise* — which was always a single-site finding wearing the clothes of a law. What does survive is the mechanism: tempo entrains movement and arousal, and that is physiology, not fashion. A 128 BPM track drives arousal in 2026 for the same reason it did in 1982. Generations changed; the autonomic nervous system did not.

The honest conclusion is therefore uncomfortable for anyone selling a prescription, and it is the premise of this paper:

> *There is no transferable formula for what a given store should play. The correct tempo depends on format, hour, staff shift and audience, and it can only be established in the estate where it will be applied. What a retailer needs is not a prescription but an instrument — one that can set tempo deliberately, deliver it reliably to every room, and prove afterwards what each room actually received.*

The rest of this paper describes that instrument, and the evidence that it works as specified.

3 The other audience: staff

Almost all published work on retail music studies the *shopper*. This is a strange asymmetry. The shopper is exposed for twenty minutes. The sales assistant is exposed for eight to ten hours, in the same room, usually with no daylight and no openable window.

The conditions those staff work in are themselves a measured problem. Controlled studies of sealed, mechanically ventilated interiors show that cognitive performance degrades measurably as indoor CO_2 rises within the range routinely found in occupied commercial buildings. Satish et al. [8] reported statistically significant declines in decision-making performance at concentrations previously assumed to be harmless. Allen et al. [9] found substantially higher cognitive function scores in enhanced-ventilation conditions across nine functional domains. A shopping-centre unit at 16:00 on a Saturday is precisely such an environment.

Separately, the tempo–arousal relationship is well characterised in the exercise-science literature, where music tempo is used deliberately to modulate arousal and perceived exertion [14]. The mechanism — entrainment of movement and physiological arousal to an external rhythmic driver — does not stop applying because the room is a shop rather than a gym.

We therefore treat the retail floor as having *two* populations with different exposure profiles and different needs, and we schedule for both. The afternoon lift that keeps a sales assistant upright and engaged in a sealed, artificially conditioned room is not the same programme as the one that slows a Saturday-morning shopper down. A system that cannot distinguish between them is not a music system; it is a radio.

We are explicit about the boundary of this claim. The published work establishes that tempo modulates arousal, and that sealed environments degrade alertness. We do not claim to have measured a causal effect of notFM programming on staff physiology. What we have built is the instrument that would make such a study possible: a room whose acoustic input is known, deliberate and logged to the second.

4 System design

4.1 A catalogue with measured tempo

notFM does not license a general music catalogue and hope its tempo is appropriate. The catalogue is produced and measured for this purpose. The figures below are given as an *illustration* of what a measured catalogue looks like at one point in time — they are the current state of one production library, not a specification, a target, or a recommended shape for anyone else's.

Catalogue property	**Measured**
Published tracks	1,338
Distinct artists	1,049
Total programme duration	71.2 hours
Genres	34
Tempo range	60–167 BPM
Mean tempo	114 BPM
Modal tempo band	120–129 BPM (474 tracks)

Every track carries a measured BPM value, not a genre label used as a proxy for energy. That is the point of the table — not the specific numbers in it, but that the numbers exist at all. Tempo becomes a *controllable* parameter rather than a descriptive one.

In this particular library the distribution happens to be weighted toward the 90–129 BPM region, which holds 888 of 1,338 tracks. That weighting reflects where this operator's programming spends its day; another operator serving different formats would and should build a different shape. The principle is that a control system needs resolution where it operates. Which range that is, is a decision, not a constant.

4.2 Selection as a specified function

The engine that chooses the next track is not a shuffle. It is a total function of station state, stated once and implemented once. At time t for station S, the selected track is

$$\tau^*(t) \;=\; \arg\min_{\prec}\; E_k(t), \qquad k \;=\; \min\big\{\, j \in \{1,2,3,4\} \;:\; E_j(t) \neq \emptyset \,\big\} \tag{1}$$

where $E_1 \supset$ the strictest eligibility set and E_4 the loosest. Eligibility narrows the pool; the ordering $\prec$ then picks deterministically:

$$\tau \prec \sigma \;\iff\; \big(n_S(\tau), \lambda_S(\tau), \mathrm{id}(\tau)\big) \;<_{\mathrm{lex}}\; \big(n_S(\sigma), \lambda_S(\sigma), \mathrm{id}(\sigma)\big) \tag{2}$$

n_S is the track's play count *on that station*, λ_S its last play time. Least-played first, then longest-rested. The four eligibility layers exist so that a strict filter can never produce silence: if the strictest set is empty, constraints are relaxed in a defined order rather than the stream stopping. Permanent client exclusions — "this song must never air in our stores" — are the one constraint preserved at every layer, including the last-resort one.

4.3 Programming the day

Tempo is scheduled by hour and day of week through a rotation pattern: a fixed repeating sequence of energy categories which the engine follows exactly. A morning pattern and an 18:00 pattern are different objects, assigned to different hours, and the change takes effect at the hour boundary rather than after the current buffer drains.

Two refinements matter operationally. First, each category carries a computed repeat interval

$$I_c \;=\; n_c \cdot \frac{|P|}{|P|_c} \cdot \bar{d} \tag{3}$$

so an operator can see, before committing, how often a given track will recur — and a floor of $\frac{1}{4}I_c$ is enforced so the displayed figure is the figure the broadcast actually honours. Second, a spread layer prevents tracks from anchoring to the same clock position every day: without it, a deterministic engine phase-locks to the working day, and one shift hears one slice of the catalogue for months while another slice they never hear at all.

5 The control plane

A specification is only as good as the last metre of its delivery. The difference between notFM and a streaming account is not the catalogue; it is that the room reports back.

5.1 The in-store player is an application, not a receiver

Each location runs a browser-based player rather than a passive stream client. It holds a heartbeat with the platform every 60 seconds and is treated as offline after three missed beats — long enough to survive a page reload or a brief network fault, short enough that a genuinely dead player surfaces within minutes.

The player is a two-way device. It reports:

- session start and end, and accumulated listening seconds — counted only while audio is actually playing, so a connected but paused shop does not accrue time;
- volume state, including a distinct signal for *muted* versus *turned down*, with the level attached;
- loss of the live stream and the switch to locally cached offline music, and the return to live;
- confirmation that a scheduled announcement or advertising spot actually played, per location and per timestamp;
- staff input: per-track ratings on a five-point scale, questionnaire responses, and requests for a new announcement.

The offline layer deserves emphasis because it is where most in-store systems fail silently. When the line drops, the player continues from a local cache rather than going quiet, and it records that it did so. A shop that lost its connection at 11:20 and ran on cache until 11:47 is a fact in the log, not an absence in it.

5.2 HQ: the chain's own view

The head-office portal is not a reporting export. It is a live operational surface, in the chain's own language, showing its own estate:

- which branches are playing, right now, and which are not;
- listening hours per branch and per day, with the anomalies raised rather than buried — a branch that muted its sound eleven days ago is a line item, not a number in a total;
- a live activity feed: branches coming online, volume interventions, campaign spots confirmed across N branches, staff ratings as they arrive, with a periodic roll-up of how many locations are listening against how many are silent;

- the ability to schedule announcements and campaign audio to specific branches, and to see them confirmed.

The commercial significance is that the loop closes. A chain does not merely buy music; it acquires the ability to state an intent, deliver it to several hundred rooms, and read back what each room did with it. Sound joins the set of things a district manager can be held accountable for — because for the first time it is a thing that can be seen.

5.3 Why the two-way link matters for the programme itself

Telemetry is not only for compliance. The staff-facing channel feeds the programming decision directly. Per-track ratings from the floor are collected from the people who hear the catalogue more than anyone else. Volume interventions are a signal about programme fit, not just about equipment: a branch that turns the sound down at the same hour every day is telling us something specific about that hour's tempo.

This is the practical reason the system is built as an application rather than a stream. A stream can only be broadcast at a room. An application can be answered.

6 Verification

A specification that cannot be checked is a claim, not an engineering artefact. We therefore state the engine's behaviour as falsifiable predictions and test them against real broadcast history: 8,784 track plays across two live stations over ten days.

6.1 Fairness of rotation

Prediction: because play count on the station is the primary ordering key, play counts within a pool must cluster tightly. Under random selection the variance-to-mean ratio (VMR) would be approximately 1.

Pool	Tracks	Min plays	Max plays	VMR
Station A, Default	261	8	9	**0.003**
Station A, Heavy	13	109	111	**0.004**
Station A, Medium	21	37	39	**0.009**
Station B, Heavy	24	37	39	**0.008**
Station B, Default	196	7	8	**0.029**

Measured VMR of 0.003–0.029 means selection is between 34 and 333 times more regular than chance. In a 261-track pool, every single track had been played either 8 or 9 times. This is not a statistical tendency; it is the signature of a deterministic rule executing correctly across thousands of decisions.

The commercial consequence is concrete: a retailer's catalogue is used uniformly rather than collapsing onto a familiar core. Nothing is worn out and nothing is stranded.

6.2 Constraint compliance

Prediction: repeat protection must hold except where the pool is genuinely exhausted, in which case the engine is specified to relax rather than go silent.

Station	Constraint	Adjacent pairs	Violations	Compliance
A	track, 90 min	2,588	0	100.0 %
A	artist, 45 min	2,597	0	100.0 %
B	track, 140 min	2,373	14	99.4 %
B	artist, 60 min	2,419	2	99.9 %

The 16 exceptions on Station B are not defects; they are the relaxation layer behaving as specified, and the rate quantifies exactly how often the pool was too small — six occurrences per thousand.

6.3 Time-of-day spread

Prediction: the spread layer should prevent clock anchoring. Measured as the entropy of each track's distribution across hours, normalised to the maximum attainable, the median was 0.986 on Station A ($n = 191$ tracks) and 0.940 on Station B ($n = 74$). A value of 1 is perfectly even across the trading day.

6.4 Scheduling accuracy

Prediction: a category's share of broadcast should match its share of the schedule. Observed deviation was at most 2.1 percentage points across 4,559 plays; on the dominant category of the second station, 0.5 points.

7 What this means operationally

The verification above is not an academic exercise. Each result maps to a question a retail operations team has to answer:

- **"Is our catalogue being used, or are we hearing the same forty songs?"** — VMR 0.003. Every track, uniformly.
- **"Will a customer in for twenty minutes hear a repeat?"** — 100 % compliance with a 90-minute repeat window on a healthy pool.
- **"Does the morning shift hear the same music as the evening shift?"** — entropy 0.99; the catalogue rotates across the clock.
- **"If I change the 18:00 programme, when does it take effect?"** — at 18:00, not five tracks later.
- **"Can you prove any of this?"** — the tables above are derived from broadcast logs, not from a specification document.

Deployed across a chain, the system reports per-branch listening hours, connection state, volume interventions, spot playout and per-track airplay. A head office can see which branch muted its sound at 14:00 on a Tuesday, and for how long. Sound stops being the variable nobody can report on.

8 What we do not claim

Credibility in this field is damaged by overreach, so we state the limits plainly.

- We have **not** run a controlled trial of notFM programming against sales in our own client estate. The sales effects cited in §2 are the published literature's, obtained in other settings, and should be read as the reason to control tempo — not as our results.

- We have **not** measured physiological or cognitive outcomes in staff. §3 states an environmental problem and a plausible mechanism; it does not report an intervention study.
- The verification in §5 measures **delivery**, not outcome. It proves the system plays what it was told to play. That is a precondition for any outcome claim, and it is the part the industry currently cannot demonstrate.

We regard the third point as the substantive contribution. Any vendor can assert an effect on sales. Very few can show you the log.

9 Roadmap

The instrument now exists; the studies it enables do not yet. Three are in preparation:

Tempo–footfall coupling. Pairing per-branch airplay logs with point-of-sale timestamps to estimate the association between programmed tempo and transaction pace within branch and hour — a within-store design, using the branch as its own control.

Staff-reported alertness. A short in-player questionnaire, already deployed for other purposes, extended to periodic alertness self-report, paired with programme tempo. Self-report is weak evidence individually and adequate evidence at scale.

Programme provenance. Every track in the catalogue carries a cryptographic audio fingerprint and a hash-chained certificate, timestamped against an external authority, so a client can verify exactly what aired and that it was licensed to air.

References

[1] Milliman, R.E. (1982). Using Background Music to Affect the Behavior of Supermarket Shoppers. *Journal of Marketing*, 46(3), 86–91.

[2] Milliman, R.E. (1986). The Influence of Background Music on the Behavior of Restaurant Patrons. *Journal of Consumer Research*, 13(2), 286–289.

[3] Areni, C.S. & Kim, D. (1993). The Influence of Background Music on Shopping Behavior: Classical Versus Top-Forty Music in a Wine Store. *Advances in Consumer Research*, 20, 336–340.

[4] North, A.C., Hargreaves, D.J. & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology*, 84(2), 271–276.

[5] Oakes, S. (2003). Musical Tempo and Waiting Perceptions. *Psychology & Marketing*, 20(8), 685–705.

[6] Garlin, F.V. & Owen, K. (2006). Setting the tone with the tune: A meta-analytic review of the effects of background music in retail settings. *Journal of Business Research*, 59(6), 755–764.

[7] Kämpfe, J., Sedlmeier, P. & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, 39(4), 424–448.

[8] Satish, U. et al. (2012). Is CO_2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO_2 Concentrations on Human Decision-Making Performance. *Environmental Health Perspectives*, 120(12), 1671–1677.

[9] Allen, J.G. et al. (2016). Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers. *Environmental Health Perspectives*, 124(6), 805–812.

[10] Song-length analysis of 6,618 Billboard Hot 100 entries, 1959–2025. Decade averages: 4:25 (1990s), 4:03 (2000s), 3:42 (2010s), 3:18 (2020s).

[11] Washington Post (2024). Pop songs are getting shorter in the era of streaming and TikTok.

[12] Chartmetric (2025). Are songs actually getting shorter? Charting tracks now average close to three minutes, some thirty seconds shorter than in 2019.

[13] How streaming has changed music (2025). On the pay-per-play model: a stream is counted after roughly thirty seconds, rewarding the number of plays rather than total listening time.

[14] Karageorghis, C.I. & Priest, D-L. (2012). Music in the exercise domain: a review and synthesis. *International Review of Sport and Exercise Psychology*, 5(1), 44–65.

Hakan Emren — Founder, notFM • notfm.app

Measurements were taken from the production catalogue and from broadcast logs over the period 28 July – 7 August 2026. The selection engine is specified in full in an internal engineering document; the specification and the verification method summarised here can be discussed with prospective clients under appropriate terms.