Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2582–2591
https://doi.org/10.54941/ahfe1008483

Acoustic Latency and Timekeeping in Large Co-Located Ensembles: A Human Factors Analysis of Marching Band Coordination

Sean Ge

Westwood High School, Austin, TX, USA

ABSTRACT

Sound propagates at approximately 343 m/s, imposing delays of 175–292 ms between performers separated by 60–100 m. Marching ensembles manage these delays through selective listening, predictive timing, and visual conducting. This paper develops a document-informed systems and task analysis and a reproducible computational audit of five synthetic formations, each containing 100 modeled performers, at quarter-note tempi of 120–200 BPM. It distinguishes simultaneous sound production, alignment to locally heard references, and coherent arrival at a specified audience position. The audit enumerates 4,950 unique performer pairs per formation and evaluates nearest-neighbor delays, benchmark exceedance, geometric connectivity, and audience-arrival dispersion. Mean nearest-neighbor delays are 5.3–6.7 ms, whereas 56.4–87.1% of all pairs exceed a 25 ms comparison benchmark and maximum pair delays range from 67.9 to 254.2 ms. Simultaneous emission produces 48.9–124.2 ms of arrival dispersion at the reference listener. A specified college back-hash battery requires a 93.8 ms lead over the front ensemble, equivalent to 1.13 sixteenth notes at 180 BPM. A geometric derivation explains when listening back naturally compensates distance and when lateral routing introduces additional delay. These results characterize propagation constraints rather than establish behavioral failure thresholds. They support explicit listening assignments, staged rehearsal, audience-specific calibration, and formation-audit tools, while identifying priorities for subsequent field validation.

Keywords: Marching ensembles, Acoustic latency, Sensorimotor synchronization, Human factors, Temporal coordination

INTRODUCTION

Marching performance couples instrumental action, locomotion, visual monitoring, and shared musical timing across an unusually large workspace. At 343 m/s, a 60 m acoustic path introduces 175 ms of delay and a 100 m path introduces 292 ms. At 180 BPM these delays occupy 52–87% of a quarter-note beat. A performer may therefore receive a physically accurate but temporally outdated signal from a distant colleague. Coordination depends on deciding which cues to use and what kind of alignment to achieve.

Practitioner guidance already addresses this problem through local listening zones, designated pulse sources, and depth-dependent timing offsets

Received September 26, 2026; Revised November 2, 2026; Accepted November 9, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

(Madden, 2004; Bough, 2011). Recent field research documents deliberate microtiming adjustments in marching percussion (Lookenbill, 2025). The present contribution is a transparent geometric audit that connects those practices to measurable propagation constraints. It asks how formation geometry changes potential auditory links, how source synchrony differs from audience synchrony, and how these distinctions can guide rehearsal and drill evaluation.

RELATED WORK AND ANALYTICAL SCOPE

Sensorimotor synchronization involves anticipation and correction of timing errors, rather than an independent reaction to every event (Repp and Su, 2013). Joint musical action further requires selective attention and adaptation to other performers (Keller et al., 2014). Empirical phase-correction models distinguish the influence of individual partners from the physical availability of their signals (Wing et al., 2014). Consequently, a distance matrix describes potential transmission paths, not the listening network actually used by an ensemble.

Networked music provides comparison values, but results depend on task and strategy. Chafe et al. (2010) observed graded tempo changes across one-way delays of 3–78 ms. Bartlette et al. (2006) found differing adaptation strategies in two instrumental duos, with pronounced disruption at delays of 100 ms and above. Although a 25 ms target appears in network engineering (Lakiotakis et al., 2019), it is not a universal biological boundary. Here, 10 and 40 ms provide sensitivity landmarks around that benchmark.

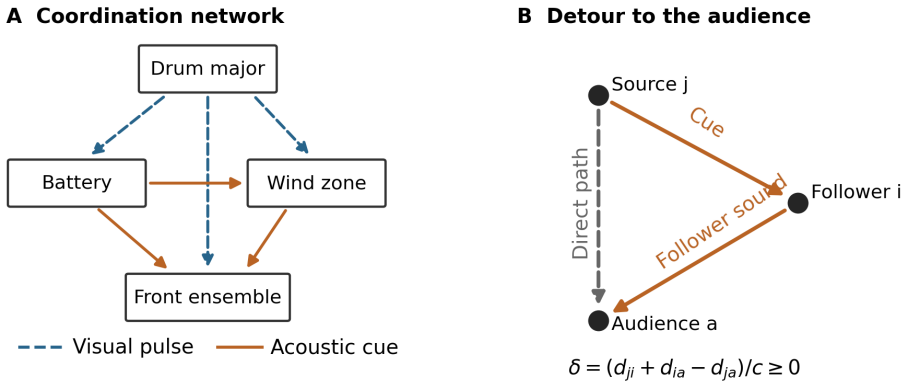


Figure 1: Analytical coordination model. A shows candidate visual and acoustic links; assignments are selected by passage rather than all followed simultaneously. B shows why a lateral listening path can add delay at the audience. Arrows depict information direction, not measured coupling strength.

Visual gestures supply predictive kinematic information (Luck and Sloboda, 2009), while expertise and stimulus motion affect competition between visual and auditory rhythms (Hove et al., 2013). Light requires only about 0.33 microseconds to cross 100 m; visual perception and motor preparation

remain noninstantaneous. The visual channel reduces distance-dependent transport disparity without eliminating human timing error.

This study combines an analytical task decomposition with deterministic simulation. It contains no new participant observations, interviews, or recordings. Its formations illustrate spatial mechanisms; they are not sampled competitive drill charts. Existing practitioner calculators already relate field distances to musical subdivisions (Blakley, n.d.). The contribution here is the joint treatment of network exposure, audience alignment, and their task implications.

SYSTEMS AND TASK MODEL

The ensemble is modeled as a spatial sensorimotor network. Performers generate sound, attend to selected references, predict future pulses, and adjust action while moving. The drum major distributes a visible pulse; its relationship to the battery may itself be rehearsed. The director evaluates the combined result from a listening location. Published guidance motivates these roles (Madden, 2004; Bough, 2011; Miller, 2016); the decomposition below is an analytical model rather than a validated account of every ensemble.

Three events must be distinguished: the emitted onset at an instrument, the arrival of a reference at another performer, and the arrival of sound at an audience location. Let p_i denote a source position, e_i its emitted onset, c sound speed, and d_{ij} the Euclidean distance between positions. One-way propagation and arrival at receiver a are:

$$\tau_{ij} = \frac{\left\| p_i - p_j \right\|}{c}, \quad t_{ia} = e_i + \tau_{ia} \tag{1}$$

Equal emitted onsets do not imply equal audience arrivals. Define the range operator $R(v) = \max(v) - \min(v)$, and let τ_a be the vector of source-to-audience travel times. The following bound follows from subadditivity of the range operator applied to $\tau_a = (e + \tau_a) - e$:

$$R(\boldsymbol{e}) + R(\boldsymbol{e} + \tau_a) \leq R(\tau_a) = D_a \tag{2}$$

Thus simultaneous emission produces an audience spread D_a, while exact arrival alignment requires an emission spread of the same magnitude. The schedule $e_i = -\lambda\tau_{ia}$, for $0 \leq \lambda \leq 1$, divides that spread into λD_a at the sources and $(1 - \lambda)D_a$ at the audience. A shared visual pulse can support either schedule, provided performers learn their intended offsets.

Why Listening Back Sometimes Compensates Distance

Consider an ideal steady-state relation in which follower i aligns its emitted pulse to the sound heard from source j, without an additional phase lag. The follower's audience arrival relative to the source is:

$$\delta_{j \to i,a} = \frac{d_{ji} + d_{ia} - d_{ja}}{c} \geq 0 \tag{3}$$

The triangle inequality makes this residual nonnegative. It vanishes only when the follower lies on the straight segment between source and receiver. This formalizes the straight-line listening criterion described by Miller (2016): back-to-front listening can implement audience compensation, whereas lateral routing adds path length. A multi-hop chain replaces d_{ji} with the sum of its edges. This is a geometric phase relation, not a claim that humans react with zero latency; prediction and learned offsets can modify the result.

COMPUTATIONAL METHOD

Geometry and Acoustic Assumptions

The field model spans $x = -45.72$ to 45.72 m between goal lines and $y = 0$ to 48.768 m from front to back sideline. All 100 sources in each formation have height $z = 1.5$ m. The reference listener is $a = (0, -20, 10)$ m. Straight-line propagation uses $c = 343$ m/s, constant across direction and frequency. Wind, reflections, attenuation, instrument directivity, amplification, and onset envelopes are excluded. The model isolates transport time rather than synthesizing an acoustic recording.

Table 1: Synthetic formation definitions with 100 sources in each.

Formation	Construction	Spacing or Radii
Compact block	10 columns × 10 rows	2 yd × 2 yd
Wide block	20 columns × 5 rows	5 yd × 2 yd
Deep block	5 columns × 20 rows	2.5 yd × 2.5 yd
Double arc	2 semicircles × 50 points	30 yd and 34 yd
Split blocks	2 grids of 10 × 5 points	2.5 yd × 2.5 yd

For a grid, $x_k = c_x + [k - (n_x - 1)/2]d_x$ and $y_l = c_\gamma + [l - (n_\gamma - 1)/2]d_\gamma$, with zero-based integer indices. For each arc, $(x, y) = (r \cos \theta, 6 \text{ yd} + r \sin \theta)$, where $\theta = j\pi/49$ for $j = 0, \ldots, 49$. One yard equals 0.9144 m. All grids are centered at depth 24.384 m; split-block centers are at $x = \pm 30$ yd. These constructions vary width, depth, curvature, and separation as illustrative scenarios rather than reconstructions of observed drill.

Metrics and Verification

Each formation supplies 4,950 unique unordered pairs. For every source, the nearest-neighbor measure is the minimum nonzero pair delay. We report its mean and maximum, pair-delay distributions, and the percentage of pairs strictly above 10, 25, and 40 ms. We also count connected components in undirected graphs retaining links at or below each benchmark. These graphs describe geometric access, not measured attention.

Audience dispersion D_a is the range of travel times under simultaneous emission. Ideal compensation advances each source by its excess travel time relative to the nearest source. Quarter-note tempi of 120, 140, 160, 180, and 200 BPM convert delays into musical units; a sixteenth lasts 15,000/BPM ms.

Checks verify pair symmetry, diagonal zeros, pair counts, field bounds, and equal compensated arrivals. No inferential tests are used because the coordinates are constructed, not sampled observations.

RESULTS ACROSS FIVE FORMATIONS

Every source has a nearest neighbor within 6.7 ms, but whole-formation pair delays are substantially larger. The compact block already has 56.4% of pairs above 25 ms. The wide block has the largest maximum delay, 254.2 ms, and the highest exceedance proportion, 87.1%. Splitting the ensemble preserves short within-block paths while creating long inter-block paths. These are properties of the selected geometries, not estimated frequencies in competitive practice.

Audience dispersion follows a different ordering. The deep block produces 124.2 ms of arrival spread, compared with 69.0 ms for the wide block. Across-field separation strongly increases pairwise delays, while front-to-back depth more directly differentiates paths to the central audience target. A single formation-size score would therefore conceal the distinction between performer communication and audience alignment.

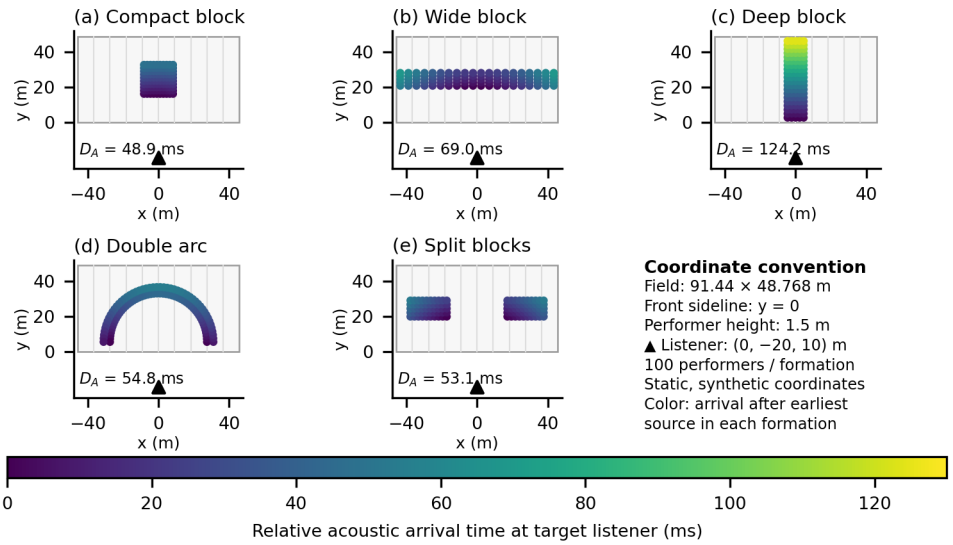


Figure 2: Five synthetic formations with 100 sources each. Shared spatial axes show field position; color encodes travel time relative to the earliest arrival at the reference listener. These are modeled onset arrivals under simultaneous emission.

Table 2: Propagation at 343 m/s and the reference audience position.

Formation	NN Mean ms	Pair Max ms	Pairs >25 ms (%)	Audience Spread ms
Compact block	5.3	67.9	56.4	48.9
Wide block	5.3	254.2	87.1	69.0
Deep block	6.7	129.4	72.6	124.2
Double arc	5.5	181.3	83.6	54.8
Split blocks	6.7	221.5	75.2	53.1

LOCAL PATHS AND GLOBAL EXPOSURE

Mean nearest-neighbor delays range from 5.3 to 6.7 ms, and the largest individual nearest-neighbor delay is 6.7 ms. The benchmark comparison supports the availability of short local paths in all five examples. It does not establish that these paths are audible, attended to, or behaviorally sufficient. In particular, instrument identity and musical role may make the closest performer an unsuitable temporal reference.

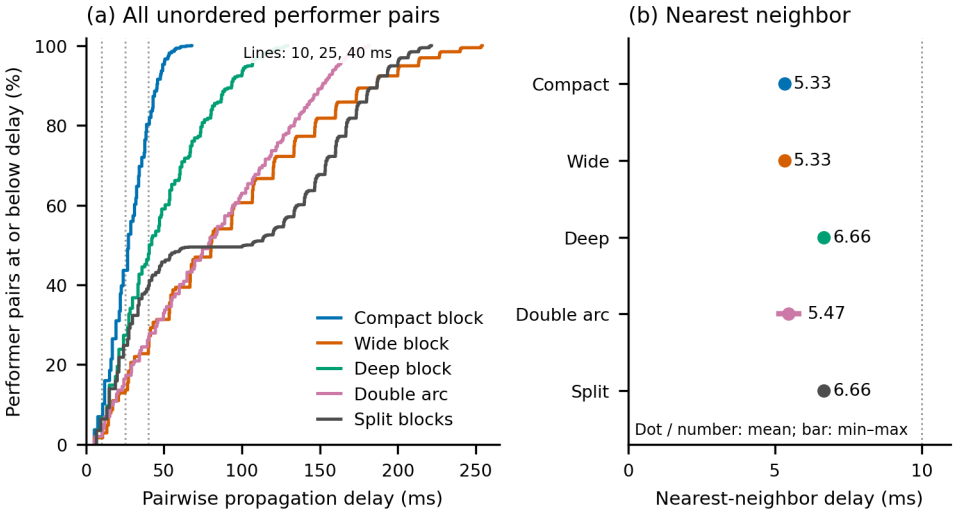


Figure 3: Local and whole-formation propagation exposure. A shows cumulative distributions of 4,950 unique pair delays per formation, with 10, 25, and 40 ms reference lines. B shows nearest-neighbor means and ranges. These are geometric quantities, not participant measurements.

Changing the screening benchmark alters the magnitude of exposure but preserves the distinction between local and extended paths. At 10 ms, 93.1–98.4% of all pairs exceed the benchmark; at 40 ms, 19.7–75.6% do. For the wide block, the proportion decreases from 98.4% to 75.6%, whereas the compact block decreases from 93.1% to 19.7%. Reporting several landmarks avoids treating an arbitrary cutoff as a categorical performance diagnosis.

Connectivity adds information that nearest-neighbor averages omit. At 25 ms, the compact, wide, deep, and arc formations each form one connected component, but the split blocks form two. At 10 ms the wide block separates into 20 columns and the double arc into two arcs. The split blocks remain disconnected even at 40 ms. A short path to somebody nearby does not imply access to a common ensemble reference through short links.

Conversely, a connected geometric graph does not establish stable collective timing. Multi-hop listening can accumulate detour, and error correction depends on directed influence and prediction. The all-pairs matrix is therefore an exposure map from which an instructor could examine candidate assignments; it is not a proposal that every performer should couple to every other performer.

TEMPO AND AUDIENCE COMPENSATION

Physical delays remain unchanged across tempi. The wide block's 254.2 ms maximum spans 50.8% of a quarter note at 120 BPM and 84.7% at 200 BPM. At 180 BPM, the deep block's 124.2 ms audience spread equals 1.49 sixteenth notes. These conversions communicate the scale of geometric offsets without predicting whether a particular rhythmic texture will remain perceptually coherent.

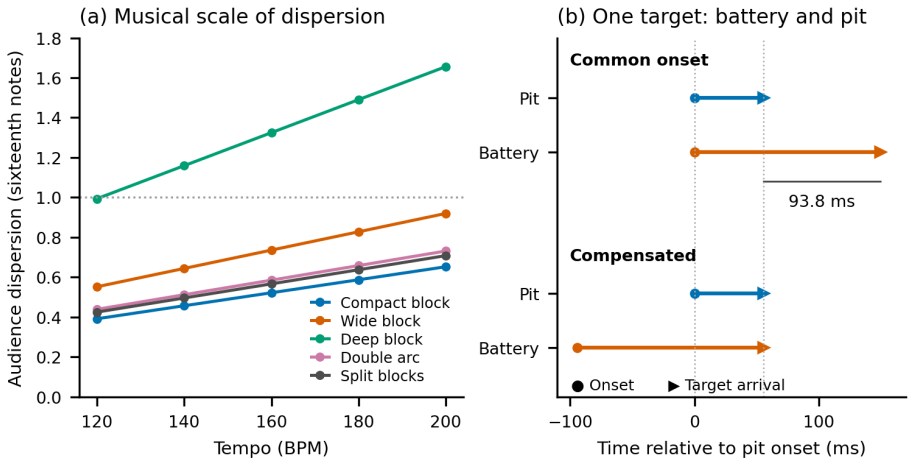


Figure 4: Musical magnitude and ideal compensation. A expresses audience-arrival dispersion in sixteenth-note units across tempi. B shows emitted onsets and target arrivals in the battery–pit example. Exact alignment follows algebraically from the schedule; it is not an observed intervention effect.

Battery and Front Ensemble Example

Place a battery source at (0, 30.48, 1.5) m, corresponding to the far college hash in a 160 ft wide field with hashes 60 ft from each sideline (NCAA, 2024). Place the front-ensemble source at (0, −3, 1.5) m. Their distances to the reference listener are 51.191 and 19.007 m. The battery must emit 93.831 ms earlier for equal arrival: the difference in audience path lengths divided by sound speed. At 180 BPM this is 1.126 sixteenth notes, so an exact sixteenth is only an approximation.

The direct battery-to-front-ensemble delay is 97.609 ms. If the front ensemble follows the heard battery pulse with ideal zero added phase lag, its audience arrival is 3.778 ms late. The small residual is the path detour in Equation 3, arising here because the audience is elevated. The required compensation is therefore listener-specific and is not identical to source-to-source delay.

Sensitivity to Environment and Receiver Position

A linear approximation to dry-air temperature dependence (Benson, 2021), $c(T) = 331.3 + 0.606T$ m/s at 0, 20, and 35°C changes all path times relative to the 343 m/s baseline by +3.5%, −0.1%, and −2.7%. Across 27 receivers

with $x \in \{-20, 0, 20\}$, $y \in \{-10, -20, -30\}$, and $z \in \{1.5, 10, 20\}$ m, wide-block dispersion spans 60.6–126.0 ms and deep-block dispersion 107.2–127.8 ms. These are scenario ranges, not confidence intervals.

IMPLICATIONS FOR REHEARSAL AND DRILL DESIGN

The audit shifts diagnosis from a single instruction to "play together" toward specifying the reference, target, and expected timing relation. Table 3 translates the analytical task model into rehearsal decisions. The recommendations are hypotheses grounded in geometry and published practice; their instructional effectiveness has not been tested.

Table 3: Analytical task allocation and proposed rehearsal checks.

Task	Decision to Rehearse	Failure to Diagnose
Establish pulse	Name the temporal source and intended phase offset.	Conflicting references
Maintain local timing	Select an audible source with a useful musical role.	Nearby but irrelevant cue
Read visual timing	Learn gesture prediction and preserve sightlines.	Late cue interpretation
Change formation	Rehearse when and how the reference changes.	Unclear handoff
Align at the audience	Calibrate from the target location and nearby seats.	Local correction worsens arrival

A practical rehearsal progression begins with short-path coordination, adds the actual formation while stationary, then introduces motion and reference changes. This isolates whether a breakdown follows the instrumental task, propagation geometry, or movement demands. Bough (2011) recommends stationary rehearsal and listening from multiple vantage points; Miller (2016) describes explicit reference assignments and geographically restricted listening. The audit makes the path-length consequences of those choices inspectable.

Temporal-source placement should be evaluated against the desired audience target. Moving a reference toward the geometric center can reduce long performer-to-reference paths, but that does not automatically minimize audience dispersion. A rear source may support forward listening when followers lie along paths toward the audience. For laterally displaced zones, visual referencing or a rehearsed phase offset may be more appropriate than unqualified listening back.

A director-oriented tool should import drill coordinates, display candidate directed listening links, and show both pair delay and audience arrival. Its essential controls are receiver position, tempo, sound speed, and reference assignment. It should report disconnected zones, identify source changes, and compare alternative schedules. Threshold colors must indicate selected benchmarks rather than label performers as correct or incorrect. Dynamic drill would additionally require updates over time and a record of when cue responsibility changes.

LIMITATIONS AND VALIDATION PRIORITIES

Five regular formations cannot establish population-level claims about competitive ensembles. Constant population controls the pair denominator but not footprint, spacing, or musical role. These are scenario contrasts, not isolated causal effects. Results should be recomputed for actual drill and venue geometry before specific instructional use.

The model includes only direct acoustic travel. Reflections, masking, wind gradients, radiation patterns, and amplification can alter which onset is perceptually dominant. Instrumental attack and audiovisual timing also vary. A visible beat is not necessarily interpreted identically by all performers, and a 25 ms communication benchmark is not an onset-detection threshold. Consequently, neither maximum delay nor benchmark exceedance is a validated predictor of ensemble quality.

The selective, document-informed task model is not a systematic review or observational analysis. Workload, attention, expertise, and coupling strength are unmeasured. Static snapshots omit receiver movement during propagation and changing cue availability; dynamic extensions should model these effects and transitions between sources.

Exact compensation is defined for one receiver. At another receiver b, a schedule that aligns at a leaves arrival spread $R(\tau_\beta - \tau_a)$. A complete audience cannot generally be synchronized by one set of offsets. Design should therefore specify a listening region and acceptable residuals, rather than imply a unique optimum for every seat.

Field validation should first record synchronized close-microphone onsets and audience-position arrivals in stationary formations, with measured coordinates and weather. A subsequent within-ensemble comparison could contrast ordinary instruction with explicit geometry-informed reference assignments using counterbalanced passages. Outcomes should distinguish emitted-onset spread, audience-arrival spread, tempo drift, and reported task demands. Ensemble or rehearsal block should be the inferential unit; thousands of geometric pairs are not independent participants.

CONCLUSION

Short nearest-neighbor paths coexist with substantial long-range delays in five synthetic formations. Audience dispersion follows a different geometric pattern from pairwise exposure. The audit formalizes the tradeoff between simultaneous emission and aligned arrival and identifies when hierarchical listening compensates or adds delay. It provides a reproducible basis for selecting references and planning field validation.

Data and Research Transparency

No human-participant data or identifiable recordings were collected for this study. The authors acknowledge the use of Grammarly for assistance with translation from Chinese to English during the preparation of this manuscript. The authors have reviewed and edited the final text and assume full responsibility for its content and accuracy.

REFERENCES

Bartlette, C., Headlam, D., Bocko, M. and Velikic, G. (2006) 'Effect of network latency on interactive musical performance', Music Perception, 24(1), pp. 49–62. doi: 10.1525/mp.2006.24.1.49.

Benson, T. (2021) 'Speed of sound'. NASA Glenn Research Center. Updated 13 May. Accessed 26 September 2026. Available online.

Blakley, P.R.F. (n.d.) 'Speed of sound in marching band: A sound delay calculator for the football field'. Accessed 26 September 2026. Available online.

Bough, T. (2011) 'Improving marching percussion', The Instrumentalist, June. Accessed 26 September 2026. Available online.

Chafe, C., Cáceres, J.-P. and Gurevich, M. (2010) 'Effect of temporal separation on synchronization in rhythmic performance', Perception, 39(7), pp. 982–992. doi: 10.1068/p6465.

Hove, M.J., Iversen, J.R., Zhang, A. and Repp, B.H. (2013) 'Synchronization with competing visual and auditory rhythms: Bouncing ball meets metronome', Psychological Research, 77(4), pp. 388–398. doi: 10.1007/s00426-012-0441-0.

Keller, P.E., Novembre, G. and Hove, M.J. (2014) 'Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination', Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1658), 20130394. doi: 10.1098/rstb.2013.0394.

Lakiotakis, E., Liaskos, C. and Dimitropoulos, X. (2019) 'Improving networked music performance systems using application-network collaboration', Concurrency and Computation: Practice and Experience, 31(24), e4730. doi: 10.1002/cpe.4730.

Lookenbill, Z. (2025) 'Groove on the field: Microtiming "feels" and rhythmic synchronization in marching percussion' [Poster]. 47th Annual Conference of the Texas Society for Music Theory, Baylor University, Waco, TX, 1 March. Available online.

Luck, G. and Sloboda, J.A. (2009) 'Spatio-temporal cues for visually mediated synchronization', Music Perception, 26(5), pp. 465–473. doi: 10.1525/mp.2009.26.5.465.

Madden, J.T. (2004) 'From Carnegie Hall to the 50 yard line: Approaching concert and marching ensembles with the same "vision"' [Clinic handout]. 58th Annual Midwest Clinic, Chicago, IL, 15 December. Available online.

Miller, W. (2016) 'In-depth read: Timing isn't on your side!', Dr. Ward Miller, 12 September. Accessed 26 September 2026. Available online.

NCAA (2024) 2024 Football Rules and Interpretations, Rule 1-2-1 and Appendix D. Indianapolis: National Collegiate Athletic Association. Available online.

Repp, B.H. and Su, Y.-H. (2013) 'Sensorimotor synchronization: A review of recent research (2006–2012)', Psychonomic Bulletin & Review, 20(3), pp. 403–452. doi: 10.3758/s13423-012-0371-2.

Wing, A.M., Endo, S., Bradbury, A. and Vorberg, D. (2014) 'Optimal feedback correction in string quartet synchronization', Journal of the Royal Society Interface, 11(93), 20131125. doi: 10.1098/rsif.2013.1125.