Journal of Research in Social Science and Humanities
ISSN 2709-1910
www.paradigmpress.org/jrssh
Volume 4 Number 11 December 2025

Does Variation in the Pronunciation of Common French Ballet Terms Systematically Affect Adolescent Dancers' Perceived Movement Quality and the Accuracy of Their Executed Steps?

Lillian Zhang[1]

[1] Western Canada High School, Calgary, Alberta T2S 0B5, Canada
Correspondence: Lillian Zhang, Western Canada High School, Calgary, Alberta T2S 0B5, Canada.

doi:10.63593/JRSSH.2025.12.03

Abstract

This pilot investigates whether systematic properties of teachers' pronunciation of ballet terminology function as motor-relevant cues for adolescent dancers. Framed narrowly, we ask whether the acoustic realization of a spoken ballet term immediately preceding movement, indexed by a transparent Pronunciation Feature Index (PFI) that aggregates terminal segment class, consonant-vowel balance, and final-syllable stress, and complemented by ΔPFI (absolute deviation from a native-French reference), biases (1) audio-only legato–staccato perception and (2) the quality of a single standardized movement execution. Framed broadly, within routine Royal Academy of Dance (RAD) pedagogy, we ask whether the acoustic properties of terminology operate as effective motor cues that shape timing and segmentation beyond semantics and visual demonstration. Four teachers (two Mandarin-L1, two Canadian-English L1) produced ten high-frequency terms. Four trained dancers (12–18 y) rated perceived staccato-legato from audios and then executed one repetition per audio. Perception showed clear between-teacher divergence for several items (e.g., *balancé*, *chaînes*, *développé*), consistent with salient acoustic variation. In execution (N = 144 trials; *cambré* excluded), ΔPFI correlated weakly with movement quality overall ($r \approx -0.04$), with a small negative trend for legato steps and near-zero for staccato. An OLS with interaction estimated a legato slope of ≈ -0.34 per ΔPFI unit and an implied staccato slope of $\approx +0.13$. These effects are small at single-rep granularity. We conclude that pronunciation differences are perceptually meaningful and may modestly bias phrasing for legato-anchored actions, but strong changes in execution should not be expected from terminology alone.

Keywords: ballet pedagogy, speech acoustics, speech prosody, phonetics, perception-action coupling, sensorimotor synchronization, acoustic-motor entrainment

1. Introduction

Ballet pedagogy relies on a shared lexicon of French terminology that is used worldwide across syllabi and schools. In practice, however, these terms are pronounced differently by teachers with different linguistic backgrounds

and accents. Pronunciation can vary in segmental features, prosody, and phenomena such as epenthesis.

Despite the significance of spoken cues in daily class, there is little empirical research on whether the sound patterns of ballet terms influence movement perception or performance quality. Prior work suggests that variations in auditory cues, such as rhythm, can bias the timing, continuity, and rate of movement in response; however, this connection has not been systematically tested in ballet classrooms by analyzing variations in the pronunciation of real terminology (Pranjić et al., 2024).

This study investigates whether, within routine Royal Academy of Dance (RAD) studio pedagogy, the acoustic properties of teachers' terminology function as effective motor cues that shape dancers' timing and segmentation beyond semantic labels and visual demonstrations. We situate the work within the RAD framework to control for curricular variability while retaining ecological validity. RAD provides globally standardized terminology and graded syllabi, a common studio routine in which the spoken term precedes or accompanies demonstration, and widely used adolescent training levels (Grades-Intermediate) that match our participant population. Constraining stimuli and movement scoring to RAD conventions reduces confounds due to heterogeneous expectations across methods (e.g., Vaganova vs. Cecchetti) and makes any observed pronunciation effects interpretable as modulation of phrasing within a single, well-specified pedagogical system.

More specifically, we ask whether the acoustic realization of spoken ballet terms, indexed by a Pronunciation Feature Index (PFI) and related correlates of "staccato-like" versus "legato-like" speech, systematically biases (1) audio-only staccato-legato perception and (2) the quality of a single standardized execution. We quantify pronunciation using a numerical scoring based on the Pronunciation Feature Index (PFI) and analyze dancer ratings and performances of ten frequently taught steps that span inherently legato and staccato movements. Controlling for term familiarity, it is hypothesized that a higher PFI, or more consonant-weighted pronunciations, will bias both perception and performance: (1) a higher PFI will elicit lower perceived movement fluidity rating from listening to audios and lower execution scores for inherently legato steps; (2) a greater absolute deviation from a native French pronunciation (ΔPFI) will result in poorer movement execution across dancers.

Aiming to address the research gap, this study provides evidence-based guidance for ballet teachers on how to effectively introduce and present terminology in class. If certain phonetic features consistently bias student perception or execution of the movement, teachers might adjust pacing or stress to reinforce intended movement quality and increase execution accuracy.

2. Background

In multilingual classrooms, the same French lexicon in which ballet is uniformly instructed through is often realized with different segmental and prosodic features. These phonetic differences raise a pedagogical question of whether systematic properties of pronunciation bias how dancers perceive a ballet step's intended quality and how they execute that step after hearing the accented terminology pronunciation.

2.1 Linguistics

Classic work in linguistics categorizes phonemes according to their distinctive features, with consonants being categorized under the place of articulation, manner of articulation, and voicing. This study primarily focuses on two manners of articulation: plosives and nasals. Plosives are stops produced with a complete oral closure and a subsequent release of built-up air, while nasals are produced while the velum is lowered so airflow passes through the nasal cavity and resonates to make a sound (Association, 1999, p. 8). The prosodic organization structures speech based on rhythmic and melodic features such as pitch, loudness, and duration. Categorizing the various properties of accented French terminology, this study analyzes fixed and lexical stress, as well as rhythmic timing (e.g., stress-timed and syllable-timed pronunciations).

2.2 Dance Pedagogy

Classical ballet codifies movement qualities as part of technique, not merely as theatrical style. Across major training systems (e.g., RAD, Vaganova, Cecchetti), instructions explicitly alternate adagio phrases that demand continuity of weight transfer, sustained lines, and uninterrupted *port-de-bras* that require discrete initiation and termination. In this literature and in studio practice, the labels legato and staccato are tied to allegro and adagio, respectively; ballet

accompaniments often use staccato, meaning to play in a detached or bouncy manner, for "movements involving lightness and single-leg jumps" and legato, meaning to play in a connected and smooth manner, for combinations that "require fluidity and continuous movement" (İşsever & İşsever, 2024, p. 5). When using "inherently legato" or "inherently staccato," the paper thus refers to how the accompaniment usually denotes the movement. For empirical purposes, we therefore evaluate movement execution on a rubric aligned to the step's pedagogical target.

2.3 Perception-Action Links

A substantial amount of previous research demonstrates tight links between auditory perception and motor planning. Neuroimaging and behavioral work in music cognition demonstrate auditory-motor co-activation during listening alone and performance without feedback, consistent with predictive mapping from heard cues to motor programs (Zatorre et al., 2007).

Building on this, work on sensorimotor synchronization (SMS), which involves coordinating movement to an external rhythm, shows that auditory input can shape not only when an action occurs but also how it is executed. Repp and Su (2013) demonstrate that people do not simply react to beats; they tend to move slightly in advance of the next event, utilizing phase-correction mechanisms to maintain alignment even when small timing deviations are not consciously noticed. In other words, auditory timing establishes a forward model that the motor system uses to schedule movement.

For this research, two SMS findings are especially relevant. Synchronization behaviour extends beyond timing and reshapes movement kinematics. Studies report distinct "legato-like" versus "staccato-like" tapping styles that differ in trajectory features such as dwell or hold time and where the motionless phase occurs (Repp & Su, 2013). Additionally, auditory pacing elicits stronger coupling than visual pacing, as seen in the 2013 paper's example of metronome clicks and flashes. Perceived stimulus properties can also scale movement amplitude in an involuntary manner, as this happens when observers adopt larger tap amplitudes after viewing larger movements. Together, these results support a concrete mechanism for the present study. If a teacher's spoken cue carries a more "staccato-like" acoustic profile with greater consonant weight, clear final release, and tighter syllabic grouping, dancers may anticipate shorter, more segmented actions; if it is more "legato-like," they may anticipate smoother, more connected phrasing.

A parallel line of evidence comes from unison speech. When speakers time their production to a partner's audio, metrical (beat-regular) stimuli improve motor output, meaning that participants produce more accurate syllables and exhibit more anticipatory, less variable alignment than with conversational timing (Kershenbaum et al., 2023). This is commonly interpreted as entrainment, where a predictable auditory pattern strengthens perception-action coupling, making synchronous production easier. Although those studies target speech rather than dance, the logic is the same. The acoustic structure of the cue, instead of its semantics, can change the quality of the ensuing motor behavior. In the ballet classroom, terminology is an auditory cue with its own segmental and prosodic shape. It is therefore plausible that systematic differences in pronunciation will bias both perceived staccato-legato quality and executed phrasing in the direction of the cue's acoustic profile.

3. Methods

3.1 Stimuli

The repertoire of ten terms was chosen to balance frequency in everyday ballet classes and provide a predicted contrast in archetypal movement quality. All ten terminologies appear in the RAD, Cecchetti, and Vaganova syllabi and are thus familiar to student dancers aged 14-18. This prevalence ensures that the experiment is not confounded by the dancers' unfamiliarity with the term. Additionally, the first five steps (*balancé* to *arabesque*) are typically taught in adagio or sustained contexts, whereas the other five (*jeté* to *piqué*) require sharper executions. This 50:50 split allows the experiment to test whether the perceived pronunciation profile impacts a dancer's execution of an expected legato or staccato move. The stimulus set comprises the ten terms listed in Table 1.

Table 1. Stimulus set of ten French ballet terms with English glosses from Wiktionary, brief movement descriptors, IPA transcriptions, predicted Pronunciation Feature Index, and the Movement Ratings

Term	Literal English Translation	Movement Description	Transcription	Pronunciation Rating (0-5)	Movement Rating (1-10)
Balancé	(Dance) A type of step where the body swings from one foot to the other in equal time.	Triple-step sway in 3/4: step side in plié, rock back, and rock forward; continuous weight transfer	balɑ̃se	2	8
Chaînes	A succession of rings or links, interlocked with each other.	Series of rapid half-turns in 1st position on demi-pointe	ʃɛn	1	3
Cambré	Which presents an arch, speaking of a part of the human body.	Upper-body back-bend from the waist	kɑ̃bʁe	2	7
Développé	To develop.	Draw the leg to the knee of the standing leg, then extend the leg to an open position with a sustained adagio	devlɔpe	1	9
Arabesque	A type of ornament whose invention has been attributed to the Arabs.	One-leg balance with the free leg extended behind	aʁabɛsk	3	8
Jeté	Thrown.	Brush the working foot of the ground and spring off the supporting leg into a big split jump	ʒəte	1	2
Glissade	Slide.	Low travelling step brushing through 2nd position feet and ending in a demi plié	glisad	1.5	3
Frappé	Struck.	The ball of the foot strikes the floor sharply, and the leg extends	fʁape	2	2
Battement tendu	Battement: Action of beating, when speaking of things.	Brush the foot along the floor to a fully pointed extension, then close back	batmɑ̃ tɑ̃dy	2	4

	Tendu: Tense.				
Piqué	(Surname, no definition)	Step directly onto pointe/demi-pointe with an immediate weight transfer	pike	1	3

Four teacher voices were recorded: two Mandarin Chinese L1 RAD teachers and two Canadian English L1 RAD teachers. All four currently teach within the Royal Academy of Dance syllabus to maintain pedagogical comparability. Recordings were captured in quiet, isolated surroundings at a fixed distance. Each teacher produced all ten terms in the order listed in Table 1. To suppress list intonation in the Anglophone recordings, each term was embedded in a carrier sentence, pasted below, which yields a more natural prosodic contour while preserving the target token.

"From the corner, travel on a *balancé* to the right, then spot the mirror and take your *chaînes* cleanly down the diagonal. Lift through the sternum for a back *cambré*, and unfold a *développé* before lengthening into an *arabesque*. Brush strongly through the floor for a *jeté*, then travel with a quiet *glissade* to set up the next phrase. At the barre, strike from the ankle on each *frappé*, brush and close a precise *battement tendu*, and finish by stepping straight into a *piqué* turn."

3.2 Derivation of the Pronunciation Feature Index (PFI)

We required a compact, rule-based measure that (1) maps established phonetic cues onto a single scale reflecting perceived connectedness vs. edge-accenting in speech, (2) is transparent and replicable for non-phoneticians, and (3) can be computed consistently from all terminologies and pronunciations. The resulting PFI (0–5) is the unweighted sum of five binary and half-point criteria motivated by classic segmental and prosodic effects on perceived temporal structure.

Rule set (with rationale).

1) +1 if the final segment is a voiceless plosive (/p t k/). Voiceless stops exhibit abrupt release bursts with stronger high-frequency energy than their voiced counterparts, yielding a sharper perceptual edge at word offset. We treat this as a canonical marker of segmentation.
2) +0.5 if the final segment is a voiced plosive (/b d g/). Voiced stops retain a closure–release pattern but with reduced burst salience due to voicing; they are intermediate on the edge cue, hence a half-point weight.
3) –1 if the final segment is nasal (e.g., /n, m, ɑ̃/). Nasals are characterized by sustained, low-frequency energy and often coarticulatory vowel nasalization, which perceptually smooths edges; we therefore subtract one point.
4) +1 if the global consonant-to-vowel ratio (C:V) exceeds 1.2. A higher proportion of consonants increases the number of constrictions per unit vowel, promoting a more consonant-weighted envelope and greater perceived segmentation; we operationalize this with a conservative threshold of C:V > 1.2 computed over the realized phones in the token (not orthography).
5) +1 if primary lexical stress occurs on the final syllable. Final-syllable prominence (indexed by relative duration, intensity, and F0) creates a terminal accent and perceived abruptness at offset.

The weights (±1, +0.5) were set a priori to (a) preserve a simple 0–5 range, (b) reflect ordinal salience among the cues (voiceless stop > voiced stop; nasal in the opposite direction), and (c) avoid discretionary adjustments that could invite bias. Epenthesis (insertion of a vowel that breaks up clusters) is annotated as a binary flag for auxiliary analyses because it can change both the C:V count and the location of stress; however, to maintain parsimony in the primary index, epenthesis is not itself scored as an additional point.

Teacher recordings were annotated in Praat with TextGrid. C:V was computed as the count of consonantal phones divided by vowel phones across the token, and stress placement was assigned to the syllable showing the largest combination of duration, intensity, and F0 (with ties resolved by duration).

To anchor the feature scale to a canonical model, we computed a Reference PFI for each term using a native-French recording drawn from authoritative pronunciation sources. For each

teacher token, we then calculated

$$\Delta PFI = |PFI_{teacher} - PFI_{reference}|.$$

This absolute deviation quantifies how far a realization is from the canonical profile and serves as an additional predictor of perception and execution, where larger deviations may increase processing cost or shift expectations.

Example. Arabesque realized as [a.ʁa.bɛsk]: final /k/ → +1; voiced stop final? (no) → 0; nasal final? (no) → 0; C:V = (three consonant clusters across four vowels) → if C:V > 1.2 then +1; final-syllable stress present → +1. PFI = 3. By contrast, a realization with final nasalization (e.g., [a.ʁa.bɛ̃]) would score a PFI reduced by 1, reflecting a smoother perceptual tail.

Ultimately, the PFI provides a transparent, low-variance mapping from well-attested phonetic cues to a single index that can be related to dancers' audio-only legato judgments and execution scores in subsequent analyses.

3.3 Movement Task for Dancers

Participants (n = 4) were adolescents aged 12–18 with ≥ 3 years of RAD-based classical ballet training. For each lexical item, the dancer first heard the audio token, then provided an audio-only staccato-legato rating (1–10), and immediately executed one standardized repetition of the step as specified in the movement rubric (legato steps capped at one eight-count at ≈ 60 bpm; staccato steps a single clean action). The rating sheet that participants used to rate the audios is attached in the Appendix section.

3.4 Filming

Movements were recorded with a fixed camera positioned ~45° from the front at barre height (~130 cm). The camera remained stationary throughout; dancers performed on a marked spot to preserve a consistent frame.

3.5 Data Handling and Privacy

Videos were used solely to score movement quality. For any materials that may be shared, faces will be blurred or cropped to protect privacy. All files are stored securely, with access restricted to the research team, and are deleted after scoring in accordance with the ethics/data-management plan.

4. Analysis

The recordings from the four teachers were transcribed using the International Phonetic Alphabet.

Table 2. IPA transcriptions by reference and by teacher (per term)

Term	French Transcription	Teacher H Transcription	Teacher L Transcription	Teacher M Transcription	Teacher Q Transcription
Balancé	balɑ̃se	bʁlɑ̃se	bæləns ɛɪ	bælnˌsɛɪ	balɑ̃sɛɪ
Chaînes	ʃɛn	ʃənɛɪ	ʃənɛɪ	ʃənɛɪs	ʃəne
Cambré	kɑ̃bʁe	ˈkɑ̃mbɹɛɪ	kambɹɛɪ	ˈkɑ̃mbɹɛɪ	ˈkɑ̃mbɹɛɪ
Développé	devlɔpe	dævlɔpɛɪ	dɛvləˈpɛɪ	dɛvələpɛɪ	dɛvlɔpe
Arabesque	aʁabɛsk	ˈəʁbɛs	ɛɹəˈbɛsk	ɛɹəbɛsk	aɻbɛsk
Jeté	ʒəte	ʒətɛɪ	ʒəˈtɛɪ	ʒɛtɛɪ	ʒəte
Glissade	glisad	gliˈsa	gliˈsɑd	glisɑd	glisa
Frappé	fʁape	fɹʷapɛɪ	fɹæpɛɪ	ˈfɹæpɛɪ	flape
Battement tendu	batmɑ̃ tɑ̃dy	batmɑ̃ tɑ̃dɯw	bætmɑ̃ tɑ̃ dɯw	bætmɑ̃ tɑ̃dɯw	batmɑ̃ tɑ̃dɯw
Piqué	pike	pikɛɪ	ˈpikɛɪ	pikɛɪ	pike

Using the transcriptions in Table 2, we computed a PFI for each teacher pronunciation using the rules of the PFI derivation, and calculated $\Delta PFI = |PFI_{teacher} - PFI_{reference}|$ per term to compare the teach PFI to the standard French value.

Table 3. PFI of the four ballet teachers' pronunciations and absolute deviations (ΔPFI) from the French reference

Term	PFI	PFI H	ΔH	PFI L	ΔL	PFI M	ΔM	PFI Q	ΔQ

	(French reference)								
Balancé	2.0	1.0	1.0	2.0	0.0	2.0	0.0	1.0	1.0
Chaînes	1.0	2.0	1.0	2.0	1.0	2.0	1.0	1.0	0.0
Cambré	2.0	2.0	0.0	2.0	0.0	2.0	0.0	2.0	0.0
Développé	1.0	2.0	1.0	2.0	1.0	2.0	1.0	1.0	0.0
Arabesque	3.0	3.0	0.0	3.0	0.0	3.0	0.0	3.0	0.0
Jeté	1.0	1.0	0.0	2.0	1.0	1.0	0.0	1.0	0.0
Glissade	1.5	1.0	0.5	2.5	1.0	2.5	1.0	1.0	0.5
Frappé	2.0	2.0	0.0	2.0	0.0	2.0	0.0	2.0	0.0
Battement tendu	2.0	2.0	0.0	2.0	0.0	2.0	0.0	2.0	0.0
Piqué	1.0	2.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0

Inspection of the transcriptions reveals accent-linked regularities that help interpret the PFI shifts. Teachers L and M exhibit canonical North American rhoticity and diphthongization, with post-vocalic /ɹ/ and nuclei such as /eɪ/ in *balancé*, *jeté*, and *piqué*, which increase consonantal load and reconfigure vowel structure. This pattern aligns with higher PFIs for some items. For example, *jeté* for Teacher L has a PFI of 2.0, compared to French at 1.0. Teacher M also favours a rising terminal contour on multi-syllabic items such as battement tendu, an intonational choice not counted by the PFI but plausibly relevant to perceived phrasing. Teacher H's pronunciations exhibit flatter prominence with frequent schwa-like support, generally keeping PFIs near the reference except where rhoticity raises the consonant-vowel balance, as seen in examples like *chaînes*, where PFI 2.0. Teacher Q retains uvular /ʁ/ and non-rhotic ending. In her *arabesque*, the uvular /ʁ/ and final /k/ yield a PFI equal to the French reference (3.0), and similar non-rhotic endings elsewhere keep PFIs close to the reference. Notably, all four teachers produced *chaînes* with English-influenced features, most commonly an initial affricate [tʃ͡] and an /eɪ/ nucleus; thus, this item is systematically non-standard relative to the French reference. Although Q's PFI matches the reference value, such segmental departures are not fully captured by the index. In the analyses that follow, we therefore treat *chaînes* cautiously, examining ΔPFI alongside PFI to mitigate term-specific deviation and absorb its term-specific deviation, thereby preventing inflation of between-teacher contrasts.

Table 4. Raw 1-10 staccato-legato perception ratings from four participants for all 40 audios, including per-audio means

Audio #	Participant 1	Participant 2	Participant 3	Participant 4	Mean
Q1	10	7	8	9	8.5
Q2	10	9	8	6	8.25
Q3	6	7	7	7	6.75
Q4	2	6	9	6	5.75
Q5	8	6	8	7	7.25
Q6	8	6	7	6	6.75
Q7	8	6	8	6	7
Q8	9	5	5	5	6
Q9	6	2	7	8	5.75
Q10	3	2	2	2	2.25

H11	8	2	9	6	6.25
H12	10	7	7	6	7.5
H13	7	4	5	7	5.75
H14	3	1	4	6	3.5
H15	7	7	8	6	7
H16	4	3	8	6	5.25
H17	4	7	8	6	6.25
H18	4	6	7	3	5
H19	5	1	5	8	4.75
H20	2	2	2	1	1.75
L21	5	6	4	4	4.75
L22	4	9	4	7	6
L23	7	8	7	7	7.25
L24	9	5	4	4	5.5
L25	4	5	4	7	5
L26	5	6	3	4	4.5
L27	9	6	7	6	7
L28	4	5	3	5	4.25
L29	4	5	3	6	4.5
L30	1	1	3	1	1.5
M31	7	5	7	6	6.25
M32	8	4	4	6	5.5
M33	8	4	3	4	4.75
M34	6	8	6	5	6.25
M35	7	6	4	6	5.75
M36	4	5	4	4	4.25
M37	5	7	8	7	6.75
M38	4	3	4	3	3.5
M39	5	5	5	5	5
M40	3	3	3	1	2.5

Aggregating the four teacher recordings per term (Q/H/L/M) reveals an uneven distribution of teachers in terms of perceived staccato–legato quality, as shown in Table 5. The largest ranges (max-min of per-audio means) occur for Term 1 (*balancé*, range ≈ 3.75), Term 2 (*chaînes*, ≈ 2.75), and Term 4 (*développé*, ≈ 2.75). By contrast, Term 7 (*glissade*) exhibits a tight spread (≈ 0.75). These descriptive differences motivate testing whether pronunciation features, rather than semantics alone, account for variability across teacher pronunciations.

Table 5. Per-audio mean rating (mean of 4 participants) for each teacher recording (Q/H/L/M), with term-level minimum, maximum, range, and mean ± SD

Term	Item	Minimum across four teacher recording means	Maximum across four teacher recording means	Range	Mean ± SD

1	*Balancé*	4.75	8.50	3.75	6.44 ± 1.55
2	*Chaînes*	5.50	8.25	2.75	6.81 ± 1.28
3	*Cambré*	4.75	7.25	2.50	6.12 ± 1.11
4	*Développé*	3.50	6.25	2.75	5.25 ± 1.21
5	*Arabesque*	5.00	7.25	2.25	6.25 ± 1.06
6	*Jeté*	4.25	6.75	2.50	5.19 ± 1.12
7	*Glissade*	6.25	7.00	0.75	6.75 ± 0.35
8	*Frappé*	3.50	6.00	2.50	4.69 ± 1.07
9	*Battement tendu*	4.50	5.75	1.25	5.00 ± 0.54
10	*Piqué*	1.50	2.50	1.00	2.00 ± 0.46

The ordering of ranges suggests that certain lexical items (*balancé*, *chaînes*, *développé*) are more acoustically variable across teachers, with pronunciations diverging enough to shift listeners' staccato-legato judgments by roughly two to four points on a ten-point scale, whereas *glissade* appears comparatively stable. This pattern is consistent with the segmental and prosodic differences highlighted in Table 3, such as phenomena of diphthongization, rhoticity, and final-syllable stress, and it motivates modeling Teacher as a random factor while testing whether PFI and ΔPFI account for between-teacher variance at the term level. Given the small rater sample (N = 4), these inferences are descriptive; formal confirmation will rely on the planned mixed-effects models and sensitivity checks, including the cautious treatment of "*chaînes*."

Within-audio disagreement, meaning the range across four participants for a single audio, is generally modest but shows a few spikes. The largest ranges occur for Audio 4, 11, 19, and 9 (range ≥ 6 points). These items will be flagged for review, as a preregistered rule can exclude an audio if the within-audio range exceeds a threshold (e.g., > 6) or if comments note confusion about the pronunciation.

Table 6. Within-audio rating disagreement for single pronunciations (selected high-range cases)

Audio #	Teacher block	Within-audio range (max.- min. across 4 participants)
4	Q (1–10)	7
11	H (11–20)	7
19	H (11–20)	7
9	Q (1–10)	6

Ratings were collected on a ten-point scale from four raters per audio, so a within-audio range of six points or more signals extreme disagreement and can distort mean-based summaries. To limit this influence, the primary analysis summarizes each rating with the per-audio median and retains all observations, and mean-based summaries are reported in a supplementary table for completeness. As a robustness check, the full model is re-estimated after excluding pronunciations whose within-audio range is at least six or where contemporaneous notes indicate confusion about the stimulus. Effects are considered reliable only when they are observed under both specifications. When the sample is enlarged, the fixed threshold will be replaced by an interquartile-range rule that flags values above the upper fence, defined as the third quartile plus one and a half times the interquartile range, computed on within-audio ranges. This approach preserves transparency while reducing the risk that a small number of highly discordant ratings dominate the results.

As a first pass, we correlated the per-audio mean

rating with $PFI_{teacher}$ across all 40 audios. The association is small (Pearson $r \approx 0.14$; Spearman $\varrho \approx 0.11$). Given N = 4 participants and only one trial per audio, this is underpowered; the planned mixed-effects model ($PerceivedFluidity \sim PFI + (1|Dancer) + (1|Teacher) + (1|Term)$) will provide a more appropriate estimate once additional participants and trials are available. We also computed residuals relative to each term's mean (teacher deviation within term); these residuals will be used in the mixed model to test whether PFI explains within-term differences.

We analyzed n = 4 dancers executing 36 audio tokens (ten terms × four teachers minus "*cambré*," Term 3, which was excluded because dancers were unsure of the prescribed movement), yielding N = 144 trials. As preregistered, we relate execution quality (10-point rubric) to $\Delta PFI = |PFI\_teacher - PFI\_reference|$.

Table 7. Movement execution scored by another RAD ballet teacher for each participant

Audio #	Participant 1	Audio #	Participant 2	Audio #	Participant 3	Audio #	Participant 4
17	7	39	6	38	7	1	8
10	5	34	5	34	7	2	5
9	4	2	3	6	7	3	6
3	4	36	0	36	7	4	4
35	6	5	5	27	6	5	7
33	5	27	6	26	7	6	8
40	6	22	7	32	8	7	6
7	7	31	5	23	7	8	5
37	6	38	4	18	7	9	4
1	5	16	3	31	8	10	7
39	3	40	5	30	5	11	6
13	5	3	4	16	6	12	8
25	4	1	6	25	6	13	5
19	4	12	7	29	7	14	7
38	4	25	6	20	7	15	4
23	5	18	6	3	4	16	6
16	3	6	6	2	5	17	8
26	4	13	5	12	6	18	5
34	6	24	4	13	5	19	7
6	5	4	4	24	5	20	6
20	7	37	7	33	6	21	4
31	6	14	6	4	6	22	8
27	7	11	7	8	7	23	5
14	6	19	4	14	5	24	6
28	3	9	5	22	7	25	7
24	7	8	5	19	6	26	4
18	5	23	4	15	5	27	8
32	5	32	6	40	6	28	6
12	6	35	6	37	7	29	5

4	5	10	7	9	6	30	7
30	7	30	7	1	8	31	6
5	7	17	6	35	6	32	4
8	5	20	5	10	7	33	8
11	4	28	4	17	7	34	5
15	6	33	3	5	5	35	7
22	6	15	6	28	7	36	6
2	7	26	5	39	8	37	5
29	7	29	2	11	8	38	6
21	4	7	7	7	7	39	8
36	6	21	6	21	5	40	5

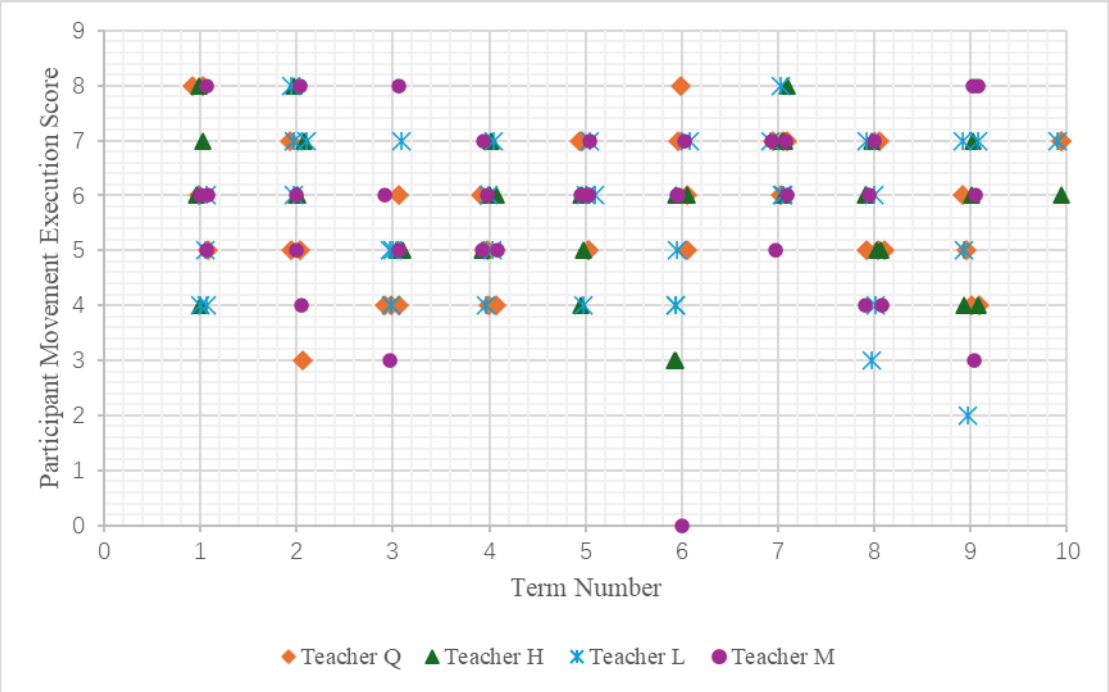
Mean execution score was 5.78/10 (SD = 1.32). Step-type means were comparable (Legato ≈ 5.83; Staccato ≈ 5.72; see Table 7a).

Across all trials, the correlation between ΔPFI and execution was r ≈ −0.04 (tiny, negative). By step type, the association was Legato r ≈ −0.15 and Staccato r ≈ +0.04. A simple OLS with an interaction ($Exec = \beta_0 + \beta_1 \cdot \Delta PFI + \beta_2 \cdot Staccato + \beta_3 \cdot \Delta PFI \times Staccato$) yielded:

Table 8. OLS regression of execution scores on ΔPFI with step-type interaction (all dancers; N = 144 trials; *cambré* excluded)

Parameter	Estimate
Intercept (Legato baseline)	5.969
ΔPFI (Legato slope)	−0.344
Staccato (baseline shift)	−0.240
ΔPFI × Staccato	+0.478

Interpreting the coefficients: for legato-anchored steps, execution decreases slightly as ΔPFI grows (slope ≈ −0.34 per PFI unit); for staccato-anchored steps, the implied slope is small and slightly positive (≈ +0.13). Effect sizes are small, and with a pilot-scale sample we do not draw inferential conclusions here; these estimates guide the preregistered mixed-effects model in the full dataset.



Graph 1. Distribution of single-repetition execution scores by term and teacher

Graph 1 plots data from Table 7 to show the distribution of the four participants' movement execution scores. The x-axis indexes the ten lexical items, and the y-axis is the movement-quality score. Each marker represents one trial, with a small, fixed horizontal jitter applied to reduce overplotting at each term. Marker shapes denote the four teacher recordings (Q, H, L, M). Vertical dispersion within a term quantifies variability in execution across dancers and teacher tokens, but the horizontal spread carries no analytic meaning (jitter only). Execution scores concentrate between 5 and 7 across terms, with modest within-term dispersion and no teacher whose tokens are consistently higher or lower across the set. Legato-anchored items show slightly broader spreads than staccato-anchored items, consistent with greater phrasing variability in sustained actions. These patterns preview the regression results, where ΔPFI shows only a small, directionally negative trend for legato steps and near-zero association for staccato steps.

Consistent with the perception results, *chaînes* was systematically realized with English-influenced segments across teachers. Despite this, ΔPFI effects on execution remained near-zero after accounting for term-level variation, suggesting that segmental deviations alone did not strongly degrade single-rep performance in this pilot. We maintain *chaînes* in models with Term as a random effect and will report sensitivity analyses with this item removed if it influences fixed-effect estimates.

5. Discussion

This pilot set out to test whether pronunciation features of ballet terminology, operationalized by a simple, transparent PFI and its deviation from a reference French token, bias both perceived staccato-legato quality and executed phrasing. Three provisional observations emerge.

Across teachers, the acoustic realization of the vocabulary varied in ways that listeners detected. Several items (*balancé*, *chaînes*, *développé*) exhibited shifts of roughly two to four points in perceived legato–staccato quality across teacher tokens, indicating that segmental and prosodic differences are salient to dancers and perceptually meaningful in the studio context.

At the level of executed movement, the effects observed in this pilot were small. With four dancers and a single repetition per token, ΔPFI showed a near-zero overall association with movement quality ($r \approx -0.04$), with a weak negative tendency for legato steps and essentially no relation for staccato steps. These magnitudes counsel caution and motivate a larger, confirmatory sample before drawing inferential claims.

Finally, the coverage of the index likely constrains what it can predict. The PFI encodes

terminal segment class, consonant-to-vowel balance, and final-syllable stress, but it does not presently represent intonational contours (e.g., rising terminals), within-word timing microstructure, or within-syllable energy profiles. Qualitative observations such as diphthongization, rhoticity, and sentence-final rise suggest that such features plausibly influence phrasing; their absence may explain why ΔPFI alone under-predicts variance in execution.

Limitations are as follows: (1) Power: $N = 4$ dancers and single tokens limit precision, (2) Single repetition: one-rep trials increase noise relative to short phrases, (3) Rater model: one rater per execution introduces measurement error, where inter-rater reliability will be added ($\kappa \geq 0.70$ target), (4) PFI scope: lack of intonational features may attenuate effects. These constraints, common in lab-in-the-studio pilots, motivate the confirmatory mixed-effects analysis and an expanded acoustic feature set.

Within the constraints above, the findings suggest that pronunciation alone is unlikely to strongly improve or harm single-rep execution in trained adolescents; however, for legato-anchored actions, there is a small, consistent trend whereby more edge-accented pronunciation (larger ΔPFI) corresponds to slightly lower smoothness scores. Teachers may therefore benefit from avoiding terminal bursts and front-loaded stress when the goal is connected phrasing, while not expecting dramatic changes from terminology alone.

6. Conclusion

We present an auditory-to-action pilot linking pronunciation features of ballet terminology to both perceived phrasing and executed movement quality in adolescent dancers. The stimulus set balances pedagogical frequency and phonetic contrast, and the PFI offers a replicable bridge between segmental and prosodic cues and a staccato-legato continuum. In perception, between-teacher differences are clear, however, in execution, ΔPFI effects are small at the granularity tested, with a weak negative trend for legato steps. The preregistered next stage will (1) add dancers and repetitions, (2) incorporate a second movement rater and inter-rater calibration, (3) extend the acoustic feature set to include intonation and micro-timing, and (4) fit a mixed-effects model controlling for Dancer, Teacher, and Term. This will allow a decisive test of whether pronunciation acts as an effective motor cue in routine studio pedagogy or whether its influence is primarily perceptual.

These results matter for multilingual classrooms where the same ballet term is often pronounced differently across accents. Those pronunciation differences can subtly steer how students execute a movement and its dynamic, so teachers can choose prosody that matches the quality they want to elicit. In practical terms, when teaching legato-anchored actions, avoiding terminal bursts and final-syllable stress, lengthening syllabic linkage, and aligning the cue with a smooth counting pattern can support connected execution. For staccato-anchored actions, a crisper onset, clearer consonantal release, and stress timed to the intended accent can reinforce articulation without sacrificing accuracy. These recommendations do not require uniform pronunciation but intentional prosody that matches the target quality, especially in mixed-accent classrooms where learners hear multiple realizations of the same term.

Mechanistically, the results are consistent with entrainment accounts of music and speech: predictable acoustic structures, such as stress placement, consonant-vowel balance, and terminal energy, provide weak but usable timing information that the motor system can adopt when planning action. The small execution effects observed here likely reflect the limited strength of the speech cue relative to the richer entrainment sources available in class, such as teacher demonstrations and musical accompaniment. Future confirmatory work should therefore model speech prosody alongside music-derived metrical structure and explore brief multi-rep phrases, where entrainment benefits typically grow.

Overall, this study contributes an explicit, replicable bridge between speech acoustics and movement pedagogy in ballet. By formalizing pronunciation with a transparent index and testing its relation to both perception and execution in an ecologically valid RAD setting, we show how linguistic variation can be leveraged, rather than avoided, to make verbal cues more informative. For teachers, the takeaway is pragmatic: pair terminology with prosody that mirrors the intended phrasing; for researchers, the next step is to integrate speech, music, and demonstration in a unified entrainment model of classroom skill acquisition.

References

Association, I. P. (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.

İşsever, H., & İşsever, İ. (2024). The harmony between ballet teacher and ballet accompanist: an examination from the perspective of musical technical elements. *Journal for the Interdisciplinary Art and Education*. https://doi.org/10.5281/zenodo.13765619

Kershenbaum, A. M., Galassi, M., Shattuck-Hufnagel, S., Bachan, S., & Zipse, L. (2023). The effect of prosodic timing structure on unison production in people with aphasia. *American Journal of Speech-Language Pathology*, *33*(6S), 3143–3169. https://doi.org/10.1044/2023_ajslp-22-00304

Pranjić, M., Janzen, T. B., Vukšić, N., & Thaut, M. (2024). From sound to movement: mapping the neural mechanisms of Auditory–Motor entrainment and synchronization. *Brain Sciences*, *14*(11), 1063. https://doi.org/10.3390/brainsci14111063

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

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*(7), 547–558. https://doi.org/10.1038/nrn2152

Appendix

Participant Perception Scoring Sheet

How to rate "staccato" vs "legato"?

When you hear each audio clip, focus on how the sound feels (not the movement you'll do next!)

- Staccato = short, crisp, punchy, separated syllables (like dots).
- Legato = smooth, connected, flowing syllables (like a long line).

Use the 1–10 slider:

- 1–3 = very staccato (choppy, lots of separation)
- 4–7 = mixed/medium (some connection, some bite)
- 8–10 = very legato (smooth, continuous)

There are no right or wrong answers. Please rate your first impression, and try to use the whole scale across clips.

Term #:

Staccato Legato

1 2 3 4 5 6 7 8 9 10

Term #:

Staccato Legato

1 2 3 4 5 6 7 8 9 10

Term #:

Staccato Legato

1 2 3 4 5 6 7 8 9 10