Effort

Junxiu Tang[1] **Hanna Renedo**[1] **Yifang Wang**[2] **Dashun Wang**[1]
[1]Kellogg School of Management, Northwestern University, Evanston, IL 60201, USA
[2]Department of Computer Science, Florida State University, Tallahassee, FL 32301, USA
{junxiu.tang,hanna.renedo,dashun.wang}@kellogg.northwestern.edu
yifang.wang@fsu.edu

Abstract

Effort returns the trace of a model learning from human dance to the body. The training log becomes both music and a sequence of Laban Effort words. The music conditions the trained model and accompanies a human dancer, while the words guide their improvisation as choreographic prompts. The generated dance and the improvised dance share one training trace and one timeline, but take shape through different forms. As movement passes from bodies into data, through training, and back into bodies, the work asks how agency is distributed across dancers, data, models, and artistic interpretation. Who, or what, is making the dance happen?

1 From bodies to data and back

When a machine learns to dance from human bodies, what happens between the recorded dancers and the dance it produces? *Effort* begins with human dance and returns to a human body through a machine. An open-source music-conditioned diffusion model, Lodge [1], is trained on FineDance [2], a motion-capture dataset performed by twenty-seven dancers credited as a group. During training, the changing loss and gradients leave a numerical trace. We take this trace out of the engineering background and turn it into two forms. It becomes music, and it becomes a sequence of Laban Effort words [3] such as *Float*, *Press*. The music conditions the trained model and accompanies a human dancer, while the words become a score for their improvisation. In contemporary dance and performance practice, a "score" is an open choreographic instruction or framework for a performance. A generated body, drawn as pale continuous lines from the model's own motion, and a living body then move within the same temporal structure, and each meets the training trace through a different representation. The generated body is drawn as pale continuous lines because the model outputs joint positions over time, and the lines follow those positions directly. Motion capture reduced the dancers to points in space, and the drawing repeats that reduction. Across the video, the training curves and Effort words cross the frame while the two bodies separate and overlap. The work asks how agency is distributed when bodies, data, training, models, and interpretation all take part in one cycle.

2 A training trace becomes a score

Existing music-driven dance-generation systems, such as Bailando [4] and EDGE [5], treat training primarily as a means to obtain a generative model and generated motion. *Effort* instead treats the numerical trace of training itself as artistic material and translates it into music, Laban Effort words, and movement. The log records how the model changes as it learns statistical patterns from human movement. It documents the transformation from the dancers' original motion to the generated dance. Lodge [1] is the most advanced music-to-dance model available with open training code and a public dataset, which made it the strongest model we could choose. We trained it from scratch so the trace covers the full training run, from the first batch to the last, and because it is small enough to train on

*Human dance performance and improvisation by Qianni Wang.

40th Conference on Neural Information Processing Systems (NeurIPS 2026). Creative AI Track.



Figure 1: Composite image from *Effort*, combining the training trace, sonification, Effort words, model-generated skeletal movement, and human improvisation within a single visual field.

one GPU in about two days, logging every batch was practical. FineDance [2] is the motion-capture dataset Lodge was built for. Its twenty-seven professional dancers were recorded across twenty-two genres, so the model learns from many bodies and many styles, and the trace records that range entering one set of weights.

A change-point detector divides the trace into fourteen phases, and an artist-designed mapping gives one Laban Effort word [3] to each. Laban's factors describe how a movement is weighted, timed, and directed. We paired each factor with the training quantity that plays the same role in optimization. The gradient norm is the force of each update on the weights, so we read it through *Weight*. Its volatility within a phase shows the evenness of the updates, so we read it through *Time*. The strength of the temporal trend shows the steadiness of a phase's direction, so we read it through *Space*. We set each threshold from the run's own statistics, so each word describes a phase relative to the whole run. We chose the pairing as an analogy and applied it to every phase.

A second artist-designed mapping turns the same trace into a musical composition whose data-driven parameters and phase structure can be reproduced from the log. We then re-perform that composition with Suno [6], a generative music service, which primarily changes its timbre and production while preserving the source melody and structure. One cover is heard in the video, and a second cover of the same score conditions the model, so the human dancer and the model share one temporal environment. Their movements emerge from different forms of instruction. The model generates a skeletal dance from the music, while the human dancer hears it, interprets the Effort words, and translates words and sound into embodied interpretation.

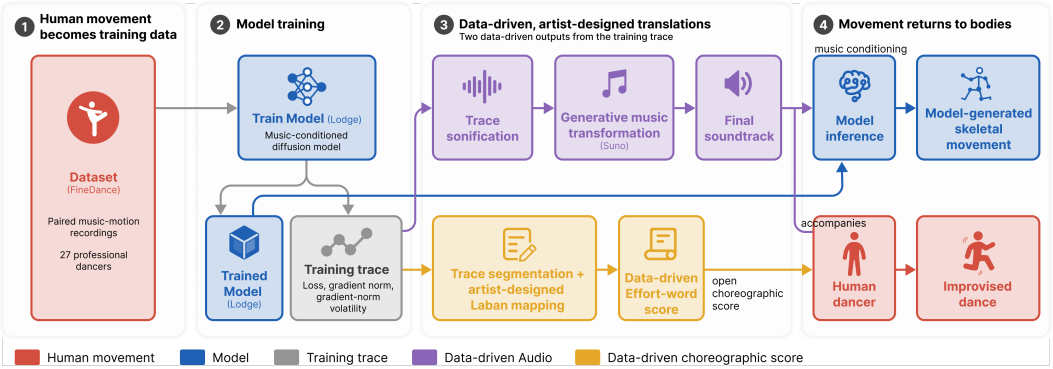


Figure 2: The implemented process. (1) FineDance provides paired music and motion recordings from professional dancers for training. (2) The training process produces both a trained model and a training trace composed of loss, gradient norm, and gradient-norm volatility. (3) Artist-designed procedures translate the trace into a soundtrack and Laban Effort score. (4) The soundtrack conditions the model and accompanies the human dancer, while only the dancer receives the word score.

3 What makes the dance happen?

The cycle begins in dance, and the trace returns to dance. Motion capture turned the movement of twenty-seven dancers into numbers, and training turned those numbers into weights. Those movements now shape all of the model's generated motion, and the dataset credits the dancers as one group. Returning the training trace to a living body runs that path backward, and the vocabulary of that reading, *Weight*, *Time*, and *Space*, comes from dance as well. An Effort word does not prescribe a single movement. For the human dancer, it passes through their training and embodied judgment before becoming movement. Viewers may still perceive intention or expressiveness in the generated movement, although the model itself has no expressive intention. *Effort* asks what makes movement happen when action is distributed across dataset performers, training data, the model, artist-designed mappings, and the dancer's embodied judgment. Each of these elements changes the conditions of the dance, and the dance needs all of them together. Much like Sol LeWitt's Wall Drawings [7], *Effort* separates the production of an instruction from its bodily realization.

The model's training trace, read through an artist-designed mapping, produces the Effort-word score, while the dancer interprets that score through a movement that cannot be fully specified in advance. Laban used the word "Effort" for the inner impulse behind a movement. The title *Effort* gives that word to a training run. The run's trace is visible, and the impulse behind it spreads across the whole cycle. A body becomes data. The data becomes a model. The model's training trace becomes music and words, and the music and words become movement again. That movement may itself be recorded, circulated, and returned to data in the future. After this circuit, whose dance are we watching? Who, or what, is making it happen?

Code and data. The code and the derived artifacts are available at `https://github.com/ahugh19/Effort`, and the appendix follows the references. FineDance is used under its own license and remains with the dataset authors.

4 Author Bios

Junxiu Tang is an interdisciplinary researcher across computer science, digital media art, and social science. His research studies how people contribute to, read, and act on data. His art practice makes these processes material, spatial, and perceptible, while examining what is lost, invented, or redistributed through them. His works have been exhibited at the SIGGRAPH Art Gallery, IEEE VIS Arts Program, and Information+ Conference. Homepage: https://ahugh19.github.io/.

Hanna Renedo is an interdisciplinary visual artist and designer at the Center for Science of Science and Innovation at Northwestern University. Her research-based practice explores the intersections of science, visual art, and technology through artists' books, information design, and experimental image-making. She is a recipient of the Caxton Book Arts Grant, and her work is housed in several special collections, including the School of the Art Institute of Chicago and the Bainbridge Island Museum of Art. She earned a BA in biology from Bowdoin College and an MFA in Visual Communication Design from the School of the Art Institute of Chicago.

Yifang Wang is a tenure-track Assistant Professor in the Department of Computer Science at Florida State University. Her research lies at the intersection of Data Visualization, Human-AI Collaboration, and Human-Computer Interaction. She develops novel VIS and HAI techniques and tools that support general data-driven tasks and domain-specific applications in areas such as the science of science and innovation, science policy-making, public health, and more. She also experiments with data-driven art, aiming to invite reflection, provoke emotion, and encourage public discussion around technologies and social issues. Her artworks have been exhibited in the United States, the United Kingdom, Australia, South Korea, and Hong Kong. Homepage: https://wangyifang.top/.

Dashun Wang is the Kellogg Chair of Technology and a Professor at the Kellogg School of Management and the McCormick School of Engineering at Northwestern University. Dashun is a recipient of multiple awards for his research and teaching, including the AFOSR Young Investigator Award, Poets & Quants Best 40 Under 40 Professors, Junior Scientific Award from the Complex Systems Society, Erdos-Renyi Prize, Thinkers50 Radar 2021, and more.

References

[1] Ronghui Li, Yuxiang Zhang, Yachao Zhang, Hongwen Zhang, Jie Guo, Yan Zhang, Yebin Liu, and Xiu Li. Lodge: A coarse to fine diffusion network for long dance generation guided by the characteristic dance primitives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[2] Ronghui Li, Junfan Zhao, Yachao Zhang, Mingyang Su, Zeping Ren, Han Zhang, Yansong Tang, and Xiu Li. Finedance: A fine-grained choreography dataset for 3d full body dance generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10234–10243, 2023.

[3] Rudolf Laban and F. C. Lawrence. *Effort*. Macdonald and Evans, London, 1947.

[4] Li Siyao, Weijiang Yu, Tianpei Gu, Chunze Lin, Quan Wang, Chen Qian, Chen Change Loy, and Ziwei Liu. Bailando: 3d dance generation by actor-critic gpt with choreographic memory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11050–11059, 2022.

[5] Jonathan Tseng, Rodrigo Castellon, and Karen Liu. Edge: Editable dance generation from music. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 448–458, 2023.

[6] Suno, Inc. Suno. Generative music service, https://suno.com, 2026. Accessed 2026.

[7] Sol LeWitt. Doing wall drawings. *Art Now: New York*, June 1971.

A Model training and dataset

Effort uses Lodge [1], a two-stage coarse-to-fine diffusion network that generates dance motion conditioned on music, trained from scratch for this project. The trace analyzed here comes from the Global module, the coarse stage of the network. Music enters the model as 35-dimensional handcrafted audio features, and motion is represented as a 139-dimensional feature vector over 22 joints. The trace therefore records the optimization history of this specific run. The training data is FineDance [2], a motion-capture dataset of paired music and full-body 3D dance recordings performed by twenty-seven professional dancers, used in its standard training split. Training ran for 3,500 epochs on a single NVIDIA H100, about 51 hours of active GPU time. The released configuration records the optimizer settings: Adan with a learning rate of 4×10^{-4} held constant, batch size 128 with gradient accumulation of three for an effective batch of 384. We disabled gradient clipping, so the logged gradient norm is the raw norm. A logging callback writes one row per batch, giving 155,232 rows over 52,516 optimizer steps. Each row records the step, epoch, total and component losses, gradient-norm statistics, the learning rate, and binary indicators for loss spikes, plateaus, and exploding or vanishing gradients, and every indicator is false in every row.

The artwork uses three signals from the log: training loss, mean gradient norm, and gradient-norm volatility. Volatility is measured two ways. At the phase level, it is the standard deviation of the gradient norm over the phase's downsampled rows, and the ratio of that standard deviation to the phase mean drives the Effort mapping in Section C. At the per-batch level, the log stores a sign-change rate, the fraction of gradient coordinates whose sign flips between consecutive batches, and the sonifier in Section B uses it for rhythm.

A change-point detector divides the run into fourteen contiguous phases. The per-batch log is downsampled to every fifteenth row in acquisition order, and the PELT algorithm from the `ruptures` library is run on the smoothed loss and on the gradient norm with a radial-basis-function cost, a penalty of 10, a merge window of 200 rows, and a minimum segment length of 500 rows, all in downsampled rows. The two sets of breakpoints are merged and the minimum-length constraint is enforced, which yields fourteen phases. Table 2 lists the per-phase statistics. The run opens with a phase of large, unstable gradients, followed by thirteen lower-amplitude phases of gradual convergence in which the mean loss falls from about 0.116 to about 0.0098. The artistic transformations described below operate on this continuous training history.

B Translating the training trace into music

The first translation turns the per-batch training log into a musical score through an artist-designed, rule-based procedure. Given the same log and settings, the procedure regenerates the same symbolic score. The score is written for a mechanical player piano with a small chamber ensemble, in D natural minor, in 4/4, at 70 beats per minute. The player piano carries the melody, strings and a low synthesizer drone support it, and celesta, harp, contrabass, and string pad enter occasionally. Four quantities from the log control four musical parameters, each normalized against this run's own range.

1. **Register.** The gradient norm sets the register of the melody and accompaniment. Larger gradients move the material upward.
2. **Rhythmic density.** The gradient sign-change rate sets the number and spacing of note onsets. Stable intervals give sparse, sustained notes, and volatile intervals give denser subdivisions and syncopation.
3. **Tremolo and chromatic inflection.** Bars with a change in loss above the 82nd percentile across the piece receive a string tremolo and unresolved chromatic neighbor tones. The lines are otherwise diatonic.
4. **Spectral brightness.** The gradient level sets the brightness of the synthesizer pad, so the pad darkens after the high-gradient opening.

Data controls register, rhythmic density, the accents, and the phase durations. The key, the harmonic progressions, the instrumentation, and the five-part form are artist-designed.

The fourteen training phases are laid out in training order on one linear clock. Each phase receives bars in proportion to its length, with a floor of two bars to keep every phase audible. Music time

therefore follows training time at bar-level resolution. Over this clock the instrumentation follows a fixed five-part shape, a piano-led opening, a build, a short caesura, a denser passage, and a decay as the layers drop out. The score ends on an open fifth, D and A, and its final four bars slow from 70 to about 38 beats per minute. The Effort words in Section C are indexed by the same fourteen phases, so the two translations share one clock.

C Translating the training trace into Laban Effort words

The second translation gives one Laban Effort word to each of the fourteen phases through the three Effort factors of *Weight*, *Time*, and *Space* [3]. The mapping is an artist-designed convention applied to the statistics of this run.

For phase p, let μ_p be its mean gradient norm, σ_p the standard deviation of the gradient norm, and τ_p its temporal trend, the slope of a least-squares fit of the gradient norm over the phase. Let m_μ and m_τ be the medians of μ_p and $|\tau_p|$ over the fourteen phases. *Weight* is strong when $\mu_p \geq m_\mu$ and light otherwise. *Time* is sudden when $\sigma_p \geq \mu_p$ and sustained otherwise. *Space* is direct when $|\tau_p| \geq m_\tau$ and indirect otherwise. *Weight* and *Space* compare a phase with the whole run, and *Time* compares a phase's variability with its own mean. The three binary factors give the eight Effort words in Table 1.

Weight	Time	Space	Effort
light	sustained	indirect	*Float*
light	sustained	direct	*Glide*
light	sudden	direct	*Dab*
light	sudden	indirect	*Flick*
strong	sustained	direct	*Press*
strong	sustained	indirect	*Wring*
strong	sudden	direct	*Punch*
strong	sudden	indirect	*Slash*

Table 1: Artist-designed mapping from the three binary Laban Effort factors to eight Effort words. Each training phase is assigned one combination.

The first phase, with large and unstable gradients and a steep downward trend, is strong, sudden, and direct, the Effort *Punch*. It is the only sudden phase, since every later phase has a standard deviation below its mean. The run then passes through *Press*, settles into *Float* and *Glide* through its low-gradient middle, and ends in *Press* and *Wring* as the gradient rises again. Table 2 lists the full assignment. The resulting Effort sequence is a choreographic interpretation constructed through our selection of signals, segmentation, thresholds, and factor mapping. In the performance the dancer receives each word as an open score. The word sets the quality of a movement, and the pose and trajectory remain their choice.

D Suno re-performance and temporal alignment

The rule-based score fixes the melody, harmony, form, phase durations, and tempo. For the video we passed its rendering to Suno [6], a generative music service, in its Cover mode. This step is a generative re-performance outside the deterministic mapping, and Suno can return a different cover on each request, so we keep the selected outputs as fixed artifacts.

The presented cover was generated with model V5.5 as an instrumental, with Audio Influence 85%, Style Influence 15%, and Weirdness 5%, from the following prompt. We generated six candidates with the same prompt and settings and selected the take closest to the source player-piano motif and its restrained dynamic. The file is used as Suno delivered it.

> COVER FAITHFULLY — preserve the uploaded melody, chord changes, tempo (keep the final ritardando), and section structure. Keep the dynamics restrained from start to end: quiet, chamber-scale, mezzo-piano at most. The "climax" is only a slight thickening of texture, then it thins again. A small, cold ensemble: a mechanical player piano stating the melody with precise, unhurried articulation; one solo cello answering low and dry; a faint, even eighth-note pulse in muted low strings, felt more than heard; sparse high violin harmonics for air. Wide space between phrases, generous silence, cool and exact intonation. D minor,

Phase	Start step	Epoch	μ_p	σ_p	τ_p	Weight	Time	Space	Effort
1	0	0	2.741	5.017	-1.6×10^{-2}	strong	sudden	direct	*Punch*
2	2593	172	0.343	0.273	-2.3×10^{-4}	strong	sustained	direct	*Press*
3	5610	374	0.235	0.175	$+9.4 \times 10^{-6}$	light	sustained	indirect	*Float*
4	8704	580	0.233	0.200	-9.8×10^{-5}	light	sustained	direct	*Glide*
5	11260	750	0.248	0.198	$+1.0 \times 10^{-4}$	light	sustained	direct	*Glide*
6	15019	1001	0.272	0.170	$+3.2 \times 10^{-5}$	light	sustained	indirect	*Float*
7	18292	1219	0.283	0.180	-1.4×10^{-5}	light	sustained	indirect	*Float*
8	21155	1410	0.313	0.245	$+1.3 \times 10^{-4}$	light	sustained	direct	*Glide*
9	25016	1667	0.341	0.204	$+5.3 \times 10^{-5}$	light	sustained	direct	*Glide*
10	28902	1926	0.370	0.218	$+6.5 \times 10^{-5}$	strong	sustained	direct	*Press*
11	33530	2235	0.386	0.221	$+4.6 \times 10^{-5}$	strong	sustained	indirect	*Wring*
12	36394	2426	0.398	0.247	$+2.2 \times 10^{-5}$	strong	sustained	indirect	*Wring*
13	43374	2891	0.379	0.209	$+1.3 \times 10^{-5}$	strong	sustained	indirect	*Wring*
14	45931	3062	0.370	0.207	$+3.2 \times 10^{-6}$	strong	sustained	indirect	*Wring*

Table 2: The fourteen phases of the training run. Start step is the first logged optimizer step of each phase. Epoch is the approximate epoch at that boundary. The phases are contiguous, so a phase runs to the next phase's start. The remaining columns give the mean gradient norm μ_p, the gradient-norm standard deviation σ_p, the temporal trend τ_p, the three binary factors, and the resulting Effort word. The median of μ_p is 0.342 and the median of $|\tau_p|$ is 4.9×10^{-5}.

> approximately 70 BPM, 4/4. The ending slows and settles on a bare open fifth, unresolved, fading into a dry, cold room tail. Mood: wintry, elegant, austere, lucid, still; mechanical calm rather than passion. Instrumental only, close-miked modern classical recording, high detail, quiet mastering.

The cover changes the timing of the score, so we align the visual materials to it with chroma-based dynamic time warping, a monotone map from score time to cover time. The plotted curves, bar and phase boundaries, and Effort-word changes all follow the timing of the audio in the video.

The dance is conditioned on a second cover of the same score, arranged in the style of Kunqu opera accompaniment. The model produces more coherent motion from this cover than from the sparse texture of the presented cover, and its movement carries the stillness and austerity closest to the finished work. This cover keeps the key, phrase structure, and five-part form of the score at close to the same tempo, and differs from the presented cover in instrumentation and surface character. The generated dance is aligned to the presented cover with the same warping. It was generated with model V4.5, Audio Influence 85%, and Weirdness 0%, from the following prompt.

> COVER, REARRANGED IN THE STYLE OF KUNQU OPERA ACCOMPANIMENT. Keep the source's slow tempo, minor-key character, and melodic contour. Replace the Western orchestration with traditional Kunqu instrumentation. D minor with pentatonic flavor, approximately 65 BPM, 4/4. Production: bamboo flute (qudi or dizi) carrying the melodic line, sheng (mouth organ) sustained drones, pipa or sanxian plucked accents on phrase boundaries, soft guzheng arpeggios, light wooden percussion (wooden clappers and paigu on downbeats), small bells. Subtle wordless female vocalise in Kunqu falsetto style — breathy, soft, ornamented — hovering as background, sparse, only in climax and outro. Natural temple-hall reverb. Mood: meditative, classical Chinese, sacred lament, slow and flowing. Suitable as score for Kunqu and classical Chinese stage choreography. Mostly instrumental.

E Reproducibility and retained artifacts

The release at `https://github.com/ahugh19/Effort` contains the per-batch training log, the phase summary, the symbolic score and its mastered render, the two Suno covers, the alignment timeline, and the code for each deterministic step. The pipeline runs in the following order. Steps 1 and 5 are outside the release, and their outputs are kept as fixed files.

1. Train Lodge on FineDance and write the per-batch log.
2. Segment the log into fourteen phases.

3. Generate the symbolic score and, optionally, a dry preview render.
4. Assign one Effort word to each phase.
5. Generate and select the Suno cover.
6. Align the score to the presented cover by dynamic time warping.
7. Render the curve, word, dancer, and skeleton layers on the aligned timeline.

The exact parameters of steps 2, 3, 4, and 6 are recorded in the released scripts. The mastered render stays a fixed file because mastering happened outside the package, and the covers stay fixed files because the service returns a different result on each request. The FineDance recordings remain with their authors [2] under the dataset license, and the released material is provided under the license stated in the release README.