Parallel Harmonic Exploration for Musical Accompaniment with MODULO

Anonymous Author(s)

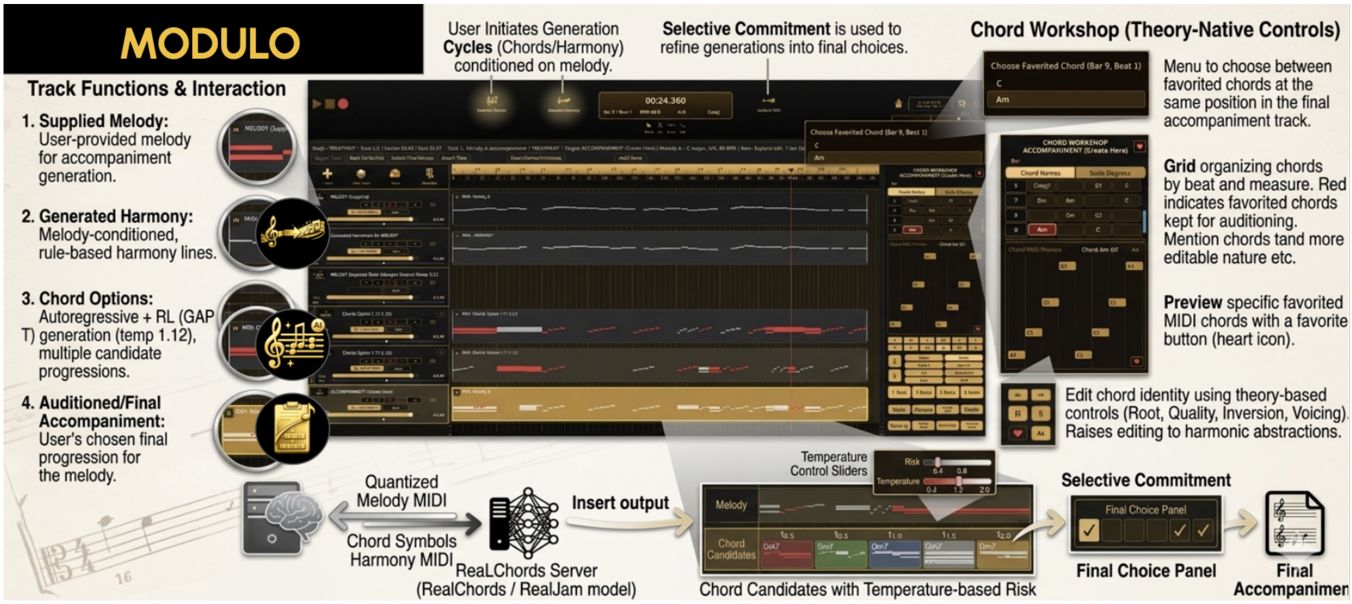


Figure 1: The Modulo interface. A melody track (top) anchors the timeline; generated accompaniment candidates appear as parallel tracks synchronized beat-for-beat. Musicians solo tracks to audition, mark favorites, and assemble the final accompaniment through selective commitment.

Abstract

Prompt-based music generators produce finished audio that musicians cannot structurally edit, while conventional DAWs offer full control but require iterative recording and meticulous MIDI editing. We present Modulo (Musician-Owned DAW for User-Led Orchestration), a generative audio workstation (GAW) that embeds structured harmonic co-creation into the DAW timeline via three mechanisms: (1) a chord generation system backed by an RL-tuned transformer producing multiple customizable accompaniment candidates as synchronized MIDI tracks; (2) the Chord Workshop, a graphical editor for chord-level refinement of root, quality, inversion, and voicing; and (3) a harmony engine applying key-aware rules at configurable intervals. A within-subjects study showed that Modulo halved time to first satisfactory accompaniment, tripled unique chord roots, nearly quadrupled unique chord qualities, and sustained 2.8× longer post-satisfaction refinement, with unanimous preference over typical DAW workflows.

CCS Concepts

• **Human-centered computing** → **HCI design and evaluation methods**; **Interactive systems and tools**.

Keywords

human-AI; co-creation; music; DAW; harmony

ACM Reference Format:

Anonymous Author(s). 2026. Parallel Harmonic Exploration for Musical Accompaniment with MODULO. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST '26).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Composing a chordal accompaniment is rarely a one-shot task. Musicians typically cycle through numerous alternatives, comparing against melody, rearranging fragments, and refining harmonic details. The dominant tools for accompaniment creation do not support this workflow well. In conventional DAWs, harmony construction is mediated through note-level MIDI editing, making even simple chord substitutions or voicing changes unnecessarily laborious. At the other extreme, prompt-based music generation systems such as Suno, Udio, and MusicGen [7] produce polished audio quickly, but offer no way to inspect or revise individual musical decisions such as chord choices, a limitation visible in how users develop elaborate prompting workarounds to compensate for the lack of direct editing [5, 6]. This leaves a gap between *generation* and *musical control.*

Prior work suggests that musicians want AI to contribute material without displacing authorship [15], and that steering tools

UIST '26, Detroit, MI, USA
2026. ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

are essential to maintaining creative ownership [16]. In accompaniment composition, the central challenges are obtaining chords and navigating a space of plausible harmonic directions while retaining the ability to compare, edit, and commit at musically meaningful levels of abstraction. We argue that this is fundamentally an interface problem as much as a modeling problem. If generated harmony is to function as co-creative material rather than a finished answer, it must be embedded in a workflow that supports exploration and deliberate choice.

Modulo (Figure 1) contributes to the recent emergence of generative audio workstations but focuses on harmonic ideation and musically meaningful abstraction over generative one-shot features. Rather than replacing the DAW with an external generator, Modulo treats generation as structured material for in-context comparison and refinement.

This paper makes three contributions:

- **Modulo, an original, standalone GAW** for harmonic co-creation built around parallel exploration, selective commitment, and in-context audition of generated accompaniment candidates.
- **Chord Workshop, a direct-manipulation interface** that raises editing from MIDI notes to harmonic structure, including root, quality, inversion, voicing, and rhythmic feel.
- We report a **within-subjects study with 25 participants** showing that this design substantially accelerates time to first satisfactory accompaniment while broadening harmonic exploration and increasing sustained refinement.

2 Related Work

Prompt-based generation and co-creative tools. Transformers like MusicGen generate audio from text or melody conditioning [7], but outputs cannot be musically edited. Large-scale analysis of Suno and Udio reveals elaborate prompting workarounds that compensate for this lack of direct control [5], and musicians prefer modular chord-level suggestions over completed audio tracks in order to preserve creative ownership [14]. AI-steering tools such as sliders and multiple alternatives increase ownership [16], and musicians most value AI-generated material as fragments to selectively adopt [15, 23]. Co-creative music interfaces have explored tactile [17], gestural [22], neurosymbolic harmony [3], live improvisation [2], and mashup paradigms [20], but none embed harmonic editing within a full DAW workflow. DAW-integrated AI has addressed mixing [25] and melody-aware notification [26], not harmonic composition.

Chord accompaniment and musician agency. ReaLchords uses RL to fine-tune a transformer for real-time chord accompaniment [27]; its jamming interface prompted users to request DAW integration [21]. Musicians want AI as a tool with ownership over "the moments instead of the whole" [15], and professional creators value constrained randomness over autonomous generation [13]. DAWs are recognized as valid sites for interactive music innovation [19], yet no existing system integrates chord generation with theory-aware editing inside the DAW timeline. Modulo fills this gap by wrapping RL-tuned models in a batch-generation paradigm with structured refinement tools.

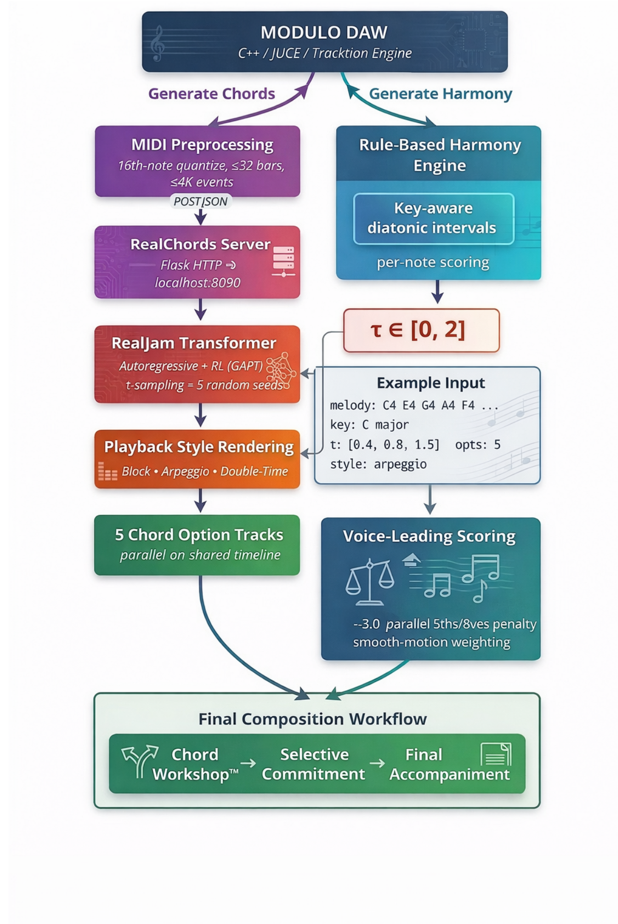


Figure 2: System architecture. *Generate Chords* (left) sends preprocessed MIDI melody to a local ReaLChords server [21]; returned chord symbols are rendered with user-selected playback styles. *Generate Harmony* (right) applies key-aware voice-leading rules at five interval strategies. Both feed into the Chord Workshop for refinement and selective commitment.

3 System Design

Modulo is built in C++ with JUCE [12] and on top of Tracktion Engine [24], providing standard DAW functionality (multi-track recording, mixing, EQ, VST/AU hosting) alongside generative harmonic tools. Generative features beyond chords and harmony exist in Modulo but are outside the scope of this study. Figure 2 illustrates the architecture.

3.1 Generate Chords: Parallel Harmonic Ideation

The *Generate Chords* dialog produces multiple accompaniment candidates conditioned on the melody and key. Modulo hosts a local Flask server (port 8090) wrapping the ReaLJam agent [21, 27], a decoder-only autoregressive transformer fine-tuned with reinforcement learning using the Generative Adversarial Post-Training

(GAPT) variant. Before inference, a copy of the MIDI melody is preprocessed to meet model input constraints: quantized to sixteenth-note frames (4 per beat) to normalize timing inconsistencies from live performance, trimmed to 32 bars to match the model's training context, and thinned to 4,000 events to stay within the model's sequence-length limit. These steps apply only to the inference copy; the user's original melody on the timeline is never modified, and generated chords are placed on separate tracks aligned to the original timing.

Temperature-controlled sampling. Temperature $\tau \in [0, 2]$ scales the softmax logits:

$$p(c_t \mid c_{<t}, \mathbf{m}) = \frac{\exp(z_{c_t}/\tau)}{\sum_{c'} \exp(z_{c'}/\tau)} \tag{1}$$

Low τ produces conservative diatonic progressions; high τ tends to yield more chromatic alternatives. Up to five temperature columns run simultaneously with different seeds, spanning the conservative-to-chromatic range so that users can compare harmonic risk levels side by side.

Playback styles. The model returns chord symbols, and Modulo renders these as **block** (simultaneous), **arpeggio** (sequenced by pitch), **block & arpeggio mixture**, or **double-time** (re-attacked at twice the harmonic rhythm for block or arpeggio style).

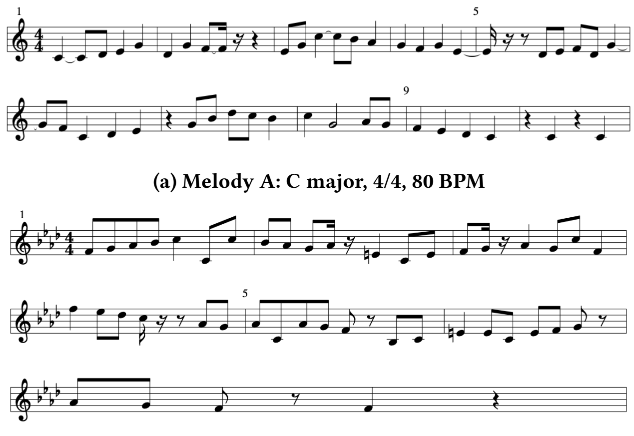
Each candidate appears as a synchronized MIDI track (Figure 1). Users audition by soloing, mark **favorites**, then use the **Final Choice Panel** to select one chord per beat slot and commit the result. This *selective commitment*—choosing individual chords rather than accepting a full generation—ensures every harmonic decision reflects deliberate authorial choice, consistent with findings that musicians value fragment-level adoption over wholesale AI output [15] and that multiple alternatives increase creative ownership [16].

3.2 Chord Workshop™

The Chord Workshop (Figure 1, right) is a direct-manipulation editor operating on harmonic abstractions rather than MIDI notes. It presents the accompaniment as labeled chord cells with controls for root (12 chromatic buttons), quality (8 types), inversion, octave, beat duration, and playback style. A toggle switches the interface between absolute names ("Dm7") and Roman numerals ("ii^7") for wider accessibility; chords are highlighted in red after the favorite button is pressed. A user who wants to try a substitution in their accompaniment composition simply presses a button rather than identifying which pitches to alter in the piano roll.

3.3 Generate Harmony: Rule-Based Voice Leading

The *Generate Harmony* engine creates up to five simultaneous harmony lines, each following a distinct interval strategy so that the musician can audition contrasting voicings side by side. Given a melody note at scale degree d in scale $S = \{s_0, \ldots, s_6\}$, four fixed-interval strategies select a target degree d': **diatonic 3rd above** (d+2 mod 7), **6th below** (d−2 mod 7), **perfect 5th above** (d+4 mod 7), and **contrary motion** (alternates between 3rd above and 6th below depending on melody direction to maximize voice separation). A fifth **adaptive** strategy selects greedily between the



(a) Melody A: C major, 4/4, 80 BPM

(b) Melody B: F minor, 4/4, 70 BPM

Figure 3: Study melodies in contrasting keys.

diatonic third and sixth for each note by scoring both candidates:

$$\text{score}(h) = w_{\text{int}} \cdot I(h, p) + w_{\text{sm}} \cdot |h - h_{\text{prev}}| + w_{\text{ctr}} \cdot \mathbb{1}[\text{contrary}] \tag{2}$$

where $I(h, p)$ rewards consonant intervals following perceptual consonance hierarchies [11] (thirds: +3.2, sixths: +3.0, fifths: +1.3) and penalizes dissonance (−1.8); $w_{\text{sm}} = -0.10$ penalizes large leaps to encourage stepwise motion; and $w_{\text{ctr}} = +1.1$ rewards contrary motion between melody and harmony. Parallel perfect fifths and octaves incur a −3.0 penalty, enforcing a standard counterpoint prohibition [1]. These weights were hand-tuned by the first author during iterative listening tests, following the rule-based harmonization paradigm established by Ebcioğlu [8]. For each melody note, both candidates are scored and the higher-scoring pitch is selected greedily, yielding a linear-time algorithm. Both major and natural minor scales are supported.

4 User Study

We conducted a within-subjects study comparing the full Modulo system (treatment) against the same DAW without generative or Chord Workshop tools (control).

4.1 Participants and Design

Twenty-five musicians (6 beginner, 8 intermediate, 7 advanced, 4 expert in piano, composition, and DAW experience) were recruited through university music department mailing lists and professional industry contacts (study determined exempt by institutional review board). Each participant completed two accompaniment tasks per condition, yielding 50 paired observations. Condition order was counterbalanced (11 treatment-first, 14 control-first) to control for practice and novelty effects. Tasks used contrasting melodies (Melody A: C major, 80 BPM; Melody B: F minor, 70 BPM; Figure 3). Instructions: *"Build a chordal accompaniment you are satisfied with. Press 'Mark First Satisfied' when you would be happy to keep the result, then continue refining if you wish."*

The *first satisfaction* marker separates initial harmonic search from subsequent refinement, providing a richer decomposition than total time alone. Participants thought aloud throughout [9].

After each condition they completed the NASA-TLX (all 6 subscales, unweighted, on a 5-point scale) [10], the User Engagement Scale Short Form (UES-SF, 12 items) [18], custom Likert-type items, and open-response questions, followed by a direct comparison survey.

4.2 Measures

Temporal. Time to first satisfaction (T_{sat}): seconds to the first-satisfaction marker. *Post-satisfaction refinement*: $T_{\text{ref}} = T_{\text{total}} - T_{\text{sat}}$.

Harmonic breadth. Unique root count $|R|$, *unique quality count* $|Q|$, and Shannon entropy over root-motion intervals:

$$H_{\text{root}} = -\sum_{i} p(d_i)\log_2 p(d_i) \tag{3}$$

where d_i are semitone distances between consecutive chord roots. While $|R|$ counts how many distinct roots appear, H_{root} captures the diversity of *movement patterns* between them: a progression that always moves by the same interval (e.g., ascending fourths) yields high $|R|$ but low entropy, whereas varied root motion produces higher entropy.

Exploration. Candidates favorited $|F|$ and *final-choice commits* $|C|$: deliberate harmonic decisions with no control-condition analogue.

5 Results

We report paired t-tests (two-tailed) with Cohen's d_z, the within-subjects effect size defined as $d_z = \bar{\Delta}/\text{SD}_{\Delta}$ ($d_z > 0.8$ is conventionally "large"). All six paired comparisons survive Benjamini–Hochberg correction at α=.05. Table 1 and Figure 4 summarize findings.

5.1 Quantitative Results

Faster initial satisfaction. Modulo halved the time to first satisfactory accompaniment (d_z=2.54). The effect held across skill levels. For example, an intermediate pianist with intermediate DAW experience dropped from 167 s to 64 s; an advanced pianist with advanced DAW experience from 117 s to 79 s.

Broader harmonic exploration. Unique roots tripled and qualities nearly quadrupled (Table 1), with narrow confidence intervals confirming consistency across all participants. To illustrate:

Control (4 roots, 2 qualities): C | Am | F | G | C
Treatment (5 roots, 7 qualities): Cmaj7 | Gsus4 | Am7 | Dm11 | G7 | Csus2 | Adim | Fmaj7 | C | Dm7 | D7 | C | C

Control accompaniments typically used three diatonic triads (I–IV–V); treatment incorporated seventh chords, suspensions, and diminished substitutions.

Sustained post-satisfaction engagement. Total task time was only modestly longer (d_z=0.49), but post-satisfaction refinement nearly tripled (d_z=1.62). Modulo compresses the harmonic *search* phase while expanding voluntary *refinement.*

5.2 Subjective Ratings

Treatment was rated higher on every survey item (Table 2). Custom Likert-type items assessed perceived exploration breadth, speed to satisfaction, refinement freedom, sense of control, utility for accompaniment, intent to adopt, and output quality. NASA-TLX subscales captured cognitive workload. All participants unanimously preferred treatment for real songwriting.

5.3 Qualitative Themes

We performed reflexive thematic analysis [4] on think-aloud transcripts and open-ended survey responses across all users, coding iteratively to identify three themes.

Expanded harmonic vocabulary. The majority of participants remarked that treatment exposed them to chords they would not have found alone. An intermediate pianist noted: *"I got access to chords I did not know existed."* Even the advanced pianist used temperature strategically: *"I favorited the most interesting and risky chords first, but I also generated blocks with lower temperatures so I could easily switch to simple chords if need be."*

Chord Workshop editing. The Chord Workshop was consistently cited as a central benefit: *"I could change, invert, change quality, add, delete chords without having to use manual piano-roll MIDI events and decipher notes."* Another advanced subject described it as "absolutely insane, so much musical diversity." In control, note-level editing was burdensome: *"Single MIDI events are taxing to find and interpret, let alone edit"*.

Exploration as enjoyment. Participants did not stop at first satisfaction: *"Satisfaction was reached really fast upon generation. It was just a fun creativity game beyond that."* Another participant noted: *"I was satisfied with the first option; it was composition-ready. I just got to take time doing my favorite part of the music: exploring alternatives."*

6 Discussion

Theory-derived abstraction. The 3× increase in roots and 3.7× in qualities (d_z=3.36 and 4.30) reflects the reduced cost of exploration when the interface operates at the level of harmonic structure rather than individual MIDI notes. Król et al. found that musicians valued AI-generated material for "the moments instead of the whole", selectively adopting fragments rather than accepting full outputs [15]. Modulo's selective commitment mechanism directly supports this pattern. Kamath et al. found that professional sound designers regard variety as "a two-edged sword," preferring constrained exploration over unconstrained generation for task-oriented creative work [13]. Similarly, Louie et al. showed that steering tools such as sliders and multiple alternatives are "fundamental to one's notion of being a creator," and that without such controls, perceived ownership erodes [16]. Modulo's temperature columns and Chord Workshop provide analogous steering at the harmonic level.

Two-phase engagement. Modulo compresses the search phase (−50%, d_z=2.54) while expanding refinement (+184%, d_z=1.62). Think-aloud data suggests this reflects productive enjoyment, not difficulty: participants described post-satisfaction editing as "a fun creativity game" and "my favorite part of the music". Smith and Freeman define *perceived creative autonomy* as users' interpretation of how much creative agency they retain when working with an AI system [22]. Modulo's structured controls appear to sustain this perception by letting musicians shape every chord rather than passively accepting generated output.

Table 1: Results across 50 paired observations (N=25 × 2 melodies). p-values are Benjamini–Hochberg adjusted for 6 comparisons. $|F|$ and $|C|$ have no control-condition analogue (structurally zero); means and CIs are reported descriptively.

Metric	Control M	Treatment M	Δ	$t(49)$	p_{adj}	d_z	95% CI$_\Delta$
Time to first satisfaction (s)	162.5	80.6	−81.9	−17.99	<.001	−2.54	[−90.8, −72.9]
Post-satisfaction refinement (s)	61.4	174.1	+112.7	11.45	<.001	1.62	[93.4, 132.0]
Total task time (s)	223.9	254.8	+30.9	3.47	.002	0.49	[13.4, 48.3]
Unique chord roots $\|R\|$	3.0	9.8	+6.7	23.75	<.001	3.36	[6.2, 7.3]
Unique chord qualities $\|Q\|$	2.2	8.4	+6.1	30.39	<.001	4.30	[5.7, 6.5]
Root motion entropy H_{root}	1.2	1.4	+0.2	3.15	.003	0.45	[0.1, 0.3]
Candidates favorited $\|F\|$	0.0	15.7	+15.7	–	–	–	[15.1, 16.3]
Final-choice commits $\|C\|$	0.0	14.4	+14.4	–	–	–	[13.7, 15.0]

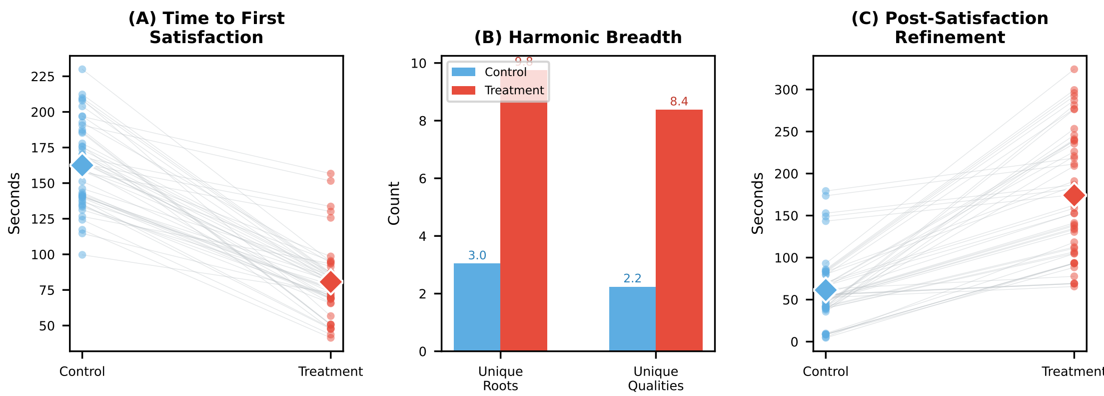


Figure 4: Key results. (A) Time to first satisfaction: treatment halves search time (p<.001). Diamonds mark means. (B) Harmonic breadth: 3× more roots, 3.7× more qualities. (C) Post-satisfaction refinement: 2.8× longer voluntary engagement.

Table 2: Subjective ratings (1–5 Likert), averaged across all 25 participants (N=25). Treatment was rated higher on every item.

Item	Ctrl	Treat
Explore multiple directions	1.00	5.00
Get to satisfactory quickly	1.88	5.00
Refine without settling early	1.08	5.00
Felt in control	2.12	4.80
Useful for accompaniment	2.60	4.64
Would use in real music-making	3.44	5.00
Satisfied with quality	1.96	5.00
NASA-TLX: Mental demand	4.32	1.76
NASA-TLX: Effort	4.48	1.36
NASA-TLX: Frustration	3.88	1.00

Limitations and future work. This study evaluated only the piano workflow with two fixed melodies in contrasting keys. Longitudinal deployment, additional keys and tempos, and multi-instrument workflows—including the audio-to-MIDI transcription pipeline already built into Modulo but outside the scope of this study—would strengthen generalizability. The sample of 25 participants, while sufficient for the large effects observed, limits demographic breadth. The think-aloud protocol may have altered natural behavior by encouraging more deliberate reflection [9]. Subjective ratings were collected immediately after each condition; delayed or ecological assessments could reveal whether preferences persist over sustained use. Finally, this study operated in the symbolic (MIDI) domain. While Modulo has features that extend beyond this domain, they were not tested within this study. Extending the generation and editing paradigm to audio-domain accompaniment for waveform-level co-creation with similar selective commitment is a natural next step.

7 Conclusion

Modulo demonstrates that embedding structured harmonic co-creation directly within a DAW can meaningfully impact how musicians build accompaniment. By enabling parallel comparison, selective commitment, and refinement, the system halves search time, triples harmonic breadth, and sustains engagement well beyond initial satisfaction. These gains are both substantial and consistent across participants. These results point toward a broader implication: the interface mediating interaction between musician and generative model is a primary site of innovation that can meaningfully determine creative outcomes.

References

[1] Edward Aldwell, Carl Schachter, and Allen Cadwallader. 2019. *Harmony and Voice Leading* (5th ed.). Cengage Learning.

[2] Misagh Azimi and Mo H. Zareei. 2025. Live Improvisation with Fine-Tuned Generative AI: A Musical Metacreation Approach. In *Proc. New Interfaces for Musical Expression (NIME)*.

[3] Lancelot Blanchard, Cameron Holt, and Joseph A. Paradiso. 2025. AI Harmonizer: Expanding Vocal Expression with a Generative Neurosymbolic Music AI System. In *Proc. New Interfaces for Musical Expression (NIME)*.

[4] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[5] Luca Casini, Laura Cros Vila, David Dalmazzo, Anna-Kaisa Kaila, and Bob L.T. Sturm. 2025. Data-Driven Analysis of Text-Conditioned AI-Generated Music: A Case Study with Suno and Udio. *arXiv preprint arXiv:2509.11824* (2025). Submitted to TISMIR.

[6] Wan Chong Choi and Chi In Chang. 2025. Suno AI: Opportunities and Challenges of AI-Generated Music and Lyrics in Secondary Music Education. (2025). Preprint.

[7] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and Controllable Music Generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 36.

[8] Kemal Ebcioğlu. 1988. An Expert System for Harmonizing Chorales in the Style of J.S. Bach. *Computer Music Journal* 12, 3 (1988), 43–51. doi:10.2307/3679959

[9] K. Anders Ericsson and Herbert A. Simon. 1993. *Protocol Analysis: Verbal Reports as Data* (revised ed.). MIT Press.

[10] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139–183.

[11] David Huron. 2001. Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception* 19, 1 (2001), 1–64. doi:10.1525/mp.2001.19.1.1

[12] JUCE. 2024. JUCE: Cross-Platform C++ Application Framework for Audio Software. https://juce.com. Version 8.0.12. Originally by Julian Storer; maintained by PACE Anti-Piracy, Inc..

[13] Purnima Kamath, Fabio Morreale, Priambudi Lintang Bagaskara, Yize Wei, and Suranga Nanayakkara. 2024. Sound Designer–Generative AI Interactions: Towards Designing Creative Support Tools for Professional Sound Designers. In *Proc. CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642040

[14] Yewon Kim, Cheng Zhi Anna Huang, Yun-Jia Jin, and Carrie Jun Cai. 2025. Amuse: Human-AI Collaborative Songwriting with Multimodal Inspirations. In *Proc. CHI Conference on Human Factors in Computing Systems*.

[15] Szymon J. Król, Ondřej Černý, and Nick Bryan-Kinns. 2025. Exploring the Needs of Practising Musicians in Co-Creative AI Through Co-Design. In *Proc. CHI Conference on Human Factors in Computing Systems*.

[16] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie Jun Cai. 2020. Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models. In *Proc. CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3313831.3376739

[17] Sandy Ma and Charles Patrick Martin. 2025. Touching Wires: Tactility and a Quilted Musical Interface for Human-AI Musical Co-Creation. In *Proc. New Interfaces for Musical Expression (NIME)*.

[18] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A Practical Approach to Measuring User Engagement with the Refined User Engagement Scale (UES) and New UES Short Form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[19] Yunyu Ong, Robert Sazdov, and Andrew Johnston. 2024. Opening DAWs to Interactive Music – Making an Orchestra out of Soloists. In *Proc. New Interfaces for Musical Expression (NIME)*.

[20] Raghavasimhan Sankaranarayanan, Nitin Hugar, Qinying Lei, Thomas Ottolin, Hardik Goel, and Gil Weinberg. 2023. Mixboard – A Co-Creative Mashup Application for Novices. In *Proc. New Interfaces for Musical Expression (NIME)*.

[21] Alexander Scarlatos, Yusong Wu, Ian Simon, Adam Roberts, Tim Cooijmans, Natasha Jaques, Cassie Tarakajian, and Anna Huang. 2025. ReaLJam: Real-Time Human-AI Music Jamming with Reinforcement Learning-Tuned Transformers. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA)*.

[22] Jason Brent Smith and Jason Freeman. 2025. Adaptation and Perceived Creative Autonomy in Gesture-Controlled Interactive Music. In *Proc. New Interfaces for Musical Expression (NIME)*.

[23] Minhyang (Mia) Suh, Emily Youngblom, Michael Terry, and Carrie J. Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proc. CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445219

[24] Tracktion Software Corporation. 2025. Tracktion Engine: Open-Source DAW Engine. https://github.com/Tracktion/tracktion_engine. Version 3.2.0. Dual GPL v3/Commercial license. Open-sourced November 2018.

[25] Soumya Sai Vanka, Jean-Baptiste Rolland, and György Fazekas. 2025. Demonstrating Diff-MSTC: A Controllable and Context-Aware AI System for Multitrack Mixing in DAW Cubase. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA)*. doi:10.1145/3706599.3721167

[26] Alexander Wang, David Lindlbauer, and Chris Donahue. 2024. Towards Music-Aware Virtual Assistants. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*. doi:10.1145/3654777.3676416

[27] Yusong Wu, Tim Cooijmans, Kyle Kastner, Adam Roberts, Ian Simon, Alexander Scarlatos, Chris Donahue, Cassie Tarakajian, Shayegan Omidshafiei, Aaron Courville, Pablo Samuel Castro, Natasha Jaques, and Cheng-Zhi Anna Huang. 2024. Adaptive Accompaniment with ReaLchords. In *Proc. International Conference on Machine Learning (ICML)*.

A Selected Participant Case Studies

This appendix presents detailed data for five think-aloud participants (identified for anonymity as P001–P005), selected to span beginner-to-advanced skill levels and illustrate individual variation within the aggregate trends reported in Section 5.

A.1 Per-Task Quantitative Results

Table 3 presents the comparative unique roots and qualities, as well as time deltas to first satisfaction and refinement.

A.2 Submitted Chord Sequences

Tables 4 and 5 present the final submitted chord sequences for each participant in both conditions. Control sequences are notably shorter and harmonically simpler; treatment sequences incorporate extended qualities (maj7, sus4, dim, m7) and chromatic motion.

A.3 Participant Quotes

Table 6 collects representative quotes from P001–P005, organized by study phase. "TA" denotes concurrent think-aloud verbalization during the task; "Survey" denotes post-condition written responses.

Table 3: Task-level metrics for P001–P005 across both melodies and conditions. T_{sat} = time to first satisfaction (s); T_{ref} = post-satisfaction refinement (s); $|R|$ = unique roots; $|Q|$ = unique qualities.

		Melody A (C major)				Melody B (F minor)			
ID	**Cond.**	T_{sat}	T_{ref}	$\|R\|$	$\|Q\|$	T_{sat}	T_{ref}	$\|R\|$	$\|Q\|$
P001	Control	164	179	4	2	197	83	3	3
	Treatment	99	218	4	8	157	191	4	4
P002	Control	212	86	4	3	178	53	3	2
	Treatment	87	299	7	8	87	141	6	4
P003	Control	143	36	3	3	117	39	3	3
	Treatment	69	240	7	7	41	209	6	8
P004	Control	159	63	2	2	209	10	3	3
	Treatment	84	78	7	6	57	105	7	5
P005	Control	100	40	3	2	115	5	7	2
	Treatment	73	105	7	8	72	93	7	7

Table 4: Submitted chord sequences for Melody A (C major, 80 BPM). $|R|/|Q|$ = unique roots / unique qualities.

ID	Control	Treatment
P001	C | Am | F | G | C ($\|R\|$=4, $\|Q\|$=2)	Cmaj7 | Gsus4 | Am7 | Dm11 | G7 | Csus2 | Adim | Cmaj7 | C | Dm7 | D7 | C | C ($\|R\|$=4, $\|Q\|$=8)
P002	C | F | Cmaj7 | G7 | F | G | C | F | G | F | E ($\|R\|$=4, $\|Q\|$=3)	Fmaj7 | Csus4 | A♯maj7 | Am | Edim | Dm7 | C | Am | Am7 | Dsus2 | Dm | Gdim | G | Am | D7 | G7 | Cdim ($\|R\|$=7, $\|Q\|$=8)
P003	G | Cmaj7 | Fsus4 | F ($\|R\|$=3, $\|Q\|$=3)	Cmaj7 | Am | G7 | B♭ | Am | B♭ | C | Dm | Em7 | Ddim | Fdim | Am7 | Dsus4 | G | A♯dim | Edim | Dm7 | C ($\|R\|$=7, $\|Q\|$=7)
P004	C | C | F | G | Fmaj7 | C | Fmaj7 | C ($\|R\|$=2, $\|Q\|$=2)	Am | Gm | Cm7 | F | F | C | C | Bm7♭5 | Am | E | Dm7 | E7 | Cmaj7 | Dm7 | F ($\|R\|$=7, $\|Q\|$=6)
P005	Am7 | E | A ($\|R\|$=3, $\|Q\|$=2)	F | C | G | F | C7 | F | C | F | Dm7 | Fmaj7 | Fsus4 | A♯dim | Am7 | Em7 | Csus2 | Em | G7 | Fmaj7 | C ($\|R\|$=7, $\|Q\|$=8)

Table 5: Submitted chord sequences for Melody B (F minor, 70 BPM). $|R|/|Q|$ = unique roots / unique qualities.

ID	Control	Treatment
P001	Fm | C | B♭m | C7 | Fm | C | Fm ($\|R\|$=3, $\|Q\|$=3)	D♭ | Cm | C7 | Fm | C7 | Fm | D♭ | D♭ | Fm | F7 | Fm7 | E♭ | Fm ($\|R\|$=4, $\|Q\|$=4)
P002	F | B♭m | C | Fm ($\|R\|$=3, $\|Q\|$=2)	Fm | C7 | D♭ | D♯7 | Cm | Cm | D♭ | B♭m | B♭m7 | E♭7 | A♭7 | F7 ($\|R\|$=6, $\|Q\|$=4)
P003	C | Fm | Ddim | Fm ($\|R\|$=3, $\|Q\|$=3)	Fm | Fsus2 | Gsus7/C | C7 | C | Fdim | A♯m7 | Fm7 | Asus4 | B♭m | B♭m7 | E♭7 | Cm7 ($\|R\|$=6, $\|Q\|$=8)
P004	F | Fm | C7 | D | C | Fm ($\|R\|$=3, $\|Q\|$=3)	Fm | C♯m7 | Gm7 | Fm | Fm7 | E♭dim | Fm | B♭m | C7 | D♭ | A♭ ($\|R\|$=7, $\|Q\|$=5)
P005	F | Fm | A♭ | Fm | C | C | E♭ | C | E♭ | C | A♭ | D♭ | A♭ | D♭ | A♭ | A♭ | B♭ | A♭ | C | G | F | C | Fm | F ($\|R\|$=7, $\|Q\|$=2)	Fm | C7 | G♯maj7 | B♭m7 | C | A♯m | D♯dim | Fm | B♭m | A♭7 | Bo7 | Cm | D♭ | Fm | C | Cmaj7 ($\|R\|$=7, $\|Q\|$=7)

Table 6: Selected participant quotes by study phase. TA = think-aloud during task; Survey = post-condition written response.

ID	Source	Quote
Treatment Condition		
P001	Survey	"The Chord Workshop. I could change, invert, change quality, add, delete chords without having to use manual piano-roll MIDI events and decipher notes."
P001	Survey	"Nothing was really frustrating. If anything, I am frustrated I haven't had a tool like this before."
P001	Survey	"The treatment gave me cool rhythms, new chords I've never heard of, and the generated chords gave me so many options. I explored music theory in a way I have not before."
P002	Survey	"I got access to chords I did not know existed. I also liked the arpeggios, because I'm not good at playing those myself."
P002	Survey	"[Most frustrating:] Deciding which chords to go with in the end, there were so many great options!"
P002	Survey	"You get a full melody+harmony chords composition within seconds! Then it's just refining work from there. More time on exploring the music, less time grinding!"
P003	TA	"I was satisfied with the first option I listened to both times; it was composition-ready for sure. I just then got to take time doing my favorite part of the music: exploring alternatives."
P003	Survey	"The menu for adding in different temperatures and different rhythms is absolutely insane. So much musical diversity there, and it explores wildly cool chords. There's no monotony in harmony nor rhythm."
P003	Survey	"I like exploring new sounds and chords. And as a professional musician, it is those unorthodox chords that make a song a hit."
P004	Survey	"All the chords generated actually worked. I also got to hear new sounds and chords I did not know about."
P004	TA	"It is so much faster and easier, and I can explore so many more options. I also learn so much from it, and can apply these new chords to my future work."
P004	TA	"I loved being able to toggle through scale degrees. I learned that way, so it was super helpful."
P005	TA	"Satisfaction was reached really fast upon generation. It was just a fun creativity game beyond that."
P005	Survey	"I favorited the most interesting and risky chords first, but I also generated blocks with lower temperatures so I could easily switch to simple chords if need be."
P005	Survey	"I actually equally loved Chord Generation and Harmony Generation. Both help me expand upon the given melody so much. The Chord Workshop is incredible. It's tailored to creating high-level music with interesting chords, harmonies, and rhythms."
Control Condition		
P001	TA	"Single MIDI events are taxing to find and interpret, let alone edit."
P001	Survey	"Not having the AI assistance was really frustrating. I had been exposed to so many cool chords in the treatment version, but now I couldn't remember."
P002	TA	"I was frustrated to be unable to really know what chords to play. I couldn't get any help, so I just stuck with 1, 4, and 5 chords."
P003	Survey	"I could not think of chords outside of the normal 1, 4, 5, 6 minor, and occasional diminished combos I happen to know. And once I solidified my first recording, I didn't really feel like changing it."
P004	Survey	"Not being able to know which chords I had played, and trying to get ones that were correct for the key and melody. I got very little help from what was available, as a novice."
P005	Survey	"Nothing frustrated me, since I am a professional pianist. This was not too hard. But I had more fun with the other version."