Scaffolding Symbolic Music Generation: An Evaluation of Two Generative MIDI Tools for Novices

Ge Liu*
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada

Griffin Page
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada

Keon Ju Maverick Lee
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada

Paul Triana
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada

Tao-Tao He
Vanderbilt University
Nashville, TN, USA

Martin E. Malandro
Department of Mathematics and Statistics
Sam Houston State University
Huntsville, TX, USA

Philippe Pasquier
School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, Canada

Abstract

The rapid advancement of generative artificial intelligence has introduced new systems for symbolic music generation (SMG). While expert-focused evaluations have reported positive usability and acceptance for some SMG systems, their effectiveness in supporting original composition for novices remains underexplored. This mixed-methods study evaluates Composer's Assistant 2 (CA2), integrated into the REAPER digital audio workstation (DAW), and the web-based Calliope system with 75 undergraduate sound-design students who had varied musical backgrounds but generally limited formal training in music theory and composition. Participants used the systems to create original music for a moving-image assignment. Quantitative results indicate low usability and creativity support across both tools, with System Usability Scale scores (CA2 $M = 42.89$; Calliope $M = 43.82$) and Creativity Support Index scores (CA2 $M = 49.41$; Calliope $M = 41.40$). These findings contrast with prior reports of high usability in expert-focused evaluations of similar systems (e.g., MMM-Cubase) and the relative success of retrieval-based sound design tools such as Audio Metaphor (AuMe). Qualitative analysis further reveals a recurring difficulty in translating creative intent into symbolic controls, often leading users to adopt a repetitive "generate-and-filter" workflow. We suggest that SMG systems for novices require stronger scaffolding than those designed for expert users. In particular, novice-facing interfaces may benefit from higher-level semantic steering, targeted post-generation manipulation, faster comparison across generated alternatives, and tighter integration with existing production workflows.

1 INTRODUCTION

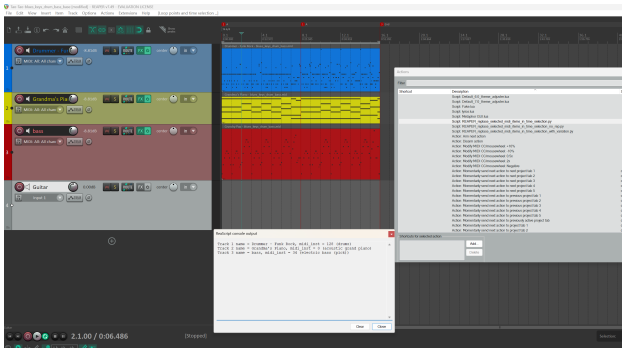
The rapid advancement of artificial intelligence has significantly transformed creative practices across various domains, prompting considerable scholarly interest in the ways these tools support and reshape artistic workflows (Zhou and Lee, 2024). While text and image generation have been widely adopted, the integration of AI into sound design presents distinct challenges, particularly concerning temporal complexity and the abstraction of audio parameters (Thorogood, 2021). For novice practitioners, the learning curve in sound design is notoriously steep; learners must concurrently master technical proficiency, such as signal processing and mixing, while cultivating their fundamental creative voice. Although musical composition encompasses a wide range of practices, it generally requires some degree of understanding of music theory. In symbolic music generation workflows, users must additionally understand MIDI-based representations and editing conventions in order to provide meaningful input, interpret generated material, and refine the results.

Recent evaluations of generative systems, such as MMM-Cubase and Audio Metaphor (AuMe), have demonstrated significant potential in augmenting creative workflows (Tchemeube et al., 2023; Liu et al., 2025a). Specifically, AI-assisted sound design through AuMe has been shown to enhance novice workflow efficiency by 21.83% (Thorogood and Pasquier, 2013; Liu et al., 2025a). However, a distinct research gap persists at the intersection of user expertise and generative modality. While MMM-Cubase reported high efficacy, its study cohort was primarily com-

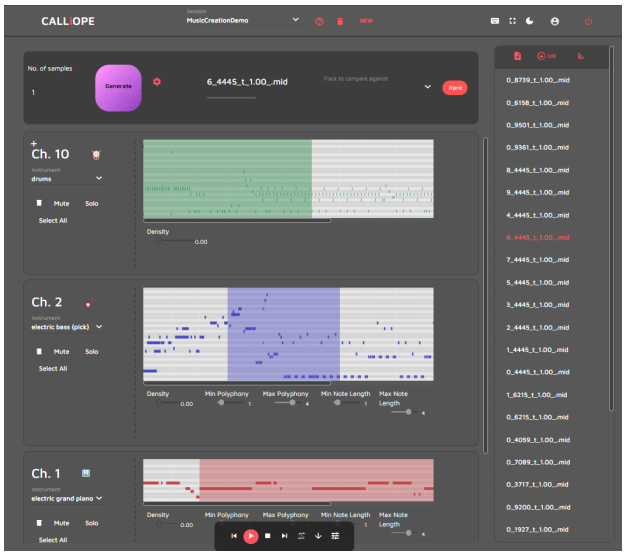
*Correspondence: ge_liu@sfu.ca

posed of hobbyist experts and professional composers (Tchemeube et al., 2023). Conversely, while AuMe successfully scaffolded soundscape composition for novices, its application was limited to sample-based audio retrieval and arrangement (Liu et al., 2025a). This distinction is important because sound-effect design and symbolic music composition involve different materials, representations, and production practices (Chion, 2019; Collins, 2008). In particular, SMG workflows require users to engage not only with music-theoretical concepts, but also with MIDI-based symbolic structures and editing processes. Specifically, this study focuses on undergraduate students enrolled in IAT 340: Sound Design at Simon Fraser University, who were asked to compose original music using SMG tools as part of a course assignment. This population had varied musical backgrounds and experience with sound design and audiovisual intention, but generally had limited formal training in music theory and composition.

Extending Liu et al.'s work on AI-assisted sound design for novices (Liu et al., 2025a), this study examines whether similar scaffolding benefits can be observed in symbolic music generation through a comparative analysis of Composer's Assistant 2 (CA2) (Malandro, 2024) and Calliope (Tchemeube et al., 2022). Although both systems operate as context-aware generators that infer musical content from existing MIDI tracks within a context window, they represent different approaches to workflow integration in Human-AI Interaction. CA2 operates as a DAW-integrated plug-in in REAPER, prioritizing workflow fluency by allowing users to generate content that fits seamlessly within the project's temporal structure. In contrast, Calliope functions as a web-based interface, offering the advantage of immediate use without the need for software installation. By evaluating these different workflow approaches, we examine how their respective affordances and sources of friction were experienced by novices in symbolic music composition.[1]



(a) Composer's Assistant 2: DAW-integrated interface within REAPER.



(b) Calliope: Web-based standalone interface.

Figure 1: The two AI-assisted MIDI generation tools evaluated in this study: (a) CA2, representing a DAW-integrated workflow, and (b) Calliope, representing a standalone web-based workflow.

2 RESEARCH QUESTIONS

Prior studies have shown the potential of AI-assisted tools to support novice sound-design workflows (Thorogood and Pasquier, 2013; Liu et al., 2025a). However, symbolic music generation (SMG) introduces a distinct set of cognitive and technical demands because users must work with both music-theoretical concepts and MIDI-based symbolic representations. This study therefore examines how SMG systems support novice sound-design students in creating original music, as well as the factors that shape usability, perceived creative support, future adoption, and authorship attribution.

Primary RQ: How do SMG systems support or constrain novice sound-design students in composing original music for audiovisual work, particularly when translating creative intent through music-theoretical concepts and MIDI-based representations?

To guide this investigation, we propose three supplementary research questions:

RQ1 Usability & Perceived Creative Support: To what extent are SMG systems perceived as usable and supportive of novice creative work?

RQ2 Acceptance & Adoption: What factors influence the acceptance and future adoption of symbolic music generation tools among novices?

RQ3 Perceived Authorship: How do novices attribute authorship of the musical outputs created with SMG systems?

3 RELATED WORK

Recent work on AI-assisted music composition has increasingly shifted from model capability toward the question of how generative systems can be meaningfully integrated into human creative workflows. Recent surveys and interaction-focused studies argue that generative music systems should be evaluated not only by musical output quality, but also by user experience, interaction design, acceptance, and their role within creative practice (Lerch et al., 2025; Hu et al., 2025). This shift is particularly important for symbolic music generation (SMG), where users must interact with discrete musical structures such as notes, bars, tracks, and MIDI parameters rather than working with rendered audio or clip-based material.

[1] A de-identified sample collection of participants' musical outputs is available through the Microsoft OneDrive supplementary folder managed by Simon Fraser University.

3.1 Symbolic Music Generation and Computer-Assisted Composition

SMG systems operate on symbolic musical representations such as MIDI, enabling generation, continuation, and infilling at the level of notes, bars, and tracks. Prior work has explored the analogy between music and language, motivating the adaptation of sequence-based and Transformer-based methods to symbolic music generation (Le et al., 2025). In interactive composition settings, however, the challenge is not only whether these models can generate plausible music, but whether users can steer them toward intended creative outcomes.

Early work on novice-AI music co-creation introduced the concept of "AI-steering" through systems such as CoCoCo, where users guide generation through melodic sketches (Louie et al., 2020). However, this work was primarily focused on monophonic melodic material. More recent systems and models, including the Multitrack Music Machine (MMM), Calliope, Composer's Assistant 2, MIDI-GPT, and MIDI-RWKV, expand the design space toward multitrack symbolic infilling and controllable generation (Ens and Pasquier, 2020; Tchemeube et al., 2022; Malandro, 2024; Pasquier et al., 2025; Zhou-Zheng and Pasquier, 2025). Many of these systems can condition generation on musical context or attributes such as note density, polyphony, rhythm, instrumentation, and selected tracks or measures. These controls enable detailed steering, but they may also increase the interaction burden for novices by requiring users to interpret music-structural concepts, translate high-level creative intentions into symbolic specifications, and evaluate or revise the resulting MIDI.

The MMM-Cubase evaluation demonstrated that DAW-integrated multitrack generation can be positively received by experienced composers, reporting strong usability and acceptance while still identifying limitations around controllability and predictability (Tchemeube et al., 2023). Because that evaluation involved hobbyist experts and professional composers, it provides an important expert-oriented point of comparison for the present study.

A more recent novice-focused study evaluated MMM-Cubase v2 and Calliope with a mixed-source sample of 98 novice participants recruited from art, music, and research communities. The study reported generally positive usability, user experience, and future-use intentions, while still identifying concerns around depth of control, model transparency, musical diversity, and skill learning (Tchemeube et al., 2026). The present study complements that work by examining CA2 and Calliope in an undergraduate sound-design course, where participants composed music for self-selected audiovisual projects and where tool allocation was partly shaped by installation and workflow constraints. Differences between the studies should therefore be interpreted in relation to system version, task structure, participant context, and study design rather than as direct performance differences.

3.2 Controllability, Semantic Steering, and Workflow Integration

A recurring challenge in generative music systems is the gap between high-level creative intention and low-level system controls. Zang and Zhang describe this as an "Interpretation Gap" in text-to-music generation, whereby users' semantic intentions may not translate reliably into musical outputs (Zang and Zhang, 2024). A similar gap appears in symbolic systems when users must express their intentions through parameters such as note density, polyphony, or selected tracks and measures. Although these controls are musically meaningful, they may be unintuitive for novices who think in terms of mood, genre, scene, emotion, or narrative function.

Recent user studies of text-to-music systems reinforce this concern. The PAGURI study found that generated audio did not always meet users' expectations, while participants requested greater controllability and editability, including the ability to modify specific sections of a generated track rather than regenerate the entire output (Ronchini et al., 2025). These findings are relevant to SMG even though the modality differs: both text-to-music and MIDI-infilling tools require users to translate creative intentions into system-readable controls and then refine the resulting material through iterative selection and editing. Related affective-input systems further suggest that novice-facing generative music interfaces may benefit from controls grounded in emotion or other accessible descriptors. For example, EmoGen uses emoji-based inputs to support emotional expression through generative music, illustrating one alternative to purely parameter-based steering (Guo et al., 2026).

Prior research on human–computer co-creativity has conceptualized creative interaction as an iterative process in which human and computational agents alternate or divide creative tasks (Kantosalo and Toivonen, 2016). In generative music, the Interactive Generative Music Environment (IGME) similarly supported exploration through a familiar score-editing interface and the iterative development of generated musical material (Hunt et al., 2020). These perspectives suggest that controllability may depend not only on exposing additional parameters, but also on enabling users to explore alternatives, preserve useful material, and translate creative judgements into actionable revisions.

Workflow integration is also central to user experience. DAW-integrated systems such as MMM-Cubase and CA2 reduce transfer friction by embedding generation within an existing production environment, whereas browser-based tools such as Calliope lower installation barriers but require users to move MIDI files between systems. Prior co-design work with practising musicians similarly emphasizes that musical AI tools should reflect how users develop, evaluate, and integrate ideas within their existing practices (Krol et al., 2025b). For novices, technical setup, file management, and unfamiliar MIDI-editing procedures may introduce additional demands that interrupt creative exploration.

3.3 Acceptance, Authorship, and the Novice Gap

Beyond usability, adoption of AI-based creativity support tools depends on more than whether a system can produce usable material. In the broader context of AI-based creativity support tools, adoption has been linked to perceived usefulness, effort expectancy, interactivity, personalization, and broader ethical or social concerns (Liu et al., 2025b). More music-specific adoption studies similarly suggest that perceived usefulness, perceived ease of use, trust, content quality, perceived enjoyment, perceived AI risk, and identity-related concerns can shape users' willingness to adopt generative AI for music composition (Zhou et al., 2025; Liu and Mao, 2026). These factors are particularly relevant for novice users, whose willingness to adopt a

system may be shaped not only by its outputs, but also by whether the interaction feels understandable, controllable, and creatively meaningful.

Creative ownership is another unresolved issue in AI-assisted music composition. Recent work on deep learning-based music variation suggests that musicians' sense of ownership can be strengthened when AI systems depend on the user's own musical input and support the transformation of human-originated ideas rather than replacing them (Krol et al., 2025a). First-person accounts of working with AI-generated musical material also show how generated outputs can be appropriated and reinterpreted within creative practice, rather than simply accepted as finished compositions (Bryan-Kinns et al., 2024). This has direct implications for novice-facing SMG tools. When users have limited opportunities to contribute their own musical material, steer generation, or meaningfully refine the output, they may attribute a greater share of the creative contribution to the system.

Taken together, prior work suggests a gap between technically capable SMG systems and the scaffolding required for novice creative use. Existing systems offer increasingly sophisticated controls, but these controls often presuppose musical knowledge, DAW fluency, or confidence in iterative editing. The present study addresses this gap by evaluating two contrasting SMG interfaces, CA2 and Calliope, with novice sound designers composing original music in a pedagogical setting.

4 METHODOLOGY

Following Creswell's framework for mixed-methods research design, we combined standardized quantitative measures with qualitative analysis of open-ended user responses (Creswell and Creswell, 2018). This design was appropriate because the study examines not only measurable usability and creativity support outcomes, but also participants' subjective experiences of control, authorship, and workflow friction. The study was conducted in a naturalistic pedagogical setting within an undergraduate sound design course (IAT 340) at Simon Fraser University. Embedding the evaluation in an assessed course workflow allowed us to examine participants' experiences under authentic instructional and technical constraints.

4.1 Participants

The participant group consisted of undergraduate students enrolled in IAT 340: Sound Design at Simon Fraser University during the Fall 2025 semester. Out of approximately 150 students who completed the relevant course assignments, 75 provided informed consent to participate in the post-task research survey ($N = 75$).

In this study, the term "novice" refers primarily to participants' limited formal training and prior experience in music composition and symbolic music generation, rather than to an absence of all musical experience or technical experience with audio-production software. A separate pre-course survey indicated that students had varied musical backgrounds: some had received instrumental instruction or learned music independently, while others reported no musical training. However, no response indicated formal composition lessons.

The pre-course survey also suggested that most students had beginner- or novice-level experience with digital audio workstations (DAWs), although one respondent identified as an expert DAW user while reporting no formal musical training. Because informal composition experience was not measured separately, this case should not be interpreted as evidence of no prior composition experience. Because the pre-course survey was administered separately from the research survey and included some duplicate submissions, these background data are reported descriptively rather than as precise participant statistics. Taken together, the available background information indicates that the cohort had limited formal preparation for music composition and MIDI-based generative workflows, despite variation in instrumental and audio-production experience.

4.2 Study Procedure & Tool Allocation

The study was embedded in the Assignment 2 sequence of IAT 340: Sound Design. In Assignment 2-a, participants selected a moving-image clip of approximately 50–70 seconds and created a sound-design project for it. This stage focused primarily on sound effects and audiovisual sound design and did not require original music composition, although limited use of sourced music of up to 15 seconds was permitted.

In Assignment 2-b, participants were required to compose original music for the same video clip and existing sound-design project using a designated symbolic music generation (SMG) system. To support this task, we provided a library of MIDI templates containing a range of musical styles, chord progressions, and structural frameworks. Participants could use these templates, or where appropriate their own musical material, as input for generating and refining music aligned with their creative intentions.

Composer's Assistant 2 (Malandro, 2024) was initially intended as the primary tool because of its integration with the REAPER digital audio workstation (DAW). However, a substantial number of students encountered technical barriers, particularly Python environment configuration and VST pathing issues. Calliope (Tchemeube et al., 2022), a web-based alternative, was therefore introduced for students who were unable to install or reliably use CA2.

The final distribution consisted of 51 participants who used CA2 ($n = 51$), 19 who used Calliope ($n = 19$), and 5 who used both systems ($n = 5$). Participants who used both systems were included in the qualitative analysis but excluded from the between-group quantitative comparisons because they could not be assigned to a single independent tool group.

Tool allocation was not randomized and was partly shaped by technical accessibility. The comparison should therefore be interpreted as a naturalistic, non-randomized evaluation of the two systems under authentic pedagogical and technical constraints, rather than as a controlled causal comparison of their interface designs.

4.3 Data Collection and Measures

Given that the study was located within a sound design curriculum rather than a formal music composition program, the research design did not include an independent expert assessment of musical output quality or a matched non-AI comparison condition. Course assessment was conducted separately from the research analysis and was not used as a measure of musical quality. Consequently, the study focused on participants' perceived usability, creativity support, effectiveness, future adoption, and authorship attri-

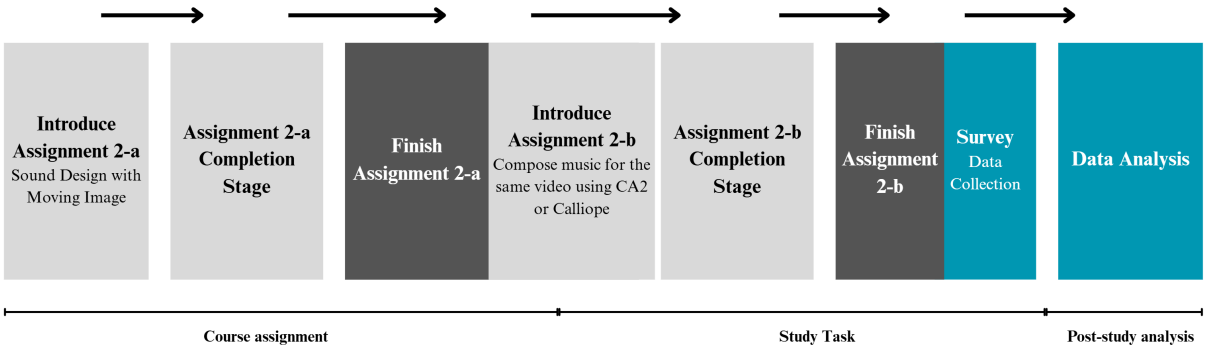


Figure 2: Overview of the study procedure: Assignment 2-a (sound design), Assignment 2-b (music composition with CA2 or Calliope), followed by the post-task survey and analysis.

bution rather than on whether SMG assistance objectively improved musical quality relative to manual composition. Representative de-identified examples of participants' musical outputs are provided in the supplementary Microsoft OneDrive folder for illustration rather than formal evaluation.

We used the System Usability Scale (SUS) to assess the participants' perceptions of the usability and learnability of each interface (Brooke, 1996). The Creativity Support Index (CSI) was used to assess perceived support for dimensions of the creative process, including Exploration and Expressiveness (Cherry and Latulipe, 2014).

In addition to these quantitative measures, the post-task survey included open-ended questions about perceived effectiveness, future adoption, authorship attribution, desired areas of control, satisfaction, and broader concerns about generative music systems. Open-ended responses were analyzed using an inductive, non-mutually exclusive thematic coding approach. One researcher read all responses, developed codes from recurring themes in the data, assigned the relevant codes to each response, and counted the number of responses associated with each theme. A response could receive multiple codes when it addressed more than one theme. The reported frequencies therefore indicate the number of responses coded under each theme and are not mutually exclusive.

4.4 Limitations

The study has several limitations arising from its naturalistic, course-based design. First, tool allocation was not randomized, and the groups were unequal in size ($n = 51$ for CA2 and $n = 19$ for Calliope). Because Calliope was introduced primarily for students who encountered difficulties installing or configuring CA2, the groups may have differed in their technical circumstances, prior experience, or other unmeasured characteristics. The comparison therefore cannot establish that differences in participant responses were caused solely by the systems' interfaces or deployment methods.

Second, participants worked with self-selected video clips that varied in pacing, genre, and emotional structure. Because the study included neither an independent expert assessment of musical quality nor a matched non-AI condition, it cannot establish whether either system improved musical or audiovisual outcomes relative to manual composition.

Third, although completing the assignment provided participants with practical exposure to MIDI-based composition and generative workflows, the study did not include pre- and post-task measures of musical knowledge, MIDI understanding, or compositional skill. It therefore cannot determine what participants learned through the process, whether any learning transferred to subsequent tasks, or whether the systems primarily supported completion of the immediate assignment. Accordingly, the term "scaffolding" in this study refers to task-level support for initiating, structuring, and refining the composition process, rather than to demonstrated long-term learning or skill development.

Fourth, the qualitative coding was conducted by one researcher. Although the coding procedure and non-mutually-exclusive frequency counting are reported transparently, the code assignments were not independently checked by a second coder.

Finally, participants were drawn from a single undergraduate sound-design course and completed the task with course-provided MIDI templates, instructions, and technical support. This pedagogical context may differ from the conditions experienced by independent learners or users in professional production settings, limiting the generalizability of the findings.

5 RESULTS

5.1 SUS (RQ1): Low and Variable Usability Scores

In alignment with established human-computer interaction (HCI) methodologies, we evaluated the usability of both generative MIDI systems using the System Usability Scale (SUS). Due to the observed unequal variances between the groups, we applied Welch's independent-samples t-tests, which do not assume equal variances.

Both Composer's Assistant 2 and Calliope received low mean SUS scores of approximately 43/100 ($M_{\text{CA2}} = 42.89$, $SD = 17.40$; $M_{\text{Calliope}} = 43.82$, $SD = 18.90$). According to Bangor et al. (2008), SUS scores below 50 fall within the *Not Acceptable* range, indicating substantial usability concerns. The AuMe evaluation reported a higher mean SUS score of 64.33 ($SD = 13.93$). Welch's independent-samples t-tests indicated that the observed SUS scores were lower in the present samples than in the AuMe sample (AuMe–CA2: $t(92.61) = 7.28$, $p < .001$; AuMe–Calliope: $t(23.49) = 4.42$, $p < .001$), while CA2 and Calliope did not differ significantly from each other ($t(30.08) = -.19$, $p = .85$). These tests quantify dif-

ferences between the observed sample means. However, because the studies involved different participants, creative tasks, systems, and procedures, they do not isolate causal system effects or establish that AuMe is inherently more usable. Nevertheless, both SMG systems fell within the *Not Acceptable* SUS range and presented substantial usability challenges for the novice participants in the present study.

Table 1: Thematic Analysis of User Pain Points

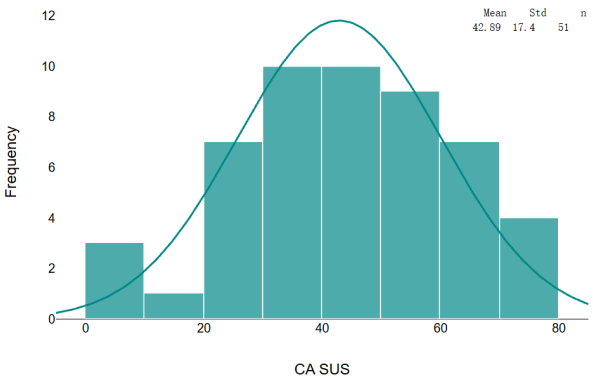
Target System	**Identified Pain Point**	**Coded Responses (n)**
CA2 (DAW)	Mood/Genre Control	12
	High Randomness	9
	Repeating Notes	5
	Matching Instruments	2
Calliope (Web)	Workflow Transfer Friction	4
	UI/UX Navigation	4
Both Tools	Uncontrollable Outputs	1
	Regeneration Issues	1

Note. Frequencies represent the number of valid open-ended responses coded under each theme. Codes were not mutually exclusive, and a single response could contribute to multiple themes. Some participants did not answer the relevant question or provided responses that were too limited, unclear, or unrelated to be meaningfully coded; therefore, the number of analyzable responses may be lower than the total sample size. Access and stability barriers are reported separately in the text rather than included in this table: CA2 installation/setup issues ($n = 19$) and Calliope web technical failures ($n = 5$).

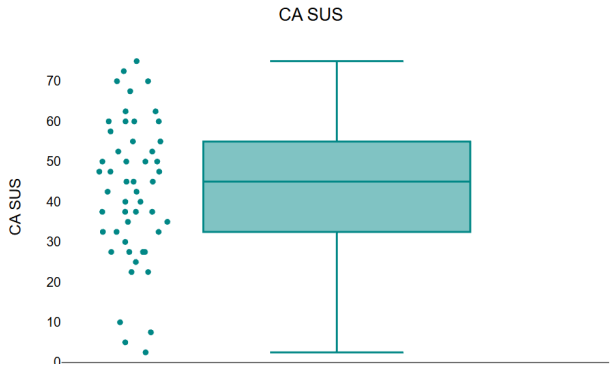
Deconstructing the Sources of Friction: The low usability scores can be understood in relation to several distinct sources of friction. As indicated by our thematic analysis of user qualitative feedback (summarized in Table 1), the reported patterns of friction differed by tool deployment method:

For Composer's Assistant 2 (DAW-Integrated): Access friction was a notable contextual barrier before creative interaction could begin. In total, 19 participants referred to setup-related difficulties, including installation ($n = 16$) and broader configuration problems ($n = 3$), such as Python environment configuration and VST pathing. Because these comments primarily concern deployment and access rather than the creative interaction itself, we report them separately from the interaction-focused pain points in Table 1. Once users reached the generative workflow, the more central issues concerned steerability and musical control: participants cited a lack of Mood/Genre Control ($n = 12$), High Randomness ($n = 9$), and Repeating Notes ($n = 5$). Qualitative feedback also revealed a "steep first, smooth after" pattern ($n = 4$). As one participant noted: "It was tedious to install but after that it was easy and straightforward to use." Seven users also explicitly described the tool as "Efficient and Easy to Use" once operational.

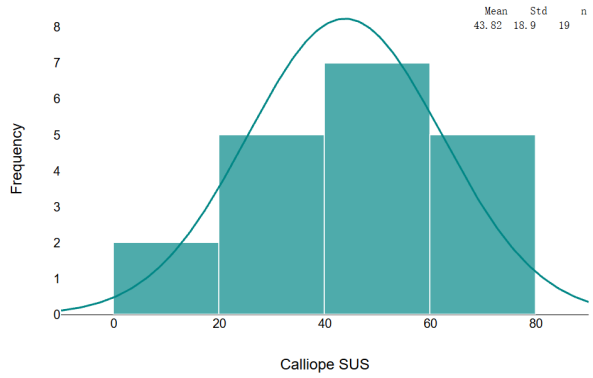
For Calliope (Web-Based): Despite circumventing local installation issues, Calliope did not receive significantly higher SUS scores than CA2 ($p = .85$). Instead, participants reported friction related to interface design, web stability, and workflow transfer. Participants also encountered access and stability friction with Calliope, including Web Technical Failures ($n = 5$). However, the main interaction-level issues were Workflow Transfer Friction ($n = 4$) and



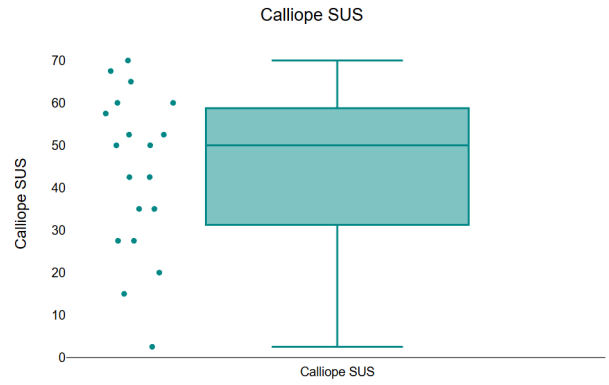
(a) CA2: SUS Score Distribution.



(b) CA2: SUS Score Boxplot.



(c) Calliope: SUS Score Distribution.



(d) Calliope: SUS Score Boxplot.

Figure 3: System Usability Scale (SUS) evaluation for Composer's Assistant 2 and Calliope, illustrating the high variance and generally low usability scores among novices across both generative interfaces.

UI/UX Navigation ($n = 4$). Here, transfer friction refers to the extra steps required to move MIDI materials between the DAW and the web-based application: exporting MIDI from the DAW, using it as input in Calliope, downloading the browser-generated MIDI, and importing it back into the DAW for playback and further editing.

For Users Evaluating Both Tools: Among the subset of participants who used both systems, two provided analyzable qualitative comments on this issue. Both comments concerned limited steerability, specifically uncontrollable outputs and difficulty discovering or using partial-regeneration workflows. Although both tools technically support selected-part regeneration, these participants did not always recognize or understand this functionality, suggesting a need for clearer onboarding and in-interface guidance. One user articulated this conceptual friction regarding generative SMG tools: "The features worked well for creating a base idea, but they felt limited when I tried to match the fast, exaggerated style of Looney Tunes."

Addressing RQ1: These findings suggest that neither system consistently offered a sufficiently low barrier to entry for this cohort. While access and stability issues shaped the initial experience of both systems, the interaction-focused findings point to a deeper problem: both systems were often difficult for participants to steer, refine, and align with higher-level creative intentions.

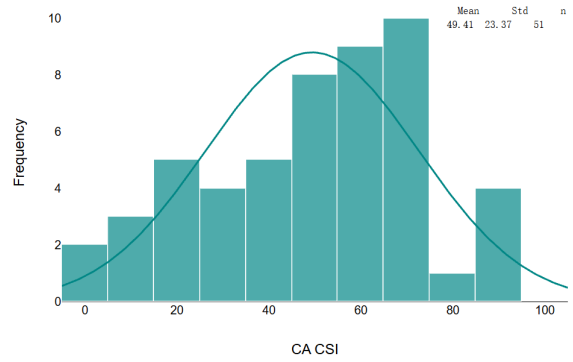
5.2 CSI (RQ1): Initial Usefulness, Limited Creative Support

To assess the creative process, we utilized an unweighted Creativity Support Index (CSI) (Cherry and Latulipe, 2014), omitting the "Collaboration" component. As illustrated in Figure 4, the overall unweighted CSI scores were CA2 $M = 49.41, SD = 23.37$ and Calliope $M = 41.40, SD = 22.18$. These scores suggest that neither system consistently provided strong perceived support for exploration, expressiveness, or effort reduction across the participant cohort. However, qualitative feedback suggests a more nuanced pattern: users often valued the tools as a rapid starting point, but struggled to transform generated material into intentional musical outcomes.

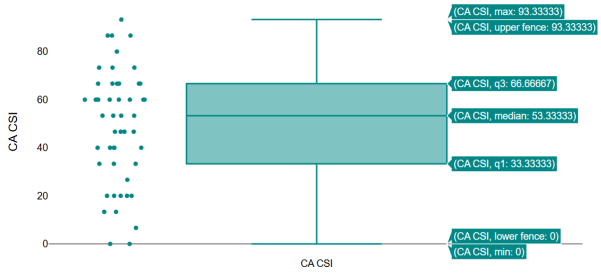
Despite the relatively low mean CSI scores, qualitative feedback indicated that some participants valued the generative systems as a rapid starting point ($n = 6$). As one participant noted: "The AI seemed pretty useful and was much easier than if I'd have to make something myself from scratch." This suggests that the generative capabilities of these tools provided an initial sense of productivity by helping some participants overcome the difficulty of starting from scratch. However, this initial usefulness did not consistently translate into sustained creative support, as participants' responses suggest that the main difficulty occurred during refinement rather than initial generation.

Initial Generation Was Easier Than Refinement
If the tools provided an initial sense of productivity, why did the overall CSI scores remain relatively low? The qualitative data suggest that a recurring bottleneck occurred during the *refinement* stage. When asked which aspects of generation required greater control, participants frequently identified both structural and semantic forms of steering, as summarized in Table 2.

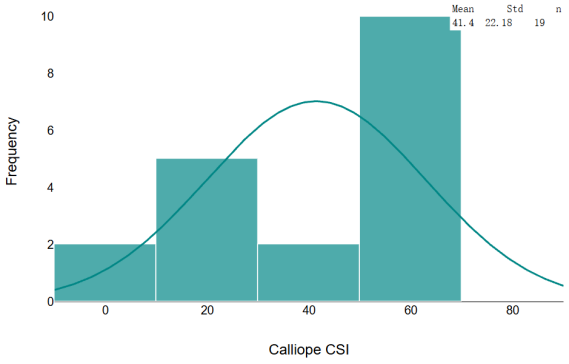
Participants primarily requested greater control over musical structure, particularly *Notes/Melody* ($n = 20$) and



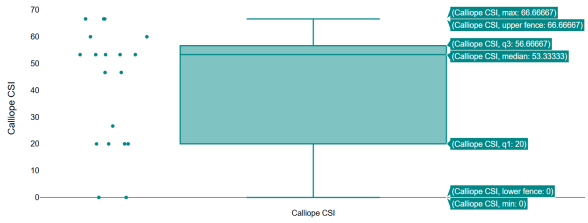
(a) CA2: CSI Score Distribution.



(b) CA2: CSI Score Boxplot.



(c) Calliope: CSI Score Distribution.



(d) Calliope: CSI Score Boxplot.

Figure 4: Unweighted Creativity Support Index (CSI) evaluation for Composer's Assistant 2 and Calliope, highlighting the overall low creativity support scores.

Table 2: Desired Areas for Increased Generative Control

Aspect of Generation	**Coded Responses** (n)
Notes/Melody	20
Prompt (Semantic Input)	16
Mood/Genre	16
Rhythm/Groove	12
Instruments	9
Editing/Adjusting	7

Note. Frequencies represent the number of valid open-ended responses coded under each theme. Codes were not mutually exclusive, and a single response could contribute to multiple themes. Some participants did not answer the relevant question or provided responses that were too limited, unclear, or unrelated to be meaningfully coded; therefore, the number of analyzable responses may be lower than the total sample size.

Rhythm/Groove ($n = 12$). The *Notes/Melody* code grouped requests for control over the sequence of MIDI notes generated by the system. Some participants described this need at the level of the resulting melody, for example: "I would like more control over the melodic contour and instrument selection, so the generated music can better match the emotional tone I want." Others described it at the level of the underlying note events, such as: "Being able to control the MIDI notes that are generated." Although expressed at different levels, both responses concerned the same underlying capability: generating and adjusting a sequence of MIDI notes so that it better matched the user's intended musical result.

Related requests concerning *Rhythm/Groove* and *Editing/Adjusting* emphasized incremental revision rather than complete regeneration. As one participant explained: "I would like to have more control over the beats of the sound; slight edits instead of having to completely redo and generate over and over again." Another participant similarly described limited freedom in choosing which parts of the MIDI template tracks should be changed.

Furthermore, a recurring conceptual disconnect emerged between users' semantic intentions and the systems' technical input controls. Many participants highlighted limitations related to *Prompt (Semantic Input)* ($n = 16$) and the inability to specify *Mood/Genre* ($n = 16$). One participant emphasized the need for semantic steering: "I would need a more detailed prompt on how the music sounds like, what the overall mood should be, instead of just what instruments are used."

Finally, the interface and workflow design added to the difficulty of refinement. Participants requested features such as the ability to "highlight the changes between different AI outputs." Rather than establishing that comparison or regeneration functions were absent, these responses suggest that support for comparing and tracking iterative changes was not sufficiently visible, understandable, or usable for some participants.

5.3 Perceived Authorship (RQ3)

When asked to characterize the final musical output and reflect on its authorship, 74 participants provided valid responses to this question. Only 2.7% ($n = 2$) categorized the output entirely as "My music." Among the remaining participants, opinions were evenly divided: 48.6% ($n = 36$) viewed the final output predominantly as "AI's work," while the other 48.6% ($n = 36$) considered it a "Collaboration" between themselves and the generative system.

Notably, within the "Collaboration" group, we identified three participants who provided their own original melodies and used CA2 primarily for harmonization and arrangement. All three categorized the final piece as a collaboration. This suggests that, even when the core melodic material originated from the participant, they still attributed a substantial creative contribution to the system's harmonization and arrangement.

The limited number of "My music" responses indicates that participants generally perceived the AI system as making a substantial contribution to the resulting composition. Participants' reported difficulties with steering and refining symbolic MIDI may help explain why many attributed the output either primarily to the AI or jointly to themselves and the system. However, categorizing the output as a "Collaboration" should not itself be interpreted as a loss of creative agency; rather, it reflects participants' recognition that both human input and system-generated material contributed to the final result.

- 48.6% ($n = 36$) viewed it as "AI's work"
- 48.6% ($n = 36$) viewed it as a "Collaboration"
- 2.7% ($n = 2$) viewed it as "My music"

5.4 Perceived Effectiveness (RQ1)

Reception was divided. For CA2, 49% of participants found the system beneficial, while 51% felt that it did not enhance their workflow. Calliope showed a similarly mixed response, with 52.63% agreement but substantial polarization, including 21.05% "Strongly Disagree" responses. Overall, the perceived utility of both generative MIDI systems was mixed and varied considerably across participants.

5.5 Future Adoption Intent (RQ2)

When assessing the likelihood of integrating these tools into their future music composition or audio production practices, participants expressed considerable uncertainty. Only 14.7% ($n = 11$) gave a definitive "Yes," while 34.7% ($n = 26$) selected "No." The majority, 50.7% ($n = 38$), selected "Maybe," suggesting continued interest tempered by current usability, workflow, and creative concerns.

This hesitation can be contextualized by participants' broader concerns regarding generative MIDI systems. While 24 participants reported having no concerns, other responses identified several issues that may influence future adoption. Twenty-seven participants mentioned concerns related to output quality, the impact of AI on human creators, or the potential restriction of personal creativity. In addition, 12 participants raised ethical or legal concerns regarding AI-generated music. These thematic categories were not necessarily mutually exclusive. Together, the responses suggest that future adoption may depend not only on improving technical usability and steerability, but also on addressing concerns related to creative contribution, authorship, output quality, and the broader implications of AI-generated music.

- Yes: 14.7% ($n = 11$)
- Maybe: 50.7% ($n = 38$)
- No: 34.7% ($n = 26$)

6 DISCUSSION

6.1 Different Results for Different Users: Comparison with the MMM-Cubase Evaluation

The low usability and creativity-support scores observed in the present study contrast with previous evaluations of related generative music systems. For instance, Tchemeube et al. evaluated MMM-Cubase, a DAW-integrated generative MIDI system supporting multitrack composition and infilling tasks related to those examined here (Tchemeube et al., 2023). Their study reported generally positive measures, including System Usability Scale (SUS) scores consistently above 70 and above-average Technology Acceptance Model (TAM) scores. However, their participant sample consisted of eight hobbyist experts and ten professional experts.

The present results also differ from the more recent novice-focused evaluation of MMM-Cubase v2 and Calliope, which reported generally positive usability, user experience, and future-use intentions among 98 novice participants (Tchemeube et al., 2026). This contrast suggests that novice status alone does not determine whether a generative music system will be experienced as usable or supportive. The studies involved different systems and versions, task structures, participant contexts, support conditions, and tool-allocation procedures and should therefore not be treated as a controlled comparison. Nevertheless, the contrasting results suggest that novice scaffolding may depend substantially on how a system is introduced and integrated into a specific creative workflow.

Taken together, these cross-study differences situate our findings within a broader discussion of how expertise, task context, and workflow design shape interaction with generative music systems. While some research characterizes AI as a "Great Leveler" for novices (**?**), our findings suggest that domain familiarity and task-specific knowledge can influence users' ability to steer complex, open-ended generative systems (Zamfirescu-Pereira et al., 2023). For the expert composers in Tchemeube et al.'s study, the minimal one-parameter interface received high usability scores, although participants still requested more granular steering control. In the present study, some participants appeared to have difficulty providing sufficiently structured musical input for the systems to mimic, extend, or accompany, which may have contributed to outputs being perceived as arbitrary.

6.2 Representational and Workflow Demands: Comparison with AuMe

Differences in representational and interaction demands become apparent when contrasting these results with the Audio Metaphor evaluation (Liu et al., 2025a). The AuMe study reported a SUS score of 64.33 and a CSI score of 66.90, both numerically higher than the scores observed in the present study. The cross-study tests reported in Section 5.1 quantify differences between the observed samples. However, because the studies involved different tasks, systems, participants, and research designs, these differences should not be interpreted as causal estimates of the systems' relative usability. Instead, the comparison provides contextual evidence for examining how different representations and workflows may shape novice user experience.

AuMe allows users to describe a scene or intended affect through natural language, retrieves and initially arranges relevant sound materials, and exports an editable REAPER project. Participants therefore worked primarily with perceptually accessible audio clips: they could listen to the retrieved material and refine it through selection, trimming, repositioning, replacement, layering, effects, and mixing. These clip-level operations do not necessarily require users to interpret an abstract musical representation. Nevertheless, AuMe was not friction-free, as participants also reported prompt misinterpretation and limited controllability.

In contrast, CA2 and Calliope generate discrete symbolic MIDI note events. Refining this material may require users to relate an auditory judgement to note-level, rhythmic, harmonic, or structural changes. This may impose greater interpretive and editing demands on novices with limited familiarity with music theory and MIDI-based representations, even though the generated material is technically editable.

The systems also introduced different forms of workflow friction. AuMe provided browser-based access and direct export to an editable REAPER project. CA2 was integrated into REAPER, but installation and configuration prevented or delayed some participants from beginning the creative task. Calliope avoided local installation but required repeated exporting, uploading, downloading, importing, and instrument assignment when moving between the browser and the DAW. Participants' responses suggest that these access and transfer requirements could interrupt creative work and reduce the practical benefits of rapid generation. The comparison therefore indicates that novice support depends not only on what a system generates, but also on whether its representations and workflow are understandable and manageable for its intended users.

6.3 From Novice Intuition to Actionable Refinement

Participants' requests for control over notes and melody suggest that some could recognize when a generated passage did not match their creative intention, yet lacked an accessible way to translate that judgement into a targeted MIDI edit.

One possible design response is semantic-to-symbolic manipulation. Rather than requiring users to identify and manually edit specific notes, measures, or parameters, a system could translate novice-accessible descriptions into targeted transformations of selected regions. For example, it might support illustrative commands such as "make this phrase less repetitive," "keep the rhythm but change the melody," or "make the ending feel more tense." Such operations could preserve unaffected material, modify melodic contour or rhythmic structure, and make the resulting changes visible. The goal would not simply be to expose more parameters, but to make refinement actionable through users' existing perceptual and descriptive vocabulary.

Calliope already supports batch generation and bar-level selection (Tchemeube et al., 2022), while both systems permit partial or selected-region regeneration in different forms. However, participants' responses suggest that these capabilities did not consistently translate into accessible creative control. The design opportunity is therefore not only to provide generation and refinement functions, but also to make them discoverable, understandable, and sufficiently low-friction for novice users. Existing generation and regeneration controls could be made more accessible through rapid auditioning and clearer guidance, and could be extended with side-by-side comparison, difference highlighting, region locking, and the recombination of

preferred material from different alternatives. Such designs could replace an unstructured generate–audition–reject cycle with a workflow in which users compare alternatives, preserve useful material, and refine the result incrementally. By making differences between alternatives visible, these interactions may also create opportunities for users to develop a better understanding of MIDI representations and symbolic musical structure. The present study did not directly evaluate these interface designs. However, participants' requests for targeted edits, semantic input, partial regeneration, and clearer comparison between outputs motivate these possibilities as directions for future research.

6.4 Perceived Authorship and Creative Agency

The representational, usability, and workflow barriers discussed above were reflected in how participants perceived authorship and creative contribution. Only 2.7% of participants categorized the output entirely as "My music," while the remaining responses were evenly divided between "AI's work" and "Collaboration." These results indicate that participants generally regarded the systems as making a substantial creative contribution, but they do not by themselves demonstrate a loss of creative agency.

Participants' reported difficulties with semantic steering and targeted refinement may help explain why many attributed a substantial share of the creative contribution to the AI. Because many participants had difficulty precisely editing generated MIDI, they often relied on repeatedly generating, auditioning, and selecting alternatives rather than making targeted changes to the musical structure. This may have reduced their opportunities to exercise creative agency by intentionally shaping the output.

However, perceived AI authorship and human creative agency are not necessarily opposites. The three participants who supplied their own original melodies retained clear creative agency over the core musical idea while using CA2 primarily for harmonization and arrangement. Their classification of the final result as a "Collaboration" suggests that they recognized meaningful creative contributions from both themselves and the system, rather than indicating that the AI had displaced their agency.

More broadly, when the system makes compositional decisions such as harmonic progression, melodic contour, and arrangement without clear opportunities for user intervention, participants may have fewer ways to align the generated material with their artistic intentions. For some participants, this may have contributed to the perception of the system as the primary composer. For others, particularly those who contributed their own musical material or meaningfully selected and refined the output, the process was understood as collaboration. The study did not directly measure participants' degree of control during each stage of harmonization, arrangement, and refinement, so agency cannot be inferred from authorship attribution alone.

The course context may also have shaped how participants interpreted the role of music and AI authorship. Open-ended responses suggest that participants differed in how they positioned music composition within their broader creative practice. Some described primarily functional uses, such as "making copyright-free background music for videos." Others placed greater emphasis on personal creative ownership; for example, one participant stated, "I wouldn't use AI to generate work because then it would be the AI's work and not mine." Several participants occupied an intermediate position, viewing AI-generated material as a starting point for ideation or subsequent development. Because the survey did not directly measure whether participants regarded music composition as an autonomous artistic practice or primarily as a means of completing a broader audiovisual task, these responses should be interpreted as exploratory evidence rather than as distinct participant categories. Future work should examine how creative orientation and identification with music-making influence perceived authorship, control, and adoption.

Behind the High Standard Deviations

A notable characteristic of these quantitative results is the high standard deviation observed ($SD \approx 17$–19). In standard HCI evaluations, such variance suggests that the user experience was not uniformly poor, but rather characterized by substantially different responses across participants. This contrasts with the expert evaluation of MMM-Cubase, where SUS standard deviations were lower ($SD = 10.0$–14.48), suggesting more consistent responses among the experienced participants in that study. This statistical spread is consistent with the heterogeneous experiences observed in the qualitative responses. Technical setup, familiarity with MIDI concepts, and attitudes toward AI-assisted creativity may have contributed to the variation, although these relationships were not directly tested.

7 CONCLUSION

This study evaluated two Symbolic Music Generation (SMG) systems, Composer's Assistant 2 and Calliope, in supporting musical composition for novices. The two systems received low mean usability scores among participants in this course-based study. While the systems provided some participants with a rapid starting point and reduced the difficulty of beginning a composition from scratch, the subsequent refinement process presented a recurring bottleneck due to limited intuitive control over the generated output.

Compared with the prior evaluation of Audio Metaphor (AuMe), the SMG workflows examined here introduced additional demands associated with foundational music theory, MIDI-based symbolic representations, and technical workflow management. Participants frequently struggled to map what they wanted to hear onto the available system controls and to provide or interpret the musical context required by the generative systems.

These recommendations are directly grounded in three recurring patterns in the qualitative findings. First, some participants valued the systems as rapid starting points ($n = 6$), while participants also encountered difficulty providing or interpreting the musical context required for generation. This motivates the use of guided templates and structural starting points that help novices begin composing without requiring them to independently construct detailed musical input. Second, requests for greater control over semantic prompts and mood or genre were among the most frequently coded themes ($n = 16$ for each). These responses support the development of mappings between higher-level creative intentions and technical music-structural controls. Third, participants most frequently requested greater control over notes and melody ($n = 20$), rhythm and groove ($n = 12$), instrumentation ($n = 9$), and editing or adjustment ($n = 7$). Their comments also identified difficulties with partial regeneration, making small

targeted edits, and comparing changes between generated alternatives, providing direct motivation for stronger post-generation refinement support.

Based on these findings, we propose that future research and system design prioritize three areas:

1. **Novice-Oriented Structural Scaffolding:** Providing higher-level musical templates or guided starting points that reduce reliance on users' ability to supply detailed or well-structured musical input.
2. **Semantic-to-Symbolic Manipulation:** Translating novice-accessible descriptions, such as emotional affect, musical genre, audiovisual function, or requests to alter a particular passage, into targeted changes to notes, melodic contour, rhythm, instrumentation, or selected regions.
3. **Accessible Structured Creative Search and Refinement:** Making existing and future capabilities for batch or parallel generation, rapid auditioning, comparison between alternatives, visible differences across generations, region locking, and selective regeneration discoverable and understandable to novice users, allowing them to preserve useful material and refine outputs incrementally.

These proposals are motivated by recurring patterns in participants' responses, but the proposed interaction designs should be evaluated directly in future studies. By connecting higher-level creative intentions with more actionable symbolic operations and structured exploration, future SMG systems may support more intentional user participation and clearer control over the respective creative contributions of the human user and the system.

AUTHOR DECLARATIONS

The authors declare no conflicts of interest. Ethical approval for the study was granted by the Simon Fraser University Research Ethics Board.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING

During the preparation and revision of this manuscript, the authors used Google Gemini and OpenAI ChatGPT for language refinement, copyediting, and assistance in reorganizing text. The authors critically reviewed and revised all AI-assisted text and take full responsibility for the final content of this publication.

REFERENCES

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In Jordan, P. W., Thomas, B., Weerdmeester, B. A., and McClelland, I. L., editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis.

Bryan-Kinns, N., Noel-Hirst, A., and Ford, C. (2024). Using incongruous genres to explore music making with AI-generated content. In *Proceedings of the 16th Conference on Creativity and Cognition*, pages 229–240. Association for Computing Machinery.

Cherry, E. and Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction*, 21(4):21:1–21:25.

Chion, M. (2019). *Audio-Vision: Sound on Screen*. Columbia University Press, New York, NY, 2 edition.

Collins, K. (2008). *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. MIT Press, Cambridge, MA.

Creswell, J. W. and Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications, 5 edition.

Ens, J. and Pasquier, P. (2020). MMM: Exploring conditional multi-track music generation with the transformer. arXiv preprint arXiv:2008.06048.

Guo, J., Lu, Z., Guo, J., Chen, W., Yuan, T., and Zhang, Y. (2026). EmoGen: An emoji-based interactive app to support emotion expression via generative music (exploring usability and emotional resonance in symbolic music interaction). In *HCI International 2025—Late Breaking Papers*, volume 16342 of *Lecture Notes in Computer Science*, pages 41–57. Springer.

Hu, X., Xing, Y., Cai, X., Zhao, Y., Cook, M., Borgo, R., and Neate, T. (2025). Designing interactions with generative AI for art and creativity: A systematic review and taxonomy. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, pages 1126–1155. Association for Computing Machinery.

Hunt, S. J., Mitchell, T., and Nash, C. (2020). Composing computer-generated music: An observational study using IGME, the interactive generative music environment. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 61–66.

Kantosalo, A. and Toivonen, H. (2016). Modes for creative human–computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the Seventh International Conference on Computational Creativity*, pages 77–84, Paris, France. Sony CSL.

Krol, S. J., Llano, M. T., and McCormack, J. (2025a). Supporting creative ownership through deep learning-based music variation. NeurIPS Creative AI Track 2025.

Krol, S. J., Llano Rodriguez, M. T., and Loor Paredes, M. J. (2025b). Exploring the needs of practising musicians in co-creative AI through co-design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery.

Le, D.-V.-T., Bigo, L., Herremans, D., and Keller, M. (2025). Natural language processing methods for symbolic music generation and information retrieval: A survey. *ACM Computing Surveys*, 57(7):1–40.

Lerch, A., Arthur, C., Bryan-Kinns, N., Ford, C., Sun, Q., and Vinay, A. (2025). Survey on the evaluation of generative models in music. *ACM Computing Surveys*, 58(4):1–36.

Liu, G., Lee, K. J. M., Thorogood, M., Anderson, C., and Pasquier, P. (2025a). AI-assisted sound design with audio metaphor (AuMe): An evaluation with novice sound designers. In *Proceedings of the 6th Conference on AI Music Creativity*.

Liu, J., Zou, J., Zhang, J., and Teng, J. (2025b). Investigating users' acceptance of AI-based creativity support tools: An empirical study from china's creative industries. *Current Psychology*, 44(16):13933–13950.

Liu, X. and Mao, Z. (2026). Micro-learning of generative AI on digital media and its adoption among college music majors: Applying the technology acceptance model with AI identity threat as a moderator. *Education and Information Technologies*, 31(9):2513–2538.

Louie, R., Coenen, A., Huang, C. Z., Terry, M., and Cai, C. J. (2020). Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13. Association for Computing Machinery.

Malandro, M. E. (2024). Composer's assistant 2: Interactive multi-track MIDI infilling with fine-grained user control. In *Proceedings of the 25th International Society for Music Information Retrieval Conference*, pages 438–445, San Francisco, CA, USA.

Pasquier, P., Ens, J., Fradet, N., Triana, P., Rizzotti, D., Rolland, J.-B., and Safi, M. (2025). MIDI-GPT: A controllable generative model for computer-assisted multitrack music composition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 1474–1482.

Ronchini, F., Comanducci, L., Perego, G., and Antonacci, F. (2025). PAGURI: A user experience study of creative interaction with text-to-music models. *Electronics*, 14(17):3379.

Tchemeube, R. B., Ens, J., Lee, K. J. M., Pasquier, P., Rolland, J.-B., Grabit, Y., and Safi, M. (2026). Novice users' evaluation of two multi-track music machines for AI-assisted music composition: Usability, user experience and acceptance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Tchemeube, R. B., Ens, J., and Pasquier, P. (2022). Calliope: A co-creative interface for multi-track music generation. In *Proceedings of the 14th Conference on Creativity and Cognition*, pages 608–611. Association for Computing Machinery.

Tchemeube, R. B., Ens, J., Plut, C., Pasquier, P., Safi, M., Grabit, Y., and Rolland, J.-B. (2023). Evaluating human-AI interaction via usability, user experience and acceptance measures for MMM-C: A creative AI system for music composition. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 5769–5778.

Thorogood, M. (2021). Developing a sound design creative AI methodology. In *Doing Research in Sound Design*, pages 224–237. Focal Press.

Thorogood, M. and Pasquier, P. (2013). Computationally created soundscapes with audio metaphor. In *Proceedings of the Fourth International Conference on Computational Creativity*, pages 1–7.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., and Yang, Q. (2023). Why johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–21. Association for Computing Machinery.

Zang, Y. and Zhang, Y. (2024). The interpretation gap in text-to-music generation models. In *Proceedings of the 3rd Workshop on NLP for Music and Audio*, pages 112–118, Oakland, California, USA. Association for Computational Linguistics.

Zhou, E. and Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, 3(3):pgae052.

Zhou, Y., Wang, Y., and Su, J. (2025). A study on factors influencing music enthusiasts' behavioral intention to use generative AI for music composition based on an expanded technology acceptance model. *Scientific Reports*, 15:44802.

Zhou-Zheng, C. and Pasquier, P. (2025). Personalizable long-context symbolic music infilling with MIDI-RWKV. arXiv preprint arXiv:2506.13001.