Novice Users' Evaluation of Two Multi-track Music Machines for Computer-Assisted Music Composition: Usability, User Experience and Acceptance

Renaud Bougueng Tchemeube*
rbouguen@sfu.ca
Simon Fraser University
Surrey, Canada

Jeff Ens*
jeffreyjohnens@gmail.com
Simon Fraser University
Surrey, Canada

Keon Ju Maverick Lee*
keon_maverick@sfu.ca
Simon Fraser University
Surrey, Canada

Philippe Pasquier*
pasquier@sfu.ca
Simon Fraser University
Surrey, Canada

Jean-Baptiste Rolland*
jb.rolland@steinberg.de
Steinberg Media Technologies GmbH
Hamburg, Germany

Yvan Grabit*
y.grabit@steinberg.de
Steinberg Media Technologies GmbH
Hamburg, Germany

Maryam Safi*
m.safi@steinberg.de
Steinberg Media Technologies GmbH
Hamburg, Germany

Abstract

The maturity of creative AI systems in the arts raises new questions regarding their integration into creative practices. The music field is no exception, and is seeing a rise in new creative AI tools, notably in music composition. We study how these systems and their design impact user's adoption. Specifically, we conducted a user study with 98 novice participants evaluating usability, user experience, and technology acceptance for two computer-assisted composition (CAC) systems: MMM-Cubase v2 and Calliope. Findings show both systems are easy to control and use, with Calliope being easier to use and more immersive. 76.9% (MMM-Cubase v2) and 72.9% (Calliope) of users report positive predicted future use, while the novel and efficient workflow contributes to lower barriers to music-making. Depth of control and model transparency remain outstanding issues while users highlight concerns over loss of musical diversity and skill learning.

Keywords

Evaluation Studies, Human-AI Interaction, Computer-Assisted Music Composition, Creativity Support, Machine Learning, Artificial intelligence, Sound and Music Computing, Interaction Design

1 Introduction

With the continued advancements in Artificial Intelligence (AI), there is growing research interest for human-AI interaction in creative domains, including drawing [15], writing [10], video game generation [20], sound design [25, 38], animation [1, 2] and computer-assisted music composition (CAC) [27]. CAC is an area of computer music research that explores technological formalisms for digitally assisting composers in the process of ideating and developing musical compositions [3]. Although a plethora of CAC systems have been developed, much of the research historically focused on "demo" systems to showcase novel algorithms [6, 34], rather than improving the affordances of their interfaces for practical scenarios. The opaque behavior of modern LLM-based CAC systems [22] continues to pose interaction challenges. As CAC systems become accessible to lay users, we seek to understand their user adoption, in particular for AI-based symbolic multi-track systems. We also consider how this may vary across user profiles and system design configurations.

To achieve this research goal, we conduct a mixed methods study which evaluates *usability*, *user experience* and *acceptance* of novice music composers for two CAC systems: MMM-Cubase v2 (MMM-C v2) and Calliope [37]. Cubase is a popular software for music composition and production. The two MMM-based CAC systems offer comprehensive, semantically meaningful, steering parameters for generation (i.e., note duration, polyphony range), support multi-track instrumentation, are agnostic to music genres, and cover two deployment configurations: embedded desktop (MMM-C v2) and online standalone (Calliope). MMM-C v2 is a design iteration over MMM-C v1 [36], a "1-parameter" generative interface of the "Multi-track Music Machine" (MMM) [18] in Cubase. Like the first iteration, it sits as a plugin within the Cubase system architecture. Our evaluation study of MMM-C v1 [36] provided insight into the level of adoption by expert users. This proposed research reports insights regarding those constructs for 98 novice users organized into two user groups: 1) *early novice* and 2) *skilled novice* users. Our key contributions are: 1) a comprehensive evaluation of the novice experience on CAC systems employing GPT-based models (in this case MMM) for genre-agnostic multi-track music creation, 2) a mixed methods-based assemblage for such evaluation, and a demonstration of its

*All authors contributed equally to this research.

NIME '26, June 23–26, 2026, London, UK

benefits in helping standardize evaluations across studies, 3) differences in usability across novice user groups and deployment configurations, and 4) an evaluation of the level of technology acceptance for two GPT-based CAC systems. Our research questions are as follows:

- RQ1: To which extent are MMM-based CAC systems such as MMM-Cubase v2 and Calliope, usable by novice users for multi-track music creation?
- RQ2: How can we characterize the user experience of novice composers, including their sense of perceived autonomy, flexibility, accessibility, and authorship, when using MMM-based CAC systems such as MMM-Cubase v2 and Calliope for multi-track music creation?
- RQ3: What is the level of technology acceptance by novice users of MMM-based CAC systems such as MMM-Cubase v2 and Calliope for multi-track music creation?

2 Related Work

We focus exclusively on studies of CAC research on some form of usability, user experience or acceptance (Table 1). Bray et al. [7] select three tools designed for CAC and analyze three specific computationally creative interface categories; direct manipulation systems, programmable interfaces and highly encapsulated systems; using respectively Jnana Live, Patter and StyleMachine systems. They conducted an investigation looking at the user experience of a single expert user using the tools and discussed the implications that encapsulating system features has on visibility of user parameters across different computationally creative scenarios.

Roberts et al. [33] develop Magenta Studio, a system bringing interactive generative music to professional music creators, combining deep learning-based music generation and direct integration with Ableton Live, a popular music software. Magenta Studio is a collection of five distinct music plugins for continuation, 4-bar generation, drum generation, interpolation, and rhythmic performance. The authors conduct a survey using mixed questionnaire with early adopters (a mix of musicians, producers and machine learning enthusiasts) evaluating the system's effectiveness and emerging experiences. From the 89 responses, they learn about the ease of using Magenta Studio's outputs in their musical work (66% easy, 24% neutral, 10% difficult), the observed response time in producing generated outputs (58% experienced little to no delay and only 6% reported significant delays), the usefulness of the system in how easy it was to achieve desirable musical effect (41% easy, 40% neutral, 19% difficult), whether the system made them feel more creative (72% more, 20% neutral, 8% less), and more productive in their creative process (93% yes).

Louie et al. [27] compare Bachdoodle [21]) with infill mask and the Cococo interface with interactive steering controls: voice lanes, semantic sliders, and generative alternatives. The within-study design of 21 novice musicians shows semantically-relevant user control features improves overall usability of the AI music generative model, namely on creative ownership (or authorship), self-efficacy and collaboration. The study measures 3 dimensions of user-AI interaction: 1) the composition experience, 2) the attitude towards AI and 3) the perception of composition. On composition experience, they record significant increase in expressing goals, self-efficacy, engagement and learning, with no significant difference on effort. On attitude towards AI, they observe an increase in controllability, comprehensibility, collaboration and trust. On perceptions of composition, they get increase in ownership, AI's vs user's contribution, completeness, and no significant difference on uniqueness of the generative outputs. The study only addresses compositions of 4-voice piano-based polyphonic western music. Thus, it is not clear how the results might generalize to multi-track generation, to other music genres, or to different system deployments.

Finally, we conduct a comprehensive evaluation of 18 expert users' adoption for MMM-C v1 [36] on three composition tasks: arrangement, variation, and original creation. They integrate MMM into Cubase [1], a popular Digital Audio Workstation (DAW) by Steinberg, by producing a "1-parameter" plugin interface named MMM-Cubase (MMM-C), which enables human-AI co-composition. They recruit the expert participants via Steinberg's pool of beta-testers from diverse demographic backgrounds and musical levels. Study results show positive usability and acceptance scores. Users report experiences of novelty, surprise and ease of use from using the system, and limitations on controllability and predictability of the interface when generating music. Findings indicate no significant differences between the two user groups.

Our study extends the literature by looking at *novice* composers using MMM-based CAC systems to compose multi-track music pieces across music genres. Our results add to our understanding of the acceptance level and HCI evaluation challenges associated with the usage of such systems.

3 Methodology

To conduct this study, we employed a *mixed methods* methodology, using both quantitative and qualitative data collection techniques to better understand usability, user experience, and technology acceptance. We adopt a convergent parallel design [11] where quantitative and qualitative data analysis are combined to triangulate findings, offering a stronger reliability to the interpretative results. We analyze the two types of data separately and then compare the results to see where they agree, differ, or complement each other. In the end, this helps come up with deeper, more meaningful insights about the findings. The measurement instruments for each construct are outlined in the sub-sections below. All the forms, assets and details are accessible on our companion webpage at https://metacreation.net/projects/novice-eval-ai-cac-2026.

3.1 Interactive CAC

MMM-C v2 and Calliope both assist at music composition by generating musical parts given the parameter settings of the user request and taking into account the existing musical context of the composition workspace. The generative capability is enabled by their use of the MMM model. MMM is an expressive GPT model for generating symbolic (music-sheet like) music which performs musical score *inpainting* [17, 18]. Inpainting [5, 16, 39] is a task which consists in automatically filling the blank of a piece of content. The task is prevalent in text, audio and music, and predominently found in image generation work. It enables effective addition and editing of content.

MMM-C v2's goal is to offer controllable music generation that embeds directly into musicians' typical work environment and flow. For our research design, we choose a CAC tool embedded in a commercial DAW to replicate that work environment for composers. Cubase, a Steinberg product, was selected over other options as it is one of the oldest and most popular DAWs, with

[1] https://www.steinberg.net/cubase

Papers	Participants	Deployment	Systems	Methodology	Tools	Constructs
Bray et al. (2017) [7]	Experts (n=1)	1 Embedded, 2 Standalone	Patter, Jnana, Style Machine (Ableton)	Within-subject study, partitioner's reflections,	Qualitative surveys	Usability, parameter design, interface opaqueness
Roberts et al. (2019) [33]	Novices (n=89)	Embedded	Magenta Studio (Ableton)	Remote unmoderated survey	Mixed questionnaire	Ease of use, usefulness, creativity support
Louie et al. (2020) [27]	Novices (n=21)	2 Standalone	Cococo, Bachdoodle	Within-subject study	Think-aloud, questionnaire, semi-structured interview	Composer experience, attitude towards AI, perceptions of composition
Louie et al. (2022) [28]	Novices (n=26)	2 Standalone	Performance RNN, Music Transformer	Within-subject studies (interfaces and AI models)	Think-aloud, questionnaire, semi-structured interview	Expression, communication, musical coherence, ownership, control, efficacy
Tchemeube et al. (2023) [36]	Experts (n=18)	Embedded	MMM-C (Cubase)	Remote unmoderated survey	Mixed questionnaires	Usability, user experience, trust, authorship, controllability, acceptance
Tchemeube et al. (2026)	**Novices (n=98)**	**1 Embedded, 1 Standalone**	**MMM-C v2 (Cubase), Calliope**	**Remote unmoderated survey**	**Mixed questionnaires**	**Usability, user experience, trust, authorship, controllability, acceptance**

Table 1: Shortlist of CAC Evaluation Studies

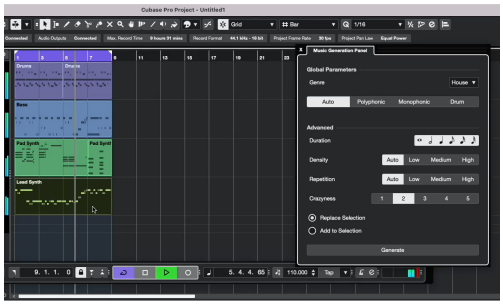

Figure 1: MMM-Cubase v2 Plugin in Cubase

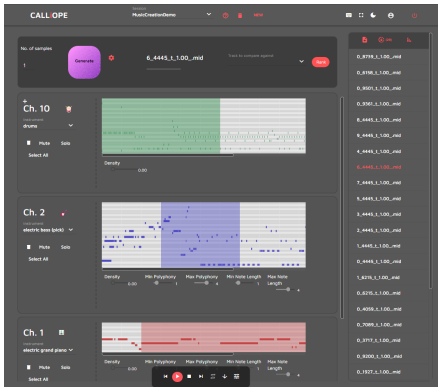

Figure 2: Calliope [37]

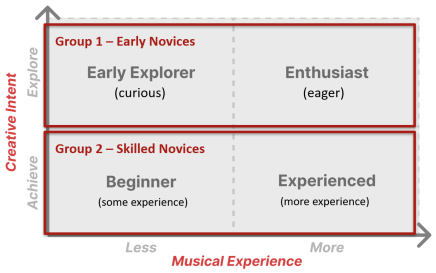

Figure 3: Novice User Profiles

a long history and a dedicated user base The user interface (UI) presents a mix of global and granular parameters to specify a generative request for MMM and steer its behavior. The global parameters are *musical genre* [Pop, House, Electronic, Jazz, and more], *musical pattern* [Automatic (Auto), Polyphonic, Monophonic, Drum]. The granular parameters controlling MMM's infilling behavior are: *duration* [Select All, Whole, 1/2, 1/4, 1/8, 1/16, 1/32], *density* [Auto, Low, Medium, High], *repetition* [Auto, Low, Medium, High], *craziness* [1, 2, 3, 4, 5, default: 3] controls the typicality of the generated lines, *replace selection vs add to selection* which decides whether the generated material replaces or not the content of the selected bars.

Calliope's goal [2] is to support web-based standalone access to music creation tools for artists, novices and lay users. The project enables access to the MMM model for any user 1) to make music without necessitating extensive composition knowledge, and 2) to access MIDI-GPT [32] without computer code. The online user interface regroups essential features for symbolic music manipulation [37]. These are MIDI file management, MIDI track visualization, MIDI playback in the browser, enabling respectively the ability to browse, display and play MIDI files. User can select bars to guide generation given a complete set of global (model-level) and local (track-specific) parameters which overlap with MMM-C v2, but offers its own original sampling controls such as *Bar per Step*, *Track per Step* and *Polyphony Hard Limit* [37]. Its workflow allows composers to tackle a broad range of compositional tasks. Calliope is released under the Apache License 2.0 for research purposes.

[2]Calliope's webpage: https://metacreation.net/projects/calliope

3.2 Participants

In the previous study, we looked at experts and professionals [36]. Here, we consider novice composers, defined as individuals who

are: non-professional users for which music composition is not their main activity, their musical training level is limited and thus cannot qualify as experts in music composition. Participants identify to one of four distinct composer profiles, broken into two user groups: the *early novices* group which contains early explorers and enthusiasts, and *skilled novices* group consisting of beginner and experienced users. Participants self-report the composer profile that best describes them. Each profile is associated with a defining statement :

- *Early Explorer*: "I am always curious about trying out new tools".
- *Enthusiast*: "I want to compose music, but never really tried before".
- *Beginner*: "I did compose some music but never finished a complete track".
- *Experienced*: "I finished at least one complete music track before". They are more likely to have experience with typical DAWs such as Cubase, but might not be as competent as Expert users.

Early novices (group 1) are thus more aspiring and curious, while skilled novices (group 2) are more developing and achieving.

3.3 Task

The participants are given one task typical in multi-track music composition: **Arrangement**; the task of producing a composition around a 16-bar long rhythmic MIDI track in a given music genre. This consists in extending the composition with additional MIDI tracks with instrumentation based on the target music genre. The user must use MMM-C v2's AI features (i.e., track in-filling, bar in-filling, polyphony controls, note duration, repetition, note density) to compose the music piece. They are provided with existing genre-specific MIDI files to select from (pop, rock, electronic and hip-hop).

3.4 Usability

Usability is the extent to which the system enables the user to effectively, efficiently and satisfiably achieve its goals. Our chosen evaluation technique for usability is remote unmoderated quantitative usability testing [4]. The participants fill three surveys: The Standard *System Usability Scale* (SUS) (5-point Likert scale) [8], The *Creativity Support Index* (CSI) [9], and a *Controllability* survey. SUS scores have a range of 0 to 100. A value < 50 is considered *unacceptable*, 50-70, *marginal*, and > 70, *acceptable*. CSI measures the extent to which MMM-C effectively support the composer's creative process. It outputs a single CSI score out of 100, with a higher score indicating greater creativity support. CSI questions (12 agreement statements, 10-point Likert scale) evaluate the following factors (2 statements per factor): Results Worth Effort, Exploration, Collaboration, Immersion, Expressiveness and Enjoyment. Our CSI measurements are adjusted to use one statement per factor (5-point Likert scale) and because our MMM-C did not involve any human collaboration, we omit the corresponding, ending up with a total of 5 agreement statements. The Controllability survey measures the tool's functional *AI-steerability*, we use two 10-point Likert scale questions and an open-text comment for complementary qualitative data.

3.5 User Experience

User experience is the quality of the experience the user has when interacting with the system. We are interested in user feedback about the software's capabilities as well as in understanding the level of:

- *Trust and perceived quality*: regarding the system's actions and the generated musical outputs
- *Authorship*: the degree to which the participants consider themselves authors of the final music piece.
- *Flexibility*: the extent to which the MMM model offers various opportunities for the participant to achieve the compositional objectives.

Thus, each participant completes four open-ended qualitative questions about their feeling of trust, perceived quality, authorship and flexibility. They then answer eight reflective questions regarding their experience and the major benefits and inconveniences of using a creative AI tool like MMM-C v2 or Calliope. We apply *thematic analysis* [23], using inductive coding to develop the resulting themes from the data. Specifically, two coders code separately and independently each survey question, then later discuss and consolidate them into a set of finalized codes from which the main themes of user experience are generated.

3.6 Technology Acceptance

Technology acceptance is understood as the potential of adoption for such systems in the future. The Technology Acceptance Model (TAM) [12] is used to evaluate the acceptance level from the participants by measured three factors: perceived usefulness, perceived ease of use, and attitudes towards usage of the system. TAM has been shown to reliably predict and explain user acceptance of information technologies [14] and is the most widely used instrument for this purpose [13]. Our TAM questionnaire consists of 12 questions (5-point Likert scale) with half (6) measuring perceived ease of use and the other half, the participant's perceived usefulness. Scores for individual questions are averaged out to obtain a single score for each of 1) *perceived ease of use (PEU)* and 2) *perceived usefulness (PU)*.

4 Study

We ran a pre-study mockup with internal participants (i.e., lab members) to test our protocol and review procedures. We then refined research materials, participant instructions and surveys to fine-tune the study process. We then recruited participants from the art, music and research communities including amateur musicians, sound design students (IAT 340 course at Simon Fraser University (SFU)), and computer science researchers in music AI from various sources including school research recruiting portals, mailing lists, and newsletters. Participants were asked to fill out an onboarding form, access the music software, complete the music arrangement task, and fill out the exit survey. They had the option to enroll for one (or both) of the systems. Study batches ran in parallel from March 15th to July 19th 2024, and took 2h to 3h of effort for a given participant, depending on their level of familiarity with such AI systems, the Cubase software, music and research processes. They then upload their results to us for review: a project file and, a final audio master for MMM-Cubase v2 users, or a final MIDI file for Calliope users. They were compensated with a choice among Steinberg's music software licenses such as Cubase Element [Audio Production] (99.99 EUR), Dorico Element [Scorewriting] (99.99 EUR) or Wavelab Element [Audio Mastering] (129.00 EUR). The sound design students received course credit bonuses only for their enrollment and effort of completion, in accordance with the school's academic research policies. A total of 98 out of the 251 who enrolled effectively took

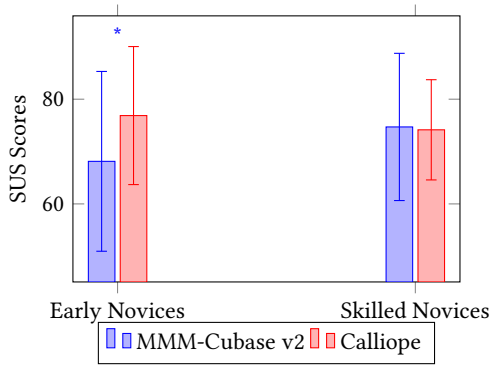

Figure 4: SUS Results per User Profile across CAC Systems

part in the study by completing the study from at least one of the systems. An anonymized sample collection of the musical excerpts by participants of MMM-Cubase v2 (audio files) and Calliope (MIDI files) can be found on the companion webpage at https://metacreation.net/projects/novice-eval-ai-cac-2026.

5 Results

5.1 Demographics

For MMM-Cubase v2, 39 participants enroll: 16 early novices and 23 skilled novices. 29 male, 9 female and 1 prefer not to say. Age groups breakdown as follow: < 20 years old (2/39), 20-29 years old (28/39), 30-39 years old (3/39), 40-49 years old (3/39), 50-59 years old (1/39) and ≥ 60 years old (2/39). Geographic distribution is North America (31), Europe (6), Oceania (1), Africa (1). 35/39 had no prior experience with Cubase and 34/39, no related experience with music-related AI tools. The **average musical experience level** is **4.88** / 10 (*SD* = 2.14). For comparison, the average musical experience level reported by a set of 18 expert users from the MMM-C v1 study is 8.1 / 10 (*SD* = 1.43) [36].

For Calliope, we have 59 participants composed of 46 early novices and 13 skilled novices. The gender spread is 37 female, 19 male and 3 non-binary. Age groups are as follow: < 20 years old (4/59), 20-29 years old (52/59), 30-39 years old (0/59), 40-49 years old (1/59), 50-59 years old (1/59) and ≥ 60 years old (1/59). 30/59 participants reported not being familiar with the DAWs, 20/59 were and 9 were unsure. 55 had no prior experience with music AI tools. 35 had reported not being familiar with or having used MIDI (symbolic) or Instrument tracks when making music. The **average musical experience level** is **4.3** /10 (*SD* = 1.92).

5.2 Usability

5.2.1 Standard System Usability Scale (SUS). Overall SUS are *M* = 72 / 100 (*SD* = 15.54) for MMM-C v2 and *M* = 76.27 / 100 (*SD* = 12.45) for Calliope. Therefore, SUS scores of MMM-C v2 and Calliope on the arrangement task are ***acceptable***. Figure 4 reports a breakdown of SUS scores per user profiles. Calliope's combined SUS score for Early Novice group is 76.87 (*SD* = 13.18) vs 68.13 for MMM-C v2 (*SD* = 17.17) ($t(60) = 2.11$, $\boldsymbol{p = .039}$, d = 0.61). Meanwhile, Calliope's combined SUS score for the Skilled Novice group is 74.15 (*SD* = 9.57) and 74.7 for MMM-C v2 (*SD* = 14.06) ($p = .901$). This indicates that MMM-C v2 as a plugin embedded in a typical DAW is well suitable and offer familiar operability for experienced and beginner users who have previously interacted with such designs to work on their compositions. P13 (Early Explorer, MMM-C v2) says: "*It feels too complicated and too professional*". Qualitative data confirm that 25.6% of novice users (10/39 codes) reported being overwhelmed by the Cubase software, its number of controls and options which is ideal for expert users, and experienced installation issues on their machine. P9 (Enthusiast, MMM-C v2) says "*The amount of features is overwhelming for me. I would love to explore them on my own but for the sake of this study, I try my best to only touch what I was required to do.*". Some users experienced lags in music playback at time on Calliope, with one user being unable to use it due to using a low-end computer device.

	MMM-C v2	**Calliope**
Early Novices	6.5 ± 2.03	6.65 ± 2.49
Skilled Novices	5.87 ± 3.09	5.77 ± 2.09

Table 2: Ease of Control across user profiles and systems

5.2.2 Controllability. Controllability scores (Tables 2 and 3) are measured on two dimensions: *ease of control* and *would like more control*. MMM-C v2's ease of control is *M* = 6.13 / 10 (*SD* = 2.70), and the would like more control is *M* = 7.72 / 10 (*SD* = 2.48). Calliope's ease of control is *M* = 6.46 / 10 (*SD* = 2.42), and the would like more control is *M* = 7.76 / 10 (*SD* = 2.24). These results indicate that novice users find both systems reasonably controllable, yet with room for improvement as indicated by the overall would like more control scores. The findings from qualitative data offer further details. With respect to ease of control, users often report MMM-C v2's interface to offer *sufficient* controls for music creation (10/29 codes) while Calliope is said to be *easy to control* (14/28 codes) and *comprehensive* in its feature set. Despite this ease, users acknowledge that more precise editing and the ability for greater specificity of requests are needed. Skilled novice users such as P35 (Experienced, Calliope) report on controllability : "*It was easy to control the system based on rhythm and density ... It would also be nice to select smaller sections than bars*". P25 (Experienced, MMM-C v2) says: "*Lots of flexibility, but lacks depth of control*". With respect to the need for more control (*would like more control*), users were clear in their request for additional musical parameters. P18 (Beginner, MMM-C v2) says "*I would not want any more parameters on the main window because that would decrease user-friendliness. I would however like an optional panel of advanced parameters*". The following suggestions for advanced parameters were extracted from the qualitative data: key, time signature, tempo, music genre mixing, the option to specify between a melody, sub-melody or a riff (P16, Beginner, MMM-C v2), control for the exact level of polyphony (as opposed to polyphony range), the pitch range, the amount of silence and text-based prompting. There were also requests for controlling musical structures and patterns in a way which reflects user's need for more nuanced generative requests (*depth of control*). P20 (Experienced, MMM-C v2) says: "*I would like to more explicitly control the chords and the rhythmic nature of the music (selecting note durations is something, but not enough; I care about rhythmic patterns, not note durations per se)*", then adds: "*I also felt like I couldn't control the formal structure, even in this simple 8-measure context*".

5.2.3 Creativity Support Index. CSI factors give us a glimpse into the participant's creative experience and the ability of the tools based on their design (e.g., parameters, workflows, deployment style) to support user's creativity for composition tasks. Across the five constructs , novice users find MMM-C v2 to be

	MMM-C v2	Calliope
Early Novices	6.13 ± 2.58	7.74 ± 2.29
Skilled Novices	8.83 ± 1.72	7.85 ± 2.12

Table 3: Would Like More Control results across user profiles and systems

Questions	*MMM-C v2* Themes	*Calliope* Themes
Trust	idea starter, repeated generation	trusted composer, limited trust
Authorship	significant, minimal	partial, substantial, limited
Flexibility	flexible, more features	highly, moderate, not enough
Challenges	difficulty steering, repeated generation	poor UI, lack of control
Features	comprehensive, enjoyable, inconsistencies	efficient and fun, enjoyable, more features
Workflow	efficient, structured, software issues	efficient, structured, convenient
Benefits/ Concerns	idea starter, efficient process, musical diversity	idea starter, convenient, efficient, musical diversity
Future Use	idea starter, maybe, no	idea starter, maybe, no
Releasable	yes, +iteration, no	yes, +controls, no
Final Thoughts	enjoyable	enjoyable, UI improvements

Table 4: Themes per question for MMM-Cubase and Calliope

best at supporting enjoyment (M = 3.82 / 5, SD = 0.94) (vs the other constructs), while Calliope is found to best at supporting *immersion* (M = 3.97, SD = 0.87) and *enjoyment* (M = 3.97, SD = 0.83). Specifically, users find *Calliope to be, on average, more immersive than MMM-Cubase v2* ($t(96) = 4.45$, $\boldsymbol{p < .001}$, $d = 0.92$). Otherwise, both systems performed quite similarly:

- Both systems support creative *Exploration* for music composition (M = 3.56, SD = 0.99, MMM-C v2) (M = 3.86, SD = 0.8, Calliope).
- *Enjoyment* remains among the best performing factors across the both systems (M = 3.82, SD = 0.94, MMM-C v1) (M = 3.97, SD = 0.83, Calliope).
- *Expressiveness* remains among the least performing factors across both systems (M = 3.36, SD = 1.11, MMM-C v1) (M = 3.54, SD = 1.04, Calliope).
- *Results Worth Effort* (M = 3.74, SD = 1.02, MMM-C v2) (M = 3.69, SD = 0.97, Calliope) indicates that users are relatively content with the result of interacting with the systems.

5.3 User Experience

Table 4 shows a breakdown themes generated for each system per survey question. It offers an overview of the state of novices' user experience for MMM-based CAC systems such as MMM-C v2 and Calliope. Out of those, we extracted the key broad themes expressed as follows:

- **Idea starter**: Novice users report high trust in both systems as *competent brainstormers* (8/28 codes, MMM-C v2; 10/24 codes, Calliope). P20 (Experienced, MMM-C v2) says "*There were also some little riffs in the bass and piano that I liked and that I would not have come up with myself, so that was potentially a source of inspiration.*". This is similar to MMM-C v1 study results regarding creative blocks [36]. P18 (Enthusiast, Calliope) says "*I think it is easy to produce music with good quality as a person who doesn't know much about music so I trust the application is giving me good pieces but don't know how experts think about it.*".
- **Repeated generation**: The model behavior and generation seemed random at times for many users. More experienced novice users found generation on both systems untrustworthy for precise control and inconsistent in outputs. This lack of *transparency*, which reduces trust in the system capabilities, also exacerbates the need for *depth of control* (15/29 codes, MMM-C v2) and for more *steering features* (7/28 codes, Calliope). P25 (Experienced, MMM-C v2) reports about challenges associated with the tools: "*The lack of depth of control. Because of this we end up having to rely on a certain amount of luck for the random result to meet our needs, this is not a big problem during the initial phase, the brain storm, but during the other stages of production, yes.*".
- **Minimal authorship**: For most novice users, the systems have more authorship as they mostly come up with the musical ideas that eventually direct the entire composition. P38 (Early Explorer, MMM-C v2) says: "*Very little because I didn't have skills to make the music all I did was move stuff around that the system created.*". P4 (Enthusiast, MMM-C v2) expressed: "*It feels like it is the generator who made everything because I didn't put much effort into making it (just pressing few buttons).*". Novice users experience having less input into the creative decisions apart from providing their preferences via the request parameters.
- **Efficient and structured workflow**: Novice users find the proposed way of composing music *novel* and *structured*. P16 (Beginner, Calliope) says about the benefits of using the system: "*The experience was very good, providing me with a brand new way to create music*". P39 (Early Explorer, MMM-C v2) comments "*I found that using the system helped me compose better. The system provided a wide array of tools that facilitated the creative process, such as generating chord progressions and suggesting harmonies. It also helped in structuring the composition more efficiently, allowing me to focus on refining the musical ideas rather than getting bogged down by technical details.*". Novice users acknowledge a need to receive guidance into the composition process and highlight the plugin's ability to support this. P3 (Experienced, MMM-C v2) says on the *Benefits* of the system: "*I don't consider myself expert in music theory neither a good composer, so this tool certainly helps me creating better and maybe more interesting melodies, and maybe even learn how to create them.*".
- **Musical diversity**: Learnability is an important concept in user experience and creative engagement. It also impacts the support of creative skill development and task proficiency. 30% (MMM-C v2) and 16.95% (Calliope) of users also expressed a sense of *diminished creative participation*. P23 (Early Explorer, Calliope) states "*... after*

	MMM-C v2	Calliope
Perceived Ease of Use (PEU)	3.64 ± 0.84	3.91 ± 0.62
Perceived Usefulness (PU)	3.59 ± 1.07	3.75 ± 0.81
Future Use (FU)	4.59 ± 1.99	4.46 ± 1.99

Table 5: Overall TAM Scores across systems

using Calliope I feel I've learned or done very little composing.". There are also reported concerns that users will become over-reliant on CAC tools and thus themselves loose their creative capacities, which will reduce overall musical diversity. P25 (Early Explorer, Calliope) shares "*I'm concerned that relying too much on Calliope might lead to a loss of creativity, as I could become overly dependent on the AI, potentially making me lazy and hindering my ability to think creatively.*". Finally, due to issues with model behavioral consistency and opacity around training data, users reported hesitation, confusion and legal concerns when it came to assessing attribution and copyright implications of their composed music pieces.

5.4 Acceptance

Overall TAM scores (Table 5) are above average. 30 out of 39 participants (76.9%) reported that they will use MMM-C v2 in the future (20.5% no, 2.6% maybe) for daily music, quick use, ideation. Similarly, 43 out of 59 participants (72.9%) reported that they will use Calliope in the future (20.3% no, 6.8% maybe) for inspiration, leisure, background music, video game or film music, or as a extra tool in the music setup. 23 MMM-C v2 participants (59%) believe the system can be realeased as is (33.3% no, 7.7% maybe). However, only 26 Calliope participants (44.1%) agree that the system is releasable as is (42.4% no, 13.6% maybe).

6 Discussion

Usability results, on par with the MMM-C v1 study (M = 73.6), indicate that both systems are easy to use and to operate (RQ1). Calliope being more usable than MMM-C v2 for early novices is likely due to that it is built for direct access from the web browser, and proposes a similar, yet simplified focused interface than traditional DAWs (i.e., Pro Tools, Logic Pro) with significant learning curve. The study also shows that MMM-based CAC tools lower barriers for novices, particularly by bypassing skill requirements for music composition tasks. P28 (Enthusiast, Calliope) says as *Final Thoughts*: "*I always wanted to produce my own music. However this is impossible without a decent amount of knowledge in music theory. Calliope helps me achieve my goal easily.*".

With respect to user experience (RQ2), although the system increases efficiency, this makes the user more dependent on system output, causing minimal authorship. Reported minimal authorship can globally be explained by limitations in user agency imposed by the interface and model behavior, coupled with the user's perceived contribution to the task and sense of creative engagement. P23 (Early Explorer, Calliope) says "*but I feel as though that act of creation is being robbed from me because I really didn't do anything more than click a few buttons.*". The frustrating interactions induce observed negative attitudes towards AI and reported concerns that do not motivate technology adoption. In contrast, in a previous study of MMM-C v1 with expert composers [36], system adoption is high and participants (expert users) report a sense of authorship, though no report that the workflow is more efficient. Otherwise, the theme of "idea starter" is consistently shared across studies.

Shortcomings observed on controllability, authorship and trust (RQ2) prevent *Human Flourishing*, a key theme of human-centered responsible AI [35]. It refers to the sustainable growth and well-being of users which must be considered when designing interactive systems including in the context of CAC tools, with creative engagement, thus creativity-support needs. User frustrations resulting from the lack of transparency in AI behavior, can result in lower user trust, limiting the user's expressivity [35] and jeopardizing the ability to make effective creative decisions. For example, P35 (Early Explorer, MMM-C v2) shares: "*There were instances where the system's suggestions felt overpowering, making it difficult to assert my own artistic direction.*". Literature [26, 29] shows that active co-creation nurtures self-efficacy. Similarly, we argue that over-reliance on AI without creative engagement and active learning (RQ2), can hinder the skill learning and competence development necessary for user creative growth and satisfaction. This is consistent with Gerlich's study [19] which shows that "*Higher AI tool usage is associated with reduced critical thinking skills*". We should thus aim to design creative processes and enable interaction affordances based on task division [24, 31] in a way that preserve user's creative self-expression (or *expressive capacity*), and open up creative avenues in human-AI collaboration. Overall, acceptance results (RQ3) indicate that although MMM-based CAC systems manage to be useful for composing music, there remains challenges such as depth of control, model transparency and unresolved societal concerns which impact user's perceived sense of future use. comment on overall tendencies independent of the interface differences.

7 Limitations

Our study focuses on novice users where most are design students located in North America, which is not a balanced representation of the true population. Although we evaluate the level of adoption of MMM-based CAC systems, the combination of usability, user experience and acceptance does not give a single metrical answer to the question. We recognized existing methodological limitations in evaluating CAC systems. This is also why we do not use the CSI score and instead compute its individual factors. Lastly, our study only summarily considers ethical considerations.

8 Conclusion

We presented comprehensive evaluations of MMM-C v2 and Calliope, two MMM-based CAC systems by novice users, which help contextualize the state of progress of AI and design of CAC tools. We conducted studies measuring and collecting novices' user feedback regarding their perceived *experience*, *usability* and technology *acceptance*. The results demonstrate the promises of co-creative interfaces to 1) *increase efficiency* of music composition processes and 2) *lower barriers of entry* for novice users to express musical ideas. Both systems show to be reliable brainstormers, satisfactorily supporting divergent thinking, and offering more efficient creative processes, despite ongoing issues regarding depth of control, model transparency, and user concerns impacting loss of creative diversity and human self-expression. Nevertheless, users consistently report the feeling of enjoyable experience and refreshed interest in music-making. As capabilities of creative AIs in synthesizing high-quality media content improve, we must rethink how we engage users with creative processes so as to retain human creativity in the loop. Despite their ubiquituousness,

prompt-based interfaces are a poor long-term design strategy for facilitating robust, scalable and effective interactions [30]. This implicates the new consideration for the human role beyond mere parameter setter and curator of generated outputs, and for expressive approaches for users to engage and interact with AI systems. Furthermore, differences in user creative profiles which affect expectations of quality and creative control means that further research is required to address key interaction design challenges for AI-based CAC systems.

Ethical Standards

This study follows standard protocols for human participant studies and was approved by the SFU's Research Ethics Board, Application ID #30000223.

Acknowledgments

We would like to acknowledge the Canadian funding bodies, the Social Sciences and Humanities Research Council (SSHRC) and Natural Sciences and Engineering Research Council of Canada (NSERC), for making this research possible. We also would like to thank Steinberg Media Technologies GmbH as our research partner for this study, and Alex Kitson for her valuable feedback.

References

[1] Omid Alemi, Jules Françoise, and Philippe Pasquier. 2017. GrooveNet: Real-time Music-driven Dance Movement Generation using Artificial Neural Networks. *Networks* 8, 17 (2017), 26.

[2] Omid Alemi and Philippe Pasquier. 2017. WalkNet: A Neural-Network-based Interactive Walking Controller. In *International Conference on Intelligent Virtual Agents*. Springer, 15–24.

[3] Gérard Assayag. 1998. Computer Assisted Composition Today. In *First Symposium on Music and Computers*. Corfu Greece: Ionian University.

[4] Carol M Barnum. 2020. *Usability Testing Essentials: Ready, Set... Test!* Morgan Kaufmann.

[5] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. 2000. Image Inpainting. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. 417–424.

[6] Renaud Bougueng Tchemeube and Philippe Pasquier. 2026. Survey and Typology of Computer-Assisted Composition Systems. *ACM Computing Surveys (CSUR)* (2026).

[7] Liam Bray, Oliver Bown, and Benjamin Carey. 2017. How Can We Deal With The Design Principle Of Visibility In Highly Encapsulated Computationally Creative Systems?. In *Eighth International Conference on Computational Creativity, ICCC, Atlanta*. 65–71.

[8] John Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. *Usability Evaluation in Industry* 189, 3 (1996).

[9] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.

[10] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 329–340. https://doi.org/10.1145/3172944.3172983

[11] John W Creswell, Vicki L Plano Clark, Michelle L Gutmann, and William E Hanson. 2003. *Advanced Mixed Methods Research Designs*. Sage, 209–240.

[12] Fred D Davis. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS quarterly* (1989), 319–340.

[13] Fred D Davis, Andrina Granić, and Nikola Marangunić. 2024. *The Technology Acceptance Model: 30 Years of TAM*. Springer.

[14] Fred D Davis and Viswanath Venkatesh. 1996. A Critical Assessment of Potential Measurement Biases in the Technology Acceptance Model: Three Experiments. *International journal of human-computer studies* 45, 1, 19–45. https://doi.org/10.1006/ijhc.1996.0040

[15] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 196–207. https://doi.org/10.1145/2856767.2856795

[16] Omar Elharrouss, Noor Almaadeed, Somaya Al-Maadeed, and Younes Akbari. 2020. Image Inpainting: A Review. *Neural Processing Letters* 51 (2020), 2007–2028.

[17] Jeff Ens and Philippe Pasquier. 2020. Flexible Generation with the Multi-track Music Machine. In *the 21st International Society for Music Information Retrieval Conference, ISMIR*.

[18] Jeff Ens and Philippe Pasquier. 2020. MMM: Exploring Conditional Multi-track Music Generation with the Transformer. *arXiv preprint:2008.06048* (2020).

[19] Michael Gerlich. 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies* 15, 1 (2025), 6.

[20] Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O Riedl. 2019. Friend, Collaborator, Student, Manager: How Design of an AI-Driven Game Level Editor Affects Creators. In *CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300854

[21] Cheng-Zhi Anna Huang, Curtis Hawthorne, Adam Roberts, Monica Dinculescu, James Wexler, Leon Hong, and Jacob Howcroft. 2019. The Bach Doodle: Approachable Music Composition with Machine Learning at Scale. *arXiv preprint:1907.06637* (2019).

[22] Shulei Ji, Xinyu Yang, and Jing Luo. 2023. A Survey on Deep Learning for Symbolic Music Generation: Representations, Algorithms, Evaluations, and Challenges. *Comput. Surveys* 56, 1 (2023), 1–39.

[23] Helene Joffe. 2011. Thematic Analysis. *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners* (2011), 209–223.

[24] Anna Kantosalo and Anna Jordanous. 2021. Role-based Perceptions of Computer Participants in Human-Computer Co-Creativity. AISB.

[25] Joshua Kranabetter, Craig Carpenter, Renaud Bougueng Tchemeube, Philippe Pasquier, and Miles Thorogood. 2022. Audio Metaphor 2.0: An Improved Classification and Segmentation Pipeline for Generative Sound Design Systems. In *Sound and Music Computing Conference (SMC)*.

[26] Stephen James Krol, Maria Teresa Llano, and Jon McCormack. 2025. Supporting Creative Ownership through Deep Learning-Based Music Variation. *arXiv preprint arXiv:2509.25834* (2025).

[27] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 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*. 1–13.

[28] Ryan Louie, Jesse Engel, and Cheng-Zhi Anna Huang. 2022. Expressive Communication: Evaluating Developments in Generative Models and Steering Interfaces for Music Creation. In *27th International Conference on Intelligent User Interfaces (IUI)*. 405–417.

[29] Jack McGuire, David De Cremer, and Tim Van de Cruys. 2024. Establishing the Importance of Co-Creation and Self-Efficacy in Creative Collaboration with Artificial Intelligence. *Scientific Reports* 14, 1 (2024), 18525.

[30] Meredith Ringel Morris. 2024. Prompting Considered Harmful. *Commun. ACM* 67, 12 (2024), 28–30.

[31] Srishti Palani and Gonzalo Ramos. 2024. Evolving Roles and Workflows of Creative Practitioners in the Age of Generative AI. In *Proceedings of the 16th Conference on Creativity & Cognition*. 170–184.

[32] Philippe Pasquier, Jeff Ens, Nathan Fradet, Paul Triana, Davide Rizzotti, Jean-Baptiste Rolland, and Maryam Safi. 2025. MIDI-GPT: A Controllable Generative Model for Computer-Assisted Multitrack Music Composition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 1474–1482.

[33] Adam Roberts, Jesse Engel, Yotam Mann, Jon Gillick, Claire Kayacik, Signe Nørly, Monica Dinculescu, Carey Radebaugh, Curtis Hawthorne, and Douglas Eck. 2019. Magenta Studio: Augmenting Creativity with Deep Learning in Ableton Live. In *7th International Workshop on Musical Metacreation (MuMe 2019)*.

[34] Bob L Sturm, Oded Ben-Tal, Úna Monaghan, Nick Collins, Dorien Herremans, Elaine Chew, Gaëtan Hadjeres, Emmanuel Deruty, and François Pachet. 2019. Machine Learning Research that Matters for Music Creation: A Case Study. *Journal of New Music Research* 48, 1 (2019), 36–55.

[35] Mohammad Tahaei, Marios Constantinides, and Daniele Quercia. 2023. Toward Human-Centered Responsible Artificial Intelligence: a Review of CHI Research and Industry Toolkits. *arXiv preprint arXiv:2302.05284* (2023).

[36] Renaud Bougueng Tchemeube, Jeffrey Ens, Cale Plut, Philippe Pasquier, Maryam Safi, Yvan Grabit, and Jean-Baptiste Rolland. 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 (IJCAI)*. 5769–5778.

[37] Renaud Bougueng Tchemeube, Jeffrey John Ens, and Philippe Pasquier. 2022. Calliope: A Co-creative Interface for Multi-Track Music Generation. In *C&C '22: Creativity and Cognition, Venice, Italy, June 20 - 23, 2022*. ACM, 608–611. https://doi.org/10.1145/3527927.3535200

[38] Miles Thorogood and Philippe Pasquier. 2013. Computationally Created Soundscapes with Audio Metaphor. In *Fourth International Conference on Computational Creativity (ICCC)*. 1–7.

[39] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2018. Generative Image Inpainting with Contextual Attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5505–5514.