Balancing Musical Co-Creativity: The Case Study of Mixboard, a Mashup Application for Novices

Ottolin, Thomas, Sankaranarayanan, Raghavasimhan, Lei, Qinying, Hugar, Nitin, and Weinberg, Gil *

[1] Georgia Institute of Technology
tottolin3@gatech.edu
[2] violinsimma@gmail.com
[3] qlei33@gatech.edu
[4] nitin.hugar@gatech.edu
[5] gilw@gatech.edu

Abstract. Recent developments in generative AI have posed a challenge for developers who attempt to maintain an effective balance between the system's generative input and user's sense of creativity and control. In this paper, we present a longitudinal study of a web/mobile application we developed, Mixboard, which allows novice music lovers to create and share personalized musical mashups in a co-creative manner. Different balances between the role of system automation and user creative input have been developed and studied over a period of two years. Findings from users studies indicate that while novices appreciate the system's AI driven automation and suggestion, they seek to expand their level of control and creative input into the final product over time. Future developments may therefore include a personalized level of control balance based on continuous assessment of user behaviour.

Keywords: Co-creativity, Musical AI, Longitudinal User Studies, Mobile Applications, Novices

1 Introduction

Systems that use Artificial Intelligence (AI) to aid in creative processes have recently increased in popularity, partly driven by OpenAI's suite of Generative Pre-Trained Transformer (GPT) releases starting in 2018 [13]. One of the main challenges facing developers of co-creative systems is how to provide automation and content in a manner that would maintain a sense of creative control and agency for the user. User satisfaction may be negatively impacted if the system prompts the user to contribute too much or too little to the creative outcome. This also comes at a time where music production and

* Thanks to Hardik Goel for his developmental work.

consumption is, or at least appears to be to novices, more widely accessible. Popular social media, like TikTok or Instagram, empower users to select and edit music and sounds to go along with their planned content. As Jenkins et al. describe, these technologies and new forms of consumption ”signals a movement toward a more participatory model of culture one which sees the public not as simply consumers of pre-constructed messages, but as people who are shaping, sharing, reframing and remixing media content in ways which might not have been previously imagined” [8].

We developed Mixboard [15] to allow music lovers to ”shape and remix” any set of songs into high-quality musical mashups, assisted by AI. A mashup, in this context, can be defined as a blend of elements from 2+ songs. Aimed at novices, the application acts as a co-creative agent that contributes to the musical decision making, rather than giving the user full control over the final outcome. The AI handles both low-level computational tasks such as source separation, segmentation, tempo and key detection, stretching, and transposition, as well as high-level artistic decisions such as selecting appropriate musical segments and suggesting compositional structures. A previous set of comprehensive user studies with the app identified a clear desire for further user control. This motivated our team to rewrite the system’s software infrastructure to provide a more effective balance between user control and system automation. In this paper, we provide a short summary of the original system, describe the new features developed to address the control balance, and present newly conducted research studies that indicate a higher level of user satisfaction and productivity while working with the app.

2 Related Works

Recent generative audio systems rely on artificial intelligence and machine learning for creation and manipulation of sound data. Certain products depend on Digital Audio Workstations (DAW), such as Avid Pro Tools [11] or iZoTope’s mixing product suite [7] to support professional musicians who are familiar with advanced musical concepts. Such systems require layered knowledge and experience with waveform editing, rendering the musical outcome to be fully dependent on the user’s abilities and talent. Conversely, applications designed for novices such as Splash Music [6], Amper [2], OpenAI’s MuseNet [14] or Jukebox [5], allow little creative input for the users in constructing the musical outcome. With these kinds of systems, the user only provides high level input such as mood, length, or style, while the AI generates the music without supporting ongoing creative input for the users. Santo et al. [16] identified that users would like a co-creative to provide some control over the output. As Tanaka et al. [17] found with their co-creative musical systems, ”The ability of the listener to distinguish his own contribution within the total resulting music is a crucial element in granting musical agency to individual users.”

For mashup applications, too, recent efforts tend to simplify the interaction design, which limits the creative expression and control of the user. MixMash [10], for example, presents users with a song proximity map but does not provide an interface for users to creatively generate full songs. Other systems such as AutoMashUpper [3] and PopMash [19] pose creative constraints, whether it is limiting the songs a user can work with or limiting the user’s creative potential by providing a overly technical user inter-

face. These systems also do not allow users to choose any song of their liking, which limits personalization and engagement. DropMix [12], on the other hand, does provide commercial songs for users to mashup. However, DropMix's song library is limited and the system does not allow the user to engage creatively in constructing the final product. Mixboard was designed to address these challenges, providing users with ongoing AI-driven creative input during the construction of their songs.

3 Web Application Overview

The first implementation of Mixboard was designed for the Web [15]. The application allows users to select any four songs from Spotify and organize them over a visual canvas. The users can drag song album art onto the canvas, positioning them over four lanes: Vocals, Instruments, Bass, and Drums. These stems have been source separated using Demucs [4]. Users can then edit the length and location of each segment by dragging and dropping segments over the canvas. The system selects the optimal key and tempo for the mashup, and stretches and transposes all songs segments to the optimal tempo and key using Elastique by Zplane [20]. It also makes high level creative suggestions such as providing templates for songs and selecting the particular audio segments for each placed segment on the canvas. In a set of comprehensive user studies [15], we found that the majority of users asked for more control over their creations. Additionally, we found that while users may have started their mashup process by leaning on the AI-powered features to select random songs with (*Choose for Me*) or determine the placement of their songs with (*Surprise Me*), no user exclusively used the AI features; this indicates that even novice users were capable and willing to explore more nuanced AI-powered features, but they still wanted to exert their own creative goals themselves.

4 iOS Application Overview

To address our initial evaluation findings, we developed a new iOS version of the app. The iOS app interface can be seen in Figure 1. A video demonstration of the application can be viewed here: `http://bit.ly/mixboard`. Three main features were added to the application in an effort to provide more control to the users, while still providing meaningful AI input. To allow users to better decipher between the components of the mashup, *Mute* and *Solo* functionalities were added to each lane. To provide users with more control over which audio segment is chosen by the AI for each section, we added a *Shuffle* function:

Mute: Turning on *Mute* for a lane will silence corresponding sound pulled from songs placed in that lane. This lets the user silence lanes while listening to live playback, enabling the user to zero in on sounds they want to highlight or remove.

Solo: Turning on *Solo* will only play sound generated from that corresponding lane.

Shuffle: After clicking a placed segment, a *Shuffle* button appears next to the *Delete* button. When *Shuffle* is clicked, the system will sort through all available relevant segments to pull another segment that matches the length of the placed segment. Given the high volume of requests for users to select specific segments from songs, the *Shuffle*

function is designed to grant the user more control choosing the segment, while still allowing the AI to make an informed decision on which audio segment would fit well.

In addition to these changes, the iOS version also prompts users to log in with their Spotify accounts, which leads to their most recently played songs to display in the *Spotify* window of the song selection window. This allows for easier and faster personalization, which was requested by many users. The iOS version also removed the *Generate* button, and replaced it with *Play/Pause* to reduce wait time in listening to a mashup.

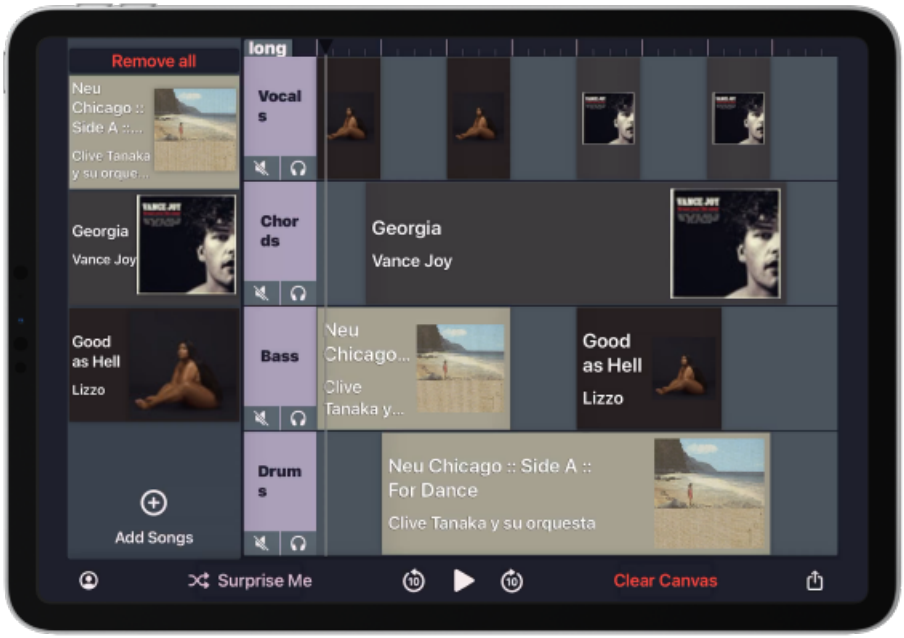

Fig. 1. The iOS version of Mixboard, rendered on an iPad 2 in Dark mode

5 Evaluation

We conducted one study on the iOS version of Mixboard using a tablet to maximize screen size. We recruited 20 participants, 16 of which participated in research of the web version. The participants ranged from 22 - 27 years of age, and no one held more than a year of professional or recreational music mixing experience. Participants were given up to 30 minutes to interact with the system; the audio and visual content of the device was recorded throughout the experiment. After the experimentation phase ended, subjects participated in a semi-structured interview and survey. The interview included questions that explicitly asked about how the participant liked and used the

three new control-granting features. The survey asked users about their experience using 20 Likert-scale questions, some of which were adapted from previous musical AI experiments [9]. Two survey questions asked participants to rank potential features in terms of how interested they were in trying the feature and how effective they perceived the features could be in helping them create better mashups. The 27 features included in this section all came from previous participants' desires or misconceptions of Mixboard; these ideas both further expanded existing functionality, e.g. lane labels to set expecations on what to hear, and generated functionality, e.g. a song recommendation system based on the selected songs' tempo or key. The survey data was aggregated to generalize findings quantitatively by assessing the measures of central tendency of this study against previous studies conducted.

6 Results

Results from the 20 Likert-scale measures are shown in Figure 2. Two major system bugs were identified during research, one of which broke *Shuffle* and the other frequently broke the *Play/Pause* button. The team was able to identify these issues and fix them after the 8th study. As such, survey means were calculated across all studies (labeled as "Study 3 Mean"), as well as specifically for participants 9-20 (labeled as "Study 3 Post-Fix Mean"). ANOVA tests were conducted on all 20 measures across these three groups, and each measure was proven to be statistically signifcant between groups. After these fixes, the iOS version of Mixboard proved to be more **consistent** ($mean(\mu) = 1.62$ *(decrease of 0.63 from previous research)*, *standard deviation* $(\sigma) = 0.8$), **well-integrated** ($\mu = 4.31, +.30, \sigma = 0.74$), and **easier to use** ($\mu = 4.69, +0.27, \sigma = 0.5$) than the web version. The iOS version also scored better in the **control** ($\mu = 3.69, +0.60, \sigma = 1.15$) and **need for more learning** ($\mu = 2.23, -0.35, \sigma = 1.02$) measures than the previous version of the system. Interestingly, the average for **automation** ($\mu = 2.85, +0.47, \sigma = 1.18$) moved closer to 3, meaning more participants "neither agreed nor disagreed" with the statement, "The system should automate more of the composition process for me." There was minimal difference in the **creative expression**, **trust**, **learnability**, and **user confidence** measures, which demonstrated that the new version's changes were not noticeably detrimental to the well-favored user experience of the system.

Technical errors impacted 9 screen recordings. The team decided to only analyze recordings that captured the full experimentation period, so 11 screen recordings were analyzed. *Mute* was the most commonly used feature, with only 2 of the 11 users observed choosing not to interact with the feature at least once during production. It is possible that returning participants were more drawn to interact with the feature given its newness; some returning participants requested this feature previously, which could have further motivated its use. In the features aspect of the survey, only 7 features received strictly positive remarks, meaning no participant stated they were "Not interested" in trying the feature, and no one believed the feature worsen the experience. Each of these features would grant the user more control and improve the quality of the final product. All 20 participants chose to use the full 30 minutes to experiment with the system, and each participant stated they would want more time with the system.

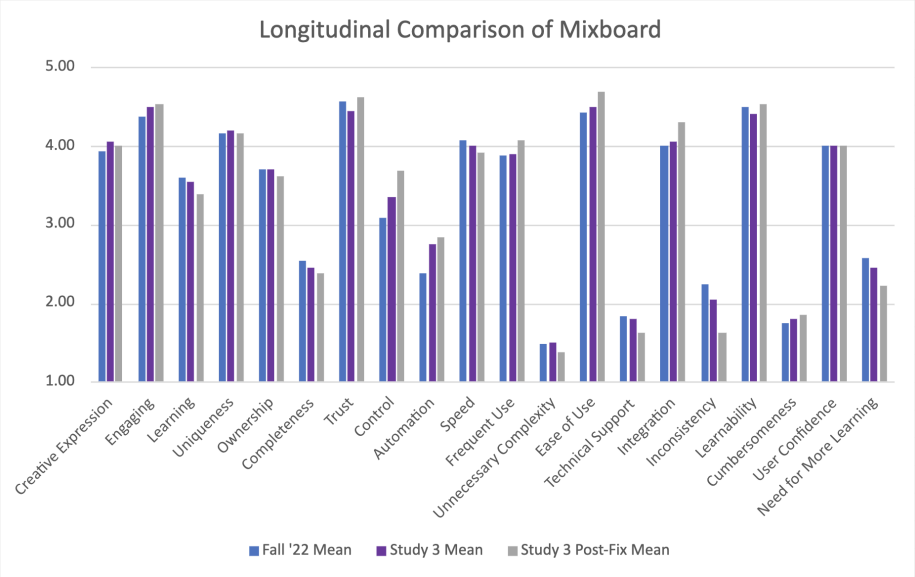

Fig. 2. Longitudinal Comparison of Mixboard

7 Observation and Discussion

7.1 Version Comparison

We asked participants "Do you think the system provides you too much, too little, or just enough control over your creations?" 8 participants stated "Just enough", and another 10 qualified their "Just enough" responses by saying they would want more control as time went on. This data paired with the improved **control** measure score indicate that this version achieves a more desirable balance between automation and user control; however, there is more to be desired. One returning participant shared, "I like that I can now change it (using *Shuffle*), but I think it'd be more of a unique experience if I could choose which part of the song. Whenever I think of a song I want, I have a specific part I want to add, not just the entire song" (P17.7). Furthermore, participants generally used *Choose for Me* and *Surprise Me*, two AI-driven features, less frequently than in the web version, which could be due to returning participants having a clearer vision of what they want to create or because their pre-existing system knowledge meant they had less to explore. Nearly all returning participants supported the transition from web to tablet, yet 7 of the 16 participants stated they felt less precise without using a mouse or bigger screen. One of these 7 participants stated she felt she had less control over her mashups in this version compared to the web, making her the only returning participant who said they lost control in a negative sense. Participants often stated *Solo* and *Mute* made it easier to identify sounds they wanted to accentuate or eliminate, which granted more control. *Shuffle* likely should remain, even if more data should be gathered around the feature when it works.

7.2 Achieving Long-Term Co-Creativity

The juxtaposition captured by our results demonstrates how difficult it is to provide a universal co-creativity balance that would be appreciated by all users in longitudinal studies. One participant stated throughout the study that his expectations had changed due to capabilities and limitations experienced in the previous study, "I can't necessarily choose the exact seconds of a track, so knowing that means I have to be very open with the vision going into this...if the system's already going to choose the parts of the track for me, then I feel like trying to put specific tracks down is in conflict with that." This reflects gradual user trust can also prepare the user for more advanced features. 13 of the 20 participants requested Mixboard's AI to expand to influence their work further; participants most commonly requested suggesting songs or placement based on what they already had selected (10 participant requests) and more information about the AI's decision making process (8 participant requests). It is worth noting that the latter request did not emerge in the first iteration of research studies, again showing how user expectations can evolve. Users were more likely to request these advanced features when they had previously participated in our studies, reaffirming Turchet et al.'s [18]finding that "personalization mechanisms (should be) based on the expertise level of the user." More control could allow these users to evolve their abilities over time, which increases the likelihood of creating works they are happy to claim as their own.

7.3 Ethical Standards

This project was developed by Georgia Tech students for academic purposes. The human subjects research was approved by the Georgia Tech Institutional Review Board. Informed consent was collected verbally and in writing at the beginning of each research study. No compensation was offered to participants. Anonymized data was stored in a secure drive only accessible to the researchers included on the IRB protocol.

The ethics of remixing and redistributing musical works will be addressed in future work of this system. Since Mixboard is not publicly available, there is minimal risk regarding copyright infringement or improper compensation for the artists. Mixboard could greatly increase the number of works that could be used to generate revenue outside of proper royalty structures, namely if a user were to use a mashup on sponsored social media content or to sell to other social media users. Furthermore, since participating in streaming music slightly increases the likelihood to participate with music piracy [1], we must be especially careful that our users understand the consequences of illegal usage of copyrighted music.

8 Conclusions and Future Work

Mixboard will continue to evolve to address the wealth of user feedback we have collected. The team plans to evaluate whether the system should intentionally scaffold learning via unlockable features or advanced tutorials. While it is clear that different users will have different expectations and different preferences, we will explore variation that would personalize the level of control based on assessing users interaction with the system.

References

1. Karla Borja and Suzanne Dieringer. Streaming or stealing? the complementary features between music streaming and music piracy. *Journal of Retailing and Consumer Services*, 32:86–95, 2016.
2. Wayne Cheng. Amper: Custom music in seconds, April 2021.
3. Matthew E. P. Davies, Philippe Hamel, Kazuyoshi Yoshii, and Masataka Goto. AutoMashUpper: Automatic Creation of Multi-Song Music Mashups. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12):1726–1737, December 2014. Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.
4. Alexandre Défossez. Hybrid spectrogram and waveform source separation. In *Proceedings of the ISMIR 2021 Workshop on Music Source Separation*, 2021.
5. Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A Generative Model for Music, April 2020. arXiv:2005.00341 [cs, eess, stat].
6. Stuart Dredge. What's the real end-game for ai music? popgun's ceo has ideas..., December 2019.
7. iZoTope team. Izotope mixing plug-ins.
8. HENRY JENKINS, SAM FORD, and JOSHUA GREEN. *Spreadable Media: Creating Value and Meaning in a Networked Culture*. NYU Press, 2013.
9. Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J. Cai. 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, Honolulu HI USA, April 2020. ACM.
10. Catarina Maçãs, Ana Rodrigues, Gilberto Bernardes, and Penousal Machado. Mixmash: a visualisation system for musical mashup creation. In *2018 22nd International Conference Information Visualisation (IV)*, pages 471–477. IEEE, 2018.
11. Mark Marrington et al. Composing with the digital audio workstation. *The singer-songwriter handbook*, pages 77–89, 2017.
12. Nick Mudry and Riley Davis. Harmonix music systems: Dropmix, April 2017.
13. OpenAI. Introducing ChatGPT, November 2022.
14. Christine Payne. MuseNet, April 2019.
15. Raghavasimhan Sankaranarayanan, Nitin Hugar, Qinying Lei, Thomas Ottolin, Hardik Goel, and Gil Weinberg. Mixboard - a co-creative mashup application for novices. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2023.
16. Luıs Espırito Santo, André C Santos, and Marcio Lima Inácio. Focusing on artists' needs: Using a cultural probe for artist-centred creative software development.
17. Atau Tanaka, Nao Tokui, and Ali Momeni. Facilitating collective musical creativity. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA '05, page 191–198, New York, NY, USA, 2005. Association for Computing Machinery.
18. Luca Turchet, Carlo Zanotto, and Johan Pauwels. "give me happy pop songs in c major and with a fast tempo": A vocal assistant for content-based queries to online music repositories. *Int. J. Hum.-Comput. Stud.*, 173(C), may 2023.
19. Baixi Xing, Xiang Zhang, Kejun Zhang, Xinda Wu, Hui Zhang, Jun Zheng, Lekai Zhang, and Shouqian Sun. Popmash: an automatic musical-mashup system using computation of musical and lyrical agreement for transitions. *Multimedia Tools and Applications*, 79(29):21841–21871, 2020.
20. Elastique pro v3 by zplane. https://licensing.zplane.de/uploads/SDK/ELASTIQUE-PRO/V3/manual/elastique_pro_v3_sdk_documentation.pdf. Accessed: 2023-01-30.