Figure 1: Google AI. http://ai.google

Figure 2: Magenta is a project within Google Brain that focuses on training machines for creative expressivity, generative personalization, and intuitive interfaces. https://g.co/magenta

Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface

Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, Douglas Eck
Google, Inc.
Mountain View, CA , 94043 USA
(cbalgemann,sherol,noerly,jessh,adarob,deck)@google.com

ABSTRACT

Creative generative machine learning interfaces are stronger when multiple actors bearing different points of view actively contribute to them. User experience (UX) research and design involvement in the creation of machine learning (ML) models help ML research scientists to more effectively identify human needs that ML models will fulfill. The People and AI Research (PAIR) group within Google developed a novel program method in which UXers are embedded into an ML research group for three months to provide a human-centered perspective on the creation of ML models. The first full-time cohort of UXers were embedded in a team of ML research scientists focused on deep generative models to assist in music composition. Here, we discuss the structure and goals of the program, challenges we faced during execution, and insights gained as a result of the process. We offer practical suggestions

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3299059

for how to foster communication between UX and ML research teams and recommended UX design processes for building creative generative machine learning interfaces.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction**; • **Interaction paradigms**; • **Collaborative interaction**;

KEYWORDS

Human-computer interaction; human-computer collaboration; expressive tools; generative music; UX research; industry research

ACM Reference Format:
Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts,, Douglas Eck . 2019. Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/kayacik.chen.noerly.holbrook.roberts.eck

INTRODUCTION

While research scientists are the primary actors developing machine learning (ML) models and system architecture, the models and user interfaces (UIs) that result are stronger when multiple actors bearing different points of view actively contribute to them. Maintaining a user-centered approach throughout the product innovation process contributes to greater success in high-tech applications [1]. Perspectives from user experience (UX) researchers and designers can enhance and expedite the process by providing knowledge about end-user needs that shape project goals, training data and model outputs, UI interactions and visuals.

In order for UXers to be active contributors to ML projects, they must have a working understanding of ML concepts. Currently, many UXers do not have the ML education that is needed in the industry, and this lack of education is hampering ML research teams' capacity to have broad user impact with their projects. To meet these needs and form a symbiotic exchange in which UXers and ML research scientists learn about each other's industry, skill set, and processes, the People + AI Research (PAIR) group at Google developed the PAIR Bungee program.

The PAIR Bungee program is a novel program method that embeds three UXers (a researcher, an interaction designer, and a prototyper) into an ML research host team for three months. During that time, UXers receive training on basic ML concepts and provide UX input and direction to jump-start the host team's ML projects. At the end of three months the UXers have new ML knowledge and ML research scientists have a greater understanding of user-centered practices.

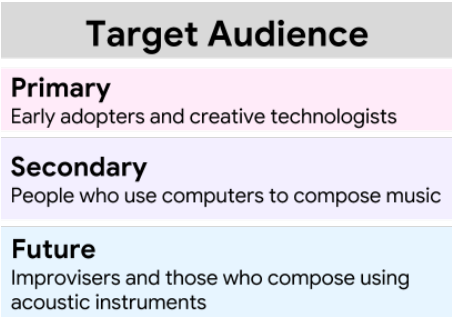

Figure 3: Early adopters and innovators are more forgiving of the functionality and design of a novel product, will not stop using a product because of minor system bugs [7], and will be able to provide concrete and actionable feedback.

Phase	Deliverable
1. Prepare	ML bootcamp for UXers Defining the target audience Defining the master timeline Design sprint
2. Develop	Testing the initial concepts Exploring concepts with users Interaction and visual design
3. Evaluate and conclude	Usability testing Graduation and next steps

Table 1: The PAIR Bungee method for developing ML driven expressive technologies over a 3 month period.

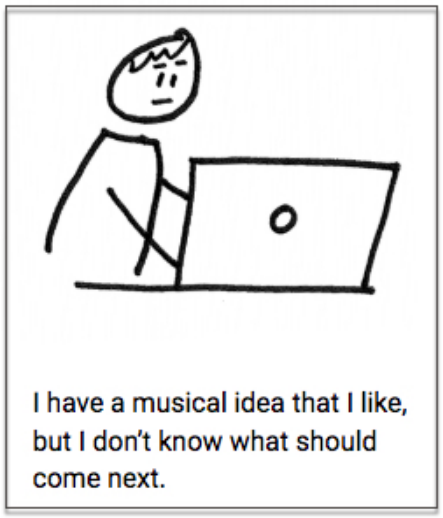

Figure 4: First frame from one of the 3-frame storyboards used to test concepts with users.

Figure 5: Concept mock developed by participatory design participants.

The first ML host team was the Magenta project team. Magenta is a research project of the Google Brain team exploring the application of deep learning as a tool to enhance creativity in music and the arts (g.co/magenta). The core team is composed of research scientists and software engineers with expertise in machine learning, but not UX. In its first two years, the project produced multiple novel generative models to aid in the creation of art and music [2–4, 10]. Nonetheless, the work did not exist in a form that was easily accessible to the target user community, limiting the team's ability to understand and catalyze the potential impact of these models. For example, early interfaces lacked graphic user interfaces (GUIs) and required users to interact via the command-line. In order for the team to realize its ultimate goal of the models being accessible to a broader spectrum of musicians that didn't have programming skills, Magenta needed to partner with UX researchers and designers to go beyond the command line interface. This case study describes the method of the Bungee program and details the experiences, challenges, and learnings of its first full-time cohort.

METHOD

In this section we outline the structure and process of the PAIR Bungee program; from preparation and development to final evaluation. Many of the program's components utilize well-known HCI and UX processes and therefore in-depth descriptions are reserved for more novel aspects of how these processes were applied in an ML + UX context.

1. Prepare

ML bootcamp for UXers: A bootcamp course was designed for UXers in the Bungee program to teach them basic concepts in ML and AI and then incorporate medium-to-high level information about the host team's current ML models and other ML work going on within Google.

Defining the target audience: It is tempting to say that a product is meant for everyone (and at Google, we are often interested in building products for billions of people). However, restricting and defining a target audience for the Bungee project expedited user research and design and made building a product in the short three-month timeframe more achievable. Prior to the start of the cohort, the project's Product Manager (PM) researched and identified the project's target audience as early adopters and innovators, those people who are ahead of the product adoption curve [8]. Figure 3 describes the primary, secondary, and future work targets.

Defining the master timeline: Mapping out a timeline is important for any project, but especially so when multiple teams are collaborating with a short three-month total timeframe. Beginning with the end, we started at the project completion date and worked backwards. Typical project management timelines were defined for user research, design, and prototyping efforts. Within the protoyping workstream, extra time was allotted for understanding the backend system and how it will work with ML model outputs.

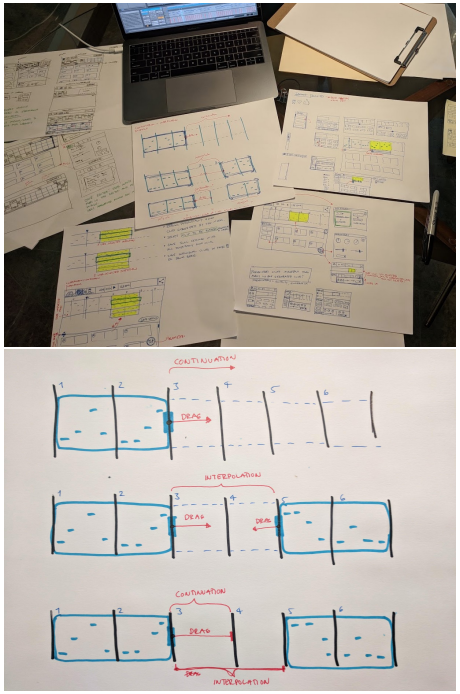

Figure 6: Design mocks for early user interfaces (above) and in-filling melodies (below).

Design sprint: We used a loose design sprint model [6] to spend time bonding in a creative setting and preparing for the rest of the project. One challenge of doing a design sprint very early in the Bungee program was that

we didn't have user research to offer user insights, mental models, and pain points, which is a typical component of a design sprint. To account for that perspective, we included three live interviews on the first day. The interviewees fit our target audience and were interviewed about their digital music composition process, delights, and pain points. Interview observers wrote *how might we* (HMW) questions, such as "HMW help a musician to overcome creative block?"

Following the user interviews, we grouped observations and HMW questions to identify patterns. The groupings resulted in nine product opportunity clusters. Then we paired up one research scientist with one UXer. Using craft supplies, each pair selected clusters to work with and created physical representations of their proposed product. Creating physical representations of ML concepts was a new experience for the research scientists. Although skeptical at first, they enjoyed the experience and they gained a new perspective on their work and the goals for the cohort's Bungee project.

The UXers reviewed the observations and HMW questions from the design sprint and chose seven concepts to pursue, based on our knowledge so far about user music composition processes, target users, and the team's general direction.

2. Develop

Testing the initial concepts: To further understand users' music composition processes and learn which concepts appealed and why, we conducted eight user interviews. Modeled after part of the Triptech method [12], we created 3-frame storyboards [7] to help participants understand our proposed concepts, see figure 4. Storyboards are particularly useful for describing ML-related concepts, as some participants may be shy or even averse to highly technical concepts. Distilling complex concepts into a simple storyboard is also a good exercise for team members to do together to get a common language and ensure that all disciplines understand the goals and complexities.

As an outcome of the interviews we narrowed the potential concepts from seven to three. These decisions were made based on participant level of interest in the concepts, how well the concepts might fit into existing music creation workflows, and whether there was an ML model already trained that would make the concept a reality.

Exploring concepts with users: We learned that faculty and students at the California Institute of the Arts (CalArts) were interested in the Magenta project and many fit the criteria of our target audience. As a result, we organized a two-hour participatory design session [11] with 24 members of the Music Technology program.

During the session, eight Google employees led small groups of participants through tasks that allowed them to warm up, ideate on specific prompts, and create a final product as a team. Each

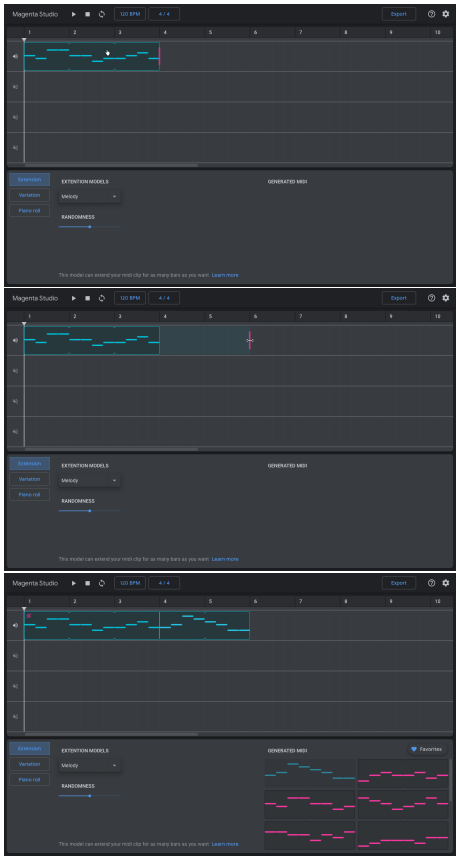

Figure 7: Example UI for a composition tool that extends musical lines. (top) The user starts with a melody they composed. (middle) The user extends the melody by a desired length. (bottom) The model generates the next part of the melody in the bottom right of the screen, like auto-completion for music.

group shared their product idea with the room, see figure 5. Based on the assets and insights from this study, the team chose two concepts that would make up the first version of the final product. Narrowing to two concepts was important to keep the initial scope of front-end development small.

In addition to providing rich qualitative data, the participatory design session was an opportunity for ML research scientists to interact with external users, some for the first time. This experience built empathy and a deeper understanding of user needs.

Interaction and visual design: Interaction and visual design ramped up quickly after the participatory design study (see figure 6). Five design considerations were critical when designing the UX of this generative ML product.

(1) User mental models: A user's mental model represents the way he or she thinks about a concept or system. We found it helpful to think about how people currently solve the problem without ML. That way, the users' mental model of the current non-ML solution might inform how to design the new ML solution and help the users adapt to it.
(2) Transparency: If the design is not transparent about how the system works, users might develop incorrect mental models, which may affect their trust in the system. With creative UIs, it's especially important to be transparent about how user data will be used so that users will feel comfortable using the product with their original material.
(3) Augmentation, not automation: Especially with ML products that produce creative outputs, it is undesirable to do everything for the user. Effort is part of the creative process and a feeling of accomplishment can only occur if a user is working alongside an ML system.
(4) Error handling: Sometimes ML models fail. Failure from a user perspective is even more likely with creative UIs because artistic taste varies. Thus, it is important to build controls into the UI that allow the user to correct failures.
(5) Feedback controls: Enabling users to give feedback when the model does not perform as expected is a powerful method to improve the models, generate personalized content, and increase user satisfaction.

3. Evaluate and conclude

Usability testing (with mocks and prototypes): As with any UI in-progress, we wanted to put designs in front of users to understand whether they were understandable and usable. The UX designer and prototyper created stimuli for participants to interact with (see figures 7 and 8). The UX researcher recruited participants within the target audience, moderated the sessions, and updated the moderator script as iterative design changes were made.

Graduation and next-steps: At the end of the three months, we were excited to share what we had accomplished. It was also important for executive stakeholders to see how we spent our time. A

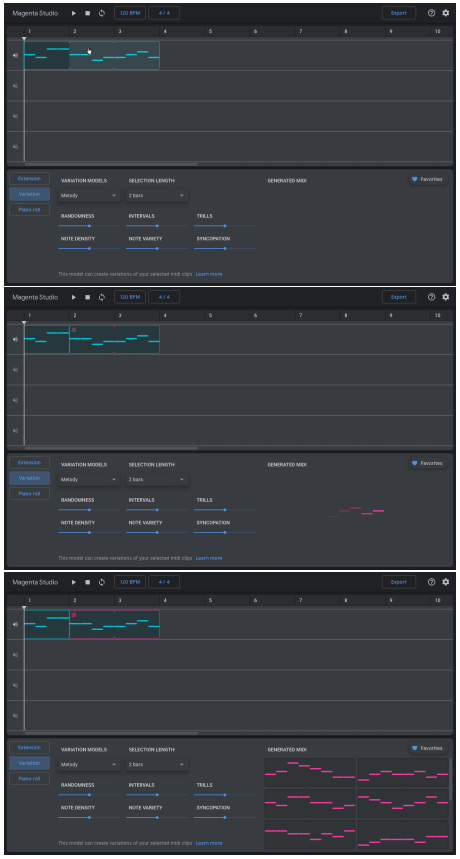

Figure 8: Example UI for a composition tool that generates variations. (top) A single melody line represented in the composition tool. (middle) The user selects a portion of this melody. (bottom) The model generates variations in the bottom right of the screen, based off of the selection.

one-hour Bungee program "graduation" meeting was scheduled in which we discussed the process and outcomes from our Bungee cohort. The host team gave feedback and corroborated our stories.

CHALLENGES

Here we outline the challenges we encountered while preparing for and during the project, including strategies for working through them.

The design sprint was one of the most productive ways we prepared for the project. However it can be challenging if ML research scientists are unfamiliar with design sprints and express skepticism over whether days spent away from current projects will be worthwhile. Host team participation is critical, however, because they have the deepest knowledge of current ML models, the most familiarity with challenges in the space, and relevant ideas for future ML models and product features. To convince skeptical host team members that their time will be well spent, openly discuss the benefits and expected outcomes of the design sprint. It may also help to draw parallels to hackathons, a similarly concentrated workshop format that they may be more familiar with.

The biggest challenge we faced during the Bungee program was a misunderstanding about the capabilities of one of the ML models. This is one of the most common issues that occur as teams are turning ML model outputs into a product. To prevent that kind of misunderstanding, time should be spent thoroughly reviewing and discussing existing and planned ML models' training data inputs, outputs, and capabilities. UXers should be able to accurately describe in their own words what models do. Surfacing design concepts early can also help facilitate discussion. As UXers describe design elements, they should describe step-by-step what they expect the model to be able to do at each point. This will help to smoke out any misconceptions.

OUTCOMES

The three concepts we evaluated during the participatory design study are defined in table 2. During that study, we learned that the first two concepts, Extension and Variation, were the most attractive to target users because of their ability to seamlessly fit within users' existing music composition processes while being useful throughout the composition process. Next we tied the concepts back to machine learning models. Extension could be achieved using musical language models such as PerformanceRNN [9] and Music Transformer [5]. Variation was achievable by applying Latent Constraints [2] to a MusicVAE [10]. In-Filling could be done with a CocoNet [4], although this was de-prioritized based on the participatory design study. (See table 3)

The Extension and Variation feedback led us to modify the existing ML models. For example, more user controls were added to the MusicVAE model such as adjusting syncopation and average interval, in addition to existing note density Latent Constraint transformations. In this way, we were able to

Action	Function
Extension	Given a short music segment, use it as a starting point for a longer piece
Variation	Given a complete music segment, generate some variations or transformations of it
In-filling	Given a short music segment, use it as a starting point for a longer piece

Table 2: The three concepts we evaluated during the participatory design study. We learned that the first two, Extension and Variation, were most attractive to target users.

Action	ML Model
Extension	Latent Constraints (Engel et.al) and MusicVAE (Roberts et.al)
Variation	Performance RNN (Simon et al) and Transformer (Huang et.al)
In-filling	CocoNet: Counter Point by Convolutional Neural Networks (Huang el.al)

Table 3: We tied the concepts back to machine learning models.

impact the ML work using UX research. Ideally we would have started UX research earlier, when the models were initially being created, to be more targeted with the model work.

Magenta Studio

Following the Bungee program, we built Magenta Studio (https://g.co/magenta/studio) a collection of models inspired by the user research and design work. Ableton Live (https:/www.ableton.com/en/live), a digital audio workstation (DAW) produced by Ableton, has an extensive user base and was very popular among study participants. Thus we decided to forge a partnership with Ableton to build a custom Magenta integration. We built four music tools using Magenta.js and packaged them individually as javascript apps using Electron (electronjs.org). The Electron apps worked stand-alone or in Ableton live via Max for Live (https://www.ableton.com/en/live/max-for-live/).

The four tools we built relied heavily on what we learned from our user research studies: **Continue** uses the predictive power of recurrent neural networks (RNN) to generate notes that are likely to follow your drum beat or melody. This can be helpful for adding variation or creating new musical material. **Generate** is similar to **Continue**, but it generates a 4 bar phrase with no input necessary. This can be helpful for breaking a creative block or as a source of inspiration for an original sample. **Interpolate** takes two drum beats or two melodies as inputs. It then generates up to 16 files which combine the qualities of these two files. It's useful for merging musical ideas, or creating a smooth morphing between them. **GrooVAE** adjusts the timing and velocity of an input drum pattern to produce the "feel" of a drummer's performance. This is similar to what a "humanize" plugin does, but achieved in a totally different way. We invite you to compare our final Ableton versions in figure 9 https://g.co/magenta/studio to the design mocks in figure 8.

DISCUSSION

As a result of the project, we identified strategies for promoting communication and productive work between UXers and ML research scientists:

- Show and tell is important for collaboration and outputs. Seeing concrete examples of each other's daily work early in the program helps the other group to understand and empathize with their new coworkers. It also sets expectations for the final output.
- Explain and understand processes and especially timelines. How long does it take to synthesize UX research findings, create the design language for a new UI, or train an ML model? Being upfront about how long key events take will make the timeline more actionable and realistic.
- Do a design sprint together. As mentioned in the design sprint section, sprinting together helps team members get on the same page about project goals and timelines, as well as share the excitement and bond over creating something amazing together.

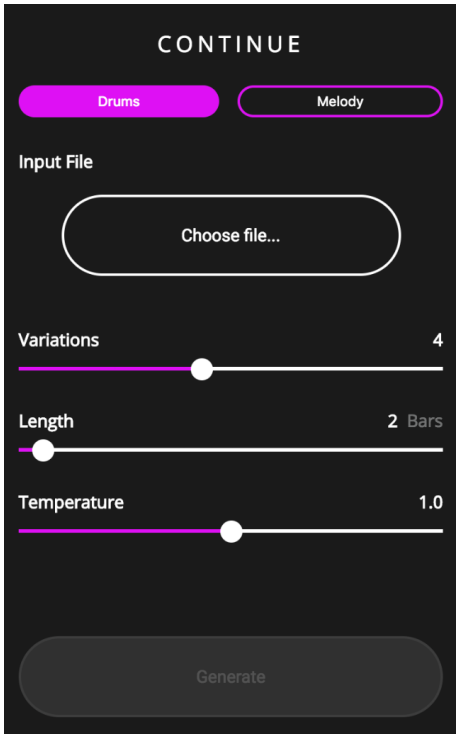

Figure 9: The Continue app from Magenta Studio. See g.co/magenta/studio **for more information. Compare to Figure 7. Keep in mind that Magenta Studio relies on Ableton Live to display and edit the generated sequences (Ableton Live is not shown here). That results in a simple Magenta Studio interface that can focus only on the controls.**

- Introduce designs early. The miscommunication that occurred around the model's capabilities may have been avoided if we had had more active discussions of the low-fidelity mocks.

Conclusion

Google has realized the importance of integrating UX more fully into the ML development processes to produce innovations with a broader reach and impact. While many UXers don't yet have an education in ML, methods such as the PAIR Bungee program will help to gain the skills needed to achieve a greater collaboration with research scientists and contribute unique UX skillsets to ML projects.

ACKNOWLEDGMENTS

We would like to thank Katherine Isbister, Curtis Hawthorne, and Ed Chi for their careful review and comments on this submission. We acknowledge Jess Holbrook and Josh Lovejoy, who created the PAIR Bungee program, and thank the many people who keep it afloat today.

REFERENCES

[1] Petra Bosch-Sijtsema and Jan Bosch. 2015. User involvement throughout the innovation process in high-tech industries. *Journal of Product Innovation Management* 32, 5 (2015), 793–807.
[2] Jesse Engel, Cinjon Resnick, Adam Roberts, Sander Dieleman, Douglas Eck, Karen Simonyan, and Mohammad Norouzi. 2017. Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. In *Proceedings of ICML 2017*. https://arxiv.org/abs/1704.01279
[3] Cheng-Zhi Anna Huang, Sherol Chen, Mark Nelson, and Douglas Eck. 2018. Towards Mixed-initiative generation of multi-channel sequential structure. https://openreview.net/forum?id=Sksy1ckvG
[4] Cheng-Zhi Anna Huang, Tim Cooijmans, Adam Roberts, Aaron Courville, and Douglas Eck. 2017. Counterpoint by Convolution. In *Proceedings of ISMIR 2017*. https://ismir2017.smcnus.org/wp-content/uploads/2017/10/187_Paper.pdf
[5] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M. Dai, Matthew D. Hoffman, Monica Dinculescu, and Douglas Eck. 2018. Music Transformer. https://arxiv.org/abs/1809.04281
[6] Jake Knapp. 2012. The Design Sprint. Retrieved September 28, 2018 from http://www.gv.com/sprint/.
[7] Rachel Krause. 2018. Storyboards Help Visualize UX Ideas. Retrieved August 26, 2018 from https://www.nngroup.com/articles/storyboards-visualize-ideas/.
[8] Geoffrey A. Moore. 2014. *Crossing the Chasm* (3rd ed.). Harper Business, New York, NY.
[9] Sageev Oore, Ian Simon, Sander Dieleman, and Douglas Eck. 2017. Learning to Create Piano Performances. In *NIPS Workshop on Machine Learning for Creativity and Design*. https://nips2017creativity.github.io/doc/Learning_Piano.pdf
[10] Adam Roberts, Jesse Engel, Colin Raffel, Curtis Hawthorne, and Douglas Eck. 2018. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. In *Proceedings of ICML 2018*. http://proceedings.mlr.press/v80/roberts18a.html
[11] Christian Rohrer. 2014. When to Use Which User-Experience Research Methods. Retrieved August 27, 2018 from https://www.nngroup.com/articles/which-ux-research-methods/.
[12] Julie Anne Seguin, Alec Scharff, and Kyle Pedersen. 2018. Triptech: A Method for Evaluating Early Design Concepts. In *Proceedings of CHI 2019*.